

A Novel Method to Improve the Speed and the Accuracy of Location Prediction Algorithm of Mobile Users for Cellular Networks

Giang Minh Duc, Le Manh, Do Hong Tuan

Abstract: Currently, mobile networks and their applications have been developed quickly. Mobile users not only request various types of information, but also demand on Quality of Service (QoS). One of the measures to improve QoS is to apply mobile users' location prediction method. Mobile users' location prediction applications include automatic bandwidth adjustment, smart handover etc. To further improve QoS, we propose a new algorithm named UMP_Add_New algorithm which helps us avoid scanning of full database again. This algorithm mines the new dataset (new transactions are added to the database). In addition, to improve the accuracy of mobile users' location prediction, we propose a data classification method by time. The experimental results show that the UMP_Add_New algorithm has implementation time less than the UMPMining traditional algorithms did. Accuracy of the prediction by our method was also improved significantly.

Keywords: Location prediction, cellular networks, Mobility prediction, Data mining, Quality of Service.

I. INTRODUCTION

Currently, with rapid development of cellular communication networks, many people use their mobile devices to search for information on the internet. Almost everyone has a mobile device such as mobile phones, mobile tablets, notebook, etc. Many people also search for information as traveling all over the world. At about 6.8 billion mobile phones are used around the world in 2013 at the rate of 96, 97% of the world population [1]. Therefore, the big question for mobile operators is how to ensure the quality of service of mobile networks to meet customer's demands.

In cellular communication networks [2], a mobile user can move from one location to another one, which is a neighbor cell in the network. When mobile users move like that, the location of mobile users will be constantly updated to Visitor Location Register (VLR) [3-5] of the system. VLR is an intermediate database to store temporary information about mobile users in the service area of Mobile Switching Center (MSC). Mobile users' location information then is transferred to home location register (HLR). The HLR is a database which is a long-term storage of mobile users' information. The movement history of mobile users is extracted from the log files and stored in the HLR of the MSC. The historical data is used to predict the mobility of mobile users.

Some researchers applied the data mining techniques and other methods with the aim of solving the problem of cell communication networks, such as mobility, disconnect, long delay time, handover, bandwidth continuously changing... However, these methods have long execution time, offline running. Therefore, to further improve the quality of service of the mobile networks, we propose new contributions as follows:

- We propose the UMP_Add_New algorithm to avoid scanning of the full database again. This algorithm mines the new dataset (new transactions are added to the database). Therefore, mobile service providers (MSPs) can supply their applications more efficiently.
- We propose a method to improve the accuracy of mobile users' mobility prediction (data classification by time).

Our paper is organized as follows. In section 2, we present related works. In section 3, we present our proposed algorithm. In section 4, we propose a new

method to improve the accuracy of mobile users' mobility prediction. The results of the experiments are presented in section 5. The conclusion is given in section 6.

II. RELATED WORKS

The techniques which mine the movement patterns of mobile users is mentioned in the article [6-9],

The articles propose the movement patterns mining algorithms of mobile users. the movement rules from these patterns, and the next location prediction of mobile users.

With experiment results of [12], the prediction accuracy is 70%, while [10] is 20% and [11] is 52%.

The algorithm in [13] applied the Apriori algorithm in grid computing, but this paper does not take into account the network topology while creating the candidates.

In [14], Byungjin Jeong applied the UMPMining algorithm to perform the decision smart handover for reducing the number of unnecessary handover in architecture Macro/Femto-cell networks.

Vincent Etter et al. [20] have built user-specific models that learn from their mobility history.

Authors developed several mobility predictors, based on graphical models, neural networks, and decision trees. Their prediction reaches an average accuracy of more than 56% (the users' next destination in Nokia MDC dataset).

Ying Zhu et al. [21] proposed a mobility prediction method. This is the first one to extract the feature data like visit location, Bluetooth, address book, call log etc.

Authors proposed a method that applied fuzzy logic technology to predict next location of mobile user. Their experimental result shows that maximum accuracy is up to 72%.

However, to evaluate efficiently the accuracy of methods, we should use a similar dataset. In addition, authors in [20], [21] hadn't given a new solution when having more new transactions.

In [22], Anastasios Noulas et al. applied two supervised learning models, based on linear regression

and M5 model trees. Authors studied the problem of predicting the next venue which a mobile user will visit by exploring the predictive power offered by different facets of user's behavior. However, this method can't apply to cellular networks due to dataset in [22] was created in user check-in data generated in Foursquare (Foursquare is one of the most popular location-based services, with more than 20 million users as of April 2012.). This data is different from data of cellular networks. In cellular networks, mobile users can move to one cell among six cells neighbor, while in Foursquare, mobile user can check-in anywhere in the world.

