Breaking the habit: Measuring and predicting departures from routine in individual human mobility

James McInerney [a,*], Sebastian Stein [a], Alex Rogers [a], Nicholas R. Jennings [a,b]

[a] University of Southampton, Southampton, SO17 1BJ, UK

[b] Department of Computing and Information Technology, King Abdulaziz University, Saudi Arabia

ARTICLE INFO

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. To address this shortcoming, we tackle two interrelated challenges. First, we develop a novel information-theoretic metric, called instantaneous entropy, to *analyse* an individual's mobility patterns and identify temporary departures from routine. Second, to *predict* such departures in the future, we propose the first Bayesian framework that explicitly models breaks from routine, showing that it outperforms current state-of-the-art predictors.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Understanding human mobility patterns is a significant research endeavour that has recently received considerable attention [1,2]. 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 [3]. 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. [4]. Given the location traces of 50,000 mobile phone users, the authors used the Shannon *entropy rate* of this data to establish, in principle, that the average predictability of a single person's location in the next hour was 93% (at the granularity of the nearest cell tower, given their location in the previous hour).

This finding inspired a number of approaches that attempt to turn this promise of predictability into reality [2,5–8]. These approaches propose to do this by learning the "structure of routine" in human mobility [9] to make predictions about future individual mobility. The implicit assumption behind these approaches is that future location behaviour follows the pattern set by historical behaviour. While this assumption is, more often than not, accurate, it is arguable whether it describes the full range of human mobility patterns. 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., travelling to another city on a weekend break, or while on sick leave). Consequently, there are contexts in which historical patterns have more to say about future mobility than others. The work described in this paper is concerned with *analysing* and *predicting* departures from routine using such contexts.

In more detail, studying and characterising these momentary transitions in predictability is important, not only for understanding human mobility patterns, but also for providing context-aware services in pervasive systems. Arguably,

---

* Corresponding author.
E-mail addresses: jem1c10@ecs.soton.ac.uk, contact@jamesmc.com (J. McInerney).

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 in-the-moment assistance that can be provided by a mobile device [10], in the areas of information, organisation and communication. For example, map applications, navigation applications recommending high quality local businesses, and messaging applications are all likely to be in higher demand when the user is out of their comfort zone. Additionally, mobile advertising is likely to be more effective when not having to change long established user mobility habits. To enable such timely assistance, we first address the task of *analysing* the user's most recent location behaviour to identify departures from routine.

Beyond this analysis of the user's present state, it is similarly important to consider *future* departures from routine. Specifically, a number of pervasive applications already notify users of information relevant to their future locations (e.g., local recommendations by Foursquare or Yelp, location-based crowdsourcing apps like TaskRabbit or FieldAgent).[1] However, these rely on the assumption that the user will be available to follow up on those notifications. As such services proliferate, an understanding of user availability over time is crucial to avoid notification overflow and to maximise impact. To this end, we also investigate the task of *predicting* future transitions to and from routine location behaviour.

Given its level of importance, surprisingly little is currently known about how the depth of routine in an individual's mobility varies with time. Existing methods typically focus on analysing routine by grouping the location history into times of the week (i.e, *time slots* of fixed duration) and measuring the fraction of instances an individual is found in their most common location per hour of the week [4], or by measuring entropy within each time slot [9]. These time slots are treated independently, and entirely ignore spatial structure (i.e., the history of recently visited locations). This gives an inaccurate picture of the current depth of routine for individuals. For example, a user transition from work to a restaurant in the centre of town might be surprising spatially (if such a transition has never been observed before), but make more sense temporally if we consider that it is Saturday night, and the user had only been in the office to meet a deadline. Extending this example, if that user then takes the train home, this might be considered a break from temporal routine (if they normally take the motorway by car on weekends), but make sense spatially when coming from town. Therefore, both types of structure are important. We address this shortcoming by providing an analysis of mobility that takes into account sequential structure.

More crucially, the current understanding of the predictability of human behaviour relies on entropy estimators, which tend to be model-free. The advantage of model-free tools is that they are applicable to new domains, and import well-established concepts (such as entropy) without requiring domain-specific modifications. However, such tools fall short when we try to go beyond analysis (of historical behaviour) and predict future levels of routine. Specifically, the incorporation of relevant information (e.g., recent locations, calendar information, whether the day is a national holiday), and domain-specific structure of the random variables are impossible without providing a model. We therefore present the first model of mobility that explicitly captures the tendency to depart from routine under a variety of temporal and situational contexts. This addresses prediction.

In more detail, this work is the first to fully 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 [11], called the *real-time entropy estimator* that provides a principled method for measuring the instantaneous predictability, or entropy, of an individual. Existing approaches use the standard Lempel–Ziv estimator [4,11,12], which gives a single summary entropy value to individuals. In contrast, our method provides a breakdown of entropy per time slot (e.g., hour, quarter of hour), which enables richer analyses.
- We apply our estimator to GPS traces from the Nokia Lausanne dataset [13], and show, for the first time, that departures from routine are correlated with mobile application use, and that applications that provide information about local surroundings (i.e., search and maps) show the strongest correlation.
- We present a new Bayesian model capable of predicting future departures from routine. This is the first mobility model that explicitly captures the tendency of users to depart from routine.
- We evaluate our estimator and model against state-of-the-art benchmarks (including an entropy measure [9] and a regularity measure [4], both of which look only at the time of week to measure depths of routine). We show that they outperform all existing approaches for inference and prediction of departures from routine.
- As a demonstration of the advantage of our approach, we present an anticipatory computing scenario in which predictions of unusual future location behaviours are used to optimise mobile push notifications to a user. Comparing our predictive model against a state-of-the-art model of temporal behaviour [14], we find that outcome utilities are significantly improved when the tendency to depart from routine is also modelled.

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 article, we describe related work in Section 2. Then, we introduce two new methods for understanding departures from routine in current and future location behaviour, respectively, in Section 3. In Section 4

---

[1] Available at www.foursquare.com, www.yelp.com, www.taskrabbit.com, and www.fieldagent.net, respectively.

we apply our methods, along with several state-of-the-art benchmarks, to the breadth of the Nokia Lausanne dataset. We evaluate both the analytical and predictive performance of all the methods introduced. Building on this evaluation, we highlight the applications of our work and present a demonstration application in Section 5. Finally, we conclude in Section 6.

2. Related work

To put our contributions into context, we now consider several threads of related work. We start with an overview existing approaches to understanding routine in individual human mobility, then look at existing ways to incorporate the user context when deciding to send mobile notifications in pervasive systems.

2.1. Understanding routine in mobility

There are several approaches relevant to analysing and predicting the depth of routine in user behaviour. Song et al. [4] focused only on summarising each individual's location behaviour as a whole. Eagle and Pentland [9] did analyse predictability in greater detail (by hour of the week) using temporal histograms, but this ignores sequential structure (i.e., the recent location history of the person). We, on the other hand, incorporate the recent history of locations when analysing predictability, which we show enables a richer historical analysis.

Since these influential works were released, many approaches to predicting individual *location* behaviour (as opposed to levels of routine) have been proposed. Such prediction is done through the use of machine learning to find the structure in individuals' location histories. The structure can either be spatial [7,15], temporal [2,16], or a combination of both [6,9]. An exception to this trend is provided by Sadilek and Kautz [17], who use correlations between the movements of users and their friends to infer the current unobserved location of users. Though they claim this as "prediction", it is not prediction in the sense that we mean here, because we are interested in knowing the *future* whereabouts of individuals, not just inferring their current (albeit, missing) location.

While the particular type of structure is usually determined by the application, models capturing both spatial and temporal structure tend to perform best (as opposed to either one alone [18]). We will therefore use such an approach to model the routine mobility of individuals. However, all the aforementioned models assume that people stick to their historical routine. In contrast, our approach explicitly incorporates the possibility that users will temporarily depart from their habits. As we shall show, this pays dividends in the analysis and prediction of routine.

