Exploring Periods of Low Predictability in Daily Life Mobility

James McInerney, Sebastian Stein, Alex Rogers, Nicholas R. Jennings
University of Southampton, Southampton, SO17 1BJ, UK
{jem1c10,ss2,acr,nrj}@ecs.soton.ac.uk

ABSTRACT

Researchers studying daily life mobility patterns have recently shown that humans are typically highly predictable in their movements. However, no existing work has examined the boundaries of this predictability, where human behaviour transitions temporarily from routine patterns to highly unpredictable states. Yet, this is arguably one of the most interesting and critical states, where users might be most in need of context-aware mobile applications. To address this shortcoming, we suggest a novel real-time estimator that can calculate an individual's *instantaneous entropy*: a measure for their momentary predictability. Applying this to a rich dataset, we show that individuals display high variance in their predictability over time. Furthermore, we demonstrate that mobile application usage patterns are correlated with instantaneous entropy, thus indicating that people use applications differently in unfamiliar situations.

1. INTRODUCTION

Understanding human mobility patterns is a significant research endeavour that has recently received considerable attention [8, 15]. Developing the science to describe and predict how people move from one place to another during their daily lives promises to address a wide range of societal challenges: from predicting the spread of infectious diseases, improving urban planning, to devising effective emergency response strategies [13]. Individuals are also set to benefit from this area of research, as mobile devices will be able to analyse their mobility pattern and offer context-aware assistance and information.

A key finding in this area was demonstrated by Song et al. [16]. Given the location traces of 50,000 mobile phone users, the authors used the Shannon entropy rate of this data to establish that the average predictability of a single person's current location (given their history of locations) was at least 80%, and was 93% on average. This finding is important because it gives us confidence that all people are highly predictable and have strong habitual elements to their daily lives.

However, very little is currently known about how this predictability varies with time. While the *average* predictability over large datasets has been well studied, no existing work has yet looked at the momentary, or *instantaneous*, predictability of an individual at a particular moment in time. Intuitively, it is expected that individuals transition through phases of relatively high predictability (e.g., during a working day at the office or while attending regular football practice on a Saturday afternoon), to sudden spikes in unpredictability (e.g., on holidays or while on sick leave).

Yet, studying and characterising these momentary transitions in predictability is important not only for understanding human mobility patterns, but also for providing context-aware services. Arguably, phases of high unpredictability are the most critical times to the user, who will, by definition, be in unfamiliar places, or in familiar places at unusual times. Novel experiences may often require extra levels of assistance that can be provided by a mobile device [17], in the areas of information, organisation and communication.

To address this shortcoming in existing research, this work is the first to explicitly investigate transient periods of low predictability in human mobility. In doing so, we make a number of contributions:

- We design a novel entropy estimator, based on the well-known Lempel-Ziv measure, called the *real-time entropy estimator* that provides a principled method for measuring the instantaneous predictability, or entropy, of an individual. Unlike existing approaches, such as plug-in estimators, context tree weighting methods and fixed and increasing window Lempel-Ziv estimators [7], our method uses only historical location traces and can be calculated in real time on a mobile device.
- We apply our estimator to GPS traces from the Nokia Lausanne dataset [12] and show, for the first time, that individuals display high variance in their instantaneous predictability. This confirms that human mobility patterns are characterised by alternating periods of high and low predictability.
- Exploiting the breadth of the Nokia dataset, we correlate the instantaneous predictability of individuals with their recorded mobile application usage. We demonstrate that application use is heavily influenced by the user's current state of predictability, and we show that it can be used to build new predictors that anticipate application use.

Taken together, our contributions offer a new perspective on the complex relationship between mobility patterns (measurable directly from a mobile phone) and broader user behaviours. This improved understanding opens the way for a new generation of mobile applications that can help the user at times of greatest need, but leave them to get on with their daily routine at other times.