To improve the quality of service (QoS) for mobile networks, in [15], authors presented two new algorithms (Find_UMP_1 and Find_UMP_2). The Find_UMP_1 algorithm reduces the runtime of the UMP Mining algorithm [12] (down 25.18% runtime) and the Find_UMP_2 algorithm reduces more runtime than the Find_UMP_1 algorithm (down 66.82% runtime). However, to improve further the quality of services, we propose a new algorithm named the UMP_Add_New algorithm. This algorithm has the runtime reduction (66.54%) in compare with the Find_UMP_2 algorithm (see experimental section 5).

Following is the algorithm that we propose:

III. THE UMP_ADD_NEW ALGORITHM

In this part, we develop the incremental algorithm [16-18] to find the frequent set from large and changable database. Firstly, we explore the formal concepts and the Galois lattice, which were given by Will (1982). R. Godin (1995) proposed his incremental algorithm to find the formal concepts and build "Hasse" chart for the concept lattice. This part surveys the relation between the concept set and the frequent set (large set), which apply Godin's concept creation algorithm.

The basic concepts are presented as follows:

Definition 1: Concept Context

Given O is a non-empty limited set of objects and I is a non-empty limited set of binary attributes. -Let R be a two-subject relation on O and I , $R \subseteq O \times I$.

A triple (O, I, R) is a concept context. Table 1 is an example of a concept context.

Table 1. A concept context

	i_1	i_2	i_3	i_4
o_0	1	1	0	0
o_1	1	1	1	0
o_2	1	1	1	0
o_3	1	0	1	1

Definition 2: Galois Connection

Given a data mining context (O, I, R) , consider two functions ρ and λ , defined as follows: $\rho: P(I) \rightarrow P(O)$ and $\lambda: P(O) \rightarrow P(I)$:

Given $S \subseteq I$, $\rho(S) = \{o \in O \mid \forall i \in S, (o, i) \in R\}$ (1)

Given $X \subseteq O$, $\lambda(X) = \{i \in I \mid \forall o \in X, (o, i) \in R\}$ (2)

Where $P(X)$ is a set of subsets of X . A pair of function (ρ, λ) is defined in such that way called Galois Connection.

$\rho(S)$ value denotes a set of transactions that have common all cells in S . $\lambda(X)$ value denotes a set of cells that have in all transactions of X .

Definition 3: Formal Concept

Given the concept context (O, I, R) and (ρ, λ) is the Galois Connections, a pair (X, S) with $X \in P(O)$ and $S \in P(I)$ such that $X = \rho(S)$ and $S = \lambda(X)$ are the formal concept of (O, I, R) , in which X is called “extent” and S is called “intent” of the concept (X, S) .

Denoted $B(O, I, R)$ is the formal concepts set, which is created from (O, I, R) .

Example 1: The formal concepts of the concept context in table 1 are: $C_1 = (\{o_0, o_1, o_2, o_3\}, \{i_1\})$; $C_2 = (\{o_0, o_1, o_2\}, \{i_1, i_2\})$; $C_3 = (\{o_1, o_2, o_3\}, \{i_1, i_3\})$; $C_4 = (\{o_1, o_2\}, \{i_1, i_2, i_3\})$; $C_5 = (\{o_3\}, \{i_1, i_3, i_4\})$; $C_6 = (\emptyset, \{i_1, i_2, i_3, i_4\})$

Property 1: The set X and the set S of the formal concept (X, S) is the closed set.

Indeed, $X = \rho(S)$ and $S = \lambda(X)$.

Where $h = \lambda \circ \rho$ and $h' = \rho \circ \lambda$.

Therefore, $X = \rho(S) = \rho(\lambda(X)) = \rho\lambda(X) = h'(X)$ (3)

and $S = \lambda(X) = \lambda(\rho(S)) = \lambda\rho(S) = h(S)$ (4)

Definition 4: Concept Lattice

Given context (O, I, R) , (ρ, λ) is the Galois connection. In $B(O, I, R)$, we define an order relation “ $\leq$ ” such that $C_1 = (X_1, S_1)$ and $C_2 = (X_2, S_2)$ then

$$C_1 \leq C_2 \Leftrightarrow X_1 \subseteq X_2 \quad (5)$$

The above definitions lead to the basic theorem of Wille to confirm that $(B(O, I, R), \leq)$ is a complete lattice.