2.2. Decision-making for mobile notifications

As a demonstration of the usefulness of modelling departures from routine, we briefly consider a ubiquitous application in Section 5.2 that notifies a user with information about future locations. To decide whether to present a notification to the user, existing work attempts to calculate either the cost or the value of notifications. The evaluation of cost is usually framed as the problem of inferring the *interruptibility* of a user (and assumes fixed value, or benefit, for each notification). For example, Iqbal and Bailey [19] and Horvitz et al. [20] evaluate the interruption cost based on the user's current activity. When the cost is high, the mobile device should refrain from (or delay) alerting the user about phone calls, reminders, or more application-specific events. Alternatively, the value of the information in the notification, in light of the user's current and future locations, can be used. For example, Amini et al. [21] predict user destinations and provide search results around the destination rather than the current location. Finally, Reddy et al. [22] use the historic mobility of crowdsourcing participants to decide whether to recruit individuals to environmental monitoring campaigns. Our scenario requires the consideration of timing as part of the value of information, and is the first to use an understanding of the tendency of the individual to depart from routine to improve the estimation of optimal timing.

3. Analysing and predicting routine

Against this background of related work, we now present our two new approaches. The first presented in Section 3.1 and is designed to perform analysis of routine (i.e., predictability of mobility). Like similar work before it, it is rooted in the framework of information theory because we can import tools (specifically, entropy rate estimators) that can help make sense of routine without much customisation.

The entropy rate estimator is simple, requires few assumptions (detailed below), and is tractable for running in real-time on mobile devices. We anticipate, however, limitations to this approach related to the fact that the entropy rate considers only sequential information in mobility and is unable to predict future departures from routine. This motivates the introduction of a more general framework of departures from routine in Section 3.2 that can include arbitrary types of structure in mobility data (e.g., temporal structure, social structure) and can yield predictive densities over future breaks from habit.

3.1. Information theory

To formalise predictability (i.e., routine) 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 $H = -\sum_i p(x_i) \log_2 p(x_i)$, where $p(x_i)$ is the probability of $x_i = X_s$, represents the entropy of the group representing the periodic time slot s (e.g., 2 pm Thursdays, 9:00–9:30 am Mondays), disregarding information related to all other time slots. However, we want to take account of conditional dependencies between time steps, i.e., that knowing a history of locations typically 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 [23].

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:

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

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 $(x_1, x_2, \ldots, x_{N-1})$:

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

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[2] [23,24]. The LZ estimator is closely related to LZ compression and prediction [25], however, it used for the arguably simpler task of quantifying the information present in a sequence.

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

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

where Λ_i is defined as the length of the shortest substring starting at position i that did not previously occur in the sequence $(x_1, \ldots, x_{i-1})$. It has been shown that the increasing window LZ estimate rapidly converges to the true entropy rate of the underlying process [24]. The estimator given by Eq. (3) assigns 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 $\log_2 |A|$ would be moving to a new location chosen uniformly from A at random. There is strong evidence that all people (in a study involving 50,000 randomly selected people) fall along a spectrum of entropy much closer to the lower extreme than the higher one [4]. However, this measure does not tell us *when* any individual is behaving unpredictably, limiting our analysis.

To overcome these limitations, we use the entropy rate as a principled way to quantify departures from routine. Reconsidering the entropy estimator given by Eq. (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., $\forall j : 1 \leq j \leq N, \Lambda_j = \Lambda_i$). This concept does not contradict 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:

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

---

[2] Other estimators, such as context tree weighting methods or plug-in estimators could be used here [24]. This choice is not fundamental to our general method.

Table 1
Summary table of symbols.

Symbol	Description
r	Transition matrix for latent states
z_n	Routine at time n
d_n	Context at time n (e.g., time, calendar)
μ	Parameters for spatial context
ω	Parameters for temporal context
x_n	Location at time n
α	μ's hyperparameter
β	ω's hyperparameter
δ	r's hyperparameter
L	Number of locations

where Γ_i is defined as the length of the shortest substring *ending* at position i that did not previously occur in the 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):

$$\hat{H}_N = \frac{N}{\sum_{i=2}^{N} \tilde{H}_i^{-1}}. \tag{5}$$

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 Eq. (2).

In Section 4, we will apply this estimator to real daily life location data to understand more about the unpredictability of mobility. The estimator provides good insight into the user's behaviour, while still being computationally tractable (even on a resource-limited mobile device).

However, this analytical technique is limited in two important respects. Firstly, it considers only sequential structure, when in fact, as we established in Section 1, routine is a multi-faceted phenomenon. For example, time of the day, weather, locations of friends, or national events could all be strong predictors of a user's whereabouts, regardless of where they have recently been. Secondly, even if the different varieties of structure could be fused in the entropy estimator technique, it would not constitute a predictive model of future departures from routine. This rules out many interesting applications, one of which we consider experimentally in Section 5.2. We therefore next provide a general framework to explicitly model an individual's departures from their location routine.

3.2. Bayesian model

The principled generative model we propose is based on existing models of mobility but is augmented, for the first time, with latent states indicating departures from routine. At a high level, the model has a set of parameters governing the structure of an individual's routine (μ and ω), another set of parameters governing how the individual departs from routine (r), and a set of latent variables indicating the level of departure associated with each location observation (z). These values are unknown beforehand. Therefore, the key challenge is to define how the hidden parameters and latent variables interact with each other, and to find a tractable inference procedure from the observed data. All the symbols we use in this model are summarised in Table 1. We adopt the notation v to indicate collections of random variables, and $v_{i,j}$ to indicate specific entries (i.e., row i, column j) in these collections.

We now detail the derivation of this model from a set of basic assumptions about routine. In our model, we explicitly incorporate the possibility that the user can choose, at any time, to depart from their historical habits. This aspect of mobility behaviour is represented as a set of binary random states z (one for each discrete time step n). We later specify the exact nature of these variables (i.e., their probability distributions), but until then, it is sufficient to simply know that they are binary representations of the categorical state of adherence (when $z_{n,0} = 1$) or departure (when $z_{n,1} = 1$) from routine at each time step n. z is incorporated into a more conventional mobility model that captures the spatial and temporal structure of user mobility that resembles the model of Etter et al. [26].

In the more conventional aspect of the location model, the sequential (i.e., spatial) aspect of mobility is achieved by placing dependencies between location observations. Like the other models mentioned above, to avoid the exponential creation of parameters, we assume the first-order Markov property,[3] stating that the probability distribution of the next location is conditionally independent of the entire location history given the current location, i.e., $p(x_n|x_{n-1}, x_{n-2}, \ldots, x_1) = p(x_n|x_{n-1})$ where x_n is a binary vector with a single 1 at index l indicating that the user was at location l at time step n, and

---

[3] Though more elaborate forms of sequential structure are certainly possible to incorporate without requiring changes to other parts of the model, e.g., a variable-order Markov chain or a nested Pitman–Yor process [7].

$p(\boldsymbol{x_n})$ is the probability of observation $\boldsymbol{x_n}$. The sequential structure of the user's mobility can therefore be represented by a transition matrix, where the row l_1 (if l_1 is the previous location) gives the probability distribution over the next location l_2. A common way to achieve this distribution is to place a multinomial distribution in each row, with Dirichlet priors [27]:

$$p(\boldsymbol{\mu}|\boldsymbol{x_n}, \boldsymbol{x_{n-1}}, \alpha) \propto \prod_{l_1}^{L} \prod_{l_2}^{L} \mu_{l_1,l_2}^{x_{n-1,l_1} x_{n,l_2} + \alpha - 1} \tag{6}$$

where $\boldsymbol{\mu}$ is the set of parameters to the Dirichlet distribution (with hyperparameter α), and $x_{n-1,l_1} = 1$ indicates that the previous location at time $n-1$ was l_1.