In the remainder of this paper, we first introduce our new entropy estimator in Section 2, then apply it to the Nokia

This material was prepared for the Mobile Data Challenge 2012 (by Nokia) Workshop; June 18-19, 2012; Newcastle, UK.

dataset in Section 3. Finally, we provide several novel application sketches made possible by this work in Section 4 and conclude in Section 5.

2. INSTANTANEOUS PREDICTABILITY

To formalise predictability in daily life mobility, we deal with a random process $X = \{X_0, X_1, \ldots, X_N\}$ which is a sequence of random variables $\{X_n\}$ indicating the location of an individual at time n. All locations are assumed to belong to alphabet A, the set of possible locations this individual could be in.

The definition of Shannon entropy given by the equation $H = -\sum_i p(x_i) \log_2 p(x_i)$, where $p(x_i)$ is the probability of $x_i = X_n$, represents the independent entropy at time step n. However, we want to take account of the fact that we expect conditional dependencies between time steps, i.e., that knowing a history of locations tells us something about the future locations of a user. To this end, the rate of new information arriving at each step in a time series, the *entropy rate*, is a fundamental measure of predictability [10].

Assuming that X is stationary and ergodic (i.e., that every subsequence of X of equal size has the same probability distribution independent of its position, and that these statistics can be discovered from a single, sufficiently long sample of X), the entropy rate exists and is given by:

$$H(X) = \lim_{N\to\infty} H\left(X_N|X_{N-1}, \ldots, X_2, X_1\right) \tag{1}$$

This is an expression of the *conditional entropy*, which is calculated from the conditional and joint probabilities of the latest observed value x_N and those of the observed history $(x_1, x_2, \ldots, x_{N-1})$:

$$H\left(X_N|X_{N-1}, \ldots, X_1\right) = \sum_{x_1,\ldots x_N \in A^N} p\left(x_1, \ldots, x_N\right) \log_2 \frac{p\left(x_1, \ldots, x_N\right)}{p(x_1, \ldots, x_{N-1})} \tag{2}$$

In practice, the conditional entropy is hard to compute for shorter time series (of lengths in the order 10^3, as we deal with here) because for any non-trivial history size, the specific combinations required to calculate $p(x_1, \ldots, x_N)$ rarely occur in the data. The solution is to use an estimator.

There are many estimators for the entropy rate, but we will focus on the class of Lempel-Ziv (or LZ) estimators, since they are known to rapidly converge to the true entropy rate and do not assume anything *a priori* about the statistics of the time series [7, 10].

The increasing window LZ entropy estimator, $\hat{H}_N$, is defined as follows:

$$\hat{H}_N := \left(\frac{1}{N}\sum_{i=2}^{N} \frac{\Lambda_i}{\log_2(i)}\right)^{-1}, \tag{3}$$

where Λ_i is defined as the length of the shortest substring starting at position i that did not previously occur in the sequence $(x_1, \ldots, x_{i-1})$. The increasing window LZ estimate rapidly converges to the true entropy rate of the underlying process.

The estimator given by Equation 3 allows the assignment of a single entropy rate to each individual, characterising their overall mobility habits. If the user has $\hat{H}_N = 0$, then their behaviour is completely regular and therefore fully predictable. At the other extreme, another user with an entropy rate as high as $log_2 \mid A \mid$ would be moving completely randomly between elements in A. There is strong evidence that all people fall along a spectrum of entropy much closer to the lower extreme than the higher one [16]. However, this measure does not tell us *when* any individual is behaving unpredictably, limiting our analysis.

Given that there are several existing algorithms for predicting people's future locations [2, 15], one solution is to use the rate of prediction failure as an indicator of entropy rate. For example, Eagle and Pentland used principle component analysis to anticipate the afternoon locations of people given their morning locations [6]. They found that prediction failures increased on Friday nights and weekends in comparison with normal weekdays.

