Synthesis of Two Solutions of Mobility Prediction Based on Data Mining Techniques

Linda Chamek
Laboratory LARI
Informatique Departement
UMMTO
Tizi-Ouzou, Algeria
chameklinda@yahoo.fr

Mehammed Daoui
Laboratory LARI
Informatique Departement
UMMTO
Tizi-Ouzou, Algeria
mdaouidz@yahoo.fr

Malika Belkadi
Laboratory LARI
Informatique Departement
UMMTO
Tizi-Ouzou, Algeria
Belkadi_dz@yahoo.fr

Mustapha Lalam
Laboratory LARI
Informatique Departement
UMMTO
Tizi-Ouzou, Algeria
lalamustapha@yahoo.fr

***Abstract*— In this paper, we present a synthesis of two solutions for mobility prediction using data mining techniques such as classification and clustering. Our solution can be implemented in a third generation network by exploiting user information (age, function, residence place, work place, etc.), the existent infrastructure (roads, etc.) and the history of displacements. Simulations carried out using a realistic model of movements showed that our strategy can accurately predict up to of 80% of mobiles' displacements and this by not knowing their history of mobility**.

Keywords - mobile network; prediction; data mining; location management.

I. INTRODUCTION

Nowadays, mobile networks have become an integral part of our daily life. Third generation networks open the way with new demands for services in multi-media and real time applications. These applications require more communication resources and higher Quality of Service (QoS) than traditional applications. However, these networks are confronted with various problems including resources wasting and signal attenuation that reduce the QoS. User mobility also generates QoS degradation and the network must deal with it.

Two functions are essential in mobile networks: location management and resource reservation. The location management locates the cell where a mobile user is in order to make a call to him. The resources reservation is intended to ensure continuity of communication when a mobile moves from one cell to another by reserving bandwidth in the cells he goes through.

The mobility of users causes performance degradation in relation to the two previous functions. For the localization, the network sends location messages (page messages) to all cells in order to locate the mobile. These messages consume a part of the bandwidth. For the reservation of resources, the network is often required to reserve resources in cells that the mobile will not cross.

Now, if the network would have information about the displacements of every mobile, and if it integrates intelligent strategies to take advantage of this information, it can anticipate their future movements with high accuracy. This way, the network can better manage its resources.

In this work, we present two solutions for mobility prediction based on data mining techniques: classification and clustering. The rest of the paper is organized as follows. In Section II, we present a state of the art of different techniques of prediction. In Section III, we present the importance of the use of data mining in mobility prediction. In Section IV, we propose a prediction solution and, in the last section, we present the evaluation of the solution and the simulation results.

II. STATE OF THE ART

Several techniques allowing prediction have already been discussed.

One of the most used techniques rests on the localization by Global Positioning system (GPS). The mobile sends its position obtained by GPS to its base station. The latter determines if the mobile is at the edge of its cell. At each reception of the position from the mobile, the system calculates the distance separating the mobile from the neighboring cells, and the shortest (near) distance is selected [1].

L. Hsu *et al.* [2] suggest a solution based on the definition of a reservation threshold. The idea is to compare the signal received by the mobile coming from the neighboring cells. If this signal is lower than the threshold, it is concluded that the mobile moves towards this cell. In [3], a map of signal power is maintained by the system. It represents the various signals recorded in various points of the cell. The authors use this map to know the position of the mobile, and to extrapolate his future position.

In [4], [5], mobility rules are generated based on a history of the movements that each mobile built and maintained during its displacement. These rules are used in the prediction process. In fact, it was observed that users tend to have a routine behavior. Knowing that, and knowing the usual behavior of the users, it becomes possible to predict the next cell which a user will visit.

Soh et al. [6] propose a technique based on the use of a multi-layer neural network in order to exploit the history of the mobile movements. The recent movements of the mobile are initially collected in order to know in which Location Area (LA) it is. A mobility model of the users is initially processed, and then it will be injected into the neural network. The mobility model represents the mobile movements history recorded in an interval of time. The movement is defined in terms of the direction taken and the distance covered.