The temporal aspect of mobility is achieved by assuming that the temporal information associated with each location observation at step n depends only on user's current location. Since the temporal aspect of presence is a mutually exclusive 1-of-V choice (where V is the number of time slots in a period, e.g., 168 h in a week), this kind of structure can also be represented as a Dirichlet distribution:

$$p(\boldsymbol{\omega}|\boldsymbol{x_n}, \boldsymbol{d_n}, \beta) \propto \prod_{l}^{L} \prod_{v}^{V} \omega_{l,v}^{d_{n,v} x_{n,l} + \beta - 1} \tag{7}$$

where $\boldsymbol{d_n}$ is the time context information (but could equally represent any other context of interest, such as weather or calendar information), and β is the hyperparameter to the Dirichlet distribution.

Combining spatial and temporal structure is straightforward, if we assume the conditional independence $p(\boldsymbol{x_n}, \boldsymbol{d_n}|\boldsymbol{x_{n-1}}, \boldsymbol{\mu}, \boldsymbol{\omega}, \alpha, \beta) = p(\boldsymbol{x_n}|\boldsymbol{x_{n-1}}, \boldsymbol{\mu}, \alpha)p(\boldsymbol{d_n}|\boldsymbol{x_n}, \boldsymbol{\omega}, \beta)$. This is also an assumption that was made by Gao et al. [18].

We now describe the derivation of the novel aspect of our model, which is the explicit incorporation of departures from routine. We make one key assumption that governs the interaction between latent states $\boldsymbol{z}$ and the other random variables in the model: that when $z_{n,0} = 1$, the user acts completely according to routine (i.e., the model becomes a standard mobility model), but when $z_{n,1} = 1$, the user chooses a location uniformly at random (N.B., $\sum_k z_{n,k} = 1$). The reasoning behind this assumption is that a user who regularly breaks from habit leaves behind structure that can be learnt by a standard mobility model. But since these breaks from routine can be predicted by a conventional model, then, by contradiction, we know that such behaviour forms part of the routine of the user. Therefore, a uniform distribution over locations accurately reflects the departures from routine that we are interested in.

More formally, the spatial structure in light of the current level of routine, $\boldsymbol{z_n}$, is modified to reflect the mixture of outcomes for different levels of routine. By inspection and considering Eq. (6) (while keeping the prior Dirichlet distribution the same):

$$p(\boldsymbol{\mu}|\boldsymbol{z_n}, \boldsymbol{x_n}, \boldsymbol{x_{n-1}}, \alpha) \propto p(\boldsymbol{x_n}|z_{n,0} = 1, \boldsymbol{x_{n-1}}, \boldsymbol{\mu})^{z_{n,0}} p(\boldsymbol{x_n}|z_{n,1} = 1)^{z_{n,1}} p(\boldsymbol{\mu}|\alpha) \tag{8}$$

where Eq. (8) is proportional to the posterior of $\boldsymbol{\mu}$ by use of Bayes' rule. Substituting Eq. (6) into Eq. (8), and using the aforementioned key assumption about a uniform distribution when $z_{n,1} = 1$:

$$\propto \left(\prod_{l_1}^{L} \prod_{l_2}^{L} \mu_{l_1,l_2}^{z_{n,0} x_{n-1,l_1} x_{n,l_2} + \alpha - 1} \right) \left(\prod_{l}^{L} \frac{1}{L^{z_{n,1} x_{n,l}}} \right) \tag{9}$$

$$= \frac{1}{L^{z_{n,1}}} \prod_{l_1}^{L} \prod_{l_2}^{L} \mu_{l_1,l_2}^{z_{n,0} x_{n-1,l_1} x_{n,l_2} + \alpha - 1}. \tag{10}$$

Incorporating $\boldsymbol{z_n}$ into the *temporal* structure requires a slightly different approach, because we also want to discover temporal structure in departures from routine. This feature will help us make new kinds of predictions about future behaviour (e.g., by answering questions such as "is the user going to be in deep routine next Tuesday at 2 pm?"). Therefore, instead of placing a uniform distribution over $\boldsymbol{d_n}$ when $z_{n,1} = 1$ (as we did with $\boldsymbol{x_n}$), we instead give departures their own temporal distribution (i.e., a row in the $\boldsymbol{\omega}$ parameter matrix) that is independent of the current location $\boldsymbol{x_n}$. However, when $z_{n,0} = 1$, the observation $\boldsymbol{x_n}$ depends on the temporal observation as before:

$$p(\boldsymbol{\omega}|\boldsymbol{z_n}, \boldsymbol{d_n}, \boldsymbol{x_n}, \beta) \propto p(\boldsymbol{\omega}|\beta) \prod_{k}^{2} p(\boldsymbol{d_n}|z_{n,k} = 1, \boldsymbol{x_n}, \boldsymbol{\omega})^{z_{n,k}} \tag{11}$$

$$= \left(\prod_{l}^{L} \prod_{v}^{V} \omega_{l,v,0}^{z_{n,0} d_{n,v} x_{n,l} + \beta - 1} \right) \left(\prod_{v}^{V} \omega_{0,v,1}^{z_{n,1} d_{n,v} + \beta - 1} \right). \tag{12}$$

To be clear, $\omega_{l,v,0}$ refers to the routine temporal parameters (for arbitrary location l and time v), while $\omega_{0,v,1}$ refers to the temporal parameters of routine departures (for arbitrary time v, ignoring the location).
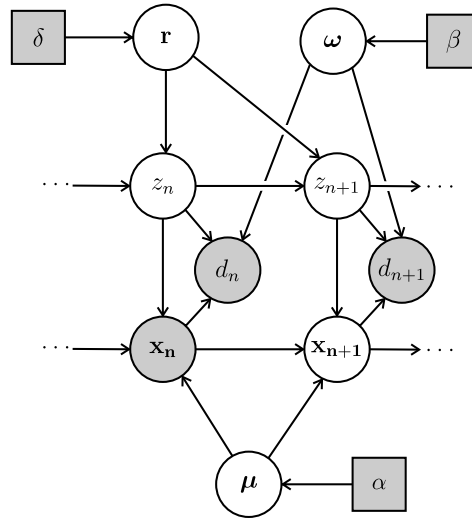
Fig. 1. A graphical representation of our model. Shaded nodes represent observed random variables and boxes represent fixed values. $\mathbf{x_{n+1}}$ is unshaded because it is assumed to be at future time step $n + 1$.

We finally consider the prior distributions responsible for the binary latent random variables $\mathbf{z}$. We assume the probability of departure at time step n is dependent on whether the user departed at the previous time step $n - 1$, resulting in a standard hidden Markov structure with 2×2 transition matrix $\mathbf{r}$:

$$p(\mathbf{r}|\mathbf{z_n}, \mathbf{z_{n-1}}, \delta) = \prod_{k_1}^{2} \prod_{k_2}^{2} r_{k_1,k_2}^{z_{n-1,k_1}z_{n,k_2}+\delta-1} \tag{13}$$

where δ is the hyperparameter to the each row of transition matrix.

This set of assumptions results in the graphical representation depicted in Fig. 1. Before we can make use of this model, we need to infer all the parameters and latent states from some historical (training) data $\mathbf{X}$ and $\mathbf{D}$. It is not possible to do this analytically, because the posterior distribution involves an intractable integral [27]. There are a variety of methods that can estimate the parameters and latent variables either by optimisation (e.g., variational Bayes, expectation–maximisation), or sampling (e.g, Gibbs sampling, or other Markov chain Monte Carlo methods). Here we choose expectation–maximisation (*EM*), specifically the *forward–backward* algorithm, because it is a simple and widely-used approach to solve this problem, but the other methods could equally be used [27].

