Mobility Prediction Based on Data mining

Rachida Aoudjit, Malika Belkadi, Mehammed Daoui, Lynda Chamek, Sofiane Hemrioui and Mustapha Lalam

Laboratoire de recherche en informatique,
Université Mouloud Mammeri de Tizi Ouzou, Tizi Ouzou, Algeria
aoudjit_rachida@yahoo.com

Abstract

This paper presents a mobility prediction technique based on one of the most used Data mining technique which is the association rules. Our solution can be implemented on a next-generation mobile networks by exploiting the data available on existing infrastructure (roads, locations of base stations ...etc.) and the users' displacements history. Simulations carried out using a realistic model of movements showed that our strategy can accurately predict up to 90% of the users' movements by knowing only their last two movements.

Keywords: *Mobile networks, Prediction, Data mining, Association rules*

1. Introduction

Nowadays, mobile networks become an integral part of our daily life. Next 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 to various problems including resources wasting and signal attenuation that reduce the quality of service (QoS). Users 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 locate 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.

User mobility causes performance degradation in relation to the two previous functions. For location management, the network use Page and Update messages to locate the cell where a mobile user is located. These messages consume a part of the already scarce bandwidth. In resource reservation, the network is often required to reserve resources in cells that the mobile will not cross. These resources will not be used even if other mobiles need them.

If the network has enough information about the mobile user displacement and if it integrates intelligent strategies to profit from this information, it can anticipate his future movement with high accuracy. Consequently, the network can better manage its resources mobilization and sharing by sending a minimum (or nothing) of location messages and reserving resources for appropriate time in the actual future cells that the mobile will visit.

In this work we present a solution for the movements prediction and location management based on the association rules. This technique is based on various characteristics of human movement, places topography and infrastructures locations.

2. State of the Art

Many studies were carried out on prediction. In [1], Zhuang et al. monitor the signal power in the base station to predict the next cell. When the signal in a new base station increases in power, the system concludes that the mobile moves towards this station. This solution was improved in [2] by adding the time factor to predict the mobile arrival time. In [3], Choi, *et al.*, estimate the users mobility according to displacements history observed in each cell. In [4], Shen et al. Developed a prediction system based on the measurement of a pilot signal, and using fuzzy logic, to take into account pilot signals interferences users random movements. In [5, 6], the authors use mobiles localization information recovered by GPS, which they provide to a Markov model to predict the future location. In [7], Soh, *et al.*, present and describe the use of a prediction technique which incorporates roads topographic information; periodically (every second for example), the mobile provides its position to its base station. An algorithm identifies the segment of the road taken by the mobile and its estimated speed. The system then predicts the future base station and estimates the expected time to reach it.

These solutions are limited because they are based on either a probabilistic model which does not completely reflect the users behavior, or on the users individual history which can be missing or insufficient. New solutions based on heuristic methods were used such as those presented in [8, 9, 10] that use neural networks for prediction. In [8], the authors adopted a structure of a cell divided in 2-tier areas (Avicinity area andan edge area). When the mobile is in a cell edges area, its coordinates are provided as input to the neural network which predicts the next cell to be visited. In [9], the authors present a system which periodically captures the connections traces. Theses traces are progressively recorded giving a history record which will be used as input to the neural network to predict the next cell to be visited.

The disadvantage of these methods is that they require a long training phase on mobile user behavior before the prediction succeeds. Moreover, the mobile user can change his behavior during the training phase or can go to a location he has never visited before, thus making the prediction ineffective. In [11], Samaan, *et al.*, present a solution which includes spatial and user contexts. The spatial context consists of a set of abstract maps describing the geographical environment in which the mobile user progresses. Places, buildings and roads which lead to these places are described in theses maps. The user context includes a set of information related to the mobile user making it possible to predict his mobility. This information is then combined using Dempster Shafer algorithm to predict the future mobile location. Even if this solution seems to provide suitable results when mobile history is lacking, it is however too constraining because it requires additional information not easy to acquire and likely to frequent change. In [12], the authors suggest a method based on information theory in which each mobile device collects a set of clues such as its position in a particular road and its stay time during its previous displacements. These clues are then processed by the current cell to predict the future cell. The use of information theory and decision trees allows choosing the most relevant clues for prediction. Its disadvantage is the need to store these clues in the mobile itself, which consumes memory, energy and bandwidth when communicating this information to the cell. In [13, 14], the authors propose a solution based on an ant system to predict the future cell a mobile will cross, based on past movements of the mobile itself and other mobiles that move in the same way. This solution provides good prediction results in environments where the mobile produce common behaviors such as cities but does not give good results when the displacements are random. [15] applies adaptive resonance theory (ART) to mobility prediction algorithms. Though it can adapt to trajectory of mobile user, it suffers from slow response to fast motion.