Theorem 1:

Given a concept context (O, I, R) , call $B(O, I, R)$ is a set of all formal concepts of (O, I, R) and “ $\leq$ ” is an order relation of concepts, then $(B(O, I, R), \leq)$ is a complete lattice in which the operation “join” and the operation “meet” are defined as follows:

$$\bigvee_{j \in J} (X_j, S_j) = (h'(\bigcup_{j \in J} X_j), \bigcap_{j \in J} S_j) \quad (6)$$

$$\bigwedge_{j \in J} (X_j, S_j) = ((\bigcap_{j \in J} X_j), h(\bigcup_{j \in J} S_j)) \quad (7)$$

Where J is an index set of all formal concepts of (O, I, R) .

If $C_1 \leq C_2$, then concept C_2 is ancestor of C_1 and C_1 is descendant of C_2 . If C_2 is a direct superior of C_1 , then C_2 is father of C_1 and C_1 is child of C_2 . We can use “Hasse” chart to perform the concept lattice. If C_1 is child of C_2 , then C_1 will have an edge join to C_2 .

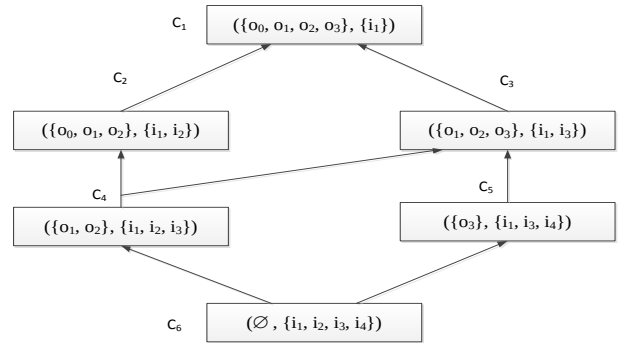


Figure 1. The “Hasse” chart of the concept lattice

Figure 1 is the “Hasse” chart of the concept lattice corresponding with the context in Table 1.

Definition 5: Frequent Concepts and Maximum Frequent Concept on “intent”.

Given the context (O, I, R) and the formal concepts set $B(O, I, R)$. Concept $C = (X, S)$ is a frequent concept on minimum frequent threshold $\text{minsupp} \in (0, 1]$ if and only if $|X|/|O| \geq \text{minsupp}$.

Denoted $FC(O, I, R, \text{minsupp})$ is the frequent concepts set on threshold “minsupp”.

Concept $C_m = (X_m, S_m) \in FC(O, I, R, \text{minsupp})$ is maximum frequent concept in “intent” if not exist concept $C_n = (X_n, S_n) \in FC(O, I, R, \text{minsupp})$ such that $C_m \neq C_n$ and $S_m \subset S_n$. If threshold $\text{minsupp} = 0.5$ then C_1, C_2, C_3, C_4 are the Frequent Concepts, in which C_4 is the Maximal Frequent Concept on “intent”.

Use algorithm to create the concept lattice to find the frequent set (large set)

This part presents the idea of Godin to build the incremental algorithm, which creates the concept lattice from the data mining context, in which there are clauses 1, 2, 3 as follows:

Clause 1:

If we insert a node $N = (\{o^*\}, \lambda(\{o^*\}))$ into lattice L , then all nodes (X, S) of lattice L , with $S \subseteq \lambda(\{o^*\})$ are updated to $(X \cup \{o^*\}, S)$.

Clause 2:

If $(X', S') = \inf \{(X, S) \in L \mid S' = S \cap \lambda(\{o^*\})\}$ and not exist $(E, S) \in L$, then they will create a new node (X', S') , that is calculated as above. The element (X, S) is called the birth element of new concept. The birth element is child of a new node.

Clause 3:

Father of node (X, S) is a new node or a repaired node (X', S') such that $(X', S') = \sup \{(X, S) \in L \mid S' \subseteq S\}$.

Based on the above results, R. Godin (1995) built the incremental algorithm to create the concept lattice from the formal context. To find the frequent set, this algorithm was improved in which the concept (X, S) of the Godin algorithm changed into $(|\rho(S)|, S)$, where $|\rho(S)|$ is the number of transactions containing the cell set of cellular networks.

The following is the UMP_Add_New algorithm applied for the mobility matrix of mobile users:

UMP_Add_New algorithm

Input: minsupp: minimum support threshold,

M_{dd}: mobility matrix,

G: mobile coverage graph

Output: New large set: L

1. If $i=1$ then //the first run
 2. $L = \emptyset$ //initially the large set is empty
 3. Find_ $L_1()$ //finding the large candidates have length-1
 4. for ($k=2; L_{k-1} \neq \emptyset; k++$) do
 5. Find_ $L_k()$ // create L_k from L_{k-1}
 6. $L = L \cup L_k$
 7. endifor
 8. Else // from the second run
 9. Find C_{inew} and L_{inew}
 10. for each ($c \in C_{\text{inew}}$)
 11. if $c \in C_i$ then // old candidate sets
 12. $\text{supptotal} = s.\text{supp} + c.\text{supp}$ // $s \in C_i$ and $s = c$
 13. $s.\text{supp} = \text{supptotal}$
 14. if $\text{supptotal} \geq \text{minsupp}$ then
 15. if $c \in L_i$ then // old large sets
 16. $l.\text{supp} = \text{supptotal}$ // $l \in L_i; l=c=s$
 17. else // $c \notin L_i$
 18. $L_i = L_i \cup c$
 19. Find_ $L_k()$
 20. endif
 21. else // $c \notin C_i$
 22. $C_i = C_i \cup c$
 23. if ($c.\text{supp} \geq \text{minsupp}$) then
 24. $L_i = L_i \cup c$
 25. Find_ $L_k()$
 26. endif
 27. endif
 28. endif
 29. Endfor
 30. Endif
 31. return L
-

We develop the incremental algorithm to find the large sets from the mobile database. The proposed algorithm is named UMP_Add_New. To avoid scanning of full database again, this algorithm

executes to mine the new dataset (new transactions are added to the database). The purpose of this algorithm is to reduce the execution time of mining the mobile user's movements. Therefore, the mobile service providers (MSPs) can supply their applications more efficiently.

While running the UMP_Add_New algorithm for the first time ($i = 1$), we find the candidate set C_i and the large pattern L_i are stored in an array. While running this algorithm for the second time or more ($i > 1$), we use the old result table and insert new results in it.

The Find_ L_1 () function (line 3) and the Find_ L_k () function (line 5, 19 and 25) are also the same as [15].

Find_ L_1 Function: find L_1

Input: minsupp: minimum support threshold;

M_{dd} : mobility matrix; G : mobile coverage graph

Output: L_1 : large set has length-1

1. $L_1 = \emptyset$
2. For each ($i \in I$ and $j \in \text{field of } M_{dd}$) //i: cell ID and it is also a column of M_{dd}
3. $S = \{s \mid s \in M_{dd} \text{ and } s_{ij} \neq 0\}$
4. For each $s \in S$
5. $s.\text{count} = s.\text{count} + 1$
6. Endfor
7. Endfor
8. $L = \{s \mid s \in C_1, s.\text{count} \geq \text{minsupp}\}$
9. $L_1 = L_1 \cup L$
10. Return L_1

Find_ L_k (L_{k-1}) Function: create L_k from L_{k-1}

Input: L_{k-1} : large set has length-($k-1$); graph G ;

M_{dd} : mobility matrix

Output: L_k : large set has length- k

1. $L_k = \emptyset$
2. For (each $X \in L_{k-1}$) do
3. For (each $Y \in L_{k-1}$ and $X \neq Y$) do
4. $S = X \cup Y$
5. $S = \{s_1, s_2, \dots, s_{k-1}, s_k\}$ // s_k : a set of cells has link with s_{k-1} of G
6. if ($|S| = k$ and $\text{SP}(S) \geq \text{minsupp}$) then
7. $L_k = L_k \cup \{S\}$
8. Endif
9. Endfor
10. Endfor
11. Return L_k

The old result table is the candidate sets C_i and the large patterns L_i . This algorithm uses the previous results and takes update to the mobility patterns as follows:

- Finding the candidate patterns (C_{inew}) from the new dataset.
- The support value is calculated for each sequence $c \in C_{\text{inew}}$, if the minsupp value is satisfied.
- Update the candidate patterns (if $c.\text{supp} \geq \text{minsupp}$) to the old candidate patterns (C_i, L_i).
- Returning the new large set: L .

IV. IMPROVING THE ACCURACY OF PREDICTION

To improve the accuracy of prediction, we perform the data classification by time of the input data as follows:

As shown in Figure 3, data from Home Location Register (HLR) are transferred to the time classification module. This data is classified into three classes as follows:

- Morning class: $0:00 \div < 12:00$
- Afternoon class: $12:00 \div < 18:00$
- Evening class: $18:00 \div < 24:00$

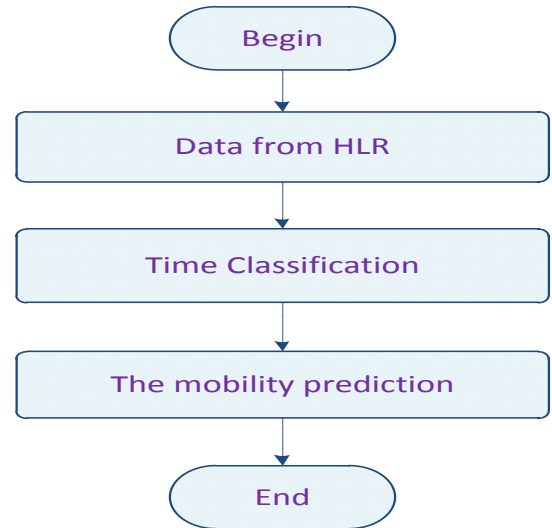


Figure 3. Input data classification flowchart

By the data classification as above, we show that the accuracy of the mobility prediction is increased significantly as in section V.2

V. EXPERIMENTAL RESULTS

V.1. The UMP_Add_New Algorithm

In this section, we consider the performance of the UMP_Add_New algorithms and compare it with the performance of the Find_UMP_2 algorithm [15] in terms of the execution time.

