

# THUẬT TOÁN DỰ BÁO VỊ TRÍ TRONG MẠNG DI ĐỘNG

Lê Mạnh<sup>1</sup>, Giang Minh Đức<sup>2</sup>

<sup>1</sup> Trường Đại học Văn Hiến

<sup>2</sup> VNPT Bình Dương

<sup>1</sup>ManhL@vhu.edu.vn

Ngày nhận bài: 02/01/2017; Ngày duyệt đăng: 28/02/2017

## TÓM TẮT

Dự báo vị trí các thuê bao di động là một trong những vấn đề quan trọng trong các hệ thống tính toán di động. Sự di chuyển của người sử dụng điện thoại di động trong môi trường di động được lưu trữ trong thanh ghi “vị trí thuê bao di động (HLR)”. Các di chuyển thuê bao di động tạo ra các mẫu di động, dùng công nghệ khai thác dữ liệu theo dõi các mẫu đó. Các mẫu dữ liệu để phát hiện vị trí được sử dụng, để cung cấp các dịch vụ khác nhau cho người sử dụng điện thoại di động. Hiện nay, một số bài báo đã đề cập các phương pháp khai thác dữ liệu di động của người sử dụng điện thoại thông minh cho các hệ di động (GSM). Trong bài báo này, chúng tôi đề xuất một vài phương pháp sẽ làm giảm thời gian để tính toán mô hình di động, phục vụ công nghệ internet di động hiệu quả hơn.

**Từ khóa:** Dự báo vị trí, mạng GSM, dự đoán sự di chuyển thuê bao di động, khai thác dữ liệu.

## ABSTRACT

### Location prediction algorithms for mobile networks

Mobility prediction is one of the important issues in mobile computing systems. The moving logs of mobile users in mobile computing environment are stored in the Home Location Registry (HLR). The generated moving logs are used for mining mobility patterns. The discovered location patterns can be used to provide various location based services to the mobile users by applying server in mobile computing environment. Currently, there are some papers writing about data mining smartphone users for mobile systems (GSM). In this paper, methods are proposed to reduce the time to calculate the cellular model, serving mobile internet technology more effectively.

**Keywords:** Location prediction, GSM network, Mobility prediction, Data mining.

## 1. Tổng quan

Ngày nay, với sự phát triển nhanh chóng của mạng truyền thông tế bào, nhiều người sử dụng các thiết bị di động cá nhân để tìm kiếm thông tin trên mạng internet. Hầu như mọi người đều có thiết bị di động như mobile phone, máy tính bảng (tablet mobile), notebook,... Ngoài ra, nhiều người tìm kiếm thông tin khi đi du lịch khắp nơi trên thế giới. Vào khoảng 6,8 tỷ mobile phones được sử dụng trên toàn thế giới vào năm 2013 với tỷ lệ 99,97% dân số toàn thế giới [15]. Do đó, mục tiêu của vấn đề là làm thế nào để đảm bảo chất lượng của dịch vụ (Quality of Service (QoS)) di động.

Trong mạng truyền thông tế bào (Cellular

communications Networks), một “mobile user” có thể di chuyển từ vị trí này đến vị trí khác (lân cận) trong vùng phủ sóng. Khi mobile user di chuyển như thế, vị trí của mobile user sẽ được cập nhật liên tục vào thanh ghi định vị tạm trú (Visitor Location Register (VLR)) của hệ thống. VLR là một cơ sở dữ liệu trung gian để lưu trữ thông tin tạm thời của mobile users trong vùng phục vụ của Trung tâm chuyển mạch di động (Mobile Switching Center (MSC)). Thông tin vị trí của mobile users sau đó sẽ được chuyển tới thanh ghi định vị thường trú (Home Location Register (HLR)). HLR là một cơ sở dữ liệu lưu trữ lâu dài thông tin của mobile users. Lịch sử di chuyển của mobile users được lấy ra từ các file

log của HLR của MSC và dữ liệu lịch sử này được sử dụng để dự báo trước đường đi tương lai của mobile users.

Nhiều ứng dụng như chăm sóc sức khỏe, sinh học, quản lý khách sạn, và quân sự cũng áp dụng tiến trình xử lý về phụ thuộc vị trí (location – dependent). Với việc dự báo trước vị trí một cách hiệu quả, có thể trả lời các câu hỏi liên quan đến các vị trí trong tương lai của các users.

## 2. Định nghĩa vấn đề

Hiện nay mạng GSM đã rất thông dụng ở Việt Nam và trên thế giới. Vùng phủ sóng của mạng GSM được chia ra những vùng nhỏ hơn gọi là “cell”. Trong mỗi “cell” của mạng GSM có một trạm thu phát sóng gọi là BTS (Base Transceiver Station) có nhiệm vụ thu và phát sóng vô tuyến đến các Mobile Users [11], [12]. Các BTS được kết nối với nhau thông qua mạng truyền dẫn quang hoặc vô tuyến. Các Mobile Users sử dụng các kênh truyền dẫn vô tuyến để truyền thông với các BTS.

Sự di chuyển của một mobile user từ một “cell” hiện tại tới một “cell” khác sẽ được ghi nhận trong một cơ sở dữ liệu gọi là VLR. Sau đó, dữ liệu từ VLR sẽ được chuyển về HLR [13], [14] đặt tại tổng đài MSC. Từ dữ liệu này chúng ta có thể lấy ra lịch sử di chuyển của mobile user để có thể dự báo trước vị trí của họ. Ta gọi những mẫu dữ liệu lịch sử di chuyển này là UAPs (User Actual Paths). UAPs là một nguồn thông tin có giá trị bởi vì sự di chuyển của mobile user bao gồm các mẫu di chuyển ngẫu nhiên và các mẫu di chuyển thường xuyên. Trong bài báo [8] đã đưa ra thuật toán khai phá các mẫu di chuyển của mobile users nhằm dự báo vị trí di chuyển kế tiếp. Trong bài báo của chúng tôi đưa ra cũng dựa theo thuật toán này nhưng được cải tiến hiệu quả hơn về thời gian chạy. Cụ thể sẽ được trình bày ở phần sau.

