Mobility Prediction Algorithms Using User Traces in Wireless Networks

Chuyen Luong[†], Son Do[††], Hyukro Park[†††], Deokjai Choi[††††]

ABSTRACT

*Mobility prediction is one of hot topics using location history information. It is useful for not only user-level applications such as people finder and recommendation sharing service but also for system-level applications such as hand-off management, resource allocation, and quality of service of wireless services. Most of current prediction techniques often use a set of significant locations without taking into account possible location information changes for prediction. Markov-based, LZ-based and Prediction by Pattern Matching techniques consider interesting locations to enhance the prediction accuracy, but they do not consider interesting location changes. In our paper, we propose an algorithm which integrates the changing or emerging new location information. This approach is based on Active LeZi algorithm, but both of new location and all possible location contexts will be updated in the tree with the fixed depth. Furthermore, the tree will also be updated even when there is no new location detected but the expected route is changed. We find that our algorithm is adaptive to predict next location. We evaluate our proposed system on a part of Dartmouth dataset consisting of 1026 users. An accuracy rate of more than 84% is achieved.

Key words: Active LeZi Algorithm, Hidden Markov Model Algorithm, Mobility Prediction.

1. INTRODUCTION

Nowadays, smart phone devices have become indispensable in human life and work. Based on embedded sensors, there are many noticeable location-based applications such as Foursquare, Gowalla, Google Places, and Shopkick to locate their friends' current location, share interesting places and photos, and make online reservations [1]. Moreover, smooth and stable connections should be achieved for streaming multimedia applications over mobile wireless network. Location prediction does not only provide a next place for location-based application but also help resource arrangement in advance to reduce the hand off latency in real time applications [2,3,4,12]. For those reasons, many researches have been carried out for location prediction.

Many techniques are proposed to predict next location based on history sequence such as Markov-based, LZ-based. In previous researches, order-2 Markov with fallback, Hidden Markov Model (HMM) and Active LeZi algorithms achieve high accuracies with an assumption that users would move regularly [2,3,8,11,12]. Instead of estimating parameters, Active LeZi algorithm builds

※ Corresponding Author : Deokjai Choi, Address: (500-757) School of Electronics and Computer Engineering, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Korea, TEL : +82-62-530-3429, FAX : +82-62-530-3439, E-mail : dchoi@jnu.ac.kr
Receipt date : May. 20, 2014, Revision date : Jun. 27, 2014
Approval date : Jul. 2, 2014
[†] School of Electronics and Computer Engineering Chonnam National University
  (E-mail : luongchuyen0789@gmail.com)

[††] School of Electronics and Computer Engineering
  Chonnam National University
  (E-mail : dvson167@gmail.com)
[†††] School of Electronics and Computer Engineering
  Chonnam National University
  (E-mail : hyukro@chonnam.ac.kr)
[††††] School of Electronics and Computer Engineering
  Chonnam National University
※ This study was financially supported by Chonnam National University, 2013.

a tree considering interesting locations as nodes and possible location contexts as branches. The transition between the previous k-history locations and a specific location is considered by using a window with variable length. Moreover, this algorithm uses the power of order-k Markov model with fallback to calculate the location probabilities for determining the next location.

In reality, a new behavior is usually established and repeated in the next time after something is done and new things appears. These changes incapacitate previous approaches since they are based on the assumption of user's regular movement [9,11]. Even though Active LeZi allows irregularity by adding new nodes, its depth of tree is deep since it considers long history information [6]. Therefore, if a huge history locations are used to create location contexts within the window in the changed behavior time, much unnecessary location information in the previous time are used for predicting the next location. Consequently, the accuracy is reduced because of the independence of the old behaviors and changed behavior. In this paper, we propose a prediction algorithm based on Active LeZi, namely ActiveLZ Update, using the fixed length window. In this approach, we divide the data sequence into subsequences collected in equal time intervals. The first subsequence is trained by Active LeZi algorithm to find the length of window k, and then the length k is fixed for the window in extracting location contexts. Based on this proposal, we can avoid using redundant location information of the previous period to extract location contexts for predicting the next location.

