Predicting the Users' Next Location from WLAN Mobility Data

Ljubica Pajevic[‖*], Viktoria Fodor[*] and Gunnar Karlsson[*]
[‖]Department of Informatics, Technical University of Munich, Munich, Germany
[*]Department of Network and Systems Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
{ljubica, vjfodor, gk}@kth.se

***Abstract*—Accurate prediction of user mobility allows the efficient use of resources in our ubiquitously connected environment. In this work we study the predictability of the users' next location, considering a campus scenario with highly mobile users. We utilize Markov predictors, and estimate the theoretical predictability limits. Based on the mobility traces of nearly 7400 wireless network users, we estimate that the maximum predictability of the users is on average 82%, and we find that the best Markov predictor is accurate 67% of the time. In addition, we show that moderate performance gains can be achieved by leveraging multi-location prediction.**

***Index Terms*—mobility prediction, trace-collection analysis, WLAN, entropy.**

I. Introduction

Human mobility analysis and prediction has been topical in diverse research areas: from ubiquitous computing to epidemiology, from transportation systems to social sciences. In networking research, being able to model and forecast human mobility helps predicting network resource availability for users in wireless networks, while novel applications that are likely to benefit from accurate mobility prediction are also emerging. One such example is mobile edge computing, envisioned as a solution to cater the increasing storage and computing demands of mobile devices by exploiting the local resources in the network and bringing them to end-users as near as possible [1]. However, to evaluate the potential gain of utilizing prediction for such applications, a solid understanding of achievable predictability of human mobility is necessary.

In this paper we evaluate the predictability of user movements in a university campus scenario by means of Markov model predictors. In summary, our contributions are:

- By using network association patterns of the users as the proxy of their locations, we find that the next visited location of the users in the analyzed trace can be predicted with 67% accuracy. We obtain this result by using a Markov predictor, specifically, the predictor of order one, which we find to be more accurate than similar predictors of higher orders.
- Adopting the theoretical framework established in [2], we estimate how predictable the mobility of the users is in general, and quantify how close our predictions are to the best achievable ones. While the average upper bound is almost 82%, the prediction results are within 20% from the expected limits for 87% of the users.
- We investigate multi-location prediction and find that the gain diminishes rapidly, as the continuous exploration of new locations significantly limits the performance.

978-1-5386-4533-8/18/$31.00

In the remainder of the paper, we first survey mobility prediction methods and previously reported results, in Section II. The theoretical background of the paper is summarized in Section III. Algorithms for mobility prediction are proposed in Section IV. We describe the analyzed dataset and pre-processing steps in Section V. In Section VI we evaluate the performance of the mobility predictors. Section VII concludes this study.

II. Related work

Our work can be positioned along two axes: mobility prediction methods in general, and the predictability of users in campus environment. Owing to their simplicity and high efficiency, Markov family predictors are the most common means for mobility prediction, utilizing the knowledge of the recent, short location history of fixed length [3–5] or variable length [6]. The second line of works utilizes naive Bayes models, with feature selection including time and location [7], or additional features extracted from communication patterns [8]. Jeong et al. [9] use non-parametric Bayesian interference to cluster users with similar mobility patterns to improve the prediction accuracy by gathering more training data. In [10, 11] location prediction is treated as a classification problem and the proposed solutions are based on supervised learning, utilizing classification methods such as decision trees, k-nearest neighbors, support vector machines, and gradient boost. Recently, more complex methods have been devised in a form of dynamic Bayesian networks [12, 13] and recurrent neural networks [14].

Markovian mobility predictors are evaluated in [2, 3, 15], with the conclusion that low order Markov predictors (order 2 and order 1 respectively) achieve highest prediction accuracy, from 70% to 85%, depending on the considered mobility trace. Common for [2, 15] are the formulation of the location sequences, accounting for every change rather than considering discretized time steps, but also the age of the datasets. As users are becoming more mobile and networks more dense, the achievable predictability of mobility may decrease. We have analyzed a recent campus mobility trace in [16], showing that user mobility is highly diverse, due to the proliferation of handhold devices. Now we use the same trace to evaluate the predictability of the next locations.