## 3. Các công trình liên quan

Cho đến nay, đã có một số nghiên cứu về quản lý sự di chuyển của user. Hầu hết các nghiên cứu tập trung vào vấn đề cập nhật vị trí [1]. Vị trí cập nhật được thực hiện bất cứ

khi nào user di chuyển từ “cell” này tới “cell” khác trong mạng để lưu vết chính xác vị trí của user. Khi một cuộc gọi đến, mạng định hướng (route) cuộc gọi tới vị trí sau cùng của mobile user [2].

Vấn đề khai phá mẫu tuần tự đã được đề cập trong [3]. Thuật toán trong bài báo này không thể ứng dụng vào bài của chúng tôi để khai phá các mẫu di chuyển, vì các thuật toán này không xem xét “topology” của mạng trong khi khai phá các mẫu di chuyển. Ngoài ra, việc khai phá mẫu tuần tự cũng được ứng dụng vào lĩnh vực dự báo trước việc truy cập của users trên Web [4], [5]. Web prefetching được định nghĩa như là việc phân phối những yêu cầu tương lai của users dựa vào những yêu cầu trước đó.

Phương pháp *Ignorant Prediction* [6] không để ý đến thông tin có giá trị trong lịch sử di chuyển của user. Để dự báo trước user đi đến “cell” nào, phương pháp này ấn định một số cells lân cận. Phương pháp này thực hiện chọn ngẫu nhiên  $m$  cells lân cận với cell hiện tại.

Phương pháp *Mobility Prediction based on Transition Matrix (TM)* [7] dự báo vị trí theo khả năng có thể xảy ra chuyển tiếp “cell-to-cell” của một mobile user được tính toán bởi sự di chuyển trước đó và sau đó ghi nhận vào một ma trận. Dựa vào cơ sở này, việc cấp phát nguồn tài nguyên được thực hiện ở  $k$  cells có khả năng nhất trong các cell lân cận. Tham số  $k$  là tham số được định nghĩa trước bởi người sử dụng.

Kết quả [8] nghiên cứu dự báo trước vị trí của mobile users sử dụng kỹ thuật khai phá dữ liệu (Data Mining). Công bố này đưa ra thuật toán khai phá các mẫu di chuyển của mobile users, các luật di chuyển từ các mẫu này, và cuối cùng là dự báo trước vị trí di chuyển kế tiếp của mobile users bằng các luật trên.

Trong thực nghiệm [8], tính chính xác của dự báo là 70%, trong khi [6] chỉ 20% và [7] là 52%.

Thuật toán trong [9] áp dụng thuật toán Apriori vào trong tính toán lưới, tuy nhiên bài báo này không tính đến topology của mạng

trong lúc tạo các mẫu ứng viên. Tạo các mẫu ứng viên trong [9] không giống tạo các mẫu ứng viên trong môi trường di động; trong mạng GSM chỉ những vị trí lân cận của cell đang xét mới có thể được đề nghị là các mẫu ứng viên.

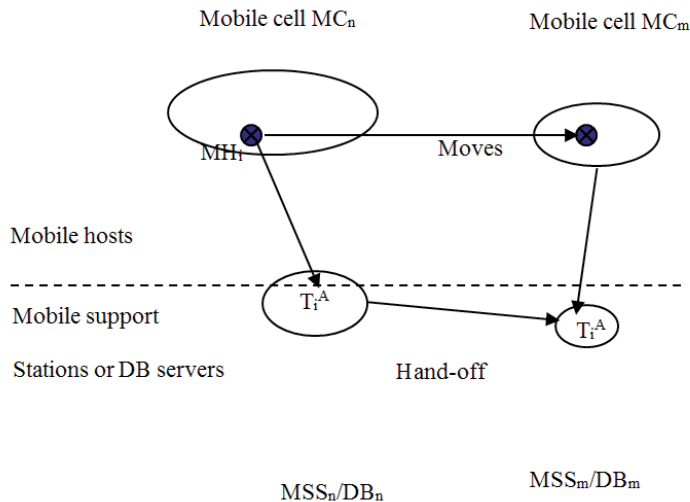
Thuật toán trong [10] giống như [8] nhưng file lưu trữ đường đi của mobile users được lưu trữ trong các nút lưới (grid node) đặt ở những vị trí khác nhau. Lưới dữ liệu (data grid) cung cấp một cơ sở dữ liệu phân tán địa lý cho lưới tính toán (computational Grid) và được thực thi bởi thuật toán gọi là KMPM (Knowledge Grid Based Mobility Pattern Mining). Bài báo này xây dựng hệ thống lưới dữ liệu (data grid system) trên một cụm các máy trạm (workstation) sử dụng nguồn mở Globus Toolkit 4.0 and MPICH-G2 (Message Passing Interface extended with Grid Services). Các thực nghiệm của tác giả được xây dựng trên các cấu hình khác nhau và thời gian tính toán được ghi nhận cho mỗi hoạt động. Khi số node tăng lên 8, thì thời gian thực hiện của thuật toán KMPM giảm đi 45%. Nói chung, nếu số node tăng lên, thì thời gian tính toán của thuật toán khai phá giảm đi.

Chúng tôi nghiên cứu dựa trên thuật toán [8],

cải tiến thuật toán này để phục vụ nâng cao chất lượng dịch vụ mạng di động.