The expectation step requires finding the expectation over each latent state $\mathbf{z_n}$. The most efficient way to find this is to perform a forwards and a backwards pass through the time steps (finding the values of $\mathbf{f_n}$ and $\mathbf{b_n}$ for all n, respectively), as is standard in the forward–backward algorithm. For our model, the equations for both can be obtained through marginalisation and the use of Bayes' rule:

$$\mathbf{f_n} = p(\mathbf{z_n}|\mathbf{X_{0..n}}, \mathbf{D_{0..n}}, \boldsymbol{\mu}, \boldsymbol{\omega}, \mathbf{r}, \alpha, \beta, \delta)$$
$$\propto \sum_{k}^{2} p(\mathbf{x_n}|\mathbf{z_n}, \mathbf{x_{n-1}}, \boldsymbol{\mu}, \alpha)p(\mathbf{d_n}|\mathbf{z_n}, \mathbf{x_n}, \boldsymbol{\omega}, \beta)p(\mathbf{z_n}|\mathbf{z_{n-1}}, \mathbf{r}, \delta)f_{n-1,k}. \tag{14}$$

Substituting Eqs. (10) and (12) into Eq. (14), while assuming the parameters $\boldsymbol{\mu}$, $\boldsymbol{\omega}$, and $\mathbf{r}$ are known (obtained from the previous iteration of the EM algorithm):

$$f_{n,0} \propto \left(\prod_{l_1}^{L} \prod_{l_2}^{L} \mu_{l_1,l_2}^{x_{n-1,l_1}x_{n,l_2}} \right) \left(\prod_{l}^{L} \prod_{v}^{V} \omega_{l,v,0}^{d_{n,v}x_{n,l}} \right) \sum_{k}^{2} r_{k,0}f_{n-1,k} \tag{15}$$

$$f_{n,1} \propto \left(\prod_{v}^{V} \omega_{0,v,1}^{d_{n,v}} \right) \sum_{k}^{2} r_{k,1}f_{n-1,k}. \tag{16}$$

The backwards pass can be calculated similarly to Eqs. (15) and (16), but instead must marginalise over the *next* latent state $\sum_{k}^{2} r_{0,k}b_{n+1,k}$, and proceed backwards from the final time step. To complete the calculation of the expectation of $\mathbf{z}$, we combine f and b in the normal way, by element-wise multiplication and normalisation.

The maximisation step finds the modes of the posterior distributions of the parameters. This can be obtained via differentiation (with a Lagrange multiplier to constrain the time context parameters ω and the rows of the transition matrix μ to sum to 1) to obtain the three closed-form solutions corresponding to the maximum *a posteriori* (MAP):

$$p(\mu_{a,b}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{z}, \alpha)_{\max} \propto \left(\sum_{n=1}^{N} z_{n,0} x_{n-1,a} x_{n,b}\right) + \alpha - 1 \tag{17}$$

$$p(\omega_{l,v,0}|\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{z}, \beta)_{\max} \propto \left(\sum_{n=1}^{N} z_{n,0} d_{n,v} x_{n,l}\right) + \beta - 1 \tag{18}$$

$$p(\omega_{0,v,1}|\boldsymbol{D}, \boldsymbol{z}, \beta)_{\max} \propto \left(\sum_{n=1}^{N} z_{n,1} d_{n,v}\right) + \beta - 1 \tag{19}$$

and $\boldsymbol{r_{\max}}$ can be obtained by the standard MAP for a transition matrix [27]. Once the parameters have been inferred, we can use them for prediction of future values of $\boldsymbol{z_{n+1}}$. Prediction requires contextual information, which we assume here to be the time contexts $\boldsymbol{d_{n+1}}$. We also have $\boldsymbol{D_{0..n}}$ and $\boldsymbol{X_{0..n}}$, the set of observations up to time step n. The expected value of $z_{n+1,1}$ for the next context is found via the marginalisation $\mathbb{E}(z_{n+1,1}|\boldsymbol{D_{0..n+1}}, \boldsymbol{X_{0..n}}) \propto \sum_k^2 p(\boldsymbol{x_{n+1}}|\boldsymbol{z_{n+1}}, \boldsymbol{x_n}, \mu, \alpha)p(\boldsymbol{d_{n+1}}|\boldsymbol{z_{n+1}}, \boldsymbol{x_{n+1}}, \omega, \beta)r_{k,1}f_{n,k}$.

We now apply our model (and the LZ instantaneous entropy estimator) to the Nokia dataset. Our model requires setting the three hyperparameters α, β, and δ. They can be used to encode prior knowledge (e.g., that someone is likely to stay in their current location, or that a certain time of day is more likely for a given location). However, in our experiments we assign uninformative uniform values to these hyperparameters (i.e., setting them all to 1) and let the model discover such patterns.

4. Real-life data analysis

To evaluate the performance of the two approaches we presented in Section 3, we test them on the Nokia Lausanne dataset, comprising global positioning system (GPS) locations, call logs and application usage for 38 people for a year recorded by their mobile phones [13]. The data consists of a series of time-stamped events (e.g., GPS readings with latitude and longitude coordinates, user usage of mobile applications, and directed message and call logs).

First, it is necessary to pre-process the data, in order to convert it to a form suitable for discrete methods. We briefly describe this process in Section 4.1. We then detail our methodology for evaluating the approaches in Section 4.2, and describe the state-of-the-art benchmarks we consider in Section 4.3. In Section 4.4, we present an evaluation of the effectiveness of all the approaches for analysis. Finally, in Section 4.5, we evaluate the predictive ability of the approaches.

4.1. Pre-processing

Pre-processing the dataset was required because all the methods we evaluate (including the benchmarks) use discrete labels indicating user presence at a location. Pre-processing also helps filter out noise from sensor error. To do this, we first derived the alphabet of locations, A, from the sequence of latitude and longitude GPS readings (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 [15,28]. We selected the online clustering method proposed by Kang et al. [28], 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 the closest element (by Euclidean distance) in A.[4] We added a special element $\Omega \in A$ for readings that were not near any significant locations (defined as being more than 1 km away from any element in A). In the entropy estimation, Ω is a special location that is always treated as a new location. Finally, as per [4], we transformed the data into windows of fixed size by selecting randomly from the set of significant locations visited during the window in proportion to the total duration at each location (e.g., in the case of 1 h windows, if the user is at work for 45 min of the window and at a cafe the remaining 15 min, the location for that hour will be selected as work or as cafe with probability $\frac{3}{4}$ and $\frac{1}{4}$, respectively). We now describe how we used this processed data to evaluate the performance of the approaches.

4.2. Experimental setup

To evaluate performance, it is necessary to first establish the ground truth. The ground truth in our problem of analysing and predicting how much an individual's mobility departs from routine should consist of a set of values: one for each time step indicating the level of routine the individual is currently in. This can be ascertained in two ways. One approach would

---

[4] The significant location extraction method we used ensured that locations were sufficiently far apart to do this unambiguously.
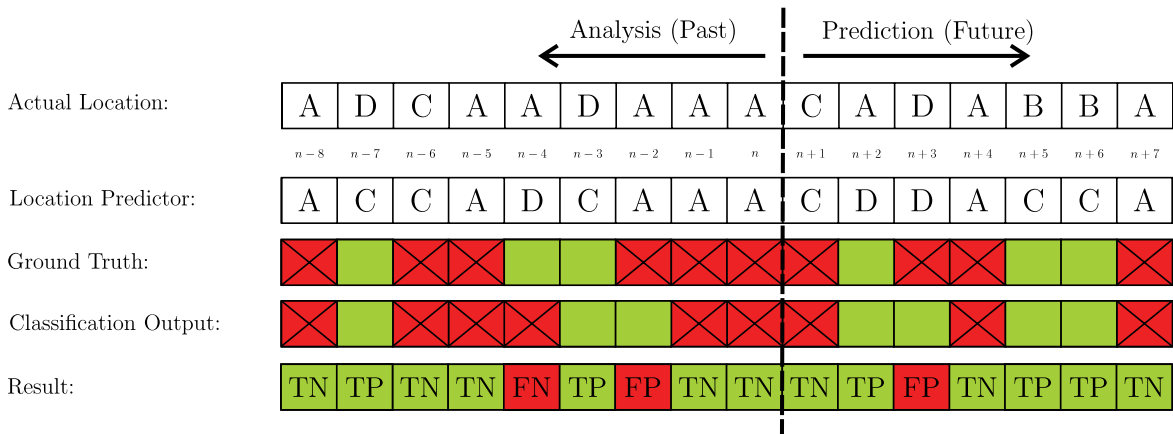
Fig. 2. Experimental setup showing how our results were derived from user locations. In the bottom row, TN, TP, FN, and FP represent true negatives, true positives, false negatives and false positives, respectively.