The rest of this paper is organized as follows. We start with a discussion of related work in Section 2. Details of proposed algorithm for inserting the new locations, creating new location patterns and predicting the next location in both regular and irregular cases are presented in section 3. The experimental results and accuracy evaluations are described in section 4. Finally, section 5 presents a conclusion.

2. RELATED WORK

Three common techniques including Markov-based, LZ-based and Prediction by Pattern Matching techniques (PPM) have been proposed and applied to Dartmouth dataset for evaluating accuracies based on the occurrence of the next location given a location history sequence.

In Markov-based technique, order-2 Markov model and HMM are proposed to predict next location. Order-2 Markov in [2,3,11] considers either two previous locations or one previous location in the unpredictable cases to predict next location. It achieves the prediction accuracy of 72% effectively for using traces with more than 1000 moves. On the other hand, HMM algorithm [12] uses the APs and the location for estimating transition and observation matrices to construct a prediction model. They perform with the traces of five users and achieve a prediction accuracy of 85%. In summary, Markov-based technique only works for regular cases because the calculation of parameter matrices requires the list of fixed locations. When the new location which has never been seen in history sequence appears, it is not considered as a member in location list. Consequently, we can't predict next location if the current location is an unseen location.

Instead of estimating parameter matrices, the LZ-based technique constructs a tree. A tree considers contexts extracting from phrases as branches. The frequency of contexts is the weights of nodes in tree. LZ-based technique uses these weights to calculate the node's probability to assign to the next location. They achieve the accuracy of 69% for LZ78 algorithm in [2,3,8]. To improve the accuracy, Active LeZi algorithm [6,7,10] based on LZ78 and LeZi Update algorithm uses a window with variable length to create contexts for reducing the loss of location information. The probability of an

order-k node is calculated by using a mixture of the lower-order weights and null weight. Unlike Markov-based technique, LZ-based considers both regular and changed cases. If a new location is appeared, it only needs to add a new node to the tree. Active LeZi algorithm still suffers from changed cases in which the tree depth depends on the length of history location sequence. It is difficult to define the k-order for deciding a prediction. Hence, the prediction accuracy will be reduced because of appearing the changed case.

Finally, PPM discovers frequent location sequences over their most significant location sequence for prediction rather than the transitions between locations. Sampled Pattern Matching (SPM) algorithm is proposed to predict next location in [2,3]. Instead of using a fixed value of context length k in order-k Markov model, SPM uses a fixed fraction to find the longest suffix that has occurred previously in the sequence. Moreover, NextPlace [9] is also proposed for user location prediction based on non-linear time series. The corresponding location sequences are selected based on the time series sequence. They achieve accuracy from 70% to 90% depending on the time interval between two consecutive predictions. In brief, the PPM technique examines the frequency location subsequences on a short time scale. It is difficult to extend predictions further because the frequency is very small.

In summary, among three techniques, Active LeZi algorithm is the most suitable one to predict the next location for both regular and irregular cases. To improve the accuracy, therefore, we propose an algorithm based on Active LeZi algorithm to overcome the drawback of Active LeZi algorithm to reduce incorrect cases.

3. PROPOSED METHOD

3.1 Problem of Active LeZi algorithm approach

Active LeZi algorithm is based on LZ78 and LeZi

Update algorithms. Pseudo-code for parsing and processing the input sequence in the Active LeZi algorithm is shown in [5, 6]. To use Active Lezi algorithm for location prediction, the researchers map a location to a symbol. History location sequences are represented string of symbols. To examine the information between the observed location and previous consecutive locations, Active LeZi uses a buffer, known as a window, to take out a location subsequence with the ending of the observed location in a history sequence. Moreover, they also use a set of possible consecutive locations, namely context, to consider the user's behavior. This context is formed from all possible location subsequence suffixes within the window. In the tree, nodes are locations in the contexts and branches are contexts. The weight of nodes are the occurrence frequency of contexts. The tree depth is the length of window chosen to be equal to of the longest location pattern. To predict the next location, the probability of each location is computed and the highest location probability is predicted as the next location.