#### 4. Phương pháp thực hiện

Vị trí của MH (mobile host) được định danh bởi ID (identity) của tế bào di động (MC-mobile cell). Trong tế bào di động mới, các MH có thể kết nối với trạm hỗ trợ di động (MSS – Mobile Support Station) mới. Agrawal và Srikant (1995) [3] có định nghĩa một giao tác gọi là “anchor transaction” được cài ở mạng cố định: ở MSS hoặc các máy chủ cơ sở dữ liệu. Các MSS này là các điểm “anchor” của các “anchor transaction”. Khi MH di chuyển vào trong các MC mới, một tiến trình “hand-over” (hoặc Hand-off) sẽ xảy ra, khi đó các “anchor transaction” sẽ được chuyển từ điểm “anchor” trước đó tới điểm “anchor” mới. Hình dưới đây thể hiện khi một MH di chuyển từ một MC<sub>n</sub> tới một tế bào di động mới MC<sub>m</sub>, tiến trình hand-over sẽ di chuyển một TiA từ một trạm hỗ trợ di động MSS<sub>n</sub> tới MSS<sub>m</sub>. Giao tác “anchor” TiA sẽ giữ vết (track) của MSS mà nó di chuyển qua, ví dụ MSS<sub>n</sub> và MSS<sub>m</sub>, do đó nó hỗ trợ “mobility” của các giao tác di động khi đi qua các tế bào di động khác [19] (Hình 1).



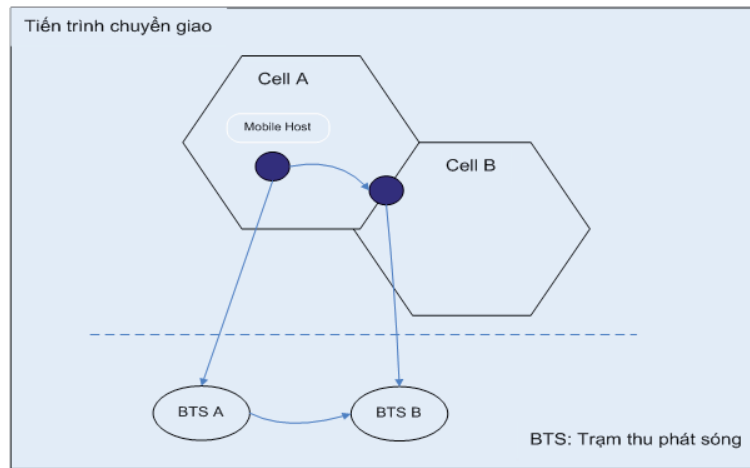
Hình 1: Nguyên lý chuyển giao kết nối internet di động

Khi các MH (Mobile Host) di chuyển từ tế bào này sang tế bào khác tiến trình hand-over thực hiện để chuyển MH sang kênh truyền mới.

**Vấn đề chuyển giao sự phục vụ thuê bao di động (Hand-over)**

- Khi người sử dụng di chuyển từ tế bào này sang tế bào khác là mạng thực hiện việc điều khiển chuyển mạch.

- Thời gian thực hiện chuyển giao nhiều nhất là 100ms.



**Hình 2: Sơ đồ mạng lưới các trạm BTS hệ di động**

***Hướng nghiên cứu để giải quyết vấn đề trong quản lý mạng internet di động***

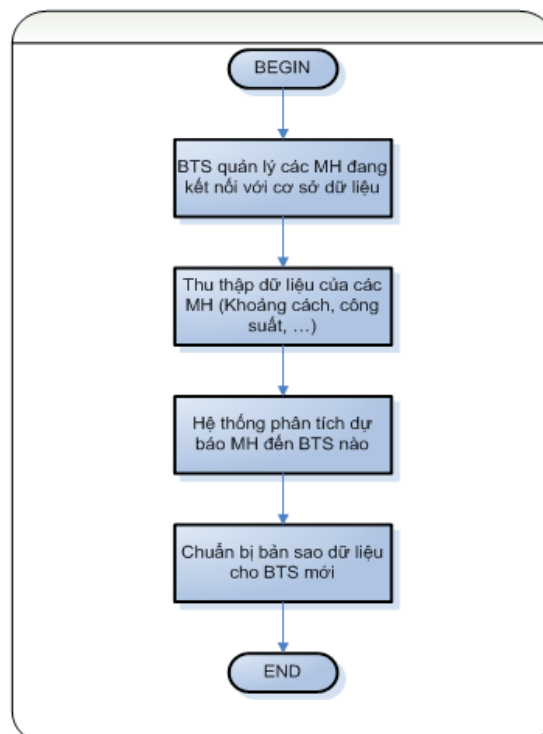
Khi Mobile Host (MH) đang di chuyển trong mạng tổ ong. Hệ thống sẽ tìm cách dự báo trước MH sẽ di chuyển đến tế bào nào trong số 6 tế bào xung quanh.

- Làm cách nào để dự báo trước MH sẽ di chuyển đến tế bào nào trong số 6 tế bào xung quanh.

- Biết được khoảng cách từ MH đến các tế bào lân cận tại từng thời điểm.

- Biết được mức thu tín hiệu của MH tại các tế bào lân cận (BTS lân cận).

- Chuẩn bị dữ liệu bản sao đến các BTS đó, không đợi đến tiến trình chuyển giao (Hand-over) hoặc bị mất kết nối (Disconnect) do MH đi vào các vùng tối (vùng không có sóng) giữa các tế bào.



**Hình 3: Lưu đồ để dự đoán vị trí trong môi trường di động**

Theo lưu đồ ở Hình 3, ta thấy rằng phương pháp thực hiện để dự báo trước vị trí của mobile users qua 4 giai đoạn sau:

### **Giai đoạn 1: Lấy dữ liệu từ logfile của HLR**

Sự di chuyển của mobile user từ cell hiện tại tới cell khác sẽ được ghi nhận trong cơ sở dữ liệu gọi là HLR. Chúng lưu trữ thường trực thông tin của khách hàng trong mạng di động. Lịch sử di chuyển của khách hàng có thể được lấy ra từ logfile của HLR.

### **Giai đoạn 2: Khai thác UAPs để được UMPs**

Giả sử, đặt tên đường đi thực tế của mobile user là UAPs (User Actual Paths). UAPs là một nguồn thông tin có giá trị vì nó chứa đựng các mẫu di chuyển thường xuyên và mẫu di chuyển ngẫu nhiên. Do đó, sử dụng UAPs, ta có thể rút ra mẫu di chuyển thường xuyên và sử dụng chúng để dự báo trước.