3. Movement Prediction and Data Mining

Users' movements are not completely random because they follow professional and social behavior. They are not completely deterministic because they obey the will of the individual. A study, carried out in the USA for better organizing public transport [16], has shown that nearly 80% of users displacements relate to work and nearly 20% relate to social or cultural reasons. During holidays, the percentage is only nearly 2%. It also showed that displacements are influenced by the infrastructures of the places (trades, highways, streets, paths, *etc.*). Displacements for work and social reasons are the most frequent and the most usual. The knowledge of the history (habits) of a user and his current location (on a road for example) could be useful to determine his probable future location. But it is also probable to determine the future location of a user even if his history is not known or if he takes an unusual displacement (holidays for example). In this case, one can use the history of his neighbors who follow the same direction. A user who is in a highway surely goes in the same direction as his neighbors.

Data mining is the process of extracting hidden knowledge or non-trivial from a large database [17]. It is a set of techniques drawn from various fields like data analysis, information theory and artificial intelligence applied to a data set to analyze and draw, either useful new information or hidden relationships between these data. The data mining applications vary widely. We cite for example the risk analysis and Marketing where the discipline is often used.

The association rules are a technique of data mining that naturally find its utility in movement prediction and location management. The association rules found hidden links between data. These links can be useful and exploited. The relationships found are like; if a customer buys bread, butter and coffee, it is likely that he also buys milk. This technique can be used in our case to find links between cells and have information of a kind; if a mobile has already crossed the cells x, y and z, It is likely that he will cross the cell t (cell covering a portion of highway for example). This information could be used for long-term prediction and making resources reservations at the appropriate time in the predicted cells.

4. Presentation of the Solution

To implement our solution, we assume an 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).

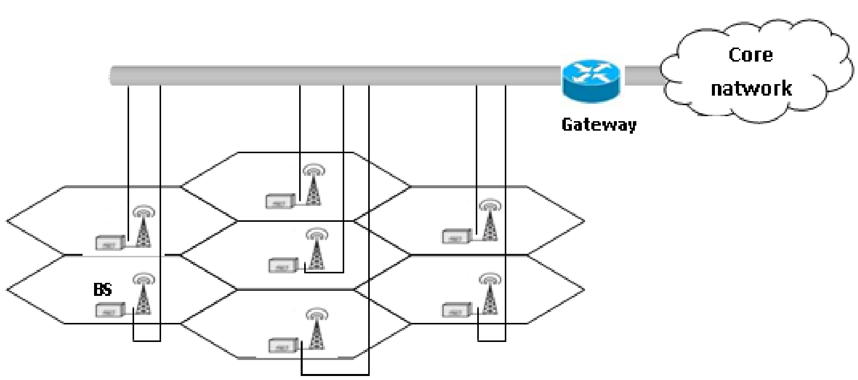

Figure 1. Architecture of a third generation mobile network

We assume that each 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).

Table 1. Structure of a history line

Mobile ID	Source cell	Destination cell	Date

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

4.1. The association rules

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.
- *... etc.*

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

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

We thus obtain a set of lists of K elements (called k-items set [18]) indicating the K last cells crossed by a given mobile. We apply then the algorithm Apriori [18] 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:

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

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

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. From the subset obtained we seek the association rules of order 3 and so on until arrived at the desired ^{2}. Subsets obtained, for example {Cell 1, Cell 4, Cell 3, Cell 8}, can constitute a location areas for a mobile and the order in which cells appear can give an indication of the longterm crossing cells which can be useful for long-term resource reservation.

5. Adjustment and Evaluation

Most data mining algorithms require training phase for adjusting their parameters. In the case of association rules it provides the best support and the best confidence. This step must be conducted using actual or 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 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 [19]. 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) as well as the ways which lead to these locations.

The simulator rests on the statistics of displacement led in the area of Waterloo [19] 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 Figure 2.

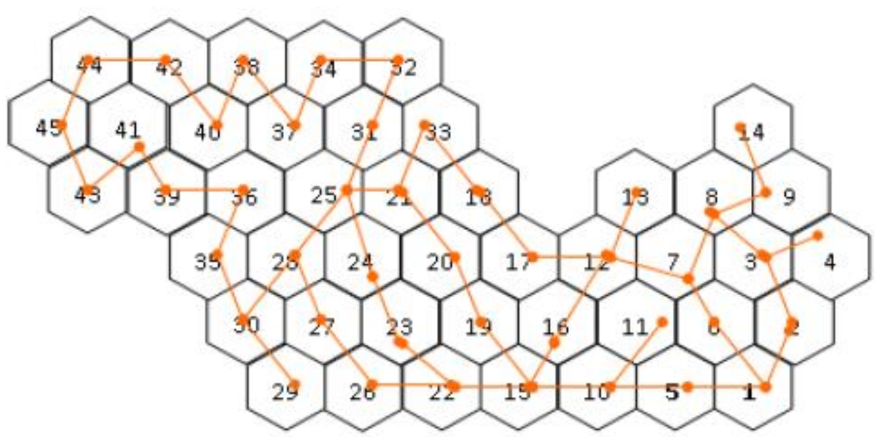

Figure 2. Cellular structure of the simulator

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 simulation.

