Mining User Mobility Rules and Location Prediction Based on Pattern Matching Similarity: A Postprint

Authors: Zhang Haitao, Jiang Jifei, Zhou Huan

Date: 2018-08-13T00:00:00+00:00

Abstract

Location prediction for mobile users can effectively facilitate spatiotemporal allocation of system resources, thereby enhancing resource utilization in mobile communication systems. Traditional location prediction methods suffer from low prediction accuracy due to unreasonable calculation methods for pattern support. To address this issue, we propose a user mobility rule mining and location prediction method based on pattern matching degree, and apply it to mobile communication systems for user location prediction using base station coverage grids as units. The method consists of three specific steps: mining user mobility patterns through graph traversal, generating user mobility rules based on the mined patterns, and conducting location prediction according to these rules. Experimental analysis using 10 batches of trajectory data for user mobility rule mining demonstrates that the proposed method extracts fewer user mobility rules with high support and confidence, exhibiting the advantage of high precision.

Full Text

A Method for Mining User Mobile Rules Based on Pattern Matching Degree and Location Prediction

Zhang Haitao[1], Jiang Jifei[2], Zhou Huan[2]
[1]College of Geographic & Biologic Information, [2]College of Communication & Information Technology
Nanjing University of Posts & Telecommunications, Nanjing 210003, China

Abstract: Location prediction for mobile users can effectively facilitate the spatiotemporal allocation of system resources, thereby improving resource utilization in mobile communication systems. Traditional location prediction methods suffer from low prediction accuracy due to unreasonable support calculation for patterns. To address this issue, this paper proposes a method for mining user mobile rules and predicting locations based on pattern matching degree, applied to mobile user location prediction using base station coverage mesh as the unit in mobile communication systems. The method consists of three steps: mining user movement patterns through graph traversal, generating user mobile rules from these patterns, and performing location prediction based on the rules. Experiments conducted on ten batches of trajectory data demonstrate that this method yields fewer user mobile rules with higher support and confidence, achieving high prediction accuracy.

Keywords: mobile user; location prediction; pattern matching degree; user mobile rule

0 Introduction

With the development of computer and communication technologies, mobile communication systems have become increasingly pervasive. Mobile computing provides the means to access any data at any time and from any location, and is regarded as a hot research area in computer science. Mobile computing exhibits characteristics such as mobility and network diversity. Current mobile communication systems, primarily 3G and 4G, serve massive user populations and support diverse data service types (e.g., video, audio, images). To guarantee service quality, more effective management of mobile user information in these systems is required. Mobile location management—namely, the storage, updating, and prediction of mobile user location information—constitutes a primary component of mobile communication system management. Among these, location prediction for mobile users can effectively support the spatiotemporal allocation of system resources, improve resource utilization in mobile communication systems, reduce system latency and call drops during handovers between base station coverage meshes, and ultimately significantly enhance the quality of service in mobile networks.

Numerous solutions for location prediction have been proposed by scholars both domestically and internationally. References [15,16] presented a method combining Gaussian Mixture Models (GMM) and Linear Mobility Models (LMM), using GMM and LMM to establish user movement models both between and within base station coverage meshes, and performed location prediction based on matching between user movement trajectories and these models. In reference [17], Jeung Hoyoung and Qing Liu et al. utilized the Apriori algorithm to process user movement trajectories and extract user movement patterns for location prediction. In references [18,19], Morzy improved the Apriori algorithm to generate association rules and later extracted frequent patterns of user movement using an improved PrefixSpan algorithm, generating user rules for location prediction from these frequent patterns.

However, these methods share a common problem: the support calculation for patterns or rules used in location prediction relies on simple counting methods, resulting in low prediction accuracy and inability to effectively address the high-precision location prediction problem. High precision is a critical technical issue for ensuring efficient system resource allocation to mobile users. Therefore, this paper proposes a method for mining user mobile rules and location prediction based on pattern matching degree.

1 Basic Principles of the Location Prediction Method

We first provide definitions relevant to the proposed method.