Giả sử rằng UAPs có dạng:  $C = \{c_1, c_2, \dots, c_n\}$ . Mỗi  $c_k$  biểu thị số ID của "cell" thứ  $k$  trong vùng phủ sóng.

**Bảng 1: Thuật toán Candidate Generation, UMP Mining chuẩn trong [8]**

#### **UMPMining()**

Vào: Tất cả các đường đi thực tế (UAPs) của users có trong  $D$ .

Giá trị của độ ủng hộ tối thiểu:  $Supp_{min}$

Đồ thị vùng phủ sóng:  $G$

Ra: Các mẫu di chuyển của users (UMPs):  $L$

1.  $C_1 \leftarrow$  các mẫu có chiều dài là 1.
2.  $k = 1$
3.  $L = \emptyset$  // khởi tạo tập các mẫu phổ biến là rỗng
4. while  $C_k \neq \emptyset$  {
5.   foreach UAP  $a \in D$  {
6.      $S = \{s \mid s \in C_k \text{ và } s \text{ là chuỗi con của } a\}$
7.     foreach  $s \in S$  {
8.        $s.count = s.count + s.supInc$
9.     }
10.   }
11.   // chọn các ứng viên thỏa  $Supp_{min}$
12.    $L_k = \{s \mid s \in C_k, s.count \geq supp_{min}\}$
13.    $L = L \cup L_k$
14.   // Tạo các mẫu ứng viên chiều dài  $(k+1)$ .
15.  $C_{k+1} \leftarrow$  CandidateGeneration( $L_k, G$ ),  $\forall c \in C_{k+1}, c.count=0$
16.    $k = k+1$
17. }
18. return  $L$

#### **CandidateGeneration()**

Input: Các mẫu phổ biến chiều dài  $k$ ,  $L_k$

Đồ thị vùng phủ sóng,  $G$

Output: Các mẫu ứng viên chiều dài  $(k + 1)$ , Candidates

1. Candidates =  $\emptyset$  // Khởi tạo tập candidates bằng rỗng
2. foreach  $L = (l_1, l_2, \dots, l_k)$ ,  $L \in L_k$  { // tìm các cell lân cận của mỗi  $l_k$  trong  $G$
3.    $N^+ = \{v \mid \text{có một cạnh trong } G \text{ như là } l_k \rightarrow v\}$
4.   foreach  $v \in N^+(l_k)$  { // cho mỗi cell lân cận này,
5.     //  $v$  tạo một ứng viên
6.      $C' = (l_1, l_2, \dots, l_k, v)$
7.     // Thêm  $C'$  tới tập các ứng viên (candidates set)
8.     Candidates  $\leftarrow$  Candidates  $\cup C'$
9.   }
10. }
11. return Candidates

**Giai đoạn 3: Tạo luật**

Từ UMPs, bây giờ có thể tạo ra tập các luật di chuyển như sau:

Giả sử, ta có một UMP  $C = \{c_1, c_2, \dots, c_k\}$ , với  $k > 1$ . Tất cả luật di chuyển có thể bắt nguồn từ một mẫu như sau:

$$\begin{aligned} \{c_1\} &\rightarrow \{c_2, \dots, c_k\} \\ \{c_1, c_2\} &\rightarrow \{c_3, \dots, c_k\} \\ &\dots \\ \{c_1, c_2, \dots, c_{k-1}\} &\rightarrow \{c_k\} \end{aligned}$$

Cho một luật di chuyển  $R: (c_1, c_2, \dots, c_{i-1}) \rightarrow (c_i, c_{i+1}, \dots, c_k)$ , hệ số confidence được tính như sau:

$$Confidence(R) = \frac{(c_1, c_2, \dots, c_k).count}{(c_1, c_2, \dots, c_{i-1}).count} \times 100$$

Bằng cách sử dụng UMPs, tất cả các luật di chuyển được tạo ra và các giá trị confidence cũng phải được tính toán. Các luật có độ confidence  $\geq conf_{min}$  sẽ được chọn.

**Giai đoạn 4: Dự báo trước vị trí của Mobile User**

Đây là giai đoạn dự báo trước mobile users sẽ di chuyển đến “cell” nào trong vùng phủ.

**Bảng 2: Tất cả các luật di chuyển**

Các luật di chuyển (Mobility rules)	
Rule	Confidence
$\{2\} \rightarrow \{0\}$	66,6
$\{4\} \rightarrow \{0\}$	50
$\{3\} \rightarrow \{4\}$	66,6
$\{5\} \rightarrow \{0\}$	50
$\{4\} \rightarrow \{5\}$	83,33
$\{3,4\} \rightarrow \{0\}$	75
$\{3\} \rightarrow \{4,0\}$	50
$\{3,4\} \rightarrow \{5\}$	75
$\{3\} \rightarrow \{4,5\}$	50

Trong Bảng 2, tất cả các luật sẽ được dùng để dự báo trước cho “user”. Giả sử “user” đi theo đường  $P = \{2, 3, 0, 4\}$  cho đến bây giờ và “user” hiện đang ở cell 4.

Theo luật trên thì có:  $\{4\} \rightarrow \{0\}$ ,  $\{4\} \rightarrow \{5\}$ ,

$\{3,4\} \rightarrow \{0\}$ , và  $\{3,4\} \rightarrow \{5\}$ . Vậy “user” có thể di chuyển đến “cell” 0 hoặc “cell” 5.

**5. Kết quả nghiên cứu****5.1. Những cải tiến các thuật toán trên:**

*Định nghĩa về ngữ cảnh khai phá dữ liệu:*

Cho tập  $O$  là tập hữu hạn khác rỗng các giao tác (UAP ID) và  $I$  là tập hữu hạn khác rỗng các cell,  $R$  là quan hệ 2 ngôi giữa  $O$  và  $I$  sao cho với  $o \in O$  và  $i \in I$ ,  $(o, i) \in R \Leftrightarrow$  giao tác  $o$  có chứa cell thứ  $i$ . Ngữ cảnh Khai phá dữ liệu (KPDL) là bộ ba  $(O, I, R)$ .