The role of the neural network is to capture the unknown relation between the last values and future values of the mobility model; that is necessary for the prediction.

The authors in [7] propose an algorithm composed of three phases. The first phase consists in extracting the movements of the mobile to discover the regularities of the inter-cellular movements; it is the mobility model of the mobile. Motility rules are extracted from the preceding model in the 2nd phase. Finally, in the 3rd phase, the prediction of mobility is accomplished by using these rules.

Capka *et al.* [8] propose a new mobility prediction algorithm. The user's behavior is represented by the repetition of some models of elementary movements. To

 ISBN: 978-1-61208-404-6

estimate the future position of a mobile, the authors propose an aggressive mobility management Predictive Mobility Management (PMM). A whole of prevision algorithms MMP (Mobile Motion Prediction) is used to predict the next position of the mobiles based on their movements' history.

The authors in [9] propose a diagram of prediction combining between two levels of prediction: global and local. The global mobility model Global Mobility Management (GMM) is given in terms of cells crossed by a mobile during its connection time. The local model Local Mobility Management (LMM) is given by using sample of 3-tuples taking into account three parameters: speed, direction and position. LMM is used to model the intra-cellular movements of a mobile, whereas the GMM is used for the inter-cellular movements by associating its current trajectory with the one of the existing mobility rules. However, the authors do not present any method allowing the discovery of these mobility rules.

A method called Dynamic Clustering based Prediction (DCP) is presented in [10]. It is used to discover the mobility model of the users from a collection containing their trajectories. These rules are then used for the prediction. The trajectories of the users are grouped according to their similarities.

Daoui *et al.* [11][12] present a technique of prediction based on the modeling of mobile displacements by an ants system. This model allows the prediction based on old displacements of the mobile and those of the other users who go in the same direction.

Chamek et al. [13] use a technique of prediction based on classification of users according to their personal profile (age, sex, place of work, etc). A new user is compared to all other users in cell and put in a class, then the history of displacements of the users in this class is used for predicting the next cell of the new user.

A technique based on clustering is presented by Belkadi *et al.* in [14]. This technique can be implemented on next generation mobile networks by exploiting the data available on the users (age, function, address, workplace, etc), existing infrastructures (roads, location of base station, etc.) and the users' displacements history. Locations areas are formed according to these different pieces of information.

III. DATA MINING AND MOBILITY PREDICTION

The mobile's displacements are often generated by socio-economic needs and are governed by the topography of the roads and infrastructures covered by the various cells of the network such as: schools, factories, supermarkets, highways, etc. The displacements related to the socio-economic needs are usual, and consequently, represent a regular aspect [15].

Information concerning a user characteristics, in other words, the profile, are also of great importance. In fact, knowing certain characteristics of a user helps us to know his future displacements with a great probability. For example, a person of an age ranging between 18 and 25 years, who is student, will probably be located in the campus one day of week. People having high incomes will most probably make their purchases in luxury shops, contrary to others who will prefer supermarkets. The profiles of mobility of these people are thus different.

Many definitions of data mining can be found, so this domain is the subject of research. Engineers, statisticians, economists, etc., can have different ideas on what this term means. We retain a definition which seems to make the compromise between various designs. We can define data mining as the process allowing the extraction of predictive latent information from wide database [16].

Classification aims to predict the class of a new user based on the class of users who are in data base. In our case, it is to predict the future cell of a user based on his last displacements and those of other users who have the same profile. The basis of this idea is that displacements of users are often regular and individuals of the same profile perform similar movements.

Clustering allows to group cells in clusters, thus forming a location area, to facilitate the search of the mobile in the network.

IV. PRESENTATION OF THE SOLUTIONS

Assume we have an architecture of third generation network composed of a set of cells. Every cell is generated by a base station. The base stations are connected to the core network wired backbone (Figure 1). We assume that the core network has personal and professional information about users such as age, marital status, occupation etc. This information may be collected when subscribing to network services. We also assume that each base station has a history of movements of mobile users.

