Volume 4, Issue 4, April 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Data Mining Approach for Mobile based location prediction by Clustering Technique

Ms. Meghana Solanki, Prof. Vidya Dhamdhere

Abstract— Mobile prediction is one of the most essential issues that need to be explored for mobility management in mobile computing systems. In this paper we propose a new algorithm for predicting the next inter cell movement of a mobile user in a personal communication system network. In first phase Apriori and Cluster algorithm are used.User mobility pattern are mined from the history of mobile user trajectories. In second phase, mobile rules are extracted from these patterns, and in the last phase, mobility prediction are accomplished by using rules. The performance results obtained in terms of precision and Recall indicate that our method can make more accurate prediction.

Keywords: Mobility Prediction, Location Prediction, Data Mining etc

I. INTRODUCTION

Data mining is a widely used technique for discovering valuable information in a complex data set. Data mining (the analysis step of the"Knowledge Discovery in Databases" process, or KDD), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Mobility prediction can be defined as the prediction of a mobile users next movement where the mobile user is travelling 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 [6]. 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 [7]. Many application areas including health care, bioscience, hotel management, and the military benefit from efficient processing of location-dependent queries. With effective prediction of location, it may also be possible to answer the queries that refer to the future positions of users.

i. **Similarity Measure**: There have been many studies on measuring the similarity between two objects. The first one is based on multiple- level hierarchical structures. The concept of set similarity is to apply Measure to calculate the similarity of two sets. Let S1 and S2 be two sets, the set similarity set similarity (S1; S2) is defined as (1). However, set similarity is not applicable to store similarity in mobile commerce.

ii. **Mobile Pattern Mining**: Here to propose the TMSP-Mine for discovering the temporal mobile sequence patterns in a location-based service environment. To propose a prediction approach called Hybrid Prediction Model for estimating an object's future locations based on its pattern information. This considers that an object's movements are more complicated than what the mathematical formulas can represent. However, there is no work consider user relations in the mobile pattern mining.

iii. **Mobile Behavior Prediction**: The mobile behavior predictions can be roughly divided into two categories. The first category is a vector-based prediction that can be further divided into two types: 1) linear models, and 2) nonlinear models. The nonlinear models capture objects' movements with sophisticate regression functions. Thus, their prediction accuracies are higher than those of the linear models. Recursive Motion Function (RMF) is the most accurate prediction method in the literature based on regression functions. The second category is a pattern-based prediction.

II. LITERATURE SURVEY

Liu et al, [3] proposed a predictive mobility management algorithm. The proposed method modeled user's movement as elementary paths, which are either circular or straight patterns. The future location of the user is found using the mobile motion prediction (MMP) algorithm. Simulations showed that the proposed algorithm had a prediction efficiency of 95%. However, the efficiency starts decreasing as the randomness of the movement increases.

Yavas et al, [4] proposed a three-phase algorithm for mobility prediction. In the first phase, user mobility traces were extracted from the historical data. Mobility rules were derived in the second phase and finally mobility prediction was done in the third phase. The efficiency of the proposed scheme was tested and compared against two other prediction methods. The proposed method performed better than the other methods.

Akoush et al, [5] proposed a mobility path prediction model and hybrid Bayesian neural network model for predicting locations on cellular networks. The proposed model is based on the probability model to represent uncertainty in the relationships learned. Markov chain Monte Carlo method is applied to N sample values of posterior weights distribution obtained from the Bayesian training. These samples vote for the best prediction. Realistic mobility patterns were used for simulation studies. Results showed that the proposed algorithm achieved higher prediction accuracy when compared to standard neural network techniques.

Sakthi et al, [6] proposed a mobility prediction algorithm based on infrequent mobility patterns. The proposed algorithm computes new mobility patterns faster and avoids scanning of full database. The computation time for datasets at various support counts were computed using the proposed method. The results proved that the mobility patterns are generated in less time using proposed method than the re-computing approach. The work proved that live prediction can be done in a grid environment based on live data captured.