III. Location prediction methodology

In this study we follow user mobility in discrete time steps and discrete locations. The locations that a user visits belong to a finite set of locations (of cardinality N) and each location is represented by a symbol x_i from the alphabet $\mathcal{A}$. Let us define

$\{X\}$ as a stochastic process that corresponds to the mobility of a user, given by the ordered set of random variables X_i, $\mathbf{X}$=$\{X_1, X_2, ..., X_n, ...\}$. The location x of the user at time n is then a realization of the process $\mathbf{X}$ generated by X_n. For a given user, the sequence of visited locations is called a *location history*, represented by a string of n location symbols and denoted by $h_n = x_1x_2...x_n$, x_i∈$\mathcal{A}$, for $1 \leq i \leq n$.

We approach the prediction task by performing *online prediction*, i.e., by recording and examining the entire history up to the current point in time, to predict the next location based on the current one and the history of visited locations. We consider domain-independent, *order-k* (*O(k)*) Markov predictors.

A. Markov family predictors

The *order-k* (*O(k)*) Markov predictor [17] assumes that the user's future location depends only on the k most recent samples from the location history. This sequence is called a *context*. Denote by X_i^j=$x_i, ..., x_j, i$≤j a substring of the location history. Then the context at the n^{th} observation is $c = X_{n-k+1}^n$ and the entire history $h_n = X_1^n$. Assuming that the observed user mobility patterns are stationary, the prediction problem can be abstracted as a prediction of the next symbol in the time series generated by a source with a stationary distribution. Hence, the probability of transition is the same wherever the context is the same. The probability that the location X_{n+1}=x_l follows the history h_n is then

$$P(X_{n+1} = x_l|h_n) = P(X_{n+1} = x_l|X_{n-k+1}^n = c) \quad (1)$$

The probability values can be represented in the form of a *transition matrix M*, such that the columns and rows of the matrix M correspond to *k*-length strings from $\mathcal{A}^k$: $P(X_{n+1}$=$x_l|h_n) = M(c, c')$ where c=X_{n-k+1}^n is the current context and c'=$X_{n-k+2}^n x_l$ is the next context. However, the matrix M is unknown and therefore the predictor needs to use an approximation $\hat{P}$ based on the current history h_n and given the current context c of length k. The probability of next symbol being x_l is then approximated by

$$\hat{P}(X_{n+1} = x_l|h_n) \approx \frac{N(cx_l, h_n)}{N(c, h_n)} \quad (2)$$

where $N(s_1, s_2)$ denotes the number of times substring s_1 occurs in the string s_2.

For the next location, the Markov predictor selects the symbol with the highest probability estimate X_{n+1}=$\arg\max_{x_l \in \mathcal{A}}[\hat{P}(X_{n+1} = x_l|h_n)]$, that is the symbol that most frequently followed the current context in the prior history. The Markov predictor makes no prediction if the current context has never been seen.

B. Entropy, entropy rate and maximum predictability

Before measuring the performance of mobility predictors on a specific set of users, a more fundamental question needs to be addressed: What is the best achievable accuracy? Or, in other words: How predictable the users are? To approach these questions, we consider the measure of *predictability* as the mean probability of correctly predicting the user's next location, given knowledge of all possible trajectories that could have led them to that point [18]. Define the *maximum predictability* Π^{max} as the highest achievable value of predictability, assuming that predictions were obtained by an algorithm that always returns the most likely next location. In notation, let h_k be one realization of the user's location history up to the time step k, $P(h_k)$ the probability of this particular sequence, and $\pi_{ML}(h_k)$ the probability of occurrence of the most likely next location. Then, Π^{max} is the expected predictability for all values of k and histories h_k:

$$\Pi^{max} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left[\sum_{h_k} P(h_k)\pi_{ML}(h_k)\right] \quad (3)$$

To assess the limits of predictability and to quantify how much the location of users can be predicted, we utilize the concept of information entropy. In information theory, the level of randomness of a process can be measured by entropy, a metric that represents the average information, provided by each realization of a random variable. Given a discrete probability distribution $\pi = \{p_1, p_2, ..., p_N\}$ of the symbols in the alphabet $\mathcal{A}$ of size N, the entropy is defined as $\mathcal{H} = -\sum_{i=1}^{N} p_i \log_2 p_i$. The entropy measures the uncertainty of predicting the user's next location by considering only the frequency of previous visits while ignoring the sequence of the visits, that is, the correlation between consecutive locations. Terms such as *non-sequential* or *temporal-uncorrelated* entropy (cf. [4]) usually refer to this definition of entropy.