*Định nghĩa về tập phổ biến:*

Cho ngữ cảnh KPDL  $(O, I, R)$  và  $S \subset I$ , độ phổ biến của  $S$  được định nghĩa là tỉ số giữa số các giao tác có chứa  $S$  và số lượng cell trong  $O$ . Độ phổ biến của  $S$  ký hiệu là  $SP(S)$  và được tính như sau:

$$SP(S) = \frac{|\rho(S)|}{|O|}$$

Với  $|\cdot|$  là lực lượng của tập hợp.

Cho  $S \subset I$  và  $\text{min\_supp} \in (0, 1]$  là ngưỡng phổ biến tối thiểu,  $S$  là một tập phổ biến theo ngưỡng  $\text{min\_supp}$  nếu và chỉ nếu  $SP(S) \geq \text{min\_supp}$ .

Ký hiệu  $FS(O, I, R, \text{min\_supp})$  là tập hợp các tập phổ biến theo ngưỡng  $\text{min\_supp}$  hay  $FS(O, I, R, \text{min\_supp}) = \{ S \in P(I) \mid SP(S) \geq \text{min\_supp} \}$

*Mệnh đề:*

Cho  $T \notin FS(O, I, R, \text{min\_supp})$ , nếu  $T \subseteq S$  thì  $S \notin FS(O, I, R, \text{min\_supp})$ .

Với  $FS(O, I, R, \text{min\_supp})$  là tập hợp các tập phổ biến theo ngưỡng  $\text{min\_supp}$ .

Tức là: tập ứng viên  $T$  không phổ biến, và  $T \subseteq S$  thì  $S$  cũng không phổ biến.

Áp dụng vào thuật toán tạo các ứng viên **CandidateGeneration()**:

Tại dòng 3:  $N^+ = \{v \mid \text{có một cạnh trong } G \text{ như là } l_k \rightarrow v\}$

Thêm vào:  $v \in L$ , tức là: cạnh  $v$  trong đồ thị  $G$  phải thuộc về tập phổ biến  $L$ .

$N^+ = \{v \mid \text{có một cạnh trong } G \text{ như là } l_k \rightarrow v \text{ và } v \in L\}$

## 5.2. Độ chính xác dự báo

### 5.2.1. Độ chính xác của việc dự báo

#### 5.2.1.1. So sánh độ chính xác của ba thuật toán

Độ chính xác của các thuật toán UPMining, Find\_UMP\_Reduce\_Complexity và Find\_UMP\_Reduce\_Transaction là như nhau.

Ta có: UAPs  $\Rightarrow$  UMPs ( $L_n$ )  $\Rightarrow$  Mobility Rules. Độ chính xác của các thuật toán là như nhau khi và chỉ khi cho kết quả tập  $L_n$  bằng nhau.

Kết quả thực nghiệm như Bảng 3.

**Bảng 3: So sánh kết quả tìm  $L_n$  của các thuật toán theo chuẩn**

$C_n$	UPMining		Find_UMP_Reduce_Complexity		Find_UMP_Reduce_Transaction		$L_n$	
	Số lượng $C_n$	Thời gian chạy	Số lượng $C_n$	Thời gian chạy	Số lượng $C_n$	Thời gian chạy		
$C_1$	351	32	351	1	351	1	348	$L_1$

C2	1488	167	1488	129	1488	129	1033	L <sub>2</sub>
C3	3340	341	3340	274	3340	5	124	L <sub>3</sub>
C4	79	8	79	6	79	1	0	L <sub>4</sub>
<b>Tổng</b>	<b>5258</b>	<b>548</b>	<b>5258</b>	<b>410</b>	<b>5258</b>	<b>136</b>		

Theo kết quả trên, ta thấy:  $L_1 = 348$ ,  $L_2 = 1033$ ,  $L_3 = 124$ ,  $L_4 = 0$  (cho tất cả 3 thuật toán: UMPMining, Find\_UMP\_Reduce\_Complexity và Find\_UMP\_Reduce\_Transaction).

#### 5.2.1.2. Độ chính xác của thuật toán cải tiến tính toán online

Định nghĩa:

Recall: số các cell dự báo đúng/ tổng các yêu cầu.

Precision: số các cell dự báo đúng/ tổng các dự báo được thực hiện.

Sự thay đổi hệ số Recall khi thay đổi giá trị min\_supp

Tập huấn luyện thứ nhất có 56.198 records.

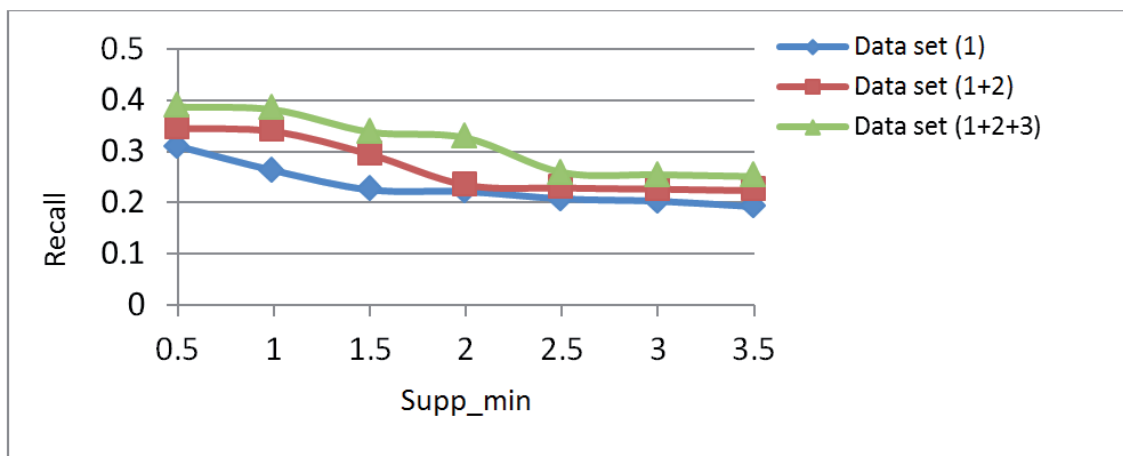
Tập huấn luyện thứ hai có  $56.198 + 68.787 = 124.985$  records

Tập huấn luyện thứ ba có  $56.198 + 68.787 + 34.895 = 159.880$  records.