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.

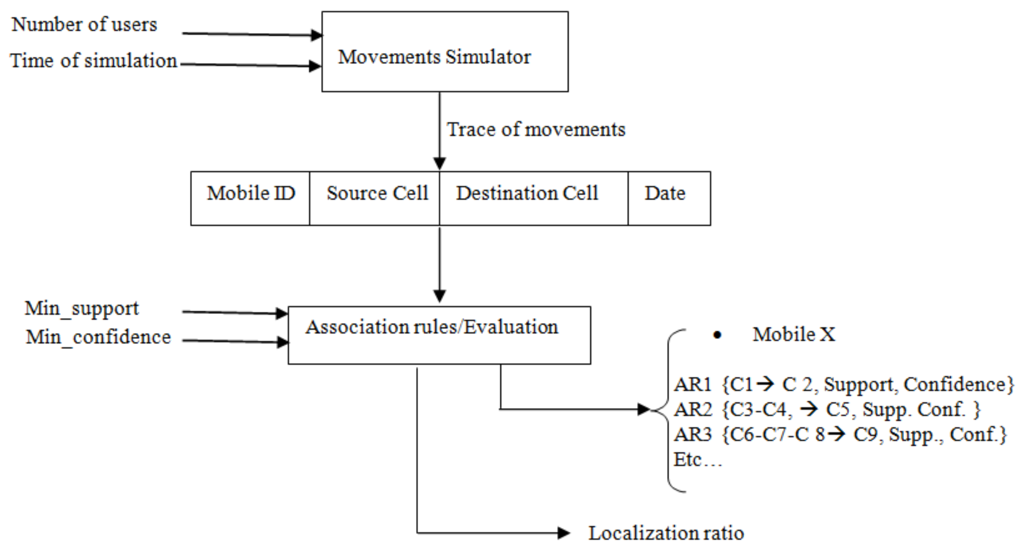

Figure 3. Simulation approach

8.1. Adjustment and evaluation of the association rules algorithm

The adjustment of the Apriori algorithm consists in determining the optimal order of association rules and the optimal confidence interval. The minimum support is fixed at 20% as in the majority of the applications. Its evaluation is carried out based on the ratio of correct localization. For an association rule cell 1, cell 2 —— > cell 3, and if the last location of a mobile is cell 2 and its precedent location is cell 1 then the mobile is probably in cell 3. The localization ratio is the ratio between the number of correct localizations thanks to the association rules and the full number of localization attempts.

We calculated the localization ratio for the association rules of order 1, 2, 3 and 4 by varying the confidence interval between 0,1 and 0,9 and we obtained the results presented in Figure 4.

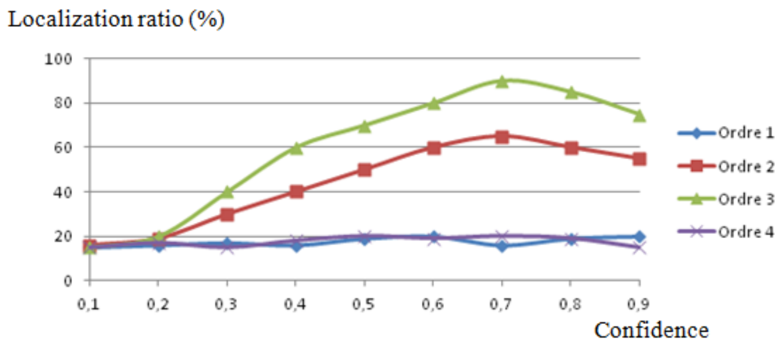

Figure 4. Localization ratio according to the confidence

The best ratio is obtained for the association rules of order 3 with a confidence of 0,7. What means that we can predict up to 90% the cell where the mobile user is knowing only the two last cells he crossed. The weak result of the rules of order 4 is probably due to the absence of this kind of rules satisfying the minimum support.

If we consider each association rule of order 3 as a location area formed of the three cells of the rule, we can create a new dynamic strategy of location management. We can also use the order of appearance of the cells in the rule to define a long-term resources reservation strategy.