Since human mobility is not random, for a realistic representation of it we cannot assume that the sequence of locations is uncorrelated. Instead of entropy, the amount of randomness in the realization of correlated random variables can be better measured by the amount of new information expected from each event, given the past events. The measure that captures how the entropy changes with the number of observations is the entropy rate, or "per-symbol" entropy. The entropy rate, $\mathcal{H}_r(\mathbf{X})$ of a stochastic process [19] can be defined in terms of the joint entropy of its n random variables as

$$\mathcal{H}_r(\mathbf{X}) = \lim_{n\to\infty} \tfrac{1}{n} H(X_1, X_2, ...X_n) \quad (4)$$

when such limit exists. For a stationary process, as it has been shown in [19], such limit always exists and equals

$$\mathcal{H}_r{}'(\mathbf{X}) = \lim_{n\to\infty} \mathcal{H}(X_n|X_{n-1}, X_{n-2}, ..., X_1) \quad (5)$$

where the conditional entropy $\mathcal{H}(X_n|X_{n-1}, ..., X_1)$ is defined as

$$\mathcal{H}(X_n|X_{n-1}, ..., X_2, X_1) = \mathcal{H}(X_n|X_1^{n-1}) \quad (6)$$

$$= -\sum_{h_n \in \mathcal{A}^n} P(h_n) \log_2 P(X_n = x_n|X_1^{n-1} = h_{n-1}) \quad (7)$$

The maximum entropy rate from Eqs. (4), (5) is linked to the upper bound of maximum predictability $\bar{\Pi}^{max}$, [18]:

$$\mathcal{H}_r(\mathbf{X}) = -\bar{\Pi}^{max} \log_2 \bar{\Pi}^{max} - (1 - \bar{\Pi}^{max}) \log_2 \tfrac{1-\bar{\Pi}^{max}}{N-1} \quad (8)$$

Formula (8) generally describes relation between the entropy rate and the maximum predictability of a symbol sequence. However, the formula takes as an input $\mathcal{H}_r(\mathbf{X})$, and does not consider that $\mathcal{H}_r(\mathbf{X})$ needs to be estimated. Therefore, the bound will depend on the quality of entropy rate estimation, as well as the number of possible next locations.

The problem of estimating the entropy rate of a stochastic process has been widely investigated in information theory, statistics, and machine learning. We adopt the often used approximation, proposed by Kontoyiannis et al. [20] for the design of lossless compression algorithms, to compute the entropy rate as

$$\hat{\mathcal{H}}_r(\mathbf{X}) \approx \left(\frac{1}{S} \sum_{i=2}^{S} \frac{\Lambda_i}{\log_2 i} \right)^{-1} \quad (9)$$

where Λ_i denotes the length of the shortest string starting at location i which previously does not appear as a continuous substring between locations 1 and $i-1$ and S is the length of the sequence. This approximation is known to provide the lower bound of the entropy rate, leading to an upper bound of the maximum predictability according to (8).

IV. Prediction algorithms

This section details the prediction algorithm and defines the performance metrics used for evaluation in Section VI.

A. Single next location predictor

We apply the method for learning and updating the estimates of transition probabilities, given by formula (2). The predictor keeps a running estimate of the transition matrix, to return the most likely next location. After observing a new location, the predictor adds this location to the list of symbols, and with each newly learned transition it expands the transition matrix. Two special situations may occur: First, when multiple locations are returned with the same estimated transition probability and the second, when the user visits a new location, and hence the current context has not been observed previously. To deal with the first case, we implement a tie-breaking method, suggested in [2], which among the ties selects the location that was *most recently* visited. In the case of a newly learned context, we distinguish two sub-cases. For higher order predictors, i.e., $k>1$ our predictor relies on the fall-back feature, recursively searching for contexts of shorter lengths until reaching $k=1$. For $k=1$, a reasonable strategy is to predict that the user's location in the next time slot is unchanged from the current one, since users tend to stay in the same location for times longer than the considered 15-minute time slots. The algorithm is summarized by the pseudo-code in Algorithm 1.