Tập dữ liệu thử có 7207 records.

Hình 4 so sánh “recall” của các luật dự báo theo ba tập dữ liệu huấn luyện khác nhau. khi tăng min\_supp thì “recall” giảm theo. Lý do là khi tăng min\_supp kéo theo số các luật dự báo được khai phá giảm đi, Do đó, số các dự báo đúng sẽ giảm theo.

Khi kích thước tập huấn luyện tăng lên thì “recall” của các luật dự báo tăng theo.



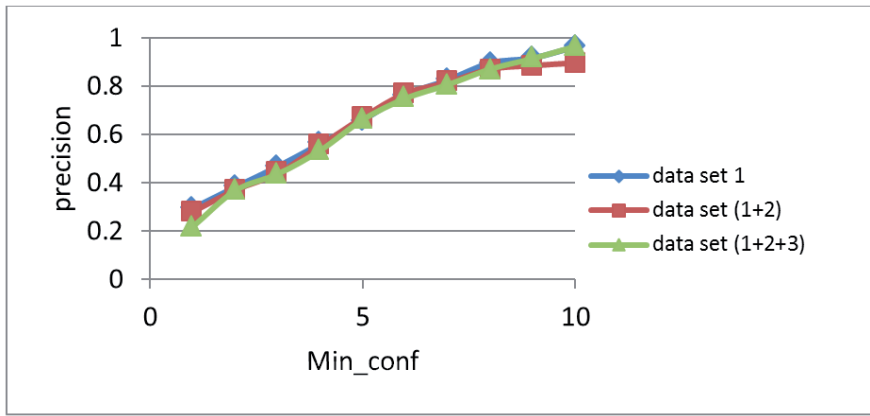
Hình 4: So sánh sự thay đổi recall theo min\_supp của ba tập dữ liệu

#### Độ chính xác (precision) của việc dự báo khi thay đổi giá trị min\_conf

Hình 5 so sánh độ chính xác của các luật dự báo theo ba tập dữ liệu huấn luyện khác nhau.

Chúng ta thấy rằng khi tăng số record của tập dữ liệu huấn luyện ( $56.198 + 68.787 + 34.895$  records) thì precision của chúng thay đổi không đáng kể.





**Hình 5: So sánh độ chính xác dự báo của ba tập dữ liệu khi thay đổi min\_conf**

**5.2.2. Nâng cao độ chính xác của dự báo**

Để nâng cao độ chính xác của dự báo, chúng tôi thực hiện phân lớp theo thời gian trong ngày của tập dữ liệu đầu vào.

Việc phân lớp chính xác tùy thuộc vào từng vùng, miền, từng quốc gia theo văn hóa xã hội và công việc của họ.

Dữ liệu từ thanh ghi HLR được đưa vào bộ phân lớp theo thời gian. Dữ liệu được chia làm ba lớp như sau:

- Lớp sáng: từ 0:00 ÷ < 12:00
- Lớp chiều: từ 12:00 ÷ < 18:00
- Lớp tối: từ 18:00 ÷ < 24:00

**Bảng 4: Tập dữ liệu phân lớp theo thời gian**

Dữ liệu không phân lớp		Dữ liệu phân lớp sáng		Dữ liệu phân lớp chiều		Dữ liệu phân lớp tối	
Training Data Set	Testing Data Set	Training Data Set	Testing Data Set	Training Data Set	Testing Data Set	Training Data Set	Testing Data Set
18662	237	18662	237	-	-	-	-
14347	1846	-	-	14347	1846	-	-
5581	2024	-	-	-	-	5581	2024

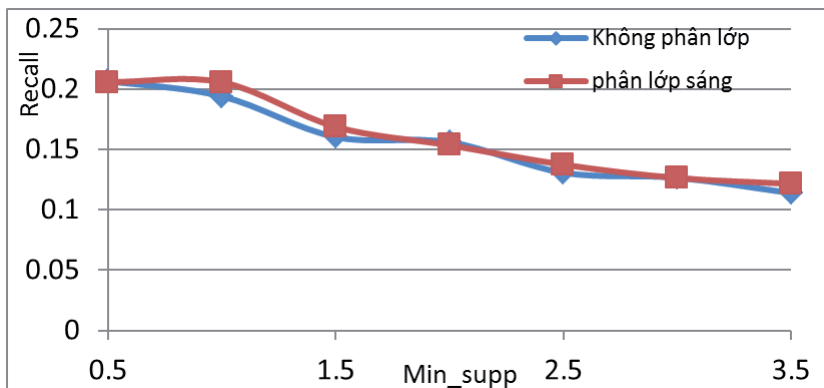
**Sự thay đổi của recall theo min\_supp**

Khi thay đổi giá trị min\_supp thì giá trị của “recall” thay đổi như sau:

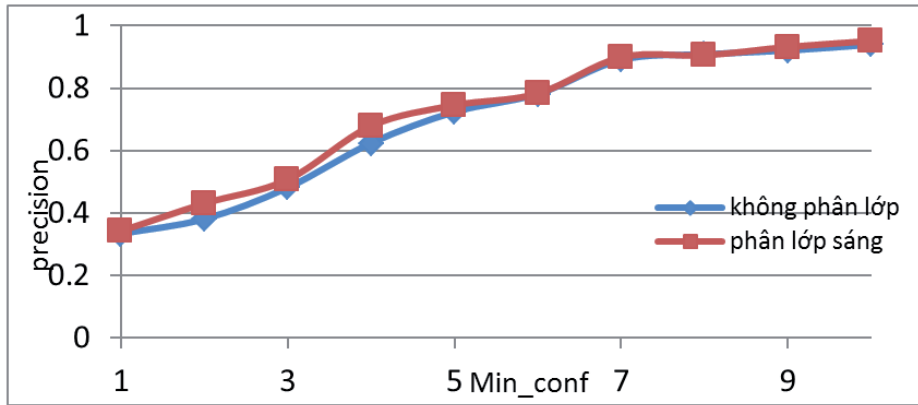
Đối với phân lớp sáng, hệ số recall có cải

thiện ở các giá trị đầu của min\_supp từ khoảng 0,5 - 1,3. Tỷ lệ nâng cao hệ số recall từ 5% - 6%.