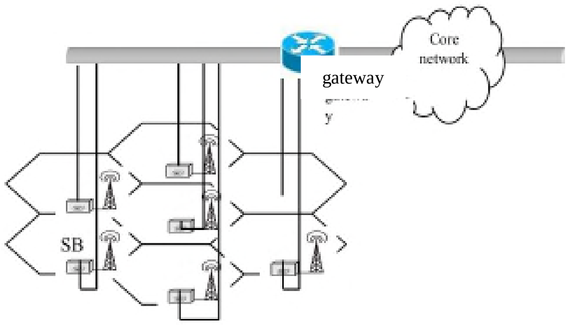

Figure 1. Architecture of a third generation mobile network

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 I). This information can be retrieved in the connections log files at each base station and stored in a database on the station itself.

TABLE I. STRUCTURE OF A HISTORY LINE

ID Mobile	Source Cell	Destination Cell	Date

A. Principle of the prediction technique based on classification

Classification is used to predict the future cell of a mobile user x. We select N mobile users who are already located in the same cell. We compare the user x to every individual y of N users using distance D [12]:

$$D(X,Y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \qquad (1)$$

where:

- x_i and y_i are values of attributes of individuals X, Y (like age, function, etc.)
- m is the number of known attributes of individual

So, we select K individuals nearest to X. Then, we take the displacement history of these K individuals. The destination cell which is the most frequented is considered the future cell prediction.

The algorithm proposed is summarized as follows:

Let I= {Y1, Y2 ,…. Yn} be the set of the N individuals being in cell C.

Entry : Let X be a new individual for whom we want to predict the future cell.

Parameter K corresponds to the number of nearest neighbors to take into account.

Parameter L corresponds to the number of history lines to take for each nearest neighbor.

Exit: the future cell to predict.

Algorithm:
For *(j = 1 to N)* ***Do***
1. Calculate the distance between Yj and X of the cell C d(X, Yj)
2. Record this distance in the vector tab
Done
3. Sort the calculated distances (the vector)
4. Select K smaller distances,
5. Select the history of K individuals closest to X
6. Determine the most frequent destination cell and return it as the future cell to predict

B. Principle of the prediction technique based on clustering

Clustering can be used to form location areas. It is a set of cells in which a mobile can as a function of its history and profile. We define the distance between two cells according to an individual X [13]:

$$D(C_1, C_2) = \sqrt{(n_1 - n_2)^2} \qquad (2)$$

with:

$$n_i = \frac{\text{Number of apparition of individual X in the cell i}}{\text{Total number of apparition s of the individual X}}$$

Then, we apply the k-mean algorithm [14] to select K homogeneous clusters. Such cluster contains a set of cells having as a common factor the frequency of visit of a mobile. The cluster having the higher number of visits is considered the appropriate location area for the mobile.

C. Localisation

The location procedure (paging) that we propose uses an intelligent paging resting on the location areas that are built by the clustering algorithm, such that:

- The first zone of localization is composed of two cells:
  * The cell in which the mobile was at the time of its last call
  * The predicted cell, by our algorithm of prediction, starting from the cell in which the mobile was at the time of its last call.
- The second zone is composed of the adjacent cells to the last cell in which the mobile was at the time of its last call by excluding the cells of the first zone of localization.
- The third zone of localization is composed of the other cells of the network.

Localization Procedure:
- Search the mobile in the first location zone
- **if** the mobile is not in the first zone **then**
  - Search it in the second location zone
- **if** the mobile is not there **then** search it in the third location zone, etc.

V. ADJUSTMENT AND EVALUATION

Most data mining algorithms require a training phase to adjust their parameters. In case of the classification, it provides the best value of two parameter K (number of nearest neighbor to take into account) and L (number of line of history to take by close neighbor).

For clustering, we found the best value of parameter M which represents the optimal number of location zone to create for the network.

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 the 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.