The output of the predictor is a sequence of predicted locations, each prediction associated with two possible outcomes: correct or incorrect. We define the predictor *accuracy* as the ratio of correct predictions and all predictions made up to the current time step. We will occasionally refer to this definition as the *per-user accuracy* to emphasize that the result was obtained for individual users, as opposed to the *per-slot accuracy* under which we will consider aggregate performance of the predictor across all users and predictions made.

B. Multi-location predictor

The accuracy of the mobility prediction can be increased by selecting more than one possible next locations, with the use of a multi-location predictor. While this reduces uncertainty, from the practical perspective it also incurs additional costs considering the predictor's memory, and more importantly additional costs for the resource allocation applications that utilize the predictor. We assume that n locations are chosen as the possible next locations, the associated cost of prediction is then n. The multi-location predictor utilizes an order 1 Markov predictor, and is therefore denoted as $O(1, n)$.

We implement the n-locations selection method as follows.

- For $n = 1$, Algorithm 1 is followed.
- Prediction for $n=2$ relies on the Markov predictor, selecting the two most probable locations.
- For $n>2$ the prediction set contains the current location, and $n-1$ other, most probable next locations.

Similarly as in the single-location setup, the multi-location predictor may find equally likely transitions, but including all those in the prediction set would exceed the required n. The predictor then returns most recently visited locations among the ties. As n increases, the number of known transitions m will become smaller than n at some prediction steps. To compensate, additional $n-m$ most frequently visited locations, denoted as *top* locations, are added to the prediction set.

Algorithm 1 Predict Next Location

```
1: procedure PredictNext(k, L)
2:     ▷ k: order of the Markov predictor, k ∈ Z+
3:     ▷ L: sequence of previous symbols, |L| ≥ k
4:     n ← |L|
5:     A ← L.UniqueMembers()
6:     C ← [L_{n-k+1}, ··· , L_n]
7:     N_c ← L.SubsequenceCount(C)
8:     p ← nil, V ← ∅
9:     if k = 0 then
10:        return L.CurrentLocation()
11:    end if
12:    for all a ← A do
13:        C' ← [C, a]
14:        p̂ ← L.SubsequenceCount(C')/N_c
15:        if p̂ > p and p̂ > 0 then
16:            p ← p̂, V ← {a}
17:        else if p̂ = p then
18:            V ← V ∪ a
19:        end if
20:    end for
21:    if |V| = 0 then
22:        return PredictNext(k-1, L)
23:    else if |V| > 1 then
24:        return TieBreakSelection(V)
25:    else
26:        return V[0]
27:    end if
28: end procedure
```

V. Mobility trace and basic statistical properties

A. Dataset collection

The mobility trace comes from the wireless network association records collected at the KTH Royal Institute of Technology during 2014 [16]. The wireless network provides coverage on one large and four smaller campuses in the larger Stockholm metropolitan area. The largest site is covered by 790 wireless access points (APs) located in 48 buildings; the smaller sites have a single building per site with 19, 22, 23 and 77 deployed APs. While APs are mainly located inside the campus buildings, most of the outdoor areas are covered as well, due to the proximity of buildings and the dense AP deployment.

B. Trace preparation and preprocessing

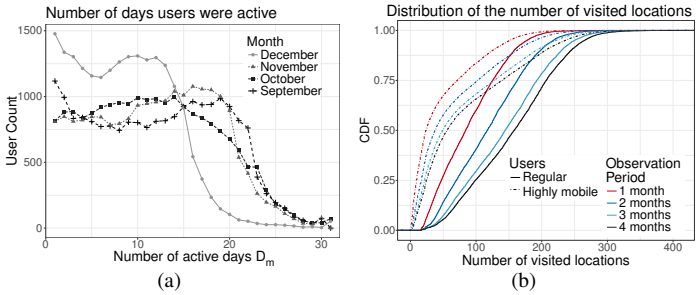
Since the raw dataset contains network association events only, the first step of the trace processing is to assign to each

entry the duration of the association. Details of the applied processing methodology can be found in our earlier study [16]. Users' trajectories are then defined by the tuples containing the access point the user was associated with, the timestamp of the association event and the estimated association time. The association time samples vary from several hours to several seconds. To capture relatively stable transitions, that could be used for the optimization of networked services like content distribution, we discretize time into 15-minute time slots and assign the location with the longest association time to each of the slots. Since the considered network is very dense the same physical location is often covered by many APs. Therefore, for the mobility trace we do not record the AP identifier, but the corresponding location, such as a given room or a corridor. We introduce an additional "offline" location to represent the user's departure from the network. This location is assigned to the slot at the end of the user's daily trace. A single offline slot is added to the trace as well, when the user is disconnected for a longer time during the day. Offline slots are however not considered under the evaluation of the predictor accuracy.

