10-2003

Mobicom Poster: Evaluating Location Predictors with Extensive Wi-Fi Mobility Data

Libo Song
Dartmouth College

David Kotz
Dartmouth College

Ravi Jain
DoCoMo USA Labs

Xiaoning He
DoCoMo USA Labs

MobiCom Poster: Evaluating location predictors with extensive Wi-Fi mobility data

Libo Song[a], David Kotz[a]
{lsong,dfk}@cs.dartmouth.edu

Ravi Jain[b], Xiaoning He[b]
{jain,xiaoning}@docomolabs-usa.com

[a]Computer Science Department, Dartmouth Colloege, Hanover, NH, USA
[b]DoCoMo USA Labs, San Jose, CA, USA

I. Introduction

A fundamental problem in mobile computing and wireless networks is the ability to track and predict the location of mobile devices. An accurate location predictor can significantly improve the performance or reliability of wireless network protocols, the wireless network infrastructure itself, and many applications in pervasive computing. These improvements lead to a better user experience, to a more cost-effective infrastructure, or both.

Location prediction has been proposed in many areas of wireless cellular networks as a means of enhancing performance, including better mobility management, improved assignment of cells to location areas, more efficient paging, and call admission control.

To the best of our knowledge, no other researchers have evaluated location predictors with extensive mobility data from real users.

In this poster we compare the most significant domain-independent predictors using a large set of user mobility data collected at Dartmouth College. In this data set, we recorded for two years the sequence of wireless cells (Wi-Fi access points) frequented by more than 6000 users.

We found that the simple Markov predictors performed as well or better than the more complicated LZ predictors, with smaller data structures.

II. Background

II.A. Domain-independent predictors

We consider here only domain-independent predictors. We are interested in *on-line predictors*, which examine the history so far, extract the current context, and predict the next location. Once the next location is known, the history is now one location longer, and the predictor updates its internal tables in preparation for the next prediction.

We implemented two families of domain-independent predictors, Order-k Markov predictors and LZ-based predictors. For detail of these predictors, please see a survey paper by Cheng, Jain, and van den Berg [1].

II.B. Accuracy metric

During an on-line scan of the location history, the predictor is given a chance to predict each location. There are three possible outcomes for this prediction, when compared to the actual location: correct, incorrect and "no prediction".

We define the *accuracy* of a predictor for a particular user to be the fraction of locations for which the predictor correctly identified the next move. Thus, "no prediction" is counted as an incorrect prediction.

II.C. Entropy metric

We believe that there are some intrinsic characteristics of a trace that ultimately determine its predictability and hence the performance of different predictors. In the results section, we compare the accuracy metric with the entropy metric, for each user.

II.D. Data collection

We have been monitoring usage on the Wi-Fi network at Dartmouth College since installation began in April 2001. By March 2003, there are 543 access points providing 11 Mbps coverage to the entire campus. We recorded the history of location changes for each of 6000 wireless cards, many for a period of months or years.

For more information about Dartmouth's network and our data collection, see our previous study [2].

III. Results

We evaluated the location predictors using our Wi-Fi mobility data.

As we step through the locations in a single trace, attempting to predict each location in sequence, we can examine the accuracy up to that point in the trace. Figure 1 shows how the accuracy metric varies