be to take detailed user surveys (perhaps obtained through experience sampling [29]). However, user surveys can often contain biases and inaccuracies [30] and they are time consuming for users to provide. Furthermore, in this case, we did not have access to the original users, and so were unable to collect such data. An alternative way is to make use of existing location predictors. Since the first aim of most location predictors is to learn the structure of habit in individual mobility, it is instructive to examine their failures. When a model predicts that an individual will be in a certain location, but gets that prediction wrong, it indicates one of two things. Either the individual departed from their routine, or the failure was due to idiosyncrasies in the predictor itself. To minimise idiosyncratic error in the ground truth, we combine the output of two state-of-the-art location predictors that use fundamentally different methods to arrive at their predictive densities.

In more detail, the location predictors we use to generate the ground truth of routine are: a variable-order Markov model (i.e., capturing a maximum of third-order dependencies, with fallback that considers lower-order dependencies according to the transitions that were seen during training) that learns only sequential structure [31], and a spatio-temporal mixture model that learns periodicities in human behaviour [14]. We extended the latter approach to include day of the week observations (in addition to time of day observations), and to automatically infer the number of clusters (significant places) by replacing the finite mixture prior with a Dirichlet process [32].

The experimental setup using this ground truth is depicted in Fig. 2. The top row shows the observed location data of an individual. The second row shows the output of a single location predictor. In this simplified illustration involving only one location predictor, the ground truth (in the third row), indicating whether the user departs from habit at each time step, is derived by comparing the output of the predictor with the actual location (in our experiments, we used majority voting). In the fourth row we show the output of the *routine approach*, which aims to classify the behaviour of the user as either being in routine (giving negative output, or red, in the diagram), or departing from it (giving positive output, or green, in the diagram). The difference between analysis and prediction is simply due to a difference in uncertainty about the user's location. Analysis deals with past behaviour, with known locations. Prediction deals with future behaviour (and location uncertainty).

Using this experimental setup, we evaluate the performance of our approaches. To put this performance in context, we also used state-of-the-art benchmarks for comparison. We next describe these benchmarks.

4.3. Benchmark approaches

The three benchmarks we use in our evaluation are:

1. *Entropy by time of week.* Eagle and Pentland analyse routine by calculating the historical entropy per time slot of the week [9]. This is done by grouping historical location observations by their time slot of the week, and calculating the Shannon entropy of each group. Analysis or prediction may be done using the d_n context. Hence, it is a purely temporal approach.

2. *Location surprise by time of week.* Song et al. use the standard increasing-window LZ estimator for most of their analyses (including the headline predictability level of 93%). This assigns a single value to an individual's entire location history, so does *not* provide the level of detail we need. However, they also provide a way of analysing location surprise for each time step. This is calculated by grouping the observations into hours of the week (like the measure from Eagle and Pentland), and finding the proportion between the most frequent location and all other observations for that group.

3. *Location likelihood.* Finally, we use a benchmark based on the likelihood of the currently observed location, with respect to all other locations previously observed at the same time of the day and week. Since higher likelihood observations correspond to being in routine, we take the complementary (i.e., $1 - v$, where v is the location likelihood) of this value as the output.

(a) 500 m radius around significant locations.　　(b) 100 m radius around significant locations.
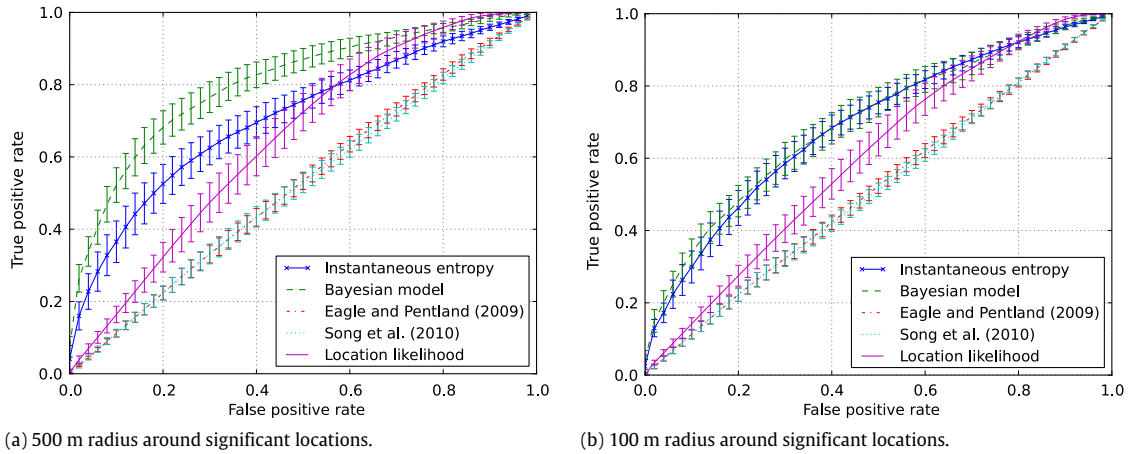
Fig. 3. ROC curves for analysis of routine with all the approaches. Error bars indicate the 95% confidence range.

We test these benchmarks, and our approaches, against the ground truth under two different scenarios. In the first scenario, the methods have access to the location of the user at time step n for classifications at time step n. This tests performance in analysis *after the fact*, which is useful for applications that require an understanding of recent levels of routine (see Section 5 for more details on applications). In the second scenario, the methods do not have access to the location at time step n. This tests predictive performance, which is useful for applications that need to know future levels of routine (again, see Section 5 for associated applications). We now describe experiments done under the first scenario.

4.4. Analysing departures from routine

Since the level of routine is a hidden state that is not present in the data, we require an approach that can provide us with a measure of routine from the location data alone. In Section 4.4.1 we evaluate all the approaches presented under this scenario. Based on these findings, we select one method (the LZ estimator) to analyse the Nokia dataset more broadly in Section 4.4.2.

4.4.1. Inference performance

In Fig. 3(a), we present the receiver operating characteristic (ROC) curves for all 5 approaches. Each ROC curve involves comparing output signal s (in this case, the output of each of the 5 approaches) against an arbitrary threshold value t to give a classification, where $s > t$ indicates a positive classification and $s \le t$ indicates a negative classification. Each classification is compared against the ground truth to determine the relationship between the rate of true positives (i.e., correctly classified departures from routine) and the rate of false positives (i.e., incorrect classifications). This comparison is repeated across the entire range of t, defined as going from 100% negative to 100% positive classifications. The results are averaged across all 38 users in the dataset.

We see that our LZ instantaneous estimator and Bayesian model perform the best out of all approaches, followed by the location likelihood benchmark. The other benchmarks (by Eagle and Pentland and Song et al.) do no better than random at analysing departures from routine. It is instructive to consider why this is the case. These benchmarks group all location behaviour (for an individual) into 168 bins, where each bin represents behaviour at a different hour of the week. This is useful for summarising behavioural trends (e.g., show high departures from routine during Friday lunchtimes, when perhaps the user likes to go to a different restaurant each time), but ill-suited for describing routine at any specific time. In this way, using these summarisation tools at the level of detail we require (i.e., specific hours of an observed location history), it is expecting too much of them. In fact, providing this level of detail was not the intention of the original authors who proposed them.

