Spatio-Temporal Routine Mining on Mobile Phone Data

TIAN QIN, WUFAN SHANGGUAN, and GUOJIE SONG, Peking University
JIE TANG, Tsinghua University

Mining human behaviors has always been an important subarea of Data Mining. While it provides empirical evidences to psychological/behavioral studies, it also builds the foundation of various big-data systems, which rely heavily on the prediction of human behaviors. In recent years, the ubiquitous spreading of mobile phones and the massive amount of spatio-temporal data collected from them make it possible to keep track of the daily commute behaviors of mobile subscribers and further conduct routine mining on them. In this article, we propose to model mobile subscribers' daily commute behaviors by three levels: location trajectory, one-day pattern, and routine pattern. We develop the model Spatio-Temporal Routine Mining Model (STRMM) to characterize the generative process between these three levels. From daily trajectories, the STRMM model unsupervisedly extracts spatio-temporal routine patterns that contain two aspects of information: (1) How people's typical commute patterns are. (2) How much their commute behaviors vary from day to day. Compared to traditional methods, STRMM takes into account the different degrees of behavioral uncertainty in different timespans of a day, yielding more realistic and intuitive results. To learn model parameters, we adopt Stochastic Expectation Maximization algorithm. Experiments are conducted on two real world datasets, and the empirical results show that the STRMM model can effectively discover hidden routine patterns of human commute behaviors and yields higher accuracy results in trajectory prediction task.

CCS Concepts: • **Information systems → Data mining**;

Additional Key Words and Phrases: Routine mining, spatio-temporal pattern, mobile phone data

ACM Reference format:
Tian Qin, Wufan Shangguan, Guojie Song, and Jie Tang. 2018. Spatio-Temporal Routine Mining on Mobile Phone Data. *ACM Trans. Knowl. Discov. Data.* 12, 5, Article 56 (June 2018), 24 pages.
https://doi.org/10.1145/3201577

1 INTRODUCTION

Mining human behaviors has always been an important subarea of Data Mining. Its main assumption is that human behaviors are dominated by some hidden patterns (Gonzalez et al. 2008). For example, Zipf's law (Powers 1998) holds for most human languages. On the one hand, these patterns provide empirical evidence to subjects that try to build the theory of human behaviors, such as linguistics, sociology, and psychology. On the other hand, patterns of historical behaviors allow us to forecast behaviors in the future, which is the foundation of various recommendation systems (Bao et al. 2015).

This work was supported by the National Natural Science Foundation of China (61572041) and the National Key Research and Development Program of China (No. 2016QY01W0200).
Authors' addresses: T. Qin, W. Shangguan, and G. Song, Room 1110, Science Building No. 2, Peking University, Beijing, 100871, China; emails: tianqin.sunny@gmail.com, {shangguanwufan, gjsong}@pku.edu.cn; J. Tang, 1-308, FIT Building, Tsinghua University, Beijing, 100084. China PR; email: jietang@tsinghua.edu.cn.

https://doi.org/10.1145/3201577

Commute routine mining has become a more and more popular topic thanks to the availability of large-scale location data collected from mobile phones. Mobile phone is able to record the daily trajectories of individuals in both time and space. For example, when a mobile subscriber calls or sends messages, he/she can be coarsely located near the cellular tower with which his/her phone is transferring data. Compared to GPS data, location data from mobile phone records are much larger in amount because mobile phones are among the few things that people carry with them wherever they go. Therefore, with mobile phone data, it has been never more possible to mine modern people's life pattern and especially commute routines (Candia et al. 2008; Liu et al. 2013; Shan et al. 2013; Yuan et al. 2012; Zheng et al. 2014).

What is routine and what is routine mining? While there are multiples ways to define the term "routine," in this article we refer to routine as the pattern that people tend to show up at certain places during certain time of a day. For example, a typical student's life might consist of two routines as "school-days" and "off-school-days." In school-day routine, the student tends to show up at classroom during daytime and dormitory at night. In off-school-days routine, the student might frequently show up at a nearby shopping mall. In different routines, people's location trajectories are subject to different pattern. The task of Mobile Routine Mining is to extract the above routine pattern from trajectories collected from mobile phone records.