The problem with using predictors to measure regularity is the bias associated with the specific prediction approach. For example, many methods use only information about recent locations, ignoring longer term correlations. Repeated patterns over larger time scales, such as visiting the gym every Monday evening, regardless of the day's activities, would be erroneously considered unpredictable.

Alternatively, metrics with no predictive component have been used to measure regularity. Song et al. considered an individual to be behaving regularly if their location matched their most visited location for that time of week [16]. Chon et al. more generally considered the top n locations rather than just the most visited location, and varied the time granularity [5]. Both these approaches capture basic long term correlations, but are susceptible to underestimating predictability even under mild transformations. For example, if a user were to shift her normal routine back an hour because of a doctor's appointment in the morning, the rest of the day could be classified as highly irregular, simply because she is not at her most visited places at the expected times.

To overcome these limitations, we use the entropy rate as a principled way to quantify departures from routine. Reconsidering the entropy estimator given by Equation 3, we introduce a modified version called the *real-time entropy estimator*. To allow a per time slot view of the entropy rate, we relate the instantaneous entropy at time i to the value of Λ_i. Specifically, the instantaneous entropy tells us what the overall entropy rate would be if the entire process X exhibited the predictability it currently has (i.e., $\forall j : 1 \le j \le N, \Lambda_j = \Lambda_i$). This concept is compatible with the assumption that X is stationary because it is measuring the properties of individual steps in the process, which together make up the stationary statistics.

The standard LZ estimator uses information about future points in the series to determine the present value of Λ_i. Therefore, it does not work in real time and has limited applicability in ubiquitous systems that provide in-the-moment assistance. To address this limitation, rather than searching forwards for the shortest substring that does not occur in the history, the real-time estimator searches backwards, truncating the search history by one step each time.

At time i, the real-time LZ estimator for the instantaneous entropy is defined as:

$$\tilde{H}_i := \frac{\log_2(i)}{\Gamma_i}, \tag{4}$$

where Γ_i is defined as the length of the shortest substring *ending* at position i that did not previously occur in sequence $(x_1, \ldots, x_{i-\Gamma_i})$. This estimate is defined such that all instantaneous entropy values can be combined to reproduce an estimator for the entire series (obtaining the original sliding window estimate, albeit with the Γ_i measure rather than Λ_i):

$$\hat{H}_N = \frac{N}{\sum_{i=2}^{N} \tilde{H}_i^{-1}} \quad (5)$$

It can be shown trivially that the reverse of a time series has the same entropy rate as the original. Therefore, $\Gamma_i \to \Lambda_i$ as $i \to \infty$ and the real time estimate also converges to the true entropy of the underlying process given by Equation 2.

In the next section we apply this estimate to real daily life location data to understand more about the unpredictability of mobility.

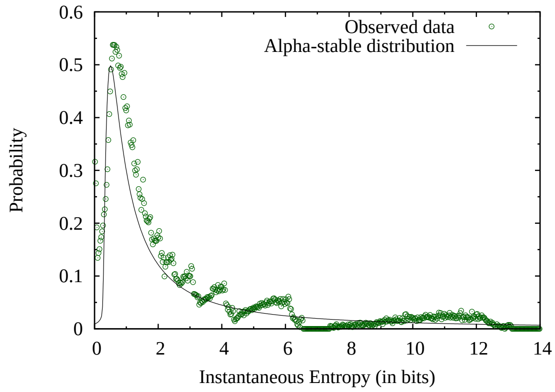


Figure 1: Probability distribution of instantaneous entropy over all hours for all users.

3. REAL-LIFE DATA ANALYSIS

To examine how the instantaneous predictability of individuals varies over time, we applied our estimator to the Nokia Lausanne dataset. This includes GPS locations, call logs and application usage for 38 people for a year recorded by their mobile phones [12]. The data consists of series of timestamped events (e.g., GPS readings with latitude and longitude coordinates, user usage of mobile applications, and directed message and call logs). As with any discrete estimator, it was necessary to pre-process the data first, in order to convert it to a form suitable for our estimator. We detail this in the following, before going on to describe the results from applying the real time entropy estimator to the processed data.

