

Location Pattern Mining of Users in Mobile Environment

Mr. Mohammad Waseem

Mrs. R.R.Shelke

Abstract — With rapid increase of technologies, we require faster medium for communication and accessing the information. For achieving this goal of connecting different users with each other or with the applications they require, first we need to find their locations. After detecting the location or location patterns we easily provide users with different services. However, mobile applications are required to operate in pervasive computing environments of dynamic nature. Such applications predict the appropriate context in their environment in order to act efficiently.

Key Words — data mining, a priori algorithm, mobile access patterns, association rule mining.

I. INTRODUCTION

In this era of wireless technology and mobile computing user needs information within few seconds or even micro-seconds. Data collected from mobile phones have the potential to provide insight into the relational dynamics of individuals [1]. Advanced data mining techniques and a different network algorithm can be combined successfully to obtain a high fraud coverage combined with a low false alarm rate [2]. Mobile mining is about finding useful knowledge from the raw data produced by mobile users consists of a set of static device and mobile device. Previous works in mobile data mining include finding frequency pattern and group pattern that builds a user profile based on past mobile visiting data, filters and to mine association rules[5]. In order to render applications and services intelligent enough to support modern users everywhere / anytime and materialize the so called ambient intelligence, information on the present context of the user has to be captured and processed accordingly. The efficient management of contextual information requires detailed and thorough data modeling along with specific processing, reasoning and prediction capabilities.

We have proposed a model that deals with location prediction of moving users. The produced model predicts the next movement of a mobile user with certain moving profile and history of movements. Thus the representation of a user can contain both spatial and temporal information.

Mobility prediction can be define as the prediction of a mobile users next movement where the mobile user is traveling between the cells of a PCS or GSM network. The predicted movement can then be used to increase the efficiency of PCSs. By using the predicted movement, the system can effectively allocate resources to the most

probable-to-move cells instead of blindly allocating excessive resources in the cell-neighborhood of a mobile user. Effective allocation of resources to mobile users would improve resource utilization and reduce the latency in accessing the resources. Broadcast program generation can also benefit from predicted mobility patterns, since the data items can be broadcast to the cell where the users are moving [3]. Accurate prediction of location information is also crucial in processing location-dependent queries of mobile users. When a user submits a location-dependent query, the answer to the query will depend on the current location of the user [4].

II. LITERATURE REVIEW

The coverage area of the PCS network is partitioned into smaller areas which are called cells. In each cell in the PCS network, there is a base station (BS) which has the capability of broadcasting and receiving information. The base stations are connected to each other via a fixed wired network. Mobile users use radio channels to communicate with base stations. The coverage area consists of a number of location areas. Each location area may consist of one or more cells but in our work we assume that each location area consists of only one cell. Base stations regularly broadcast the ID of the cell in which they are located. Therefore, the mobile users which are currently in this cell and listening to the broadcast channel will know in which cell they are now. The movement of a mobile user from his current cell to another cell will be recorded in a database which is called home location register(HLR). In addition, every base station keeps a database in which the profiles of the users located in this cell are recorded. This database is called visitor location register (VLR). Therefore, in our system it is possible to get the movement history of a mobile user from the logs on its home location register. Since mobile users may initiate calls to other users or receive incoming calls while moving in the coverage region, the ongoing calls should be transferred from one cell to another without call dropping. To avoid call dropping due to insufficient resources at the destination cell, apriori resource allocation could be employed at that cell. In our work, we collect the movement trajectories of a user in the form of $T=h(id1,t1),(id2,t2),\dots,(idk,tk)$. Here $id1$ denotes the ID number of the cell to which the user enters at time $t1$. In this record it is clear that two consecutive ID numbers must be the ID numbers of two neighbor cells in the network. After the movement history of a user is collected in

a predefined time interval in the above format, this record is partitioned into subsequences. This procedure is accomplished as follows: If the mobile user stays in a cell id_i more than a threshold value, before moving to another one id_{i+1} , we assume that his trajectory up until now id_1, \dots, id_i ends here, and at id_{i+1} a new trajectory is started. Therefore, the first subsequence is id_1, \dots, id_i . By continuing in this manner the record is partitioned into subsequences, and these subsequences are recorded to be used in our algorithm. We name the trajectories obtained by the above procedure as user actual paths (UAPs). We consider the UAPs as a valuable source of information because the mobility of the users contains both regular and random patterns [5]. Therefore by using the UAPs, we may be able to extract the regular patterns and use them in prediction.