Numerous applications of routine mining can be found in real world. In Location-Based Social Networks (Chow et al. 2010; Scellato et al. 2011), routine patterns are usually viewed as a user's feature in the real world. Along with the user's features in online virtual world, together they contribute to tasks such as user classification and user similarity measurement (Xiao et al. 2010). Besides, given a user's historical routine pattern, prediction can be made regarding his/her future location trajectory (Cho et al. 2011). If the user is predicted to appear at an entertainment area within the next few hours, recommendation of restaurant and entertainment services would be profitable. Moreover, patterns of a mobile subscriber's routine tell us what is typical of him/her, or to say, what is the expectation of his/her behaviors. Understanding the expectation allows us to measure how anomalous a location trajectory is, which makes it possible to conduct anomaly detection (Madey et al. 2006). Identification of these anomalous trajectories may find its applications in crime detection, anti-terrorism, and personal reminder systems (Banovic et al. 2016).

Despite the promising applications of routine miming, it is a challenging task because of three reasons. First, people's daily trajectories are of great variety. Even if they follow the same commute routine, variations in behaviors largely exist. Second, "routine" is never a strictly defined term. The best definition of routine is highly context dependent. Since routine mining is the basis of many downstream tasks (trajectory prediction, anomaly detection, etc.), an appropriate representation of routine pattern would greatly facilitate downstream applications. Third, human behaviors exhibit different degrees of uncertainty in different timespan of a day. Here, "uncertainty" is defined as the entropy of location distribution. For example, people locations trajectories are of low entropy at midnight because at this time most people are sleeping at home and location distribution is peaked at the place of their houses. However, trajectories at noon are typically much more erratic and uncertain. The same person could choose difference restaurants for lunch on different days, thus at this time location distribution should be more flat and of higher entropy. When mining routine patterns from location trajectories, a sensible model should emphasize on timespans with lower uncertainty, but allows a higher degree of flexibility during timespans that come naturally with high uncertainty. For instance at noon, even if the location trajectories of two days vary greatly from each other, they may still be subject to the same routine pattern because the degree of uncertainty has always been very high at this time. On the contrary, in midnight when the degree of uncertainty is low, minor deviation might suggest two days are subject to different routines.

To address these three challenges, we develop Spatio-Temporal Routine Mining Model (STRMM), aiming to mine mobile subscribers' commute routines from mobile phone data. The major contributions of this article can be summarized as follows:

(1) Proposed a novel way to understanding human commute routine. Routine behaviors are modeled as three levels: location trajectory, one-day pattern, and routine pattern.
(2) Developed the STRMM model which learns routine pattern unsupervisedly from location trajectories. The model is able to take into account different degrees of behavioral uncertainty in different timespans of a day, yielding more realistic and intuitive routines.
(3) Conducted extensive experiments on two real-world datasets, and the results show that the STRMM model is able to mine meaningful routine patterns from mobile phone data. Furthermore, when applied to trajectory prediction, the prediction accuracy by STRMM model outruns that of other two methods, which indirectly proves the effectiveness of mined routines.

The following sections are summarized as follows: In Section 2, we explore existing literatures that are related to trajectory mining and routine mining. In Sections 3–5, we introduce the proposed STRMM model. Section 3 defines five basic concepts used throughout the model, and data preprocessing methods. Section 4 explains the generative process and other properties of STRMM model. Section 5 outlines the adopted learning and inference methods for STRMM. Finally, in Section 6, we include experiments conducted on one synthetic dataset and two real world datasets.