In this algorithm the length of window depends on the length of the history sequence. If the length of history location sequence is so long, the length of window is also long. In the changed human behavior case, the unnecessary location information is used to predict the next location. Consequently, the prediction accuracy is reduced.

3.2 ActiveLZ Update algorithm

To overcome the drawback of ActiveLezi, we propose a method for updating new location or changed route, namely ActiveLZ Update. The updating process is described in algorithm 1. Order-2 Markov model and HMM discern the location information before estimating the parameters. Therefore, the size of these parameters are fixed. Moreover, new locations and changed routes are not considered. In contrast, Active LeZi and ActiveLZ Update use the context which is a neces-

sary set to update the new information. This is illustrated in line 1. Instead of using the window with variable length in Active LeZi, ActiveLZ Update uses the window with the fixed length in line 3 and line 4 to reduce the redundant information of previous locations. By the experimental researches, we recognize that when the user's life is stable, the window length is long enough to accommodate the context information sufficiently. Therefore, the window length is suitably fixed and can be set to the stable context length. Line 5 shows the way to create the contexts which are all possible suffixes within the window. Consequently, location information is effectively considered by a growing-order Markov model.

Similar to three algorithms, Active LZ Update algorithm updates the weight of location information via using the context value in the line 8. Line 10 and line 11 show the way to add new contexts and the weight of them.

Algorithm 1 ActiveLeZi Update algorithm

1: input: input sequence, context, context value
2: **for** (all observed location in the input sequence) **do**
3:      extract the k-consecutive locations with the ending of observed location to
4:      window.
5:      get all possible contexts which are all suffixes of window
6:      **for** (all possible contexts) **do**
7:          **if** (in context) **then**
8:              increase one to context value of window context.
9:          **else**
10:             add window context to context
11:             add one to this context value
12:         **endif**
13:     **endfor**
14: **endfor**
15: output: context, context value

---

4. EXPERIMENT AND EVALUATION

4.1 Dataset

A part of Dartmouth dataset [13] is used to demonstrate the compatibility of our algorithm, the ActiveLZ Update. In fact, this subdataset was collected by the movement history of 1026 users, each having approximately 1016 records and 10 Aps.

Each record has a time-stamp, an associated AP and an off-time information. To simplify the data before building the tree, we dropped the off-time field. Each AP's name is blinded in the form of SocBldg4AP7 which indicates that 4th AP in the 7th building of the Social department. The building types include Adm (Admin), Ath (Athletic), Lib (Library), Oth (Other-mainly sysadmin test APs), Res (Residential) and Soc (Social). The UNIX timestamps are modified to 'YYYY-MM-DD HH:MM:SS' format.

4.2 Experiments and Evaluation

For experiment, the dataset is equally divided into two subsets. Subset 1 is used for building the tree. Subset 2 is used for evaluating the prediction accuracy of the proposed method and other methods.

There are many kinds of window length. However, in this experiment, three window lengths are able to establish a decision tree. The first window length is the tree depth which is the maximum branch length. The second window length is the mean length of all branches and the third window length is the branch length of which the occurrence frequency is maximum. Fig. 1 shows three window's lengths of each user. In the third case, the length is two times longer than the rests. Because the length of window in the training is chosen to be equal to the length of the longest phrase, this
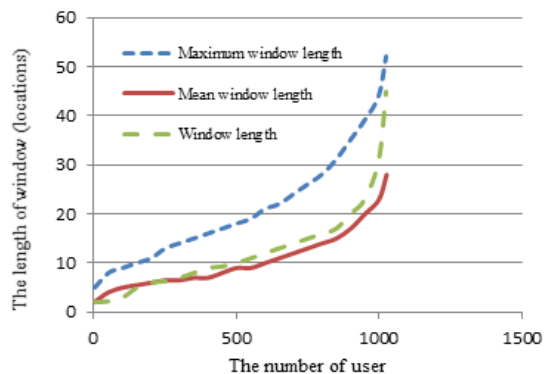