Our experimental environments are given in Table 2. Training dataset and Testing dataset used form [19].

Training dataset: the number of UAPs. Training datasets include three sets given in Table 3 ⁽¹⁾

Table 2. Experimental environments

Name	Parameter
Processor	Intel Core i3-2330M, 2.20GHz
RAM	4 GB
Operating System	32-bit
Programming language	Microsoft Visual Studio 2005
Database management system	SQL Server 2005

Table 3. Training datasets

Name	Number of transactions of mobile users
Dataset 1	56198
Dataset 2	68787

These datasets are the actual database of mobile users. The database is transformed from the User ID to the integer n ($n = 1, 2, 3 \dots$) and they cannot be decoded to protect customer information.

- The testing dataset is UAPs; it is used to evaluate the accuracy of the users' mobility prediction.

Testing dataset contains 7207 transactions of users.

The number of Base Transceiver System (BTS): 351.

- When not applying the algorithm UMP_Add_New:

Each update a new dataset, we perform as follows:

- Database (total) = database (old) + database (new)
- Running the Find_UMP_2 algorithm for database (total).

- When applying the UMP_Add_New algorithm:

We perform as follows:

- Get the old results (C_i, L_i).
- Running the UMP_Add_New algorithm for the new database and update the result with the old results $\rightarrow$ new results.

To compare the results of the two methods above, we have the actual results as follows:

- Database (old): dataset 1 (the number of records is 56 198).
- Database (new) dataset 2 (the number of records is 68 787).
- Database (total): dataset 1 + dataset 2 (the number of records is 124 985)
- The execution time is 214 seconds.

When running UMP_Add_New algorithm, the execution time is 90 seconds (down 57.94%), ss shown in Figure 4.

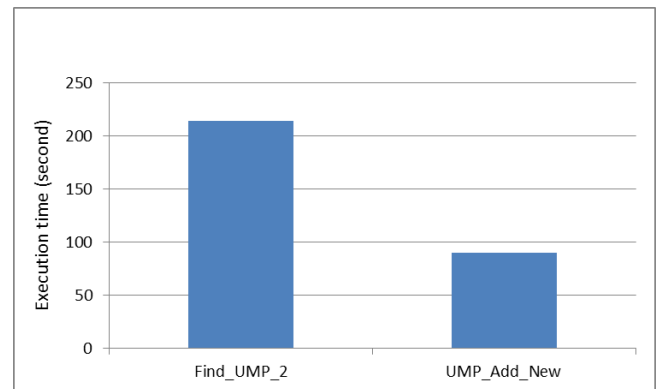


Figure 4. The execution time of two algorithms

⁽¹⁾ Appendix in References

V.2. Data classification by time

The common training dataset has 1179034 records. The testing dataset has 250403 records. They are divided as follows. After these datasets through the time classification module, they are classed as in Table 4:

Table 4. The time classification of the training and testing datasets

	Morning class	Afternoon class	Evening class
Training dataset	507030	443293	228711
Testing dataset	30464	107281	112658

After standard, the input data, we have the training dataset and the testing dataset as follows:

Table 5. The standard data

	Morning class	Afternoon class	Evening class
Training dataset	18662	14347	5581
Testing dataset	237	1846	2024

□ Changing of the recall values according to the minsupp values:

When changing the minsupp value, then the recall value changes as in Figures 5-7:

- Morning classified dataset:

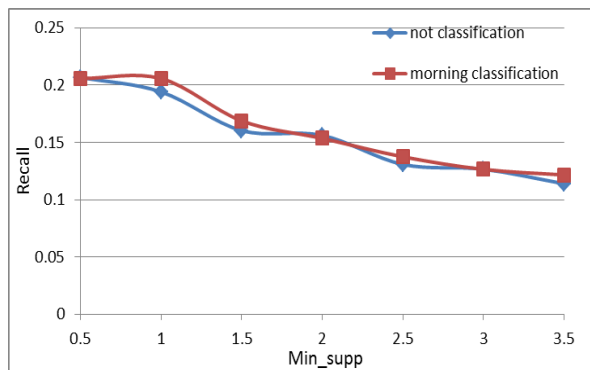


Figure 5. Compare the recall (morning class)

With the morning classification, the recall values are improved at the first minsupp values from 0.5 ÷ 1.3. (Rate increased from 5% to 6%).