We investigated the effect on performance of varying the location pre-processing parameters. Specifically, a distance threshold parameter is required that indicates the minimum distance between places (i.e., clusters below this distance are considered one place), and a time threshold parameter is required that indicates how long an individual has to spend at a location in total before it is considered a discrete place. Experimentation showed that the same broad set of findings hold, except in the case of very high spatial granularity, which we consider in more detail now.

Fig. 3(a) and (b) show the ROC performance when the distance threshold is set to 500 m and 100 m, respectively. We see that the performance of the Bayesian model is diminished by the second higher granularity, while the other approaches have a similar performance in both cases. We attribute this to the first-order Markov assumption (for locations) that only our model follows. Visual inspection of spatial plots of pre-processing confirms that reducing the threshold to 100 m (i.e., the minimum distance between significant locations, corresponding to a 50 m radius around each location) introduces many fragmented and spurious significant places that appear to refer to the same location (e.g., work or transport hub). We believe that this is caused by the influence of GPS sensor noise, which should normally be filtered out by pre-processing. This makes
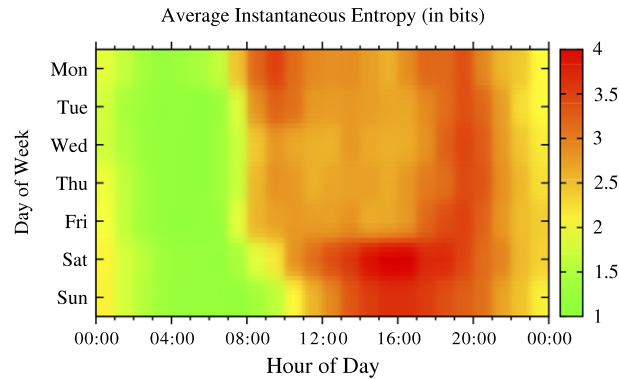
Fig. 4. Average instantaneous entropy for all users by day of the week and hour.

it harder for the transition matrix to discover structure in human behaviour. To overcome this, our general framework makes it straightforward to consider more sophisticated forms of sequential structure in the locations. In addition, dealing with sensor noise is a standard problem in probabilistic models, so if very high granularities are required (though we do not believe there is a strong case for this setting), a future extension of our model could place continuous distributions over space (in a similar way to [14]) to account for this.

We therefore conclude that our LZ estimator and Bayesian model are the best tools for analysing departures from routine in user behaviour. Since the LZ estimator considers departures from sequential structure only, we can see in Fig. 3 the benefit of additionally considering temporal structure by comparing its performance to that of the general framework (i.e., the Bayesian model that considers both sequential and temporal structure). However, the price of generality and improved performance is an increase in computational complexity. Specifically, the estimator only needs to match a sequence of locations in a tree (overall worst-case complexity of $O(N \log N)$) which, in practice, is faster than running the EM algorithm on the Bayesian model in $O(N)$ complexity per iteration over the latent variables and parameters. Given this conclusion, we next provide further analyses of the location behaviour of 38 real-world users using the LZ estimator, which we will show can give useful results even on a resource-limited mobile device. However, we shall see in Section 4.5 that the Bayesian model (the general framework for departures from routine) also has the ability to predict future departures.

4.4.2. Exploring departures from routine in the Nokia dataset

We now use our LZ instantaneous entropy estimator to explore the Nokia Lausanne dataset. In the following, we examine how instantaneous predictability varies over time and how it correlates with mobile application usage. To add some tangibility to the measure, we then discuss the outcome of detailed routine analysis of two days taken from an individual's mobility.

Plotting a heat map of the average instantaneous entropies according to local time of the week over all users yields Fig. 4. 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 8 am to 5 pm, and slightly higher levels in the evenings when users might go out to see friends.

To expand our view of behaviour when the user moves into and out of habitual location patterns, we also consider user behaviour with mobile applications, finding the probability of application use conditioned by the current instantaneous entropy. Fig. 5 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.[5] 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 more 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.

These results are in accordance with the findings by Church and Smyth [33], who used diary studies to show that "many users have a variety of information needs when they are away from their familiar contexts". However, our work goes further by providing a quantitative way to characterise the familiarity of these contexts.

To put these results in context, we created two snapshots of a single user's daily life in the Nokia dataset, illustrating both a typical and a more unusual day [34]. We also created two video animations of the location history of this user on

---

[5] The difference is statistically significant in all cases—a χ^2 test of independence rejects the null hypothesis that instantaneous entropy and application use are independent at the $p \ll 0.001$ level for all applications shown in Fig. 5.
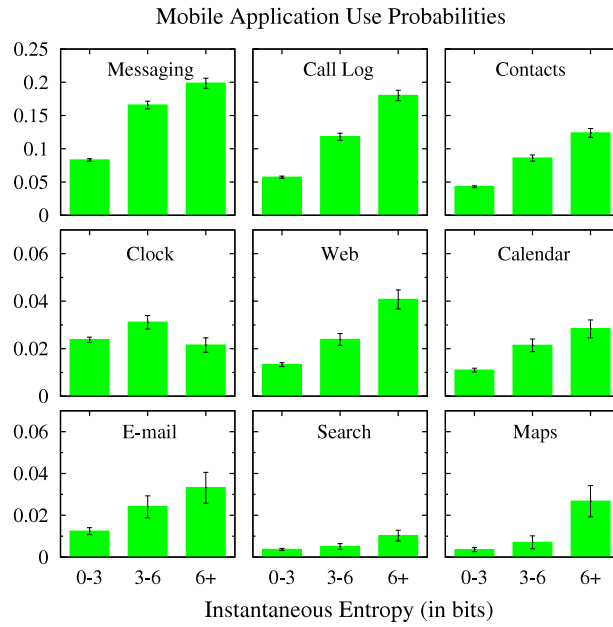
Fig. 5. Probability of using various applications in the hour as different levels of instantaneous entropy are observed (shown with 95% confidence intervals).
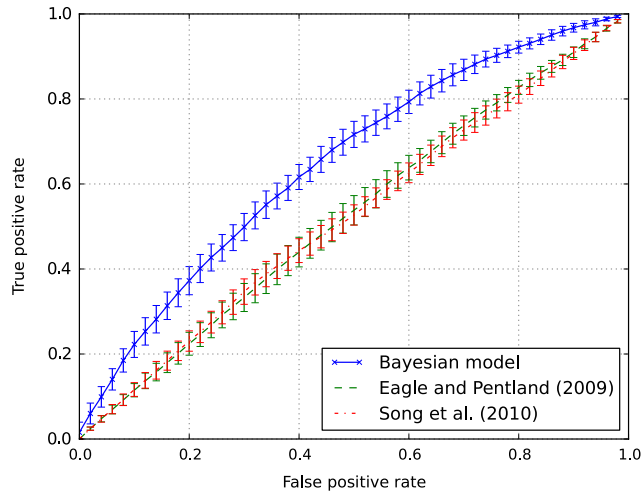

Fig. 6. ROC curves for prediction of departure from routine with the predictive approaches. Error bars indicate the 95% confidence range.

a map covering a period of three months, where the colour of the path indicates her instantaneous entropy (where red indicates departures from routine and green indicates being in deep routine). 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.

This concludes our analysis of the Nokia dataset using the LZ estimator. We next consider a different scenario, one that requires prediction of future routine behaviour.

4.5. Predicting departures from routine

In this section we evaluate the performance of all the approaches in a prediction scenario. In this scenario, less information is available than with *after the fact* analysis: although the time context is known, information about location context is restricted (because it is future behaviour).

Fig. 6 shows the performance when predicting departures from routine one hour in the future, again, using ROC curves. The results are averaged across all 38 users. We see that our Bayesian model performs best for prediction across all horizons.