2 RELATED WORK

In Farrahi and Gatica-Perez (2008a), user's daily trajectories are represented by vectors, and Support Vector Machine is leveraged to classify them into different categories of routine. Here, routines are manually selected categories, such as weekday-routine/weekend-routine. The biggest weakness of classification methods is lack of labeled data, especially among mobile phone data. Clustering methods seek to mine routines without labeled data. In Lv et al. (2013), the authors first define one-day activity as a place reference matrix, each entry $e_{i,j}$ of which represents the time duration that the user stays at the ith reference place during the jth timespan of that day. Then, one-day activities of multiple days are clustered into a few routine activities, which characterize the probability that the user is at the ith reference place during the jth timespan of the day. Both classification methods and clustering methods rely heavily on the calculation of distance, which usually treats all dimensions of location vector equally. However, in our case each part of the daily trajectories are not equal in its contribution to routine patterns. For instance, at midday, even if the location trajectories of two days vary greatly from each other, they may still be subject to the same routine pattern because the degree of uncertainty is very high in this timespan of the day. On the contrary, in midnight when the degree of uncertainty is low, minor deviation might suggest they belong to different routines.

Principal Component Analysis (PCA) method explores the use of eigenvalue decomposition in routine mining (Eagle and Pentland 2010; Jiang et al. 2012). The covariance matrix of location trajectory vectors are calculated and decomposed into eigenvectors (also called eigenbehaviors). The linear combination of "eigenbehaviors" is able to reconstruct each day's trajectory with high accuracy, and therefore the spanned subspace can be interpreted as the routine pattern. However, in real life people usually follow multiple routines, and PCA cannot model this property explicitly.

Farrahi et. al. have done comprehensive research on the use of LDA in routine mining (Farrahi and Gatica-Perez 2008b, 2010a, 2010b, 2011, 2012). For example, In Farrahi and Gatica-Perez (2008b), user's locations are recorded every half-hour and three consecutive locations plus a time identifier constitute a "word." A total of 48 words of a day are viewed as a document, on which

LDA runs and infers the latent topic of each word. Here, latent topics are interpreted as routines and in each routine there are certain words that tend to cooccur in the same document with high probability. However, routines mined by LDA can only represent partial pattern of a day because latent topics are assigned to words rather than documents. In Farrahi and Gatica-Perez (2008b), the authors use the most dominant topic (the topic with the highest probability in topic distribution) to represent the routine of the whole document. This method has its limitation in that it loses the information of other non-dominant topics in that day. In this article, we aim to directly mine the routine pattern of a whole day, integrating all the information in any time of a day. This integrated routine better supports tasks such as trajectory prediction because in most cases we have to first infer the routine pattern of the whole day based on part of the observed trajectories and then predicting future trajectory.

HMM are also widely used in the task of routine mining (Eagle and Pentland 2006; Huai et al. 2014). HMM methods assume each location in the trajectory is emitted from a hidden state and Markov property holds between consecutive hidden states. Similar to LDA, the hidden states are interpreted as routines. In Huai et al. (2014), research works further study the semi-supervised training of HMM to learn more meaningful hidden states(routines) from location trajectories. HMM based methods share the same weakness of LDA: the learnt routines only represent partial pattern of a day.

There are also works on trajectory mining that are highly related to routine mining. Trajectory mining includes data preprocessing (denoising, segmentation, compression, ROI/stay point extraction, etc.) (Deng et al. 2011; Kellaris et al. 2009; Lee and Krumm 2011; Zheng 2015) and pattern mining (Jiang et al. 2012; Ye et al. 2009; Yuan et al. 2015) (sequential patterns, trajectory clustering, etc.). This work is overlapped with the latter part. Advanced trajectory preprocessing techniques are not explored in this article because the emphasis of this article is to model/extract human routine behaviors given trajectory semantics.

This article is also related to group anomaly detection. Though having different aims, both routine mining and group anomaly detection try to extract typical pattern (or expectation) from data. For example, the MGMM model in Liang et al. (2011b) is able to learn typical topic distributions from multiple documents. Yet MGMM model uses Multinomial parameter to represent typical topic distributions, thus restricting documents of the same category to have completely same topic distributions. Liang et al. (2011a) solves this problem by introducing Dirichlet parameter, taking into account the uncertainty of topic distributions in different documents.

To overcome the limitations shown in the above traditional methods, we propose the STRMM model. The model is able to extract multiple routine patterns unsupervisedly from daily trajectories. Moreover, routines mined by the STRMM model integrate information of the whole day, overcoming the problem of "partial pattern" in LDA and HMM. Finally, the STRMM model takes into account different degree of behavioral uncertainty in different timespans of a day, yielding more realistic and intuitive routines.