- Afternoon classified dataset:

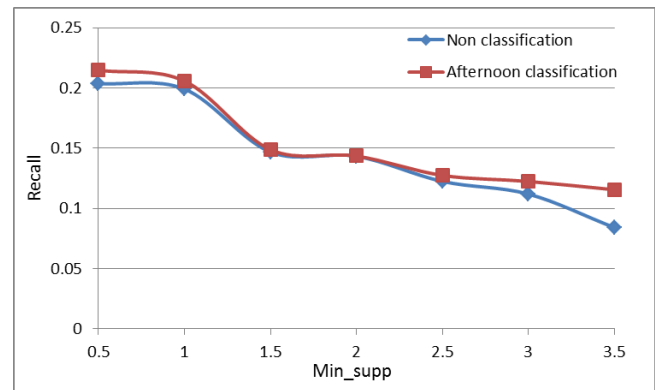


Figure 6. Compare the recall (afternoon class)

With the afternoon classification, the recall values are improved at the minsupp values from 0.5 ÷ 1.5 and from 2.5 ÷ 3.5 (rate increased from 1% to 37%).

- Evening classified dataset:

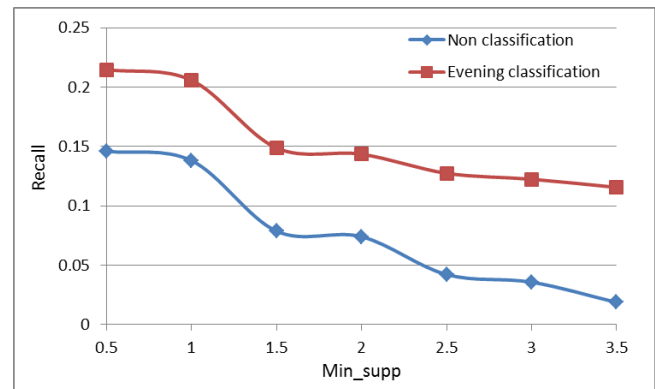


Figure 7. Compare the recall (evening class)

With the evening classification, the recall values are improved at the minsupp values from 0.5 ÷ 3.5. (Rate increased from 47% to 518%).

□ Changing of the precision values according to the minconf values:

When changing the minconf value, then the precision value changes as shown in Figures 8-10:

- Morning classified dataset:

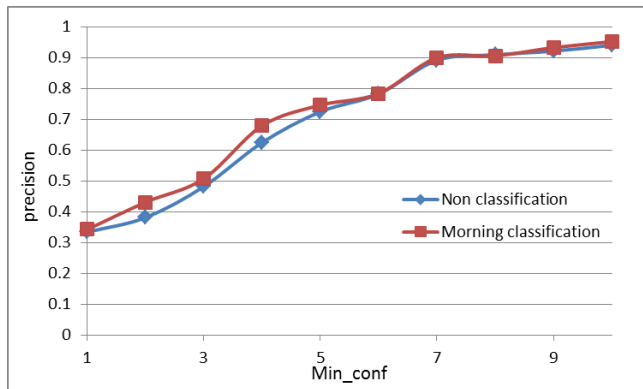


Figure 8. Compare the precision (morning class)

When the morning classification, the precision values are also improved at the first minconf values from 1 ÷ 6 (rate increased from 0.1% ÷ 13%).

- Afternoon classified dataset:

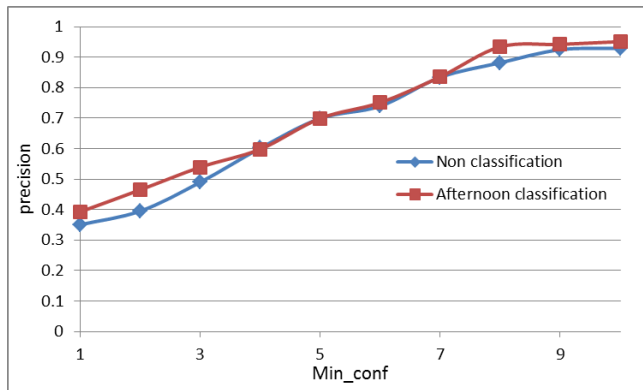


Figure 9. Compare the precision (afternoon class)

With the afternoon classification, accuracy rate increases the precision values from 0.15% ÷ 17%.