In our approach, we have chosen the activity model presented in [18]. This model is based on the work carried out by planning organizations and uses statistics drawn from five years of surveys on user displacements. It simulates a set of user displacements during a number of days. The generated displacements are based on each user's activity (work, study, etc.), the locations of these activities (house, work places, schools) as well as the ways which lead to these locations.

The simulator rests on the statistics of displacement led in the area of Waterloo [17] 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 of time. These statistics, as well as information concerning the users, such as the profile (full time employee, student, part-time employee, etc) and the infrastructures (roads, trade, stadium, etc) are recorded in the simulator database.

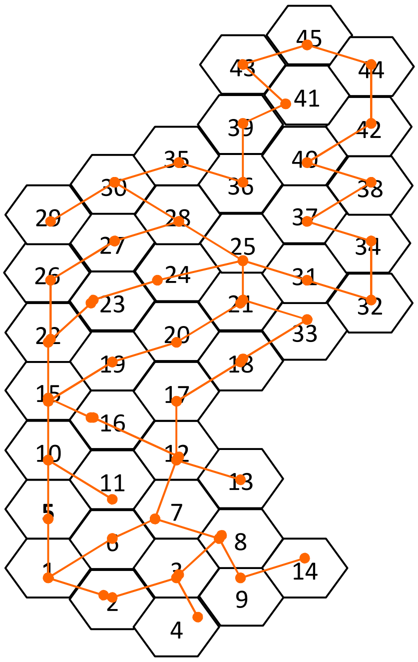

Figure 2. Cellular structure of the simulator

The area of Waterloo is divided into 45 cells, as indicated in Figure 2. According to the activity of the user, the simulator generates an activity event for a user based on the activity matrix and assigns it to a cell. It generates then the displacements relative to this activity before generating the following activity. The process continues until the end of the simulation.

D. Adjustment and evaluation of the classification algorithm

The adjustment of the classification algorithm consists of determining optimal values for the two parameters K and L. K is the optimal number of neighbors to consider, and L is the number of history lines to take for each near individual. The evaluation is done based on rate prediction which is the ratio of the number of correct predictions to the total number of attempts to predict.

Figure 3 shows the ratio prediction as a function of the parameter K, with the L value fixed to 4 (take 4 lines of history for each neighbor). The prediction ratio rises to stabilize at K = 30 with ratio prediction of 60%.

Figure 4 gives the ratio prediction as a function of the parameter L when K is fixed to 30 (the value that we get above). The prediction ratio rises with the rise of the value of L. A better ratio is obtained for L= 45 with a ratio of 70%. So, we can only keep the last 45 displacements of neighbors' users.

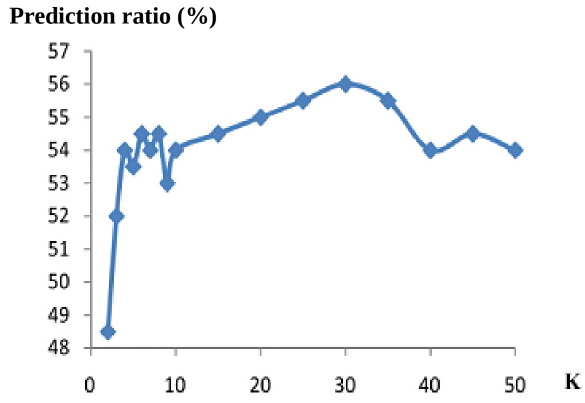

Figure 3. Prediction ratio according to parameter K (optimal number of neighbor)

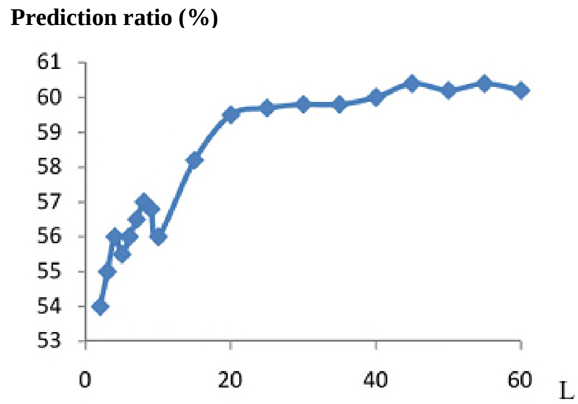

Figure 4. Prediction ration according to parameter L (number of history lines to take for each near individual)