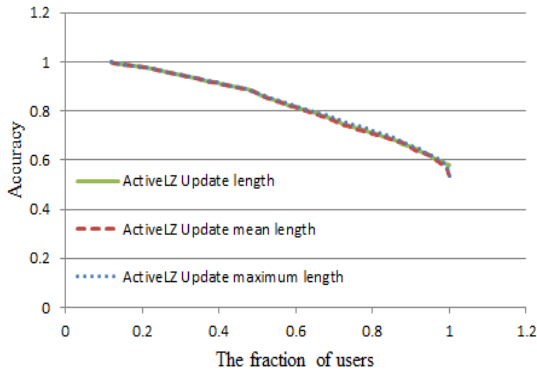Fig. 1. The three window lengths for each user.

Fig. 2. ActiveLz Update predictor with using different window length.

length depends on the length of sample length in the subset 1.

Fig. 2 shows the complementary cumulative distribution function (CCDF) for accuracy with three different window lengths. The experiments achieve average results of accuracy of over 84%. Although the window length in the third case is much longer than two other cases, its accuracy gain is insignificant. Actually, when the user's route is gradually stable and the window length is long enough to sufficiently accommodate the context information for identifying the next location, the prediction accuracy does not depend on the window length.

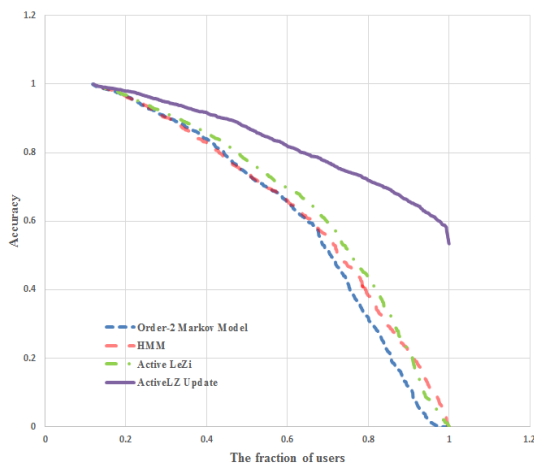Fig. 3 presents a comparison of CCDF of the ac-

Fig. 3. Accuracy of four location predictors, compared.

curacy of our proposed algorithm to other algorithms. The unpredictable case will appear when the users go to new locations or changed route because of the increase of tree depth and the number of nodes rapidly. Consequently, we achieve the accuracy of more 72% for three algorithms, Order-2 Markov, HMM, Active LeZi algorithms. ActiveLZ Update algorithm significantly improves the accuracy for predicting the next location. A fixed length suitably is chosen to create those context and update their weighs. Because of the fixed length, the length of each tree's branch length is equal to tree depth to void unpredictable cases. With using this algorithm, we not only achieve the good algorithm but also the processing time and memory are reduced because we don't need to examine the dictionary like Active LeZi algorithm.

5. CONCLUSION

We have proposed an effective algorithm, namely ActiveLZ Update, using a fixed window length to avoid incorrect cases for predicting the next location. We don't need to consider patterns in dictionary like Active LeZi algorithm to define the window length. The fixed length depends on the behavior of users. Therefore, we use Active LeZi algorithm to initialize a tree given a history location sequence. Then we examine a suitable order-k to assign to the window length for creating contexts from phrases. We achieve the accuracy of more than 84% and better result than state of the art algorithms do. In this case, changed circumstances are independent from the change of locations of student in next semesters. Because of that, some branches of tree are redundant. Thus, in the future work, the redundancies should be taken account to reject for bad predictions and reducing the memory.

REFERENCE

[1] M. Liu, "A Study of Mobile Sensing using Smartphones," *International Journal of Dis-*

tributed Sensor Networks, Vol. 2013, Article ID 272916, 2013.

[2] L. Song, D. Kotx, R. Jain, and X. He, "Evaluating Location Predictors with Extensive Wi-Fi Mobility Data," *Proceedings of INFOCOM 2004*, Vol. 2, pp. 1414-1424, 2004.