C. First look at the campus mobility

We extract a subset of the trace, spanning four months at the end of 2014—a period containing one academic semester, long enough for users' movement patterns to exhibit stationarity, but also limited to avoid seasonal changes which may be incurred by modified semester schedules.

1) Location visiting patterns: During the observed time period, 36561 unique devices were recorded, out of which 20225 were active each of the four months. We consider the latter group of user devices, denoted as *regular (WLAN) users*. We want to measure how often these users access the network, therefore for each user we count the number of *active* days when they appeared in the trace, and the number of active days per month, D_m. Fig. 1(a) shows the number of users active for D_m days in each month and we find, as expected, that users were less active during December, with D_m less than 9 days on average, than during the other months, when the average D_m count was around 12. To examine how mobile the users are, we refer to Fig. 1(b) which shows the cumulative distribution functions (CDFs) of the number of unique locations visited by regular users during the period of one to four months. The number of unique locations increases throughout the four months; after four months, the maximum and mean number of visited locations are 412 and 85.9, respectively.



(a) (b)

Fig. 1. (a) Distribution of active days for users that accessed the network during experimentation period. (b) CDF of the number of visited locations.

In the remainder of this analysis, we will focus on a selected group of regular users that exhibited higher activity, termed *highly mobile* users. A user is classified into this group if they had at least five active days each month, during which they visited 15 or more locations. The highly mobile subset contains 7379 users, who visited, on average 155.8 locations during four months; their location visiting CDFs are plotted in Fig. 1(b).

2) Entropy rate and empirical predictability: To understand first how the entropy and entropy rate of the users' location sequences depend on the length of observation, we plot the corresponding density functions (PDFs) in Fig. 2. The entropies increase with the observation period while the entropy rates show little dependence. With respect to the degree of randomness in users' movements, large entropy values at first hint at the high degree of randomness; however the entropy rate estimates, Fig. 2(b), show that the temporal order of the samples greatly reduces the uncertainty in users' next movements. We use these estimates and numerically solve (8) for $\bar{\Pi}^{max}$ for each user. The distribution of the resulting maximum predictability is shown in Fig. 2(c). With the densities peaking at around 75%, we expect that the users' movements will be highly predictable.

VI. Predictor performance evaluation

Following the procedure described in Section V-B we transform user association records into sequences of visited locations. We first allow the algorithm to obtain initial estimations of transition probabilities during the training period, upon which the predictor starts forecasting in an online manner, by updating the transition matrix at each time step with the new observation, and subsequently making a prediction for the next time step. The training period covers the first two weeks from the first day when a user was seen in the trace. In order to detect potential abrupt changes in user mobility patterns or divergence of prediction results, we take snapshots of accuracy estimations at different times, specifically after two and four months.

A. Accuracy of the O(k) predictors

Taking the first step towards investigating how the length of the context affects the accuracy, we look at the performance of predictors of different orders for a sample user's sequence. Fig. 3 shows the running accuracy of prediction for a single user, for the predictors of orders k=$\{1, 2, 3\}$. There are minor differences in the predictors' performance with asymptotic accuracy values ranging from 65.9% (k=1) to 64.6% (k=2) and to 63.7% (k=3). We can also see that the accuracy quickly converges to the asymptotic value; in this particular example after around 250 steps. Convergence, though, cannot be guaranteed: we observed that for some users the accuracy curve starts to diverge after a certain time, indicating that their mobility patterns changed during the experimentation period.

Fig. 4 shows the CDF of the prediction accuracy considering all highly mobile users. We sample the values of the accuracy after two months (history h_1) and four months (history h_2). For each of the three predictor orders, the distribution curves for the two location histories are very close to each other, and for each pair the accuracy after four months has only increased for 0.5%.