3 CONCEPTS DEFINITION AND DATA PREPROCESSING

In this section, we introduce definition of five key concepts used in STRMM model and illustrate the method for preprocessing raw mobile phone data.

3.1 Concepts Definition

POI. Point of Interest (*POI*) is a geographical point, which represents semantic meaning of the place, such as "home," "working place," "plaza," "supermarket," "on the road," and so on.

ROI. Region of Interest (*ROI*) is a region that has special meaning for individuals. This region might be user's home, or the shopping mall that user frequently visits. Different *ROIs* can be

	Time Span 1			⋯	Time Span T		
	TS1	TS2	TS3	⋯	TS τ-2	TS τ-1	TS τ
1th day	$p_{1,1}^{(1)}$	$p_{1,2}^{(1)}$	$p_{1,3}^{(1)}$	⋯	$p_{1,\tau-2}^{(T)}$	$p_{1,\tau-1}^{(T)}$	$p_{1,\tau}^{(T)}$
2th day	$p_{2,1}^{(1)}$	$p_{2,2}^{(1)}$	$p_{2,3}^{(1)}$	⋯	$p_{2,\tau-2}^{(T)}$	$p_{2,\tau-1}^{(T)}$	$p_{2,\tau}^{(T)}$
⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Dth day	$p_{D,1}^{(1)}$	$p_{D,2}^{(1)}$	$p_{D,3}^{(1)}$	⋯	$p_{D,\tau-2}^{(T)}$	$p_{D,\tau-1}^{(T)}$	$p_{D,\tau}^{(T)}$

Fig. 1. Table of trajectories.

characterize by the distribution of different types of *POIs* that exist within this region. The distribution of *POI* can provide useful information of the types of region this *ROI* represents. If most *POIs* within this *ROI* are restaurants and shopping malls, then this *ROI* is most likely a recreational region.

Location trajectory. We assume each day is divided equally into τ time slices, each of which represents $\frac{24}{\tau}$ hour. For a mobile subscriber U, we define his/her Region of Interest Set to be $ROIs = \{ROI_1, ROI_2, \ldots, ROI_C\}$. Then, location trajectory is defined as a sequence of length τ: $(p_1, p_2, \ldots, p_\tau)$, where $p_j \in ROIs$ $(j = 1, 2, \ldots, \tau)$ denotes user U's location within time slice j. The bigger τ is, the more precise location trajectory describes U's traveling behaviors. In this article, τ is usually set as 24 or 48, i.e., user's location is measured every hour/half hour. D sequences of location trajectories are used to represent user U's traveling behaviors in D days. We further divide each day into T timespans, with the kth timespan covers N_k time slices, $\sum_{k=1}^{T} N_k = \tau$. Assume we have user U's location trajectories in D days, which is shown in Figure 1. TS denotes time slice. $p_{i,j}^{(k)}$, $i = 1, 2, \ldots, D$, $j = 1, 2, \ldots, N_k$, $k = 1, 2, \ldots, T$ is user U's location in the jth time slice of the kth timespan of the ith day, $p_{i,j}^{(k)} \in ROIs$. In this article, T is set to be 8. The 8 timespans are: 0:00–7:00, 7:00–9:00, 9:00–11:00, 11:00–14:00, 14:00–17:00, 17:00–19:00, 19:00–21:00, 21:00–24:00. The reason why we chose these time slots is that we try to capture those common daily events, for example, morning work time, lunch time, and dinner time. If we aim to capture some other events, other time slots could be used.

One-day pattern. One-day pattern is the hidden rule behind each day's location trajectory. We assume all locations $p_{i,j}^{(k)}, j = 1, 2, \ldots, N_k$ in the same timespan k of day i are subject to the same Multinomial distribution $p_{i,j}^{(k)} \sim Multinomial(\theta_i^{(k)})$. This is a reasonable assumption because most people do only one type of thing in one timespan. For example, timespan 0:00–7:00 is the time where most people sleep at home, while timespan 7:00–9:00 is the time where most people travel from home to school/working place. The set $\{\theta_i^{(k)} \mid k = 1, 2, \ldots, T\}$ of length T is the one-day pattern of the ith day in all T timespans.