[3] L. Song and D. Kotz, "Evaluating Next-Cell Predictors with Extensive Wi-Fi Mobility Data," *IEEE Transactions on Mobile Computing*, Vol. 5, No. 12, pp. 1633-1649, 2006.

[4] K. H. Nguyen and W. J. Hwang, "A Mobility Control Architechture Supporting Fast Handover for Future Internet Networks," Journal of Korea Multimedia Society, Vol. 15, No. 12, pp. 1485-1491, 2012.

[5] A. Bhattacharya and S.K. Das, "LeZi-Update: An Information-Theoretic Framework for Personal Mobility Tracking in PCS Networks," *ACM/Kluwer Wireless Networks Journal*, Vol. 8, No. 2-3, pp. 121-135, 2002.

[6] K. Gopalratnam and D.J. Cook, "Active Lezi: an Incremental Parsing Algorithm for Sequential Prediction," *International Journal on Artificial Intelligence Tools*, pp. 917-930, 2003.

[7] K. Gopalratnam and D.J. Cook, "Online Sequential Prediction via Incremental Parsing: The Active LeZi Algorithm," *IEEE Intelligent System.* Vol. 22, No. 1, pp. 52-58, 2007.

[8] A. Rodriguez-Carrion, C. Garcia-Rubio, and C. Campo, "Performance Evaluation of LZ-Based Location Prediction Algorithms in Cellular Networks," *IEEE Communication Letter*, Vol. 14, No. 8, pp. 707-709, 2010.

[9] S. Scellato, M. Musolesi, C Mascolo, V. Latora, and A. T. Campbell, "Nextplace: A Spatio-temporal Prediction Framework for Pervasive Systems," *Proceedings of 9th International Conference on Pervasive Computing,* Springer, Vol. 6696, pp. 152-169, 2011.

[10] H. Fang and J. Ruan, "An Improved Position Prediction Algorithm based on Active LeZi in Smart Home," *Proceeding of Computer Science & Service System*, pp. 1733-1736, 2012.

[11] Y. Chon, H. Shin, E. Talipov, and H. Cha, "Evaluating Mobility Models for Temporal Prediction with High-granularity Mobility Data," *Proceedings of PerCom*, pp. 206-212, 2012.

[12] P.S. Prasad and P. Agrawal, "Movement Prediction in Wireless Networks using Mobility Traces," *Proceedings of Consumer Communications and Networking Conference*, pp. 1-5, 2010.

[13] D. Kotz, T. Henderson, and I. Abyzov, and J. Yeo. CRAWDAD data set dartmouth/campus (v. 2007-02-08). Downloaded from http://crawdad.org/dartmouth/campus/, February 2007.

Chuyen Luong

She received Engineering degree in School of Electronics and Telecommunications from Hanoi University of Sciences and Technology, Vietnam in 2012. She is currently studying for her MS Degree in School of Electronics and Computer Engineering, Chonnam National University, South Korea. Her research interests are mainly in the field of context awareness, pattern recognition

Hyukro Park

He received a BS degree from Seoul National University; and his MS and Ph.D. degrees from KAIST. He has been a professor at Chonnam National University since 1999. He is now in charge of the Information Retrieval Lab at Chonnam University. His research interests are in the areas of information retrieval, the Q&A system, natural language processing, database system, and data mining.

Son Do

He received BS degree in Department of Math Computer Science, University of Science, VNU-HCMC in 2011. He is currently studying for his MS Degree in School of Electronics and Computer Engineering, Chonnam National University, South Korea. His research interests are context awareness, and pattern recognition.

Deokjai Choi

He is full professor of Computer Engineering Department at Chonnam National University, South of Korea. He received BS degree in Department of Computer Science, Seoul National University, in 1982. He got MS degree in Department of Computer Science, KAIST, South Korea in 1984. He got PhD degree in Department of Computer Science and Telecommunications, University of Missouri-Kansas City, USA in 1995. His interest on research spans from context awareness, pervasive computing, sensor network, future Internet and IPv6.