E. Adjustment and evaluation of the clustering algorithm

The adjustment of clustering consists in determining the value of M corresponding to the optimal number of location zones that we need to create for the network. The evaluation is based on the number of paging messages and the update ones. For 100 days of simulation, we have varied M and calculated the total number of messages (paging and update). The results are shown in Figure 5. The optimal number of location zone is 20 with 2 to 3 cells per zone.

 ISBN: 978-1-61208-404-6

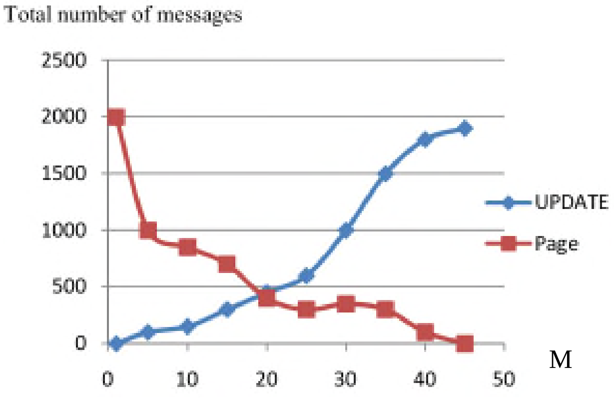

Figure 5. Page messages and updates ones in function to M (optimal number of cluster)

Next, we compared our algorithm with several algorithms, both static and dynamic. In the static strategies, cells are grouped on static location areas (0, 1, 2 and 3). The strategy static 0 contains a cell in each location area. In static 1, 13 locations areas are created, each one having 3 or 4 cells. The static 3 divides the network in 5 areas of 8 to 10 cells. In the dynamic strategy, the algorithm defined in [17] is used. The result of this comparison is illustrated in Figure 6 and Figure 7: the number of update messages and page messages are calculated during 50 days of simulation with 3, 9 and 12 calls per day.