3.1 Pre-Processing

First, we derived the alphabet of locations, A, from the sequence of latitude and longitude GPS readings, which is a process of converting continuous variables to discrete labels. For location data, this is known as the problem of finding *significant locations*, and there are several approaches [1, 9]. We selected the online clustering method proposed by Kang et al. [9], because it is computationally feasible for running continuously and in real time on a resource-limited mobile phone. It takes into account the duration of visits to a location, the frequency of visits, and the minimum distance between locations.

Given the set A, we then assigned each location in the GPS trace to a distinct element in A, using a Euclidean distance threshold of 1 km.[1] We added a special element $\Omega \in A$ for readings that were not near any significant locations. In the entropy estimation, Ω is a special location that is always treated as a new location. Finally, we transformed the data into hourly windows, selecting randomly from the set of significant locations visited during the window, in proportion to the total duration at each location (e.g., if the user is at work for 45 mins of the window and at a cafe the remaining 15 mins, the location for that hour will be selected as work or as cafe with probability $\frac{3}{4}$ and $\frac{1}{4}$, respectively).

3.2 Instantaneous Predictability Results

In the following, we first apply our new entropy estimator to the entire dataset, in order to compare it to the standard LZ estimator (Section 3.2.1). Then, we examine how instantaneous predictability is distributed (Section 3.2.2), varies over time (Section 3.2.3), and how it correlates with mobile application usage (Section 3.2.4). Finally, we discuss two examples taken from an individual's daily life (Section 3.2.5).

3.2.1 *Aggregate Entropy*

For comparison purposes, we applied both the standard increasing window estimator given in Equation 3 and our real-time variant to the entire location traces of all 38 users to find the overall entropy rate for each person. The increasing window estimate was used for a large-scale study previously done by Song et al. under similar conditions with hourly locations [16]. It gives an overall estimate of a person's predictability from their location traces. In contrast, the real-time variant gives a breakdown of predictability per time step, enabling richer analyses.

The increasing window estimate using Equation 3 was 0.71 ($\pm$0.11 with 95% confidence), which is close to the frequency peak in entropy of 0.8 found by Song et al. [16], who used the same method on their data. The similarity in rates between the two datasets confirms that the Nokia users were typical of, though slightly more predictable than, users from the larger scale study.

Our real-time estimator gave an overall average entropy of 0.84 ($\pm$0.13 with 95% confidence).[2] This shift upwards, relative to the standard increasing estimator, can be explained by the fact that searching backwards to find the value of Γ_i at every time step i reduces the history size by a small margin, whereas searching forwards does not (however, the latter of course has access to privileged information about future states). A smaller history results in a slightly higher probability of treating a previously observed sequence as novel. Hence, by Equation 4, there is an overestimation of entropy with respect to the standard estimator. However, as $N \to \infty$, both measures rapidly converge to the true entropy rate [7, 10].

3.2.2 *Distribution of Predictability*

For the rest of the analysis we focus on the instantaneous (hourly) outputs of our real-time estimator to uncover the behaviour of the users. Figure 1 shows the probability distribution of the instantaneous predictability, as given by Equation 4, for all hours and all users in the dataset. The scatter

[1]The significant location extraction method we used ensured that locations were sufficiently far apart to do this unambiguously.