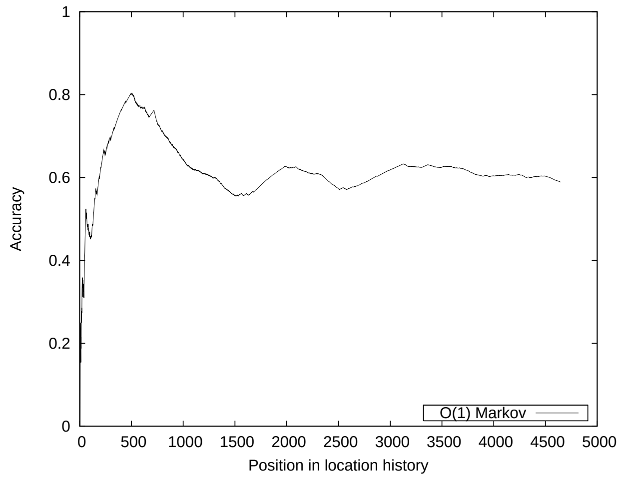

Figure 1: Prediction accuracy for a sample user

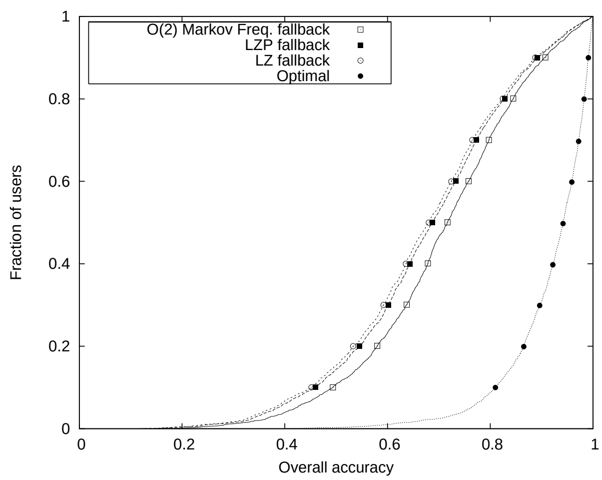

Figure 2: The best predictors, compared

over the course of one user's history, using the $O(1)$ Markov predictor. Ultimately, for comparison purposes, we define the accuracy over the entire trace, the rightmost value in this plot.

Of course, we have several thousand users and the predictor was more successful on some traces than on others. We use the accuracy of predictors in cumulative distribution function (CDF) curves to compare with each other.

In this study we examine only those users with long hisories (1000+ moves). We compare the best Markov predictors with the best LZ predictors in Figure 2. It is difficult to distinguish the LZ family, which all seem to have performed equally well. The $O(2)$ Markov predictor, with fallback to $O(1)$ whenever it had no prediction, was the best overall. It is striking that the extra complexity, and the theoretical aesthetics, of the LZ predictors apparently gave them no advantage.

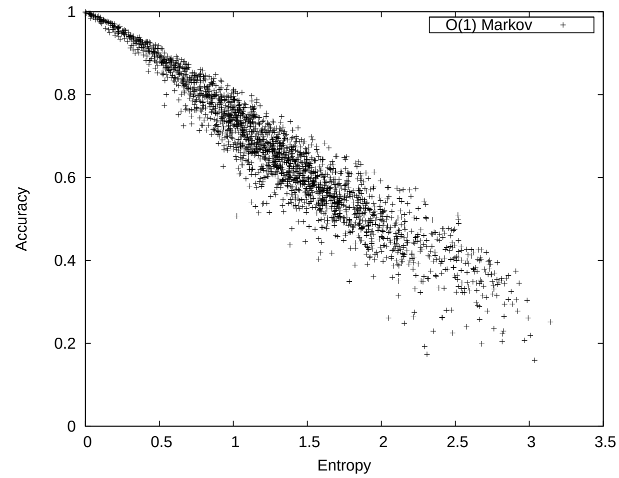

Figure 3: Correlating accuracy with entropy

We include an "Optimal" curve in Figure 2, as a simple upper bound on the performance of history-based location predictors. In our definition, the "optimal" predictor can accurately predict the next location, except when the current location has never been seen before. Although it should be possible to define a tighter, more meaningful upper bound for domain-independent predictors like those we consider here, it seems clear that there is room for better location-prediction algorithms in the future. It may be that some user traces are simply less predictable than others. In Figure 3 we compare the entropy of each user, based on the probability table built by the $O(1)$ Markov predictor, with the accuracy of the $O(1)$ predictor. The correlation is striking, and indeed the correlation coefficient is -0.95 (a coefficient of 1.0 or -1.0 represents perfect correlation). This strong correlation indicates that some users with high entropy are doomed to poor predictability.

References

[1] Christine Cheng, Ravi Jain, and Eric van den Berg. Location prediction algorithms for mobile wireless systems. In M. Illyas and B. Furht, editors, *Handbook of Wireless Internet*. CRC Press, 2003.

[2] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. In *Proceedings of MobiCom*, pages 107–118, September 2002.

This is a project of the Center for Mobile Computing at Dartmouth College, and is funded by DoCoMo USA Labs. http://cmc.cs.dartmouth.edu