.

III. LOCATION MANAGEMENT TECHNIQUES

To route a call to a mobile, the network must be able to determine its exact location without incurring excessive computation and communication costs. Cellular networks partition their coverage area into a number of location areas. Each location area consists of a group of cells as shown in the Fig.1.

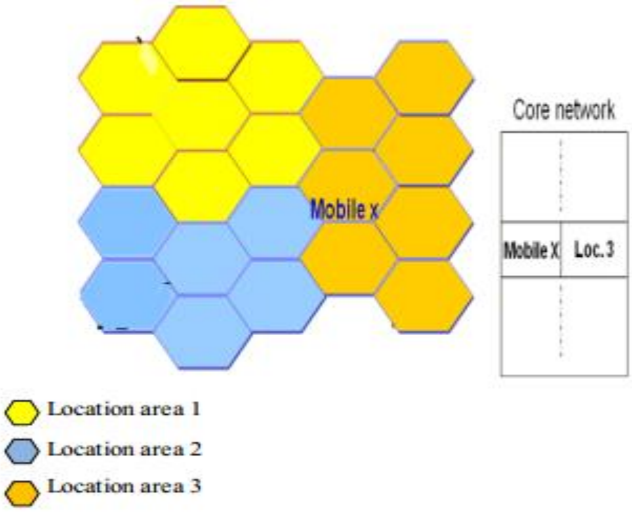


Figure 1: Grouping cells in location areas

The location management technique uses two fundamental operations [2].

• Location Update: Each mobile reports its location to the network whenever it changes a location area, so that any incoming call to that mobile can thus be routed to the correct location area.

• Paging: the network locates the mobile by polling simultaneously all cells within the location area, when an incoming call arrives. Page messages are broadcasted in the current location area of the mobile. For example, if the mobile x moves from Location area 2 to Location area 3, as shown in Fig. 1, the page message will be broadcasted to all cells forming this last location area. Generally, it is possible to use one or other of these operations. But in practice the two techniques are used conjointly to minimize the number of used messages. When the mobile is in a location area, it doesn't need to update its position, when it passes from one area to another it must send a location update message to the core network. This last knows the area where the mobile is, but does not know the exact cell. To find the cell where the mobile is, the network sends page messages to all cells constituting the area in which the mobile is localized.

This solution is valid only if the areas are well chosen. For example, if the mobile frequently moves on the periphery of the areas, a lot of update messages are sent to the network. This consumes a certain amount of limited wireless bandwidth. If the network has information on the users' profiles and the infrastructure (roads, workplace, means of transport …etc), it will be able to choose adapted location areas, capable to decrease (or to eliminate) the number of update or page messages.

ARCHITECTURE

In this Existing Architecture of a third generation mobile network formed by a set of cells. Each cell is managed by a base station that deals with mobiles that are in its range. The base stations are connected to the core network with a wired backbone (Figure 1).

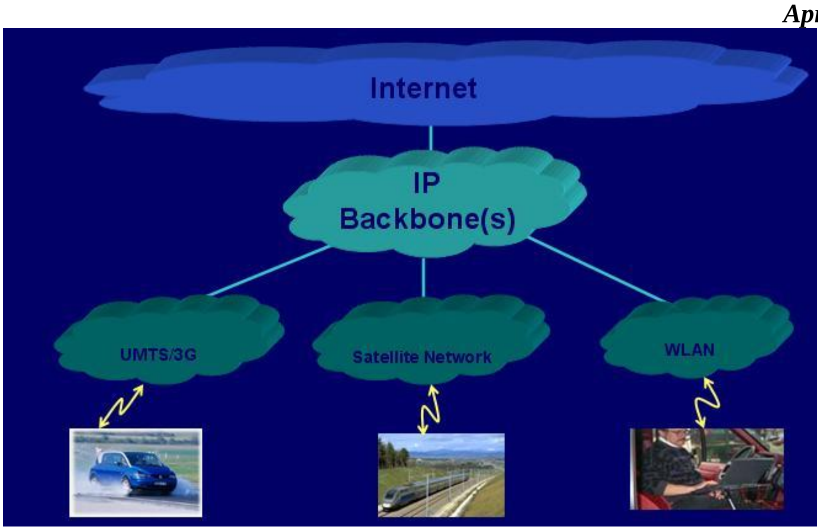


Figure 2. Architecture of a third generation mobile network

Every base station has a history of movements of mobile users. This history contains the mobile user id, the cell from which he came (Source cell), the cell to which he is moved (Destination cell) and the date of travel (Table 1).