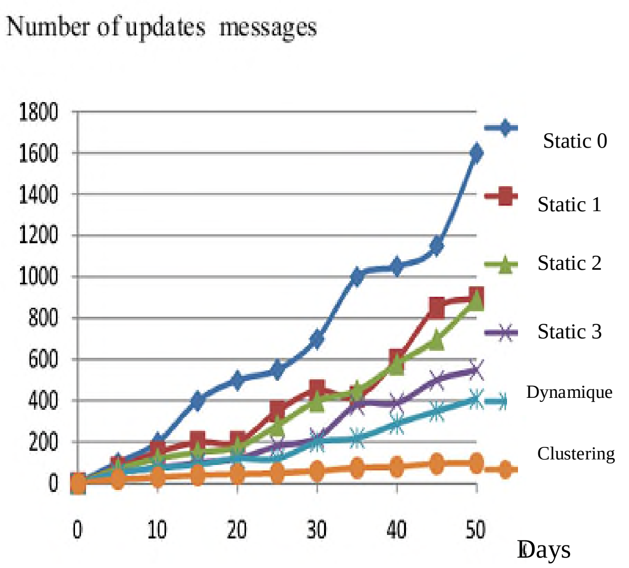

Figure 6. Number of Update messages for 6 strategies in function of days

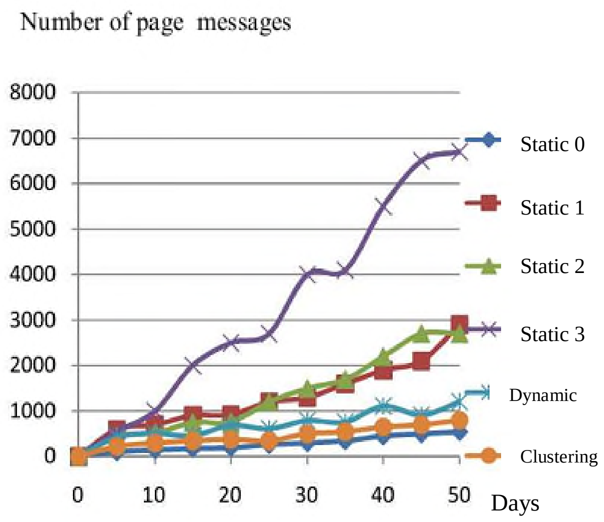

Figure 7. Number of page messages for 6 strategies in function of days

The results show that our solution produces the minimum number of page messages comparing with the other strategies.

VI. CONCLUSION

Human displacements are often caused by socio-professional needs. They are linked to existing infrastructure (roads, transport, workplace location, etc.). It is therefore possible to predict the future position by looking for links between these movements and other available information such as user profiles and the location of the infrastructure.

Due to the complexity of the characteristics of human mobility and the absence of reliable mobility rules for prediction movements, data mining can be a solution to the problem of prediction. Both techniques presented in this paper show that it is possible to predict 70% of the movements of mobile users.

REFERENCES

[1] W. Zhuang, K.C. Chua and S. M. Jiang, Measurement-Based Dynamic Bandwidth Reservation Scheme for Handoff in Mobile Multimedia Networks, 1998.

[2] L.Hsu, R. Purnadi and S.S. Peter Wang, Maintaining Quality of Service (QoS) during Handoff in Cellular System with Movement Prediction Schemes, IEEE 1999.

[3] Sunghyun Choi and Kang G. Shin, Predictive and Adaptive Bandwidth Reservation for Han-doffs in QoS-Sensitive Cellular Networks, IEEE 1998

[4] X. Shen, J. W. Mark and J. Ye, User Mobility Profile Prediction: An Adaptive Fuzzy Interference Approach, Wireless Network 6, (2000) PP 362-374.

[5] D. Ashrook and T. Staruer, Learning Significant Locations and Predicting user Movement with GPS, Proceedings of the 6th International Symposium on Wearable Computers TSWC'02.

[6] W.S. Soh and Hyong S. Kim, QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques, IEEE Communication Magazine, January 2003, PP 86-92.

[7] S.C. Liou and H.C. Lu, Applied Neural Network for Location Prediction and Resource Reservation Scheme in Wireless Network, Proceedings of ICCT, 2003.

[8] J. Capka and R. Boutaba, Mobility Prediction in Wireless Networks using Neural Networks, International Federation for Information Proceeding, IFIP2004.

[9] N. Samaan, and A. Karmouch, A Mobility Perdiction Architecture based on Contextual Knowledge and Conceptual Maps, IEEE transactions on mobile computing, Vol4, No 6 November/December 2005

[10] J. M. François and G. Leduc, Entropy-Based Knowledge Spreading and Application to Mobility Prediction, ACM CoNEXT'05 October 24-27, 2005, Toulouse, France.

[11] 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.

[12] M. Daoui, A. M'zoughi, M. Lalam, R. Aoudjit and M. Belkadi, Forecasting models, methods and applications, mobility prediction in cellular network, i-concepts press, 2010 (221-232)

[13] L. Chamek, M. Daoui, M. Lalam. Mobility prediction based on classification according to the profile Journées sur les rencontres en Informatique R2I 12- 14 juin 2011

[14] M. Belkadi , R. Aoudjit, M. Daoui, M. Lalam Mobile Localization Based on Clustering, I. J. Computer Network and Information Security, 201 3, 9 , 37 -44

[15] P. Hu and J. Young, 1990 Nationwide Personal Transportation Survey (NPTS), Office of Highway Information Management, October 1994.

[16] J. Han, M. Kamber, Data mining concepts & techniques, Morgan Kaufmann, 2 edition, 2006

[17] J. Scourias and T. Kunz, An Activity-based Mobility Model and Location Management Simulation Framework, Proc., Second ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM), 1999, PP. 61-68.