Definition 1 Let $G = \langle V, E\rangle$ represent a directed, unweighted graph structure formed by all base station coverage areas in a mobile communication system. Let $P = \langle m_1, m_2, ..., m_n\rangle$ denote a user' s movement path within base station coverage areas, i.e., a movement sequence, where each m_i represents a mesh, and consecutive meshes are spatially adjacent. The sequence database D is a set of tuples $\langle sid, T\rangle$, where sid is the ID of the current movement sequence and T represents the movement sequence.

Definition 2 Let x and y be individual characters or spaces. $\delta(x, y)$ represents the scoring function after matching x and y, defined as:

$$\delta(x, y) = \begin{cases} 0, & x = y \\ 1, & \text{otherwise} \end{cases}$$

Let A be a path generated by a mobile user during movement, using base station coverage mesh as the unit (User Path based on Mesh, UPM), and let B be a user mobile pattern (UMP) contained in A. An alignment arrangement writes A and B as strings A' and B' that may contain spaces, where $|A'| = |B'|$. After removing spaces from A' and B', we obtain A and B. For such an alignment arrangement X', the alignment scoring function is defined as:

$$\delta(X') = \sum_{i=k}^{m} \delta(A'[i], B'[i])$$

where k and m are the indices of the first and last non-space characters in B', respectively.

The alignment score value of arrangement X' is essentially the number of mismatched characters between the first and last non-space characters in B'. Therefore, the optimal alignment arrangement X' should have the minimum alignment score, i.e., the minimum pattern matching degree, which we denote as $totDist$. The incremental support of user mobile path (UPM) A for user mobile pattern (UMP) B is defined as:

$$suppInc(A, B) = \begin{cases} 1, & \text{if } A \text{ contains } B \\ 0, & \text{otherwise} \end{cases}$$

Definition 3 Let D be the database storing all user movement paths (i.e., movement sequences), and let B be a user mobile pattern (UMP) contained in database D. The support of database D for B is defined as:

$$supp(B) = \sum_{\alpha \in D} suppInc(\alpha, B)$$

Definition 4 Let $M = \langle m_1, m_2, ..., m_k \rangle$ where $k > 1$ represents a user mobile pattern (UMP) of length k. A user mobile rule (UMR) is defined as:

$$R = \langle m_1, m_2, ..., m_i \rangle \rightarrow \langle m_{i+1}, ..., m_k \rangle \quad (1 \leq i < k)$$

where the left side of the arrow is the rule head and the right side is the rule tail. The confidence of this user mobile rule is defined as:

$$conf(R) = \frac{supp(\langle m_1, m_2, ..., m_k \rangle)}{supp(\langle m_1, m_2, ..., m_{k-1} \rangle)} \times 100\%$$

Let $minconf$ be the confidence threshold. User mobile rules with $conf(R) \geq minconf$ are retained for prediction.

Definition 5 Given a user movement trajectory $P = \langle m_1, m_2, ..., m_{i-1} \rangle$ where m_{i-1} is the mesh where the user is currently located, and a user mobile rule $R = \langle a_1, a_2, ..., a_j \rangle \rightarrow \langle a_{j+1}, ..., a_k \rangle$ from the user mobile rule set R. If the head of R is contained in P and the last element of the head coincides with the last element of P (i.e., $a_j = m_{i-1}$), then R is defined as a matching rule. The sum of this matching rule' s support and confidence is calculated. Based on this definition, we can filter a set of matching rules from the user mobile rule set and sort them in descending order according to the sum of support and confidence. Let R_1 be the first matching rule in the sorted set; then a_{j+1} is the predicted mesh for user movement trajectory P.

2 Prediction Method Based on User Mobile Rules

Location prediction comprises three stages: mining user mobile patterns through graph traversal, generating user mobile rules from these patterns, and performing location prediction based on the rules.