Mobile ID	Source cell	Destination cell	Date

Table 1. Structure of a history line

This information can be retrieved in the connections log files at each base station and store them in a database of the station itself.

The association rules

Take into consideration a set of records, where every record consist of a set of items, then Support(X) is defined as the percentage of records that actually contain the item set X. an association rule is defined by the expression

$$X \longrightarrow Y, [c, s] \; [1]$$

Where,

$X \cap Y = \emptyset,$ and S=support(xUy) which corresponds to the support value of the rule,

and

$c = \frac{support(X \cup Y)}{support(X)}$ is the confidence.

The association rules can be used to create location areas. They can also be used for the long-term resource reservation because they provide a chronological order of the crossing cells. We start by sorting the displacements history of a mobile by the decreasing order of the date. From this history we generate lists in the form of Table 2 where:

- Cell 1 is the most recent cell in which the mobile is located.
- Cell 2 is the cell which the mobile had crossed before going to cell 1.
- Cell 3 is the cell which the mobile had crossed before going to cell 2.

Mobile ID	Cell 1	Cell 2	Cell 3	...	Cell K

Table 2. Structure of an element of the list for association rules

Apriori Algorithm

The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules.

Key Concepts :

- Frequent Itemsets: The sets of item which has minimum support (denoted by L_i for i^{th}-Itemset).
- Apriori Property: Any subset of frequent itemset must be frequent.
- Join Operation: To find L_k , a set of candidate k-itemsets is generated by joining L_{k-1} with itself.
- Join Step: C_k is generated by joining $L_{k-1}$with itself
- Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

Algorithm Apriori(*T*)

$C_1 \leftarrow$ init-pass(T);

$F_1 \leftarrow \{f \mid f \in C_1, f.\text{count}/n \geq minsup\}$; // n: no. of transactions in T

for ($k = 2$; $F_{k-1} \neq \emptyset$; k++) **do**

  $C_k \leftarrow$ candidate-gen(F_{k-1});

  for each transaction $t \in T$ **do**

    for each candidate $c \in C_k$ **do**

  if c is contained in t **then**

   $c.count$++;

    end

  end

   $F_k \leftarrow \{c \in C_k \mid c.count/n \geq minsup\}$

end

return $F \leftarrow \bigcup_k F_k$;

A set of lists of K elements (called k-items set) indicating the K last cells crossed by a given mobile. We apply then the algorithm Apriori to seek for association rules of order 1, 2, 3… k. An association rule of order 1 is a rule in the form CellA − − > CellB. It means that if mobile user is in the cell A it is probable that he will be in the cell B. It is obtained by searching in the lists of Table 2, the elements that satisfy a minimum support and confidence calculated as follows:

$$\text{Support(CellA-->CellB)} = \frac{\text{Nbr. of app. of Cell A and Cell B in the same list}}{\text{Total number of lists}}$$

$$\text{Confidence(CellA-->CellB)} = \frac{\text{Nbr. of app. of Cell A and Cell B in the same list}}{\text{Nbr. app. of Cell A in the lists}}$$

Confidence indicates if this rule is verified by indicating whether the right side element of the rule appears whenever the left side element appears. The support indicates if this rule
is often verified and not only in particular cases by indicating whether the left side of the rule appears sufficiently in the database.
The simulator rests on the statistics of displacement led in the area of Waterloo and recorded in the form of matrix called activity matrix indicating the probability of arrival of an activity and duration matrix indicating the probability that an activity takes a given period. These statistics as well as information concerning the users like the profile (fulltime employed, student, part-time employed ,.etc) and the infrastructures (roads, trade, stadium, etc) are recorded in the simulator database. The area of Waterloo is divided into 45 cells as indicated in Fig. 3.