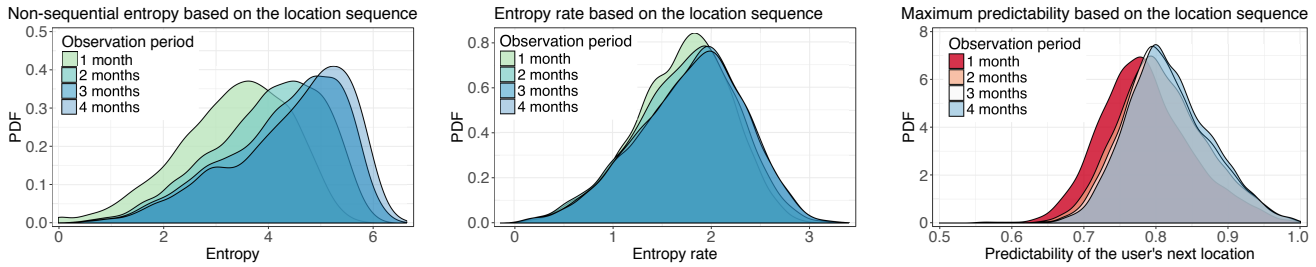


Fig. 2. Entropy and predictability measures for highly mobile users. Distributions of the users' (a) non-sequential entropies $\mathcal{H}$, (b) entropy rates $\hat{\mathcal{H}}_r$, and (c) maximum predictabilities $\bar{\Pi}^{max}$.

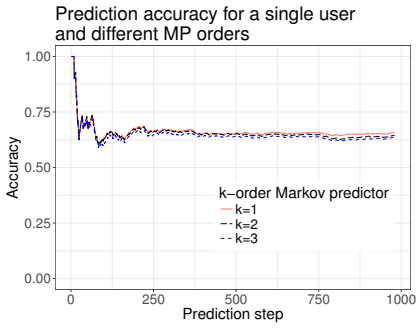


Fig. 3. Prediction accuracy for a sample user.

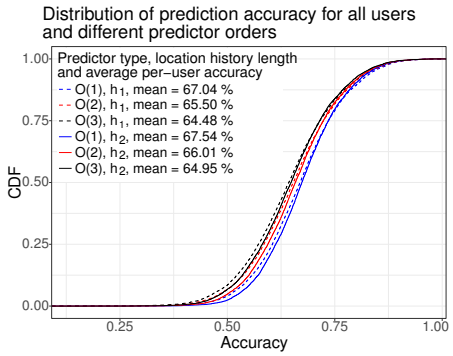


Fig. 4. CDFs of the prediction accuracy for $O(k)$ predictors, $k = \{1, 2, 3\}$.

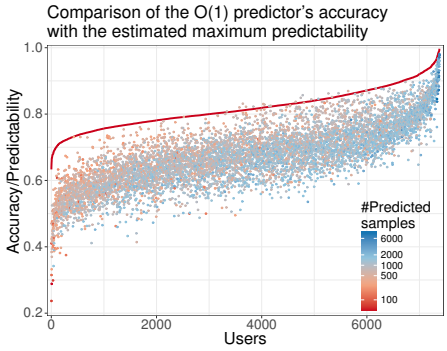


Fig. 5. Accuracy of the $O(1)$ predictor (dots) and the maximum predictability (line).

Next, as the predictor order increases, the accuracy deteriorates from the lowest to the highest order: each order incurs around 1–1.5% loss in accuracy as compared to the previous one. Overall, the $O(1)$ predictor achieves slightly better results than the other two predictors. Considering the accuracy values for each user over the entire trace duration, for most users $O(1)$ performs better than the other two predictors; for less than 2%, or 126 users, the highest accuracy is achieved by $O(2)$ and only for 12 users $O(3)$ performs best. We have also tested higher order Markov predictors (up to order 5) and observed the trend of decreasing accuracy with increased order, therefore we excluded those predictors from further analysis.

Next we compare the performance of the most accurate predictor, $O(1)$, with the estimated maximum predictability values, shown on Fig. 2(c). Fig. 5 depicts the accuracy (dots) and the predictability bound (solid line) for each individual user. We can infer that the predictor does indeed approach the expected upper bounds of predictability. The absolute differences between these two measures closely follow a normal distribution, with the mean value of 14% and the standard deviation of 5.3%. We also observe few data points above the predictability line. This counterintuitive result indicates that the accuracy *can* surpass the estimated limit of predictability, when the entropy rate estimate is not accurate due to too few samples.