We explain this by the fact that, unlike the other approaches, the model attempts to capture patterns of individual user mobility as well as changes in the levels of routine. The other benchmarks are no better than random at analysing departures from routine in specific location sequences. However, we notice that performance is diminished with respect to inference (see Fig. 3). This is to be expected, as prediction of future departures has greater uncertainty.

Finally, we consider applications made possible by an understanding of the role of routine in daily life mobility.

5. Applications of a dynamic understanding of routine

Considering the applicability of our work for pervasive computing, we first discuss several high-level application sketches in Section 5.1, before examining one of these sketches, an anticipatory mobile service that understands departures from routine, in greater depth in Section 5.2.

5.1. Application sketches

Knowledge of current and future departures from daily life location routines by individuals can be used to enhance a range of services that are currently emerging. We identify three areas where we see a direct application of our work. Firstly, mobile digital assistants (e.g., *Siri* for iPhone, Google *Now*, and *Sherpa*) perform natural language processing and service aggregation to give summarised and timely information to users, working both in the foreground and background. Our work could enhance background processing by providing better predictive power of user locations, for times when assistance is most crucial (yet when existing models are likely to fail), e.g., pre-caching maps in unfamiliar locations for faster loading times, or negotiating with services near the user's predicted destination [35]. Secondly, in mobile advertising (e.g., LivingSocial *Instant*, Groupon *Now*), our approaches could be used to allow advertisers to specifically target users that are currently *not* in deep routine and whose behaviour patterns therefore might be easier to change, given an appropriate incentive for the user. Finally, crowdsourcing physical tasks (e.g., *TaskRabbit, FieldAgent*) involves paying participants to perform tasks that are typically related to their current location, e.g., delivering groceries to a specific address near the user or taking photos of a particular location. The combination of location prediction and routine understanding would enable tasks to be assigned opportunistically, based not only on participants' current locations, but future locations and contexts.

As an exposition of exactly how our work can be applied, we now consider, in detail, the first application area we highlighted (that of a proactive mobile assistant), showing how knowledge of departures from routine can aid in the prediction of unusual future location behaviour.

5.2. Demonstration: predicting timing of unusual location behaviour

As described at a high level in Section 5.1, mobile applications like Google *Now* and *Sherpa* use personal online resources, such as emails, tweets, and calendar entries (to which the user explicitly grants access) to inform proactive mobile services. In more detail, mention of future events that appear in these resources are identified and extracted (e.g., hotel bookings, restaurant reservations, meetings with friends and colleagues). However, if no temporal information is given in the text from which the event was extracted, or if this information is vague (e.g., in messages such as "see you next week" or "let's meet at the pub after work") then this approach fails. Our work on departures from routine can address this problem of needing to identify more precise timings of future events, by assuming that such events represent a departure from routine. We believe this assumption is reasonable because most future events that are worth planning via email or social media are not likely to be everyday occurrences.

We consider a scenario in which it is known that a user is planning to attend a concert at a known location, but unknown time. The anticipatory service is required to perform three actions: (1) remind the user a suitable amount of time before the concert; (2) send directions for travel from the current location to the concert location leaving enough time for the user to get there (we assume the concert is one hour's travel away from the user's position); (3) push the entry ticket (*mTicket*) to the mobile screen just as the user arrives at the concert. For this scenario, we selected three arbitrary utility curves to reflect the desirability of their timing. They are the gamma distributions shown in Fig. 7(a), with representative shapes $0, \frac{1}{2}, 0$, and locations $1, \frac{3}{2}$, and 3, for the mTicket, Directions, and Reminder, respectively.

The experimental setup to test our model (and some benchmarks, which we detail next) with these utilities was as follows. We used the same generated ground truth for departures as in Section 4.2, extracting the location and time contexts for every classified departure from routine, for each user in the Nokia dataset. We assume that each of these departures represents the concert event, allowing us to repeat the experiment multiple times to obtain statistical significance. To select the best action a (i.e., time) for each test point, we used the definition of expected utility $\arg\max_a \sum_d U_i(a|d)p(d|x_e, \mathbf{X}, \mathbf{D})$ using three different methods.

The first method uses the probabilistic model we proposed in Section 3.2, which explicitly considers departures from routine. We also considered two benchmarks. The first was the temporal model by Cho et al. [14], which represents the state of the art in temporal prediction. We do not consider sequential prediction because it generally does not perform well for predicting behaviour more than a few hours ahead.
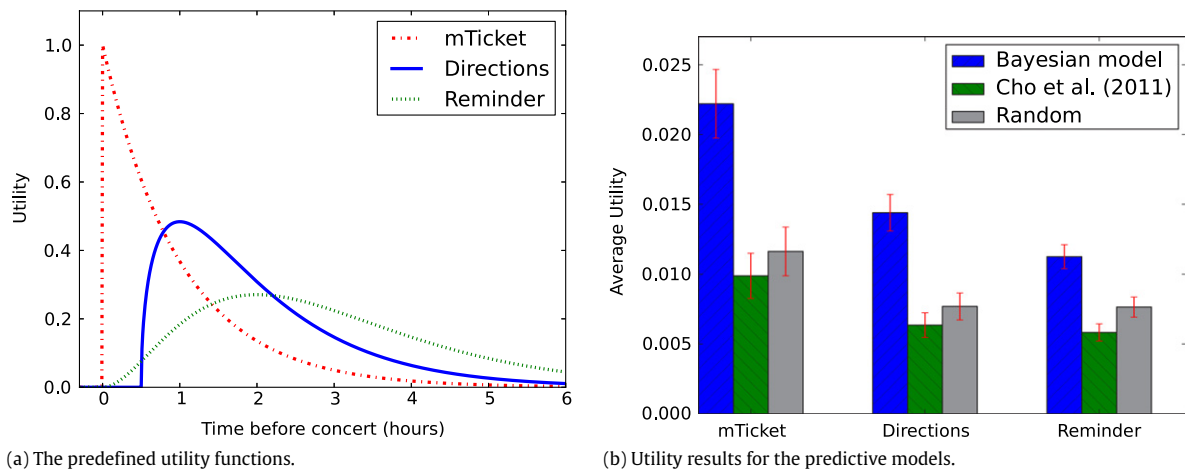
(a) The predefined utility functions.

(b) Utility results for the predictive models.

Fig. 7. Experimental results for the concert scenario. Error bars show the 95% confidence interval.

The results of the simulation are shown in Fig. 7(b). We show the average utility for each approach for the three actions (using their utility curves).[6] Looking at the differences between each approach (i.e., *within* each action), we see that our model outperforms both benchmarks. However, the absolute average utilities of all approaches is low, indicating that this is a very challenging problem. Surprisingly, the temporal model by Cho et al. does no better than random.[7] This could be because the given location information causes it to infer the *normal* time when the user is at that location, while the focus in this experiment is on location behaviours that are *not* normal. However, we expect that this will not always be the case in reality, because one would expect at least some events to belong to the routine location habits of a person. An additional reason that might play a factor is that this model was previously used to generate the ground truth of departures (along with the variable-order Markov model) so it is unsurprising that the model fails again on these data points when tested. On the other hand, we do not believe that the experiment is taking advantage of any idiosyncratic shortcomings of this specific model, as *any* model that predicts location behaviour temporally (i.e., based primarily on time features [6,16]) takes advantage of periodicities in human behaviour, so these models are likely to perform badly as well.

6. Conclusions and future work

In this article, we addressed the two interrelated challenges of (i) *analysing* a user's location history to identify departures from routine mobility, and of (ii) *predicting* such departures in the future. For the former, we extended principled concepts from information theory and presented a real-time estimator for a new metric we refer to as the instantaneous entropy. As we show in analysing a rich mobility dataset, this measure gives an understanding about a user's depth of routine at any given time, and is strongly correlated with the way they interact with the mobile devices. To deal with predicting the user's depth of routine in the future, we developed a new Bayesian model to capture this. Using established EM techniques, we showed that this model can be employed to predict departures from routine.