Tương tự, đối với phân lớp chiều và tối cũng được cải thiện.



**Hình 6: So sánh hệ số recall sau khi phân lớp sáng**



Hình 7: So sánh hệ số precision sau khi phân lớp sáng

#### *Sự thay đổi của precision theo min\_conf*

Đối với dữ liệu được phân lớp sáng: Khi thay đổi giá trị của hệ số min\_conf thì giá trị của hệ số precision thay đổi như sau:

Đối với phân lớp sáng, hệ số precision cũng được cải thiện. Tỷ lệ tăng hệ số precision từ 0,1% - 13%.

Tương tự, đối với phân lớp chiều, tỷ lệ tăng hệ số precision từ 0,15% - 17% và phân lớp tối,

tỷ lệ tăng hệ số precision từ 0,6% - 4,9%.

#### 5.3. Kết quả thực nghiệm

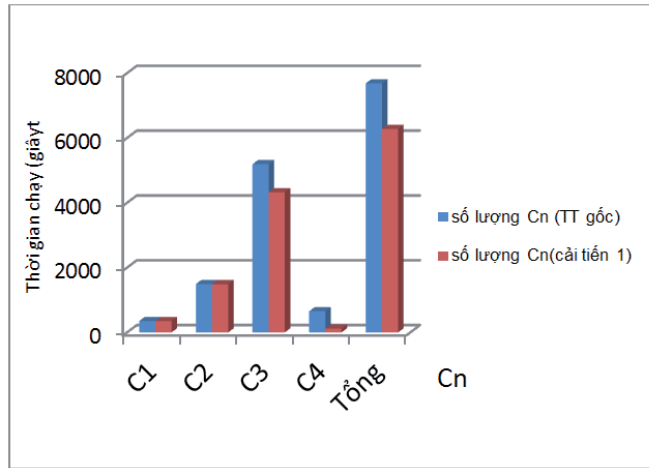
Nhóm tác giả thực hiện lấy số liệu của các trạm BTS tại Viễn thông Bình Dương, trích một phần nhỏ dữ liệu từ logfile của HLR với số trạm BTS là 351 trạm, số UAPs là 1.179.034 record. Sau khi chuẩn hóa dữ liệu số đường đi còn lại là 31.415 record.

$C_n$	Thuật toán UMPMining		Thuật toán cải tiến 1	
	Số lượng $C_n$	Thời gian chạy	Số lượng $C_n$	Thời gian chạy
$C_1$	351	33	351	32
$C_2$	1492	176	1488	167
$C_3$	5191	511	3340	341
$C_4$	651	63	79	8
<b>Tổng</b>	<b>7685</b>	<b>783</b>	<b>5258</b>	<b>548</b>

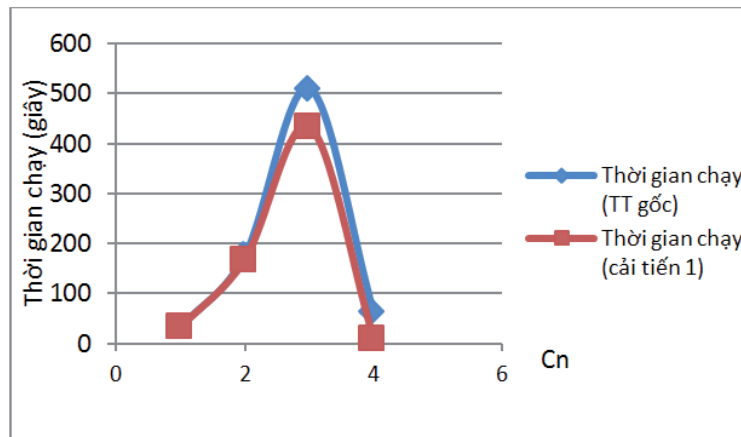
Bảng 5: So sánh kết quả thực hiện

Trong đó:

- $C_1$ : các mẫu ứng viên có chiều dài là 1.
- $L_1$ : các mẫu phổ biến có chiều dài là 1.
- .....
- $C_n$ : các mẫu ứng viên có chiều dài n.
- $L_n$ : các mẫu phổ biến có chiều dài n.



Hình 8: Biểu đồ so sánh Cn



Hình 9: Biểu đồ thời gian chạy của hai thuật toán

Tổng thời gian chạy các thuật toán với số trạm thử nghiệm:

- Theo thuật toán chuẩn [8] là 883s.
- Theo thuật toán cải tiến chỉ còn 548s.