2.1 Mining User Mobile Patterns from Graph Traversal

This stage includes three steps:

a) Convert all base station coverage meshes in the studied mobile communication system into a directed, unweighted graph structure $G = \langle V, E \rangle$ for storage and representation.

b) Collect all user mobile path (UPM) data using base station coverage mesh as the unit. Verify that consecutive mesh elements in each UPM are spatially adjacent, and store them in sequence database D.

c) Set the support threshold *minsup* for mining user mobile patterns (UMP) and use Algorithm 1 to mine UMPs of progressively increasing length from sequence database D. The algorithm uses incremental support for counting candidate patterns, applies the support threshold to obtain UMPs from candidates, and uses spatial adjacency search to generate candidate patterns of length $k+1$ from length-k UMPs.

Algorithm 1: UMPMining()

```
Input: Database D storing all user mobile paths (UPM),
       graph structure G = <V, E> corresponding to base station coverage meshes,
       support threshold minsup for mining UMPs
Output: Set L of all user mobile patterns (UMP)

1. C1 = generate length-1 candidate patterns
2. k = 1
3. L =
4. while Ck    do
5.   foreach UPM   in D do
6.     foreach candidate c in Ck where c    do
7.       c.count = c.count +  .suppInc
8.   Lk = {c | c   Ck && c.count   minsup}
9.   L = L   Lk
10.  Ck+1 = CandidateGeneration(Lk, G)  // Generate length-(k+1) candidate patterns
11.  k = k + 1
12. return L
```

Lines 5-8 in Algorithm 1 use incremental support to count candidate patterns (corresponding to Definitions 2 and 3). Lines 9-10 compare candidate pattern counts with threshold *minsup* to obtain UMPs. The embedded Algorithm 2 generates length-$(k+1)$ candidate patterns from length-k UMPs through spatial adjacency search.

Algorithm 2: CandidateGeneration()

```
Input: Length-k user mobile pattern (UMP) Lk = <l1, l2, ..., lk>,
       graph structure G = <V, E>
Output: Length-(k+1) candidate pattern set candidates

1. candidates =
2. foreach pattern L = <l1, l2, ..., lk> in Lk do
3.   N(lk) = {v | v   V,  e   E, e = (lk, v)}
4.   foreach v in N(lk) do
5.     C' = <l1, l2, ..., lk, v>
6.     candidates = candidates   {C'}
7. return candidates
```

2.2 Generating User Mobile Rules from Patterns

This stage includes four steps:

a) Retrieve all user mobile pattern (UMP) sets L generated in Section 2.1.

b) Select all UMPs with length greater than 1 to generate a series of user mobile rules (UMR).

c) Scan the UPM data in sequence database D and calculate confidence statistics for the generated UMR series according to Definition 4.

d) Compare each UMR’ s confidence with threshold $minconf$ to obtain the final UMR set R.

2.3 Location Prediction Based on User Mobile Rules

This stage includes three steps:

a) Take the current user mobile path (UPM) P for location prediction.

b) Set the number m of base station coverage meshes that the mobile user may reach next.

c) Use Algorithm 3 to predict the user’ s next location.

Algorithm 3: MobilityPrediction()

```
Input: User's current movement trajectory P = <m1, m2, ..., mi-1>,
       user mobile rule (UMR) set R,
       number m of meshes to predict
Output: Predicted base station coverage mesh set PMeshs

1. PMeshs =
2. TupleArray =
3. foreach rule R = <a1, a2, ..., aj> → <aj+1, ..., ak> in R do
4.   if <a1, a2, ..., aj>   P && aj == mi-1 then
5.     TupleArray = TupleArray   {<aj+1, conf(R) + supp(R)>}
6. TupleArray.sort()  // Sort by sum of confidence and support in descending order
7. index = 0
8. while index < m && index < TupleArray.length do
9.   PMeshs = PMeshs   {TupleArray[index]}
10.  index = index + 1
11. return PMeshs
```

Lines 3-6 identify matching rules by checking if the rule head is contained in P and if the last element of the head matches the user’ s current mesh. Line 8 sorts matching rules by support and confidence. Lines 9-11 select the top m results and return the final predicted mesh set $PMeshs$.