[2]Note that this is not simply the average instantaneous entropy, but rather the aggregate value obtained using Equation 5 (as discussed in Section 2).

plot of observed data points appears to follow a Lévy distribution [14], which has the form:

$$f(x|\gamma, \delta) = \sqrt{\frac{\gamma}{2\pi}} \left(\frac{e^{\frac{-\gamma}{2(x-\delta)}}}{(x-\delta)^{\frac{3}{2}}} \right) \tag{6}$$

To verify how closely the data follows this distribution, we ran an iterative procedure by Koutrouvelis [11]. The method estimates the 4 parameters of any alpha-stable distribution, a family that includes the Lévy, Gaussian and Cauchy distributions (but has no general closed expression for the probability density)[14]. We found the parameters[3] to be $\alpha = 0.51$, $\beta = 0.95$, $\gamma = 0.91$ and $\delta = 0.19$, which produces the curve shown in Figure 1. This supports the idea that the data follows a Lévy distribution, which is defined as having $\alpha = 0.5$ and $\beta = 1$. The curve would be more positively skewed (fitting the data points slightly better) if the entropy values were not truncated, since Equation 4 has the range $0 \leq \tilde{H}_i \leq \log_2(N)$, where N is the length of the entire user history.

We see that most of the time, the users are following their normal habits with a mode of entropy at 1.0 bit, but there is also a heavy tail of unpredictability. Heavy tailed distributions are found in many behaviours, including animal foraging displacements and aggregated human mobility [3, 4].

Since this is an aggregate distribution over all users, it is possible that the distribution for each individual follows a different form. We verified that the same distribution is present in the individual distributions too, albeit with varying probability masses near the peak and in the tail.

It is not surprising to observe the Lévy distribution in this context, however, it is interesting that no distance metric is involved in finding this distribution, since the entropy is derived from a sequence of significant locations that could all, in theory, be geographically close together. We break this distribution down according to time of week in the next section.

3.2.3 *Time of Week Predictability*

Plotting a heat map of the average instantaneous entropies according to local time of the week over all users yields Figure 2. This figure demonstrates the idea that there exist periods of high and low predictability that can last several hours. It also shows that trends about daily life can be uncovered with this type of analysis.

In particular, we can clearly see trends that match our intuitions about daily life. Weekends have the most intense levels of unpredictability, mostly in the afternoon. Weekdays show medium levels during normal working hours 8am to 5pm, and slightly higher levels in the evenings when users might go out to see friends.

3.2.4 *Mobile Phone Application Usage*

To expand our view of behaviour when the user moves into and out of habitual location patterns, also consider user behaviour with mobile applications, finding the probability of application use conditioned by the current instantaneous entropy. Figure 3 shows these probabilities aggregated over all users that have used the application at least once. This shows clearly that the probability of using almost all categories of application increases with the instantaneous entropy of the user. The first feature we notice about these probabilities is that web, map and search use show the greatest increase over normal use (with respect to other applications) in periods of highest entropy. For instance, a user in the highest state of unpredictability is approximately ten times as likely to use mapping applications than a user with a low instantaneous entropy. This supports the view that high entropy periods represent new experiences for the user, who will demand assistance about local information during these times. In contrast, clock usage appears to show a lower correlation with instantaneous entropy.

[3]Using the parameterisation given by Samorodnitsky [14].

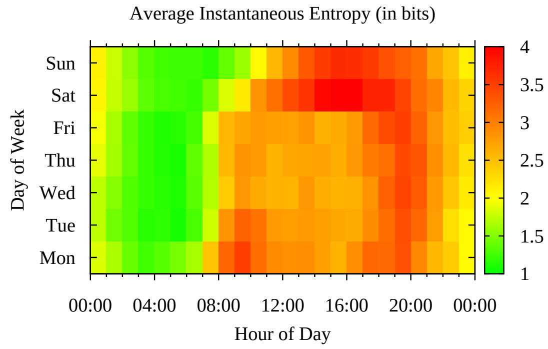


Figure 2: Average instantaneous entropy for all users by day of the week and hour of the day.