B. Performance gain from multi-location prediction

Next we investigate how much the prediction accuracy improves if multiple possible next locations can be selected. Fig. 6 shows the results for multi-location prediction with $1 \leq n \leq 10$ locations. As expected, we see that increasing n markedly improves the prediction accuracy for all users: while for 75% of users single-location prediction achieves at least 60% accuracy, adding only one additional location increases the accuracy to 70% for the same portion of users and adding four locations, $n = 5$, increases the accuracy to 75%. Considering the most predictable users, specifically the 25% users with best prediction results, the minimum accuracy increases from 73% for $n = 1$ to 75% and 81% for $n = 2$, respective $n = 5$.

To evaluate the overall performance gain achieved by predicting multiple locations, Fig. 6(b) shows the average accuracy across all users, as well as the average per-slot accuracy for all performed predictions. The latter result serves to mitigate the impact of low prediction accuracies for users with short location histories, and indeed, these values are slightly higher for all n. From the figure we can conclude that selecting more than one possible next location quickly boosts the accuracy, however the accuracy gain diminishes quickly, and the achieved average accuracy seems to settle right above 80%.

To uncover the causes of the diminishing accuracy improvements, we inspect the prediction process in more detail. One of the major challenges for online prediction is that users often visit, or *discover* new locations. For the prediction attempts in time steps when a discovery happens, the result is always incorrect. To quantify how much discovering new places contributes to predictors' inaccuracies, we consider the ratio of new locations and the total number of samples, R_{newloc}, for each user. For 40% of users, this ratio is more than 0.1, thus the highest accuracies that can be achieved for these users, with the current implementation of our algorithm, is not more than 90%. Let us explore what other properties of the users' location sequences may affect the prediction performance. Fig. 7 depicts the relation between R_{newloc} and the accuracy for n=1 and n=10. The color of the data points represents the number

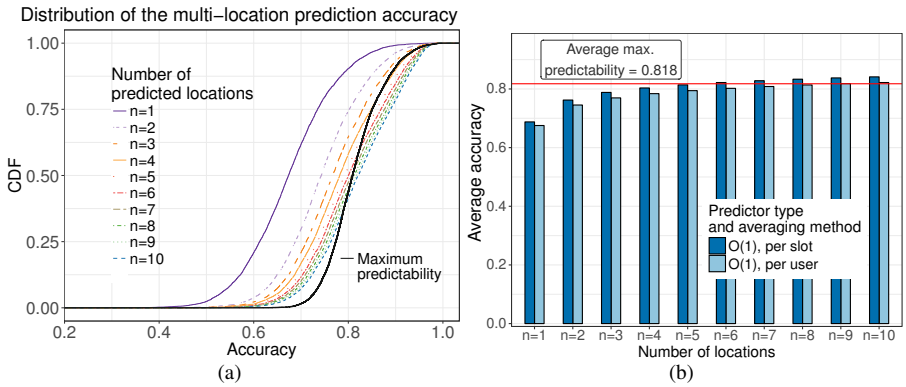


Fig. 6. Accuracy of the multi-location $O(1, n)$ predictor for $1 \leq n \leq 10$ predicted locations. (a) CDFs of per-user accuracies. (b) Bar plots of average accuracy values.

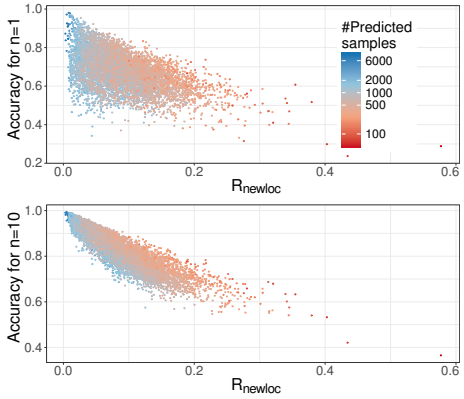


Fig. 7. Correlation between the ratio R_{newloc} and the prediction accuracy.

of samples. The plots imply that the multi-location prediction accuracy heavily depends on both R_{newloc} and the number of samples. In particular, this method of prediction is more likely to improve with n for users with low R_{newloc} and large number of samples, since such users are likely to be found in few locations, where they spend long visiting times. Comparing users with the same R_{newloc}, the predictor is likely to benefit less when the number of samples, and hence the number of observed transitions are lower. Finally, the limited predictor's performance seem to be rooted in the suboptimal selection method when numerous transitions from the given location are possible. The predictor then favors previously observed transitions—even with very low transition probabilities—over the user's top locations. This suggest a possible future improvement of the prediction algorithm, where a mix of popular next locations and top locations could be selected.