- Evening classified dataset:

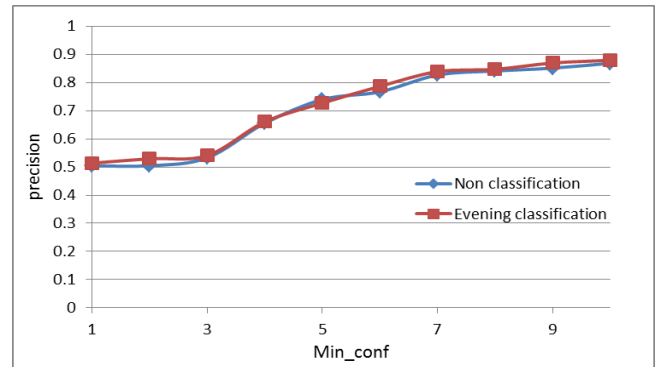


Figure 10. Compare the precision (evening class)

With the evening classification, accuracy rate increases the precision values from 0.6% ÷ 4.9%.

VI. CONCLUSION

The UMP_Add_New algorithm is used to increase the speed of the Find_UMP_2 algorithm [15] as adding new data.

Benefit of the application of the UMP_Add_New algorithm is that, the system can run “online” in real time to ensure QoS (to monitor instantaneous flow from which the mobile service providers can adjust the network bandwidth effectively implement).

In addition, we also propose a new method to improve the accuracy of the movement predicted of mobile users by the input data classification over time. The experimental results also show that the accuracy of prediction has increased significantly

REFERENCES

- [1] Wikipedia, GARTNER, "List of Countries by Number of Mobile Phones in Use," 2013.
- [2] ETSI/GSM, "Technical reports list," [http://webapp.etsi.org/key/key.asp? full list=y](http://webapp.etsi.org/key/key.asp?full_list=y) [online].
- [3] ETSI/GSM, "Home location register/visitor location register – report 11," 2010.
- [4] ALEX CABANES, *Home Location Register (HLR)*, I. S. & T. Group, Ed. IBM Blade Center, June 2007.
- [5] "HRL Look Up – Service Manual." <http://www.routomessaging.com>.
- [6] CRISTIAN AFLORI and MITICA CRAUS. “Grid implementation of Apriori algorithm. Advances in

- engineering software". *Volume 38, Issue 5*, 295-300, 2007.
- [7] GOKHAN YAVAS, DIMITRIOS KATSAROS. OZGUR ULSSOY and YANNIS MANOLOPOULOS. "A data mining approach for location prediction in mobile environments". *Data and Knowledge Engineering*, 54, 121-146, 2005.
- [8] MOHAMMAD WASEEM, R.R.SHELKE, Location Pattern Mining of Users in Mobile Environment, *International Journal of Electronics, Communication & Soft Computing Science and Engineering*, 2013, ISSN: 2277-9477, Volume 2, Issue 9
- [9] V. CHANDRA SHEKHAR RAO and P. SAMMULAL. Article: Survey on sequential pattern mining algorithms. *International Journal of Computer Applications*, 76(12):24–31, August 2013. Published by Foundation of Computer Science, New York, USA.
- [10] A. BHATTACHARYA, S. K. DAS, "Update: an information-theoretic approach to track mobile users in PCS networks," *ACM Wireless Networks* 8 (2-3), pp. 121-135, 2002.
- [11] S. RAJAGOPAL et al., "GPS-based predictive resource allocation in cellural networks," in *Proceedings of the IEEE International Conference on Networks (IEEE ICON020)*, 2002, pp. 229-234.
- [12] GOKHAN YAVAS et al, "A data mining approach for location prediction in mobile environments," *Data and Knowledge Engineering*, vol. 54, pp. 121-146, 2005.
- [13] CRISTIAN AFLORI and MITICA CRAUS, "Grid implementation of Apriori algorithm," *Advances in engineering software*, vol. 38, pp. 295-300, 2007.
- [14] BYUNGJIN JEONG, SEUNGJAE SHIN, INGOOK JANG, NAK WOON SUNG, and HYUNSOO YOON, "A Smart Handover Decision Algorithm Using Location Prediction for Hierarchical Macro/Femto-Cell Networks "in Vehicular Conference (VTC Fall), 2011 IEEE 74th, San Francisco, CA, Sept 2011, pp. 1-5
- [15] GIANG MINH DUC, LE MANH, DO HONG TUAN, "A Novel Location Prediction Algorithm of Mobile Users For Cellular Networks," *Journal on Information Communications Technology (Research and Development on Information Communications Technology)*, vol. No. 8(12), pp. 58-66, Aug. 2015.
- [16] SHIBY THOMAS et al., "An Efficient Algorithm for the Incremental Updation of Association Rules in Large Databases," in *From: KDD-97 Proceedings*, 1997
- [17] G. RAMALINGAM, THOMAS REPS, "An Incremental Algorithm for a Generalization of the Shortest-Path Problem," Technical Report # 1087, 1992.
- [18] JOHN D. KELLEHER et al., "Incremental generation of spatial referring expressions in situated dialog," in *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, Sydney, Jul. 2006, p. 1041–1048.
- [19] <http://msdn.microsoft.com/en-us/library/bb895173.aspx> (Training and Testing Data Sets – MSDN – Microsoft).
- [20] VINCENT ETTER et al, "Where to go from here? Mobility prediction from instantaneous information", School of Computer and Communication Sciences, EPFL, CH-1015 Lausanne, Switzerland, 2013.
- [21] YING ZHU YONG SUN YU WANG, Nokia Mobile Data Challenge: Predicting Semantic Place and Next Place via Mobile Data, Mobile Data Challenge 2012 (by Nokia)
- [22] <https://www.cl.cam.ac.uk/~cm542/papers/icdm2012.pdf>