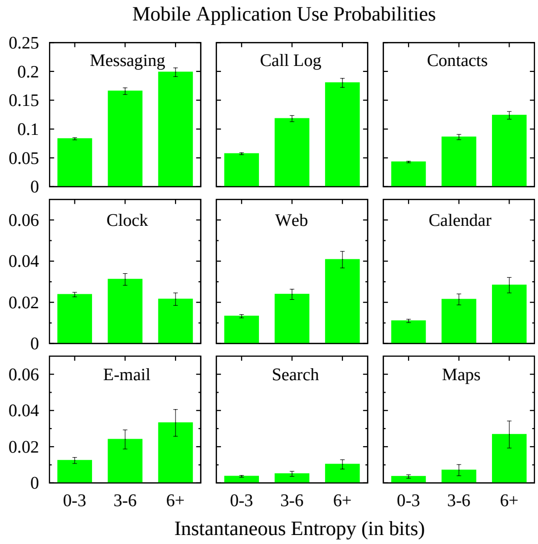
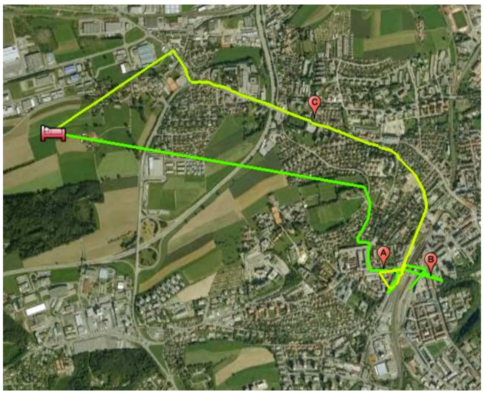


Figure 3: Probability of using various applications in the hour as different levels of instantaneous entropy are observed (shown with 95% confidence intervals).

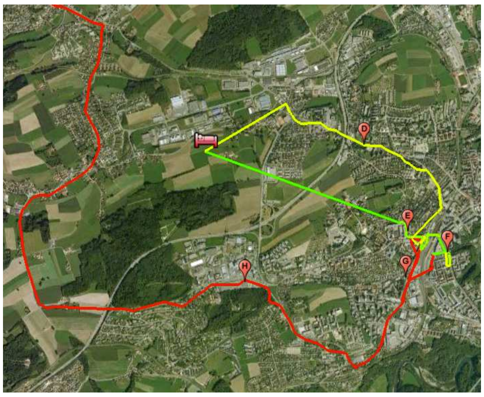
To put these results in context, we conclude this section with two snapshots of a single user's daily life, illustrating both a typical and a more unusual day.

3.2.5 *Typical Day vs. Unpredictable Day Examples*

Figure 4(a) shows a typical weekday for an individual working in Fribourg and living near Givisiez in Switzerland (the exact home location is obscured for privacy reasons). As on most working day mornings, this individual leaves her home, indicated by the bed symbol on the map, to arrive at work at around 7:45 (marked A on the map). At lunch-time,



(a) A typical day.



(b) An atypical day.

Figure 4: Location histories of a user during two days ("warmer" colours indicate higher entropy).

around 11:45, she leaves her work for a brief lunch (marked B). Finally, she finishes work at 17:40 and takes a different, but usual route back to her home (marked C). Throughout this day, the instantaneous entropy is low, indicated by the green colour, rising slightly on the way home, as her exact time of return is less predictable than her usual 7:45 arrival at work.

A much less typical day of the same individual is given in Figure 4(b), showing considerably more variance in instantaneous predictability. Interestingly, this variance is due not only to visiting new locations, but also to visiting familiar locations at unusual times.

Starting on a Thursday evening, she returns home from work at 18:00, a usual time (marked D). The next morning, she again arrives at work at 7:45 (marked E), and her instantaneous entropy remains relatively low throughout her lunch-break, which lasts from 12:00 to 13:00 (marked F). However, it begins to rise in the early evening, as she leaves work at an unusually early time of around 16:45 (marked G). Although she still remains in familiar locations at this point, moving around the city centre of Fribourg for some time, her predictability is low due to the unfamiliar time. Following this, she departs on an unusual route (marked H), which takes her outside the city and eventually to a golf course in nearby Courtepin (not shown on the map).