VII. Conclusion

We examined predictability of users' locations in a university campus scenario by analyzing traces of nearly 7400 wireless network users over a period of four months. Specifically, we focused on predicting the users' next location, given the history of their previous visits. By implementing online, low-order Markov predictors we were able to correctly predict users' next location 67% of the time. This is a reasonable result, considering the simplicity of the predictor and the empirical limits of predictability, and it will serve as the baseline for comparison with more advanced predictors based on machine learning techniques as part of our future work. Regarding the multi-location prediction, we have shown that adding a few more locations effectively increases accuracy, but the performance is limited by the frequent discoveries of new locations. For reproducibility purposes, we are planning to make the dataset used in the study publicly available.

References

[1] I. Farris, T. Taleb, A. Iera, and H. Flinck, "Lightweight service replication for ultra-short latency applications in mobile edge networks," in *Proc. IEEE ICC*, 2017.

[2] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating next-cell predictors with extensive Wi-Fi mobility data," *IEEE Trans. Mobile Comput.*, vol. 5, no. 12, 2006.

[3] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Scientific Reports*, vol. 3, 2013.

[4] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, 327(5968), 2010.

[5] L. Song, U. Deshpande, U. C. Kozat, D. Kotz, and R. Jain, "Predictability of WLAN mobility and its effects on bandwidth provisioning," in *Proc. IEEE INFOCOM*, 2006.

[6] R. Begleiter, R. El-Yaniv, and G. Yona, "On prediction using variable order Markov models," *J. Artif. Int. Res.*, 22(1), 2004.

[7] H. Gao, J. Tang, and H. Liu, "Mobile location prediction in spatio-temporal context," in *Nokia Mobile Data Challenge Workshop*, 41(2), 2012.

[8] T. M. T. Do and D. Gatica-Perez, "Contextual conditional models for smartphone-based human mobility prediction," in *Proc. ACM UbiComp*, 2012.

[9] J. Jeong, M. Leconte, and A. Proutiere, "Cluster-aided mobility predictions," in *Proc. IEEE INFOCOM*, 2016.

[10] T. Anagnostopoulos, C. Anagnostopoulos, and S. Hadjiefthymiades, "Mobility prediction based on machine learning," in *Proc. IEEE Intl. Conf. on Mobile Data Management*, vol. 2, 2011.

[11] C. Koehler, N. Banovic, I. Oakley, J. Mankoff, and A. K. Dey, "Indoor-ALPS: An adaptive indoor location prediction system," in *Proc. ACM UbiComp*, 2014.

[12] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and inferring transportation routines," *Elsevier Artif. Intell.*, 171(5–6), 2007.

[13] J. Krumm and E. Horvitz, "Predestination: Inferring destinations from partial trajectories," in *Proc. ACM UbiComp*, 2006.

[14] G. Moon and J. Hamm, "A large-scale study in predictability of daily activities and places," in *Proc. MobiCASE*, 2016.

[15] F. Chinchilla, M. Lindsey, and M. Papadopouli, "Analysis of wireless information locality and association patterns in a campus," in *Proc. IEEE INFOCOM*, vol. 2, 2004.

[16] L. Pajevic, V. Fodor, and G. Karlsson, "Revisiting the modeling of user association patterns in a university wireless network," in *Proc. IEEE WCNC*, 2018.

[17] J. S. Vitter and P. Krishnan, "Optimal prefetching via data compression," in *Proc. Symp. of Foundations of Comp. Sci.*, 1991.

[18] G. Smith, R. Wieser, J. Goulding, and D. Barrack, "A refined limit on the predictability of human mobility," in *Proc. IEEE PerCom*, 2014.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, Inc., 1991.

[20] I. Kontoyiannis, P. H. Algoet, Y. M. Suhov, and J. W. Abraham, "Nonparametric entropy estimation for stationary processes and random fields, with applications to english text," *IEEE Trans. Inf. Theory*, 44(3), 1998.