III. RELATED WORK

In mobile environment, moving log size is very large. It will increase the overhead to integrate all the moving logs into one database server. The algorithm proposed in [6] cannot be efficient for large data size. Many parallel and distributed variants of sequential apriori algorithm have been discussed in other resources [6] and [8]. In [7], grid implementation of frequent itemsets in a grid environment deals with sales transaction of a company. These algorithms [5], [6] and [7] cannot be used directly in our domain, because this algorithm does not take into account the network topology while generating the candidate patterns. The generation of candidate pattern in [9] is not same as the candidate pattern in mobile environment. In PCS, only the sequence of neighboring location of the network can be considered as the mobility pattern. We propose parallel and distributed knowledge grid mining approach based on Apriori algorithm for mining mobility patterns in mobile environment.

Mining user mobility patterns from graph traversals
 We define a user mobility pattern (UMP) as a sequence of neighboring cells in the coverage region network. The consecutive cells of a UMP should be neighbors because the users cannot travel between non neighbor cells. Indeed, UMPs correspond to the expected regularities of the user actual paths. In order to mine the UMPs from user actual paths (UAPs), sequential pattern mining [5] can be used. Sequential pattern mining has been previously used and examined in various research domains. One such work has been performed in the domain of web log mining [2,3]. In that work, sequential pattern mining is used to mine the access patterns of a user while he is visiting the
 126 G. Yavaset al. / Data & Knowledge Engineering 54 (2005) 121–146 pages of web sites. This method assumes the web pages to be the nodes and the links between these

pages to be the edges of an unweighted directed graph, G . Then, sequential pattern mining is applied to web logs by considering G . We design a new method that is convenient for our domain, by generalizing the method of [2,3] and applying it for UMP mining.

This new method employs

- a different definition of the graph G , and
- a new method for support counting, which generalizes the method presented in [2,3].

In our method, we use a directed graph G , where the cells in the coverage region are considered to be the vertices of G . The edges of G are formed as follows: If two cells, say A and B , are neighboring cells in the coverage region (i.e., A and B have a common border) then G has a directed and unweighted edge from A to B and also from B to A . These edges demonstrate the fact that a user can move from A to B or B to A directly. In example coverage region and the corresponding graph G is presented. The algorithm we have developed for UMP mining. To understand how the UMP mining algorithm works, assume that the set of candidate patterns each including k cells is found in the $(k+1)$ st run of the while loop and this set is not empty. The set of these patterns, denoted by C_k , is called length- k candidate patterns.

A priori algorithm for relationship between candidate and frequent locations used by it:-

C_k : Candidate itemset of size k

L_k : frequent itemset of size k

$L_1 = \{ \text{frequent items} \};$

for ($k = 1; L_k \neq \emptyset; k++$) do begin

$C_{k+1} =$ candidates generated from L_k ;

for each transaction t in database do

increment the count of all candidates in C_{k+1} that are contained in t

$L_{k+1} =$ candidates in C_{k+1}

with min_support

end

return $\cup_k L_k$;

Consider a database, D as shown in Table 1, consisting of 8 transactions. • Suppose min. support count required is 2 (i.e. $\text{min_sup} = 2/9 = 22\%$)

- Let minimum confidence required is 70%.

• We have to first find out the frequent itemset using Apriori algorithm.

• Then, Association rules will be generated using min. support & min. confidence.

Table 1:

UID	TIMESEMP	LOCATION
U1	T1	L1 L2 L5
U2	T2	L2 L4
U3	T3	L2 L3
U4	T4	L1 L2 L4
U5	T5	L1 L3
U6	T6	L2 L3
U7	T7	L1 L3
U8	T8	L1 L2 L3 L5

IV. EXPERIMENTAL RESULTS

Generating 4-itemset Frequent Pattern

The algorithm uses L 3 JoinL3to generate a candidate set of 4-itemsets, C4

Although the join results in $\{\{I1, I2, I3, I5\}\}$, this itemset is pruned since its subset $\{\{I2, I3, I5\}\}$ is not frequent.

Thus, $C4 = \emptyset$, and algorithm terminates, having found all of the frequent items. This completes our Apriori Algorithm.

These frequent itemsets will be used to generate strong association rules(where strong association rules satisfy both minimum support & minimum confidence).

Generating Association Rules from Frequent Itemsets

Procedure:

- For each frequent itemset “l”,generate all nonempty subsets of l.