To give further illustration over longer periods, we have created two video animations of the location history of this user on a map covering a period of three months, where the colour of the path indicates her instantaneous entropy. The first video shows all locations visited by the individual, while the second shows a more detailed view of the user's typically visited locations (we omit the map layer of the second video for privacy reasons). These can be found at `http://research.nokia.com/mdc/`. The videos suggest that low predictability patterns can be detected by our method both regionally and locally, demonstrating the generalisability of our measure.

4. PRACTICAL APPLICATIONS OF INSTANTANEOUS ENTROPY

Now that we have a quantified way to talk about the extent to which users are currently departing from routine, we briefly examine the consequences for ubiquitous mobile applications (Section 4.1) and finish with a demonstration of the real world applicability of the new estimator (Section 4.2).

4.1 Novel Mobile Application Sketches

A wide range of popular existing mobile applications could be enhanced through the use of instantaneous entropy, as it offers an insight into broader user behaviours. An example of these types of applications are digital assistants on mobile platforms that provide artificial intelligence assistance to users: currently, there is *Siri* for the Apple iPhone, *Ask Ziggy* on the Nokia Lumia and *Evi* for Android platforms. These systems all perform natural language processing and service aggregation to give summarised information to users. We believe the next step is to proactively offer assistance, such as looking up directions, travel information, recommendations of high-quality local businesses, and streamlining message sending to contacts. The instantaneous entropy could be an important feature in deciding what level of active assistance to give. In cases of low mobility predictability, the user is likely to need these services more, as was highlighted in our results. The rest of the time, there may be no need to interrupt the user.

Another example is the set of mobile applications that send vouchers directly to users' mobile phones, timed precisely to help local businesses drive footfall during off-peak times, or simply to gain additional customers. The two biggest online voucher websites in the world, Groupon and LivingSocial, have both developed mobile services along these lines, named *Now* and *Instant*, respectively. The instantaneous entropy could allow the advertiser to select potential customers who are currently not in deep routine, and whose patterns are therefore easier to change, given an appropriate discount incentive for the user.

To put these sketches in context, we next consider instantaneous predictability as a feature in anticipating application use on mobile phones.

4.2 Predicting Map Application Use

We finish with a practical application of our proposed measure. Given the correlation found in Section 3.2.4 between the current instantaneous predictability of the user and their probability of using the map application, we consider the usefulness of the former as a feature in predicting the latter. Specifically, the task is to perform binary classification on the current time slot, indicating whether or not the user will use the map application at that time. This classification could be used for mobile user interface decisions, for example to decide whether to display an easier shortcut to the map application. It could also be used to proactively pre-cache a map of the user's current location, increasing apparent speed and responsiveness of the application from the user's perspective.

On the Nokia dataset, we test the intuition that a user in unfamiliar conditions (i.e., outside their normal location habits, as indicated by a high instantaneous entropy) is much more likely to seek out geographical information about the environment from their mobile phone. We model map use very simply, comparing the current instantaneous entropy $\tilde{H}_i$ (given by Equation 4) against a threshold parameter T to decide the class label C_i at time i:

$$C_i = \begin{cases} 1 & \text{if } \tilde{H}_i > T \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The T parameter can be varied to produce the receiver operating characteristic (ROC) curve shown in Figure 5. The rate of false positives (i.e., falsely predicted map use) aggregated over all users are plotted against the rate of false negatives (i.e., correctly predicted map use). For comparison, we also show a baseline curve of a classifier that uses only the current time to classify map use (substituting hour of the day for $\tilde{H}_i$).

The results show an improvement over the baseline when using instantaneous entropy as the feature. The area under the curve was 0.69, compared with 0.63 for the baseline[4]. The class balanced accuracy (taking into account the imbalance in class sizes) was 63%, compared with 54% for the baseline.