6. Conclusion

Human displacements are often generated by socioprofessional needs (to go to work, to make purchases, *etc*.). They are related to the existing infrastructure (roads, public transport, work-places, ...*etc*.). It is thus possible to predict them by seeking the relations between these displacements and other available information like the users profiles and the site of the infrastructures.

Because of the complexity of the human displacements characteristics and in the absence of a reliable law of movements prediction, data mining can constitute a solution to the prediction. The technique presented in this article shows that it is possible to predict 90% of mobile displacements by knowing only their two last positions.

References

[1] W. Zhuang, K. C. Chua and S. M. Jiang, “Measurement-Based Dynamic Bandwidth Reservation Scheme for Handoff in Mobile Multimedia Networks”, International Conference on Universal Personal Communications, ICUPC '98, vol. 1, IEEE **(1998),** pp. 311 – 315.

[2] L. Hsu, R. Purnadi and S. S. P. Wang, “Maintaining Quality of Service (QoS) during Handoff in Cellular System with Movement Prediction Schemes”, Vehicular Technology Conference, VTC 1999 - Fall. IEEE VTS 50^{th}, vol. 4, **(1999)**, pp. 2153 - 2157.

[3] S. Choi and K. G. Shin, “Predictive and Adaptive Bandwith Reservation for Han-doffs in QoS-Sensitive Cellular Networks”, Proceedings of the ACM SIGCOMM '98 conference on Applications, technologies, architectures, and protocols for computer communication, vol. 28, no. 4, **(1998)**, pp. 155-166.

[4] X. Shen, J. W. Mark and J. Ye, “User Mobility Profile Prediction: An Adaptive Fuzzy interference Approach”, Wireless Networks, vol. 6, no. 5, (**2000)**, pp. 363-374.

[5] D. Ashrook and T. Staruer, “Learning Significant Locations and Predicting user Movement with GPS”, Proceedings of the 6^{th} International Symposium on Wearable Computers ISWC’02, (**2002),** pp. 101 – 108.

[6] M. Ni, Z. Zhong and D. Zhao, "MPBC: a Mobility Prediction-Based Clustering Scheme for Ad Hoc Networks", IEEE Transactions on Vehicular Technology, vol. 60, (**2011**), pp. 9.
[7] W. S. Soh and H. S. Kim, "QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques", IEEE Communication Magazine, (**2003)** January, pp. 86-92.
[8] S. C. Liou and H. C. Lu, "Applied Neural Network for Location Prediction and Resource Reservation Scheme in Wireless Network", International Conference on Communication Technology Proceedings, ICCT 2003, vol. 2, (**2003)** April 9-11, pp. 958 – 961.
[9] J. Capka and R. Boutaba, "Mobility Prediction in Wireless Networks using Neural Networks", In Proceedings of MMNS, (**2004)**, pp. 320-334.
[10] P. Fazio, F. De Rango and I. Selvaggi, "A Novel Passive Bandwidth Reservation Algorithm Based on Neural Networks Path Prediction in Wireless Environments", Proceedings of International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), (**2010**) July 11-14; Ottawa, Canada.
[11] N. Samaan and A. Karmouch, "A Mobility Perdiction Architecture based on Contextual Knowledge and Conceptual Maps", IEEE transactions on mobile computing, vol. 4, no. 6, (**2005)** November/December.
[12] J. M. Franois and G. Leduc, "Entropy-Based Knowledge Spreading and Application to Mobility Prediction", ACM CoNEXT'05, (**2005)** October 24-27; Toulouse, France.
[13] M. Daoui, A. M'zoughi, M. Lalam, M. Belkadi and R. Aoudjit, "Mobility prediction based on an ant system", Computer Communications, vol. 31, (**2008**), pp. 3090-3097.
[14] 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)**, pp. 221-232.
[15] T. Anagnostopoulos, C. Anagnostopoulos and S. Hadjiefthymiades, "An Online Adaptive Model for Location Prediction", Autonomic Computing and Communications Systems, vol. 23, (**2010**), pp. 1.
[16] P. S. Hu and J. Young, "1990 Nationwide Personal Transportation Survey (NPTS) Databook", Report No. FHWA-PL-94-010A, Prepared by Oak Ridge National Laboratory, Submitted to the Office of Highway Information Management, Federal Highway Administration, Washington, DC, (**1993)** November.
[17] J. Han, M. Kamber and J. Pei "Data mining concepts & techniques", The Morgan Kaufmann Series in Data Management Systems, Third edition, (**2011**).
[18] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P.p S. Yu, Z. -H. Zhou, M. Steinbach, D. J. Hand and D. Steinberg, "Top 10 algorithms in data mining, Knowledge and Information Systems", vol. 4, no. 1 (**2008**), pp. 1-37.
[19] 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.