- For every nonempty subset sof l, output the rule “s \rightarrow (l-s)”if $\text{support_count}(l) / \text{support_count}(s) \geq \text{min_conf}$ where min_conf is minimum confidence threshold.

- Back To Example:

We had $L = \{\{I1\}, \{I2\}, \{I3\}, \{I4\}, \{I5\}, \{I1,I2\}, \{I1,I3\}, \{I1,I5\}, \{I2,I3\}, \{I2,I4\}, \{I2,I5\}, \{I1,I2,I3\}, \{I1,I2,I5\}\}$.

– Lets take $l = \{I1,I2,I5\}$.

– Its all nonempty subsets are $\{I1,I2\}, \{I1,I5\}, \{I2,I5\}, \{I1\}, \{I2\}, \{I5\}$.

Generating Association Rules from Frequent Itemsets

- Let minimum confidence threshold is , say 70%.

- The resulting association rules are shown below, each listed with its confidence.

– L1: $I1 \rightarrow I2 \rightarrow I5$

- Confidence = $\text{sc}\{I1,I2,I5\} / \text{sc}\{I1,I2\} = 2/4 = 50\%$

- L1 is Rejected.

– L2: $I1 \rightarrow I5 \rightarrow I2$

- Confidence = $\text{sc}\{I1,I2,I5\} / \text{sc}\{I1,I5\} = 2/2 = 100\%$

- L2 is Selected.

– L3: $I2 \rightarrow I5 \rightarrow I1$

- Confidence = $\text{sc}\{I1,I2,I5\} / \text{sc}\{I2,I5\} = 2/2 = 100\%$

- L3 is Selected.

CONCLUSION

In this paper, we proposed a data mining method for mining user movements. We showed that the simple approach of computing by applying standard graph-matching algorithms and the DBMS primitives of grouping, sorting, and joining could be utilized to yield efficient match join operations. Moreover, a novel mining scheme was proposed to mine associated trees so that we can locate user behavior patterns.

ACKNOWLEDGMENT

It gives me; Mohammad Waseem great pleasure on bringing out paper entitled “Location Patterns Mining Of Users In Mobile Environment.”

I express my deep sense of gratitude and sincere regards to my Guide Prof. Mrs. R.R.Shelke for her timely guidance and helpful discussions has supported me immensely in selecting a new proposal and completing paper work.

REFERENCES

- [1] Nathan Eagle, Alex (Sandy) Pentland, David Lazer, “Inferring friendship network structure by using mobile phone data”. PNAS vol. 106, no. 36, pp. 15274-15278, 2009.
- [2] R. Brause, T. Langsdorf, M. Hepp, “Neural data mining for credit card fraud detection”. Tools with Artificial Intelligence, 11th IEEE International Conference, pp. 103 – 106, 1999.
- [3] Y. Saygin, O. Ulusoy, Exploiting data mining techniques for broadcasting data in mobile computing environments, IEEE Trans. Knowl. Data Eng. 14 (6) (2002) 1387–1399.
- [4] G. Gok, O. Ulusoy, Transmission of continuous query results in mobile computing systems, Inform. Sci. 125 (1–4) (2000) 37–63.
- [5] Jen Ye Goh, David Taniar, “Mobile Data Mining by Location Dependencies”. Intelligent Data Engineering and Automated Learning – IDEAL Lecture Notes in Computer Science, vol. 3177, pp. 225-231, 2004.
- [6] T. Liu, P. Bahl, I. Chlamtac, Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks, IEEE J. Select. Area Commun. 16 (6) (1998) 922–936.
- [7] Cristian Aflori and Mitica Craus. Grid implementation of Apriori algorithm. Advances in engineering software. Volume 38, Issue 5, 295-300, 2007.
- [8] R. Agrawal and J. Shafer. Parallel mining of association rules. IEEE Transaction on knowledge and Data Engineering, 8(6): 962-969, 1996.
- [9] Assaf Schuster, Ran Wolff and Dan Trock. A HighPerformance Distributed algorithm for mining association rules. Knowledge And Information Systems (KAIS) Journal, Volume 7, Number 4, May 2005.

AUTHOR'S PROFILE

	Mr. Mohammad Waseem M.E Final Year CSE, HVMP CET AMRAVATI, AMRAVATI UNIVERSITY, INDIA.
Passport Size Latest Color Photo	Prof . Rajeshri R. Shelke Associate Professor, HVMP CET AMRAVATI, AMRAVATI UNIVERSITY, INDIA.