It is clear that the performance of the instantaneous entropy classifier is not uniform over the samples. The ROC curve has a steep gradient initially, and indicates that about 45% of map uses can be correctly anticipated with a false positive rate of only 10%. Performance flattens out after that. The interpretation is that 45% of map uses in the Nokia dataset are strongly connected to periods when the user is outside their normal habits (i.e., periods of high entropy), but the remaining 55% have a very limited connection to their location habits. Perhaps the user is planning further ahead in the latter cases, creating a disconnect between location behaviour and application use. It is a subject of future work to find the right features to complement the informativeness of location habits in predicting application use.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced a new metric for measuring an individual's momentary predictability: the instantaneous entropy. We demonstrated that this measure gives a deeper understanding about the state of mind of the user at any given time, and strongly influences the way they interact with their mobile devices. In future work, we plan to use this measure to build intelligent context-aware applications to offer effective in-the-moment user assistance.

6. REFERENCES

[1] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[4] N.B. all figures take into account the reverse in output for the baseline classifier when parameter $T > 16$.

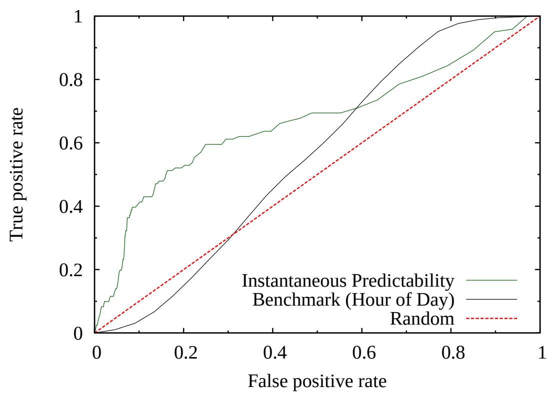


Figure 5: Receiver operating characteristics for the prediction of map application use for all users.

[2] H. Bapierre, G. Groh, and S. Theiner. A variable order markov model approach for mobility prediction. In *Workshop on Space, Time and Ambient Intelligence at the International Joint Conference on Artificial Intelligence*, Barcelona, Spain, 2011.

[3] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, Jan. 2006.

[4] M. Buchanan. The mathematical mirror to animal nature. *Nature*, 453(7196):714–716, 2008.

[5] Y. Chon, H. Shin, E. Talipov, and H. Cha. Evaluating mobility models for temporal prediction with high-granularity mobility data. In S. Giordano, M. Langheinrich, and A. Schmidt, editors, *PerCom*, pages 206–212. IEEE, 2012.

[6] N. Eagle and A. S. Pentland. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, Apr. 2009.

[7] Y. Gao, I. Kontoyiannis, and E. Bienenstock. Estimating the entropy of binary time series: Methodology, some theory and a simulation study. *Entropy*, 10(2):71–99, June 2008.

[8] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.

[9] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. In *Proc. WMASH 2004*, pages 110–118, 2004.

[10] I. Kontoyiannis, P. H. Algoet, Y. M. Suhov, and A. J. Wyner. Nonparametric entropy estimation for stationary processes and random fields, with applications to english text. *IEEE Transactions on Information Theory*, 44(3):1319–1327, May 1998.

[11] I. A. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics - Simulation and Computation*, 10(1):17–28, 1981.

[12] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Proc. Mobile Data Challenge by Nokia Workshop*, 2012. (to appear).

[13] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Life in the network: the coming age of computational social science. *Science*, 323(5915):721–723, Feb. 2009.

[14] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman and Hall/CRC, 2000.

[15] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. Campbell. Nextplace: A spatio-temporal prediction framework for pervasive systems. In K. Lyons, J. Hightower, and E. Huang, editors, *Pervasive Computing*, volume 6696 of *LNCS*, pages 152–169. Springer, 2011.

[16] C. Song, Z. Qu, N. Blumm, and A. Barabasi. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[17] J. Zhuang, T. Mei, S. C. Hoi, Y.-Q. Xu, and S. Li. When recommendation meets mobile. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pages 153–162, New York, NY, USA, 2011. ACM.