## 6. Kết luận

Trong bài báo này mục đích của chúng tôi là giảm thời gian chạy của thuật toán [8], qua các thuật toán đã cải tiến và phát triển thêm. Trong phần kết quả thực nghiệm tại các trạm BTS của Viễn thông Bình Dương, chúng ta thấy rằng: với số cell là 351, số UAPs là 31415 thì thời gian chạy của thuật toán [8] là 783 giây và thời gian chạy thuật toán đã cải tiến còn 548 giây (giảm đi 30,01%). Lợi ích của việc áp dụng thuật toán tăng cường cũng như các thuật toán cải tiến (rút ngắn thời gian chạy) là hệ thống có thể chạy "online" theo thời gian thực nhằm theo dõi được

lưu lượng tức thời, từ đó ta có thể điều chỉnh băng thông một cách hợp lý. Các thuật toán đề xuất là phù hợp với lý thuyết đã công bố trước đó, sau khi thử nghiệm thuật toán chưa cải tiến và thuật toán đề xuất thấy hiệu quả tính toán nhanh hơn, điều này dẫn đến chất lượng của dịch vụ internet di động sẽ tốt hơn. Hướng phát triển trong thời gian tới, chúng tôi sẽ nghiên cứu các thuật toán mới nhằm rút ngắn thời gian thực hiện của thuật toán ở giai đoạn 2 (giai đoạn khai phá dữ liệu UAPs để được UMPs) vì đây là giai đoạn chiếm rất nhiều thời gian. Do đó sẽ nâng cao hơn chất lượng dịch vụ (QoS) của các mạng di động, trong tương lai để phát triển các công nghệ di động mới dựa trên nền tảng điện toán đám mây. Đây là xu hướng công nghệ di động trong tương lai, điều này sẽ cung cấp các dịch vụ tối ưu cho thuê bao di động.

**TÀI LIỆU THAM KHẢO**

- [1] Akyildiz F. I., Ho S. M., Lin Y. B., 1996. "Movement-based location update and selective paging for PCS networks", *IEEE/ACM Trans. Network.* (4) pp. 629-639.
- [2] Aljadhari A., Znaiti T., 2001. "Predictive mobility support for QoS provisioning in mobile wireless environments", *IEEE J. Select. Area Communication.* pp. 1915-1930.
- [3] Agrawal R., Srikant R., 1995. "Mining sequential patterns", *Proceedings of the IEEE Conference on Data Engineering (ICDE'95)*, pp. 3-14.
- [4] Nanopoulos A., Katsaros D., Manolopoulos Y., 2001. "Effective prediction of web user accesses: a data mining approach", *Proceedings of the WebKDD Workshop (WebKDD'01)*.
- [5] Nanopoulos A., Katsaros D., Manolopoulos Y., 2003. "An data mining algorithm for generalized web prefetching", *IEEE Trans. Knowl. Data Eng.* 15 (5) pp. 1155-1169.
- [6] Bhattacharya A., Das S. LeZi K., 2002. "An information-theoretic approach to track mobile users in PCS networks", *ACM Wireless Networks* 8 (2-3) pp. 121-135.
- [7] Rajagopal S., Srinivasan R.B., Narayan R. B., Petit X.B.C., 2002. "GPS-based predictive resource allocation in cellular networks", *Proceedings of the IEEE International Conference on Networks*, IEEE ICON020, pp. 229-234.
- [8] Yavas G., Katsaros D., Ozgur Ulssoy and Yannis Manolopoulos, 2005. "A data mining approach for location prediction in mobile environments", *Data and Knowledge Engineering*, 54, pp. 121-146.
- [9] Aflori C., Craus M., 2007. "Grid implementation of Apriori algorithm. Advances in engineering software". Volume 38, Issue 5, pp. 295-300.
- [10] Sakthi U., Bhuvaneshwaran R.S., 2009. "Mobility Prediction of Mobile Users in Mobile Environment Using Knowledge Grid", *Journal of Computer Science* Vol.9, No.1. pp. 303-309.
- [11] ETSI/GSM. Home location register/visitor location register – report 11.31-32.
- [12] ETSI/GSM. Technical reports list. <http://webapp.etsi.org/key/key.asp?fulllist=y>.
- [13] Alex Cabanes (IBM Systems & Technology Group) (2007) IBM BladeCenter - Home Location Register (HLR). June 2007.
- [14] HRL Look Up – Service Manual ([www.routomessaging.com](http://www.routomessaging.com)).
- [15] Gartner, 2013. *List of Countries by Number of Mobile Phones in Use*.
- [16] Manh Le, Duc Giang Minh, 2010. "Transactions in mobile communication" Sixth International Conference on Information Technology for Education and Research in HCM City, pp. 120-126.
- [17] Lê Mạnh, Nguyễn Việt Hùng, 2012. *Theo dõi hành vi mã độc trong môi trường mạng ảo cách ly*, NXB Khoa học và Kỹ thuật, Hà Nội, tr. 443-446.

- [18] Giang Minh Duc, Le Manh, Do Hong Tuan, 2015. “A Novel Location Prediction Algorithm of Mobile User For Cellular Networks”. Các công trình nghiên cứu và phát triển công nghệ thông tin và truyền thông, pp. 58-66.
- [19] Lê Mạnh, Giang Minh Đức, Đỗ Hồng Tuấn, 2015. “Các giao tác trong thông tin di động ứng dụng thuật toán dự báo vị trí trong mạng di động”, *Tạp chí khoa học Đại học Văn Hiến*, số 9, 11/2015, tr. 50-58.
- [20] Giang Minh Duc, Le Manh, Do Hong Tuan, 2015. “Mobility Patterns Mining Algorithms with Fast Speed”. EAI ISSN 2409-0026.
- [21] Giang Minh Duc, Le Manh, Do Hong Tuan, 2016. “A Method for Mobility Management in Cellunar Networks Using Data Mining”, *Pre-Proceeding 5<sup>th</sup> International Conference ICCASA 2016*. Context – Aware Systems and Applications, pp. 31-41.
- [22] Lê Mạnh, Nguyễn Thanh Bình, 2016. *Ứng dụng thuật toán Delaunay và RQT vào bài toán quy hoạch mạng di động*, NXB Khoa học và Kỹ thuật, Hà Nội, tr. 324-329.
- [24] Giang Minh Duc, Le Manh, Do Hong Tuan, 2016. “A Novel Method to Improve the Speed and the Accuracy of Location Prediction Algorithm of Mobile User For Cellular Networks”. Các công trình nghiên cứu và phát triển công nghệ thông tin và truyền thông. Tập 2 số 10(36), pp. 113-121.