We found that our approaches (the instantaneous entropy metric and full probabilistic model) both outperformed the existing state of the art in inferring and predicting departures from routine. Instantaneous entropy can be used on resource-limited mobile devices to perform inference about routine. The probabilistic model provides a more powerful, but less efficient, approach to inferring and also predicting departures from routine. As an exposition of what is made possible with our approach, we highlighted the use of our probabilistic model in an anticipatory mobile computing scenario.

In conclusion, our tools open the way for better control over mobile notifications and assistance. To this end, we plan to explore a number of key applications in future work, and the new research challenges that these entail. Specifically, we want to ascertain the precise relationship between being in routine and being available (e.g., to follow up on a mobile offer, or to perform a crowdsourced task in the vicinity).

References

[1] M.C. Gonzalez, C.A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, Nature 453 (2008) 779–782.
[2] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, A. Campbell, Nextplace: a spatio-temporal prediction framework for pervasive systems, in: Pervasive Computing, Springer, San Francisco, CA, USA, 2011, pp. 152–169.

---

[6] Although the area under these curves is 1 in each case, the test approach can only ever take a single point on the curve, meaning that there is more utility to be had for the mTicket action versus the other actions (and the same applies to the difference between Directions and Reminder). Therefore, differences *between* the actions are not important.

[7] Analysis of variance confirms these findings to be the case to 99.9% confidence.

[3] 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, M. Van Alstyne, Life in the network: the coming age of computational social science, Science 323 (5915) (2009) 721–723.
[4] C. Song, Z. Qu, N. Blumm, A. Barabasi, Limits of predictability in human mobility, Science 327 (5968) (2010) 1018–1021.
[5] J. Wang, P. Bhaskar, Periodicity based next place predictions, in: Mobile Data Challenge by Nokia Workshop, in Conjunction with International Conference on Pervasive Computing, 2012.
[6] A. Sadilek, J. Krumm, Far out: predicting long-term human mobility, in: Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012.
[7] H. Gao, J. Tang, H. Liu, Exploring social-historical ties on location-based social networks, in: Sixth International AAAI Conference on Weblogs and Social Media, 2012.
[8] M.A. Bayir, M. Demirbas, N. Eagle, Mobility profiler: a framework for discovering mobility profiles of cell phone users, Pervasive and Mobile Computing 6 (4) (2010) 435–454.
[9] N. Eagle, A.S. Pentland, Eigenbehaviors: identifying structure in routine, Behavioral Ecology and Sociobiology 63 (7) (2009) 1057–1066.
[10] J. Zhuang, T. Mei, S.C. Hoi, Y.-Q. Xu, S. Li, When recommendation meets mobile, in: Proceedings of the 13th International Conference on Ubiquitous Computing, UbiComp '11, ACM, New York, NY, USA, 2011, pp. 153–162.
[11] A. Bhattacharya, S.K. Das, LeZi-update: an information-theoretic approach to track mobile users in PCS networks, in: Proc. MobiCom '99, 1999, pp. 1–12.
[12] B.S. Jensen, J.E. Larsen, K. Jensen, J. Larsen, L.K. Hansen, Estimating human predictability from mobile sensor data, in: 2010 IEEE International Workshop on Machine Learning for Signal Processing, MLSP, 2010, pp. 196–201.
[13] J. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.M.T. Do, O. Dousse, J. Eberle, M. Miettinen, The mobile data challenge: big data for mobile computing research, in: Proc. Mobile Data Challenge by Nokia Workshop, 2012.
[14] E. Cho, S.A. Myers, J. Leskovec, Friendship and mobility: user movement in location-based social networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 1082–1090.
[15] D. Ashbrook, T. Starner, Using GPS to learn significant locations and predict movement across multiple users, Personal and Ubiquitous Computing 7 (5) (2003) 275–286.
[16] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, N. Villar, Preheat: controlling home heating using occupancy prediction, in: Proceedings of the 13th International Conference on Ubiquitous Computing, 2011, pp. 281–290.
[17] A. Sadilek, H. Kautz, J.P. Bigham, Finding your friends and following them to where you are, in: Proceedings of the fifth ACM International Conference on Web Search and Data Mining, WSDM '12, ACM, New York, NY, USA, 2012, pp. 723–732.
[18] H. Gao, J. Tang, L. Huan, Mobile location prediction in spatio-temporal context, in: Mobile Data Challenge by Nokia Workshop, in Conjunction with International Conference on Pervasive Computing, 2012, 2012.
[19] S.T. Iqbal, B.P. Bailey, Effects of intelligent notification management on users and their tasks, in: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, 2008, pp. 93–102.
[20] E. Horvitz, P. Koch, R. Sarin, J. Apacible, M. Subramani, Bayesphone: precomputation of context-sensitive policies for inquiry and action in mobile devices, User Modeling 2005 (2005) 151.
[21] S. Amini, A. Brush, J. Krumm, J. Teevan, A. Karlson, Trajectory-aware mobile search, in: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI '12, ACM, New York, NY, USA, 2012, pp. 2561–2564.
[22] S. Reddy, D. Estrin, M. Srivastava, Recruitment framework for participatory sensing data collections, Pervasive Computing (2010) 138–155.
[23] I. Kontoyiannis, P.H. Algoet, Y.M. Suhov, A.J. Wyner, Nonparametric entropy estimation for stationary processes and random fields, with applications to english text, IEEE Transactions on Information Theory 44 (3) (1998) 1319–1327.
[24] Y. Gao, I. Kontoyiannis, E. Bienenstock, Estimating the entropy of binary time series: methodology, some theory and a simulation study, Entropy 10 (2) (2008) 71–99.
[25] A. Rodriguez-Carrion, C. Garcia-Rubio, C. Campo, Performance evaluation of LZ-based location prediction algorithms in cellular networks, IEEE Communications Letters 14 (8) (2010) 707–709.
[26] V. Etter, M. Kafsi, E. Kazemi, Been there, done that: what your mobility traces reveal about your behavior, in: Mobile Data Challenge by Nokia Workshop, in Conjunction with International Conference on Pervasive Computing, 2012.
[27] C.M. Bishop, Pattern Recognition and Machine Learning, Vol. 4, Springer New York, 2006.
[28] J.H. Kang, W. Welbourne, B. Stewart, G. Borriello, Extracting places from traces of locations, in: Proc. WMASH 2004, 2004, pp. 110–118.
[29] A. Kapoor, E. Horvitz, Experience sampling for building predictive user models: a comparative study, in: Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI '08, ACM, New York, NY, USA, 2008, pp. 657–666.
[30] H. Bernard, P. Killworth, D. Kronenfeld, L. Sailer, On the validity of retrospective data: the problem of informant accuracy, Annual Review of Anthropology 13 (1985) 495–517.
[31] L. Song, D. Kotz, R. Jain, X. He, Evaluating next-cell predictors with extensive wi-fi mobility data, IEEE Transactions on Mobile Computing 5 (12) (2006) 1633–1649.
[32] C.E. Rasmussen, The infinite gaussian mixture model, Advances in Neural Information Processing Systems 12 (5.2) (2000) 2.
[33] K. Church, B. Smyth, Understanding the intent behind mobile information needs, in: Proceedings of the 14th International Conference on Intelligent User Interfaces, ACM, 2009, pp. 247–256.
[34] J. McInerney, S. Stein, A. Rogers, N.R. Jennings, Exploring periods of low predictability in daily life mobility, in: Mobile Data Challenge by Nokia Workshop, in Conjunction with International Conference on Pervasive Computing, 2012.
[35] E. Horvitz, J. Krumm, Some help on the way: opportunistic routing under uncertainty, in: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 2012, pp. 371–380.