APPENDIX: The dataset used for experiments

- ✓ The dataset used for our experiments was chosen at around Binh Duong province of Vietnam with information as follows:
 - The number of Base Transceiver System (BTS): 351.
 - The number of mobile users: 45462.
 - Training dataset 1: 1179034 records. After the data normalization (as follows), the User Actual Paths (UAPs) dataset is 56198 records.
 - Training dataset 2: 1467884 records. After the data normalization, the UAPs dataset is 68787 records.
 - Testing dataset: 250403 records. After the data normalization, the UAPs dataset is 7207 records.
- ✓ To mine data from the HLR (Home Location Register), we perform the data normalization through four steps as follows:
 - Step 1:* exchange data from a text file into a structured data file (database)
 - Data from the Home Location Register (HLR) with the following text:

0,452028500564855,84945859880,353191034572720,2011
1001,082142,43,0985477139,MTC,848,5924,,717,17522,,
,,0,4A40EB0B11,0,2011-10-01 08:21:42

0,452022020361130,84915749135,356919030975830,2011
1001,082123,62,01234348491,MOC,,,769,1666,,712,12442
,,,1,35414F04A7,0,2011-10-01 08:21:23

...

Step 2: Linking cell_ID

In the log file retrieved from the HLR, each BTS has a cell_ID, which links into a BTS management file of the province.

Step 3: Extracting some necessary fields of this dataset for data mining.

Step 4: Filter out records that have only one cell (mobile users do not move).

AUTHORS' BIOGRAPHIES

GIANG MINH DUC



He was born on Dec 6th, 1961 in Ho Chi Minh City.

He received the B.E degree (1994) in Electrical Engineering from Ho Chi Minh City University of Technology and B.S degree (1999) in Information Technology from The University of Science - Vietnam National University,

Ho Chi Minh City. He received the M.S degree (2006) in information technology (computer science) from University of Information Technology - Vietnam National University, Ho Chi Minh City. Currently he is Ph.D.student at Ho Chi Minh City University of Technology - Vietnam National University, Ho Chi Minh City.

His research interest includes telecommunications Engineering, Information Technology (knowledge base, data mining)

LE MANH



He is a lecturer in University of Information Technology - Vietnam National University, Ho Chi Minh City. Van Hien University

He received a B.S degree (1971) in Computer Engineering from Hanoi University of Science and Technology and Ph.D. in Mathematical Foundation

of Computers and Computing Systems from Soviet Union Academy of Science (1982)

After four steps for the data normalization, we get the following database:

STT	UAP_ID	UAP
16	841224	15, 128
17	841224	249, 113
18	841225	135, 29
19	841226	63, 56
20	841226	21, 223
21	841232	8, 29
22	841232	319, 309
23	841232	212, 234, 212
24	841232	124, 142, 124, 204, 221
25	841232	29, 30
26	841232	170, 178, 170
27	841232	12, 43, 12
28	841232	54, 63, 54, 63, 54, 63, 79, 63
29	841232	19, 1, 19, 1
30	841232	282, 263
31	841232	66, 110
32	841232	297, 301
33	841232	126, 130, 126, 130, 126, 130, 126, 130, 126
34	841232	13, 11
35	841232	313, 305, 343, 295, 313, 37, 305, 335, 305, 313, 307
36	841232	4, 13, 4
37	841232	312, 114
38	841232	306, 309, 306

His research interest includes management and development of IT applications in mobile environments and computer networks.

DO HONG TUAN



PhD, Senior Lecturer, Head of Department of Electrical and Electronics, Ho Chi Minh City University of Technology - Vietnam National University, Ho Chi Minh City.

His research interest includes Smart Antennas, Mobile and Wireless

Communications, Linear and Nonlinear Microwave Circuits, Digital Image Processing.