2.4 Example Analysis

We illustrate the algorithm execution process with an example. Convert the nine base station coverage meshes shown in [Figure 1: see original paper] into the corresponding graph structure. Verify that all consecutive mesh elements in the user mobile path (UPM) data shown in are spatially adjacent, and store the results in database D.

Database storing user mobile paths
<7,6,5,0> <2,3,0,1> <1,2,3,4,5,7,0> <5,0,7,0,1> <0,1> <0,2> <0,3> <0,5> <0,7> <0,8> <1,2> <1,0> <1,8> <5,0> <5,7> <7,0> <2,1> <5,4> <2,0> <5,3> <0,1> <0,7> <5,0> <2,3> <2,3> <3,2> <5,7> <3,0> <5,6> <3,5> <7,6> <3,4> <7,5> <7,0> <7,8>

Set the support threshold $minsup = 2$ for mining user mobile patterns (UMP) and use Algorithm 1 to mine UMPs of progressively increasing length from sequence database D. The results are shown in through , where C denotes candidate pattern sets, L denotes user mobile pattern sets, and numbers indicate length.

Length-1 candidate and user mobile pattern sets
Length-2 candidate and user mobile pattern sets
Length-3 candidate and user mobile pattern sets
Length-4 candidate pattern set

The final UMP sets of lengths 1, 2, and 3 with their corresponding supports are shown in .

All user mobile pattern (UMP) sets
<0,1> <0,7> <2,3> <5,0> <5,7> <7,0> <5,0,7>

Set the confidence threshold $minconf = 50$ for mining user mobile rules (UMR). Select UMPs with length greater than 1 from to generate a series of UMRs. Scan the UPM data in and calculate confidence statistics for the generated UMRs according to Definition 4. All possible UMRs and their confidences are shown in . Applying the confidence threshold yields the final UMRs shown in .

All possible user mobile rules (UMR)
<5,0> → <0,7>

User mobile rules (UMR) meeting confidence threshold
<5,0> → <0,7>

For location prediction, take the current user mobile path (UPM) $P = \langle 3, 5, 7, 0\rangle$. Set the number of meshes to predict $m = 1$. Use Algorithm 3 to predict the user' s next location. According to , the prediction result is mesh 7.

User mobile rules (UMR) for prediction
Conf+SUPP
<5,0> → <0,7>

3 Experiments

3.1 Experimental Data

The initial data consists of request location data from telecommunications, collected every 60 seconds over a 24-hour period. A user's request location data represents the base station coverage meshes traversed during movement. Through spatial discretization, the user' s mesh sequence is converted into a spatially adjacent continuous mesh sequence. The basic information for the ten batches of trajectory data used in this paper is shown in .

Basic information of ten trajectory data batches

3.2 Performance Comparison Experiments

Our analysis of traditional prediction methods reveals that the essential difference between our proposed method and conventional approaches lies in the support calculation method. To ensure fair comparison, we designed a mining and prediction method using our framework but employing traditional support calculation (i.e., binary 0 or 1 metric). In subsequent experiments and analysis, we refer to the pattern-matching-degree-based support calculation method as the *similarity method*, and the binary 0/1 method as the *non-similarity method*.

The basic performance metrics for mining user mobile rules using both methods are shown in [Figure 2: see original paper] through [Figure 4: see original paper]. The similarity method extracts fewer rules with higher support and confidence. The rule count is lower because our method redefines how UPMs support UMPs, resulting in a smaller UMP set and consequently fewer rules. The higher average support stems from the smaller rule set. The higher average confidence occurs because the non-similarity method' s simple 0/1 metric generates numerous false strong rules.