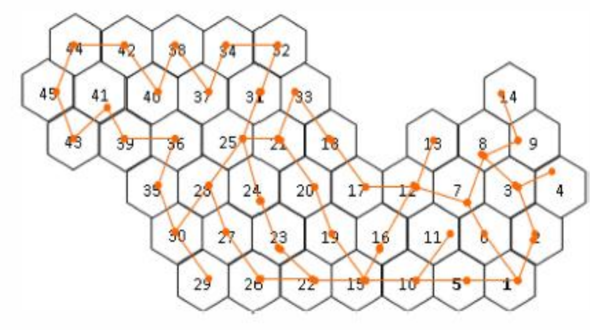
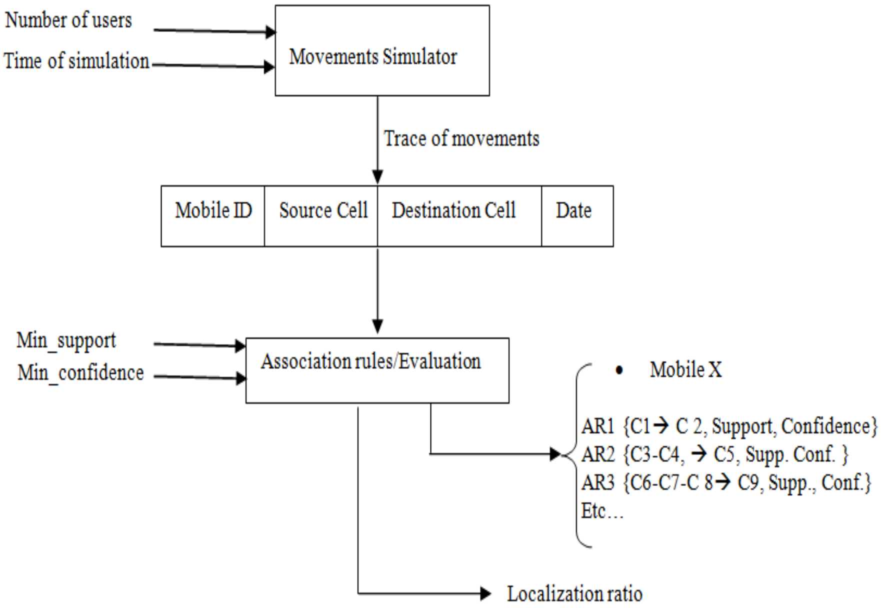


Fig. 3. Cellular structure of the simulator

SIMULATION MODEL

The simulation approach is shown in Figure 3. We use the movements' simulator to generate a trace of realistic movements according to the number of users and the simulation time (in days) provided as parameters. We then provide this trace to our algorithm of association rules search and evaluation which returns a set of rules (ARs) with the support and confidence of each rule. The rules are limited by the minimum support and minimum confidence provided to the algorithm as parameters. The algorithm calculates also the localization ratio which is the number of correct localizations by using the association rules on the number of localization attempts.



MOVEMENT PREDICTION AND CLUSTERING

The clustering can be used to group cells into location areas, as shown in Fig. 1. A location area is a restricted set of cells in which a mobile user can be located according to his profile and his history. We define the distance between two cells (Ci and Cj) relative to an individual X as follows:

$$D(C_i, C_j) = \sqrt{n_i^2 - n_j^2} \qquad (1)$$

Where:

$$n_k = \frac{\text{Number of appearances of the individual } X \text{ in cell } C_k}{\text{Total number of appearances of } X}$$

The used clustering algorithm is a variant of the K-mean algorithm [8]. First, we set the desired number of clusters and then we apply an iterative algorithm from which we obtain the set of cells forming each cluster. The clustering algorithm is described as follows:

1. Determining the number of desired clusters (the parameter K).
2. Randomly choosing K cells that we consider as centers of the K clusters.
3. Calculating the distance between each cell of the network and the centers (centers cells).
4. Assigning each cell to the nearest cluster according to the distance between this cell and the center cells by using the equation (1).
5. Recalculating the centers of the K clusters as follows:

   a. For each cell C*j* of a given cluster, calculate the value :

$$\sum_i D(C_j, C_i) \;\; with\ (j \neq i) \qquad (2)$$

this represents the sum of the distances between this cell and the other cells Ci in the considered cluster.

   b. The cell Cj with the smallest value of equation (2) is considered as a new center of the cluster.

6. Repeat steps 3, 4, 5 until the clusters become stable (the cells do not change the clusters).

ADJUSTMENT AND EVALUATION

The clustering algorithm requires training phase for adjusting its parameters. This training yields the optimal value of K that corresponds to the optimal number of location areas. This step must be conducted by using in general a sample of realistic data. In the study of the mobility management and in the absence of a real trace of mobiles displacement, we can resort to a model. The choice of a realistic mobility model is essential. This model reproduces, in a realistic way, displacements of a set of users within a network. The majority of works presented in the literature use probabilistic models (Markov model, Poisson process…etc.) which generate either highly random displacements or highly deterministic displacements which do not reflect the real behavior of the mobile users. For this we have chosen the activity model presented in [9]. This model is based on the work carried out by planning organizations and uses statistics drawn from five years surveys on users' displacements. It simulates a set of users' displacements during a number of days. Generated displacements are based on each users activity (work, study… etc.), the locations of these activities (house, work places, schools…etc) as well as the ways which lead to these locations.

IV. CONCLUSION

In this paper, we present a data mining Apriori and Clustering algorithm for the prediction of user movements in a mobile computing system. The algorithm proposed is based on mining the mobility patterns of users, forming mobility rules from these patterns, and finally predicting a mobile user_ next movements by using the mobility rules. Through accurate prediction of mobile user movements, our algorithm will enable the system to allocate resources to users in an efficient manner.

REFERENCES

[1] Sunghyun Choi and Kang G. Shin, Predictive and Adaptive Bandwith Reservation for Han-doffs in QoS-Sensitive Cellular Networks, IEEE1998
[2] X. Shen, J. W Mark and J. Ye, User Mobility Profile Prediction: An Adaptive Fuzzy Interference Approach, Wireless Network 6, (2000) PP362-374.
[3] D. Ashrook and TStaruer, Learning Significant Locations and Predicting user Movement with GPS, Proceedings of the 6th International Symposium on Wearable Computers TSWC'02.
[4] WS. Soh and Hyong S. Kim, QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques, IEEE Communication Magazine, January 2003, PP 86-92.
[5] S.c. Liou and H.C. Lu, Applied Neural Network for Location Prediction and Resource Reservation Scheme in Wireless Network, Proceedings of ICCT,2003.
[6] J. Capka and R. Boutaba, Mobility Prediction in Wireless Networks using Neural Networks, International Federation for Information Proceeding, IFIP2004.
[7] N. Samaan, and A. Karmouch, A Mobility Perdiction Architecture based on Contextual Knowledge and Conceptual Maps, IEEE transactions on mobile computing, Vo14, No 6 November/December 2005
[8] J. M. Franois and G. Leduc, Entropy-Based Knowledge Spreading and Application to Mobility Prediction, ACM CoNEXT'05 October 24-27, 2005, Toulouse, France.
[9] M. Daoui, A. M'zoughi, M. Lalam, M. Belkadi and R. Aoudjit, Mobility prediction based on an ant system, Computer Communications, , 31(2008) 3090-3097.
[10] M. Daoui, A. M'zoughi, M. Lalam, R. Aoudjit and M. Belkadi,Forecasting models, methods and applications, mobility prediction .