The prediction performance metrics for both methods are compared in [Figure 5: see original paper] through [Figure 7: see original paper]. The similarity method achieves higher average precision, recall, and F-measure. Higher precision indicates better prediction accuracy on test data. Higher recall indicates more matching rules during prediction. The F-measure (with weight $\alpha = 1$) is the harmonic mean of precision and recall, reflecting overall prediction performance. The high F-measure demonstrates the superior usability of the similarity method.

4 Conclusion

Location prediction for mobile users can effectively support spatiotemporal resource allocation and has become an important component of resource management in mobile communication systems. Prediction accuracy is a crucial performance metric. However, traditional location prediction methods cannot effectively address this challenge. This paper proposes a method for mining user mobile rules and location prediction based on pattern matching degree. By defining a matching score between user mobile paths (UPM) and user mobile patterns (UMP), we obtain the optimal alignment and design a more reasonable method for measuring UPM support for UMPs. Experimental results show that our method extracts fewer rules with higher support and confidence, achieving superior prediction accuracy.

References

[1] Dinh H T, Lee C, Niyato D, et al. A survey of mobile cloud computing: architecture, applications, and approaches [J]. Wireless Communications & Mobile Computing, 2013, 13 (18): 1587-1611.

[2] Park Y I, Jang S M. Understanding privacy knowledge and skill in mobile communication [J]. Computers in Human Behavior, 2014, 38 (9): 296-303.

[3] Hu Haiyang, Liu Runhua, Hu Hua. Multi-objective optimization for task scheduling in mobile cloud computing [J]. Journal of Computer Research and Development, 2017, 54 (9): 1909-1919.

[4] Lang Weimin. Mobile Cloud Computing: Architectures, Algorithms and Applications [M]. Beijing: Posts & Telecom Press, 2016.

[5] Kamal R. Mobile computing [M]. Oxford: Oxford University Press, 2008.

[6] Meroufel B, Belalem G. Lightweight coordinated checkpointing in cloud computing [J]. Journal of High Speed Networks, 2014, 20 (3): 131-143.

[7] Zhang Wen, Wen Yin, Wu Du. Collaborative task execution in mobile cloud computing under a stochastic wireless channel [J]. IEEE Trans on Wireless Communications, 2015, 14 (1): 81-93.

[8] Huang Guosheng. Study of handoff and mobility management in mobile IP [D]. ChangSha: Central South University, 2010.

[9] Zickuhr K. Location-based services [J]. Pew Research, 2013: 1-25.

[10] Al-Molegi A, Jabreel M, Ghaleb B. STF-RNN: space time features-based recurrent neural network for predicting people next location [C]// Computational Intelligence. 2016: 1-7.

[11] Wu Fan, Fu Kun, Wang Yang, et al. A Spatial-temporal-semantic neural network algorithm for location prediction on moving objects [J]. Algorithms, 2017, 10 (2): 37.

[12] Zheng Y, Zhang L, Ma Z, et al. Recommending friends and locations based on individual location history [J]. ACM Trans on the Web, 2011, 5 (1): 5.

[13] Meetei K P, George A. Handoff management in wireless networks using predictive modelling [C]// Proc of National Conference Communications. 2011: 1-5.

[14] Feng Zhenni, Zhu Yanmin. A survey on trajectory data mining: techniques and applications [J]. IEEE Access, 2017, 4: 2056-2067.

[15] Liu T, Bahl P, Chlamtac I. Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks [C]. IEEE J. Select. Area Commun, 1998, 16: 922–936.

[16] Xue Dan. Research on mining the user mobility patterns and location prediction [D]. Shanghai: Donghua University, 2005.

[17] Jeung H, Liu Q, Shen H T. A hybrid prediction model for moving objects [C]// Proc of the 24th IEEE International Conference on Data Engineering. 2008: 70-79.

[18] Morzy M. Prediction of moving object location based on frequent trajectories [M]. Berlin: Springer Berlin Heidelberg, 2006: 583-592.

[19] Morzy M. Mining frequent trajectories of moving objects for location prediction [M]. Berlin: Springer, 2007: 667-680.

Note: Figure translations are in progress. See original paper for figures.