Routine pattern. Given the table of D trajectories, our task is to mine R routine pattern from it. One-day pattern may vary from day to day, but routine pattern keeps what one-day patterns have in common. For example, Monday and Tuesday are all subject to the weekday routine. Then, in corresponding timespan like 9:00–11:00 on Monday and Tuesday, user is highly likely to be doing the same thing "working at company." Formally, if the ith day and the i'th day are subject to the same routine pattern, then there must be strong interrelation between $\theta_i^{(k)}$ and $\theta_{i'}^{(k)}$, $k = 1, 2, \ldots, T$. Dirichlet distribution is leveraged to characterize this interrelation: we assume $\theta_i^{(k)}$ and $\theta_{i'}^{(k)}$ are subject to the same Dirichlet distribution $\theta_i^{(k)}, \theta_{i'}^{(k)} \sim Dirichlet(\alpha_r^{(k)})$. The set $\{\alpha_r^{(k)} \mid k = 1, 2, \ldots, T\}$ is the rth routine pattern in all the T timespans.

3.2 Data Preprocessing and ROI Discovery

Data collected from mobile phone is usually in the form of "check-in" data, i.e., user is present at cellular tower X at time Y. Check-in records are collected when phone interchanges data with nearby cellular tower. Information contained in raw check-in data is limited in that it only provides coarse-grain geo-location data but lack of semantic meaning. Attaching semantic labels to raw geo-location point could prove to be helpful in extracting more intuitive and robust routines. For a mobile subscriber, multiple cellular towers' locations may share the same semantic meaning. For example, there are three cellular towers far away from each other, each neighboring a shopping mall. Depending on the downstream task, distinguishing these three cellular towers might be unnecessary because the user goes to these places only for one purpose: shopping. On the other hand, a single geo-location point can also represent multiple semantics for a user. A hot-spot in downtown could have restaurants, recreational facilities, and parking lot all in one place. Methods have been studied to extract POI (or reference places) from raw trajectory (Montoliu and Gatica-Perez 2010; Zhou et al. 2007), but they are not suitable for trajectories formed by sparse cellular towers geo-location. Therefore, we propose the following three steps to extract meaningful *ROIs* from raw check-in data.

First, we rule out noisy location points in trajectory that contribute little to user's commute routine. For example, places where user lives and works are usually essential in explaining his/her spatio-temporal routine pattern. However, there are noisy locations where user only passes through by chance, having little or no explanatory power as to either user's behavior or intention. We adopt DBSCAN algorithm to cluster cellular towers by their longitude and latitude, and exclude those outlier noisy points by setting a threshold.

In the second step, we label every non-noisy cellular tower with a semantic distribution. Typically the region covered by one cellular tower in a city is too large that it's not reasonable to label it with just one POI label. Therefore, we assign every tower with a POI-distribution. We categorize *POI* into eight groups: catering service, service for life, leisure and entertainment, education, transportation, company, accommodation, and hotel. Using *PlaceSearchAPI* provided by Baidu Map Service, we are able to count the number of eight types of *POI* facilities surrounding each cellular tower. We use the distribution of these *POI* counts as the semantic meaning of cellular towers. (e.g. if there are five schools, three hotels, and eight restaurants around $tower_i$, the label of $tower_i$ will be a vector like [cate:8, service for life:0, leisure and entertainment:0, education:5, transportation:0, company:0, accommodation:0, hotel:3])

Last, we group towers with similar semantic meaning together. Using semantic distribution (which is an integer vector of size 8) to represent each cellular tower, we use K-means algorithm to group them into several categories that most concisely capture user's intention in visiting these places (e.g., shopping). We identify cluster centers to be the Region of Interest Set (*ROIs*) and label all cellular towers belonging to a cluster with a discreet *ROI* label. User's geo-location trajectory can now be represented as a sequence τ: $(p_1, p_2, \ldots, p_\tau)$, where $p_j \in ROIs$ $(j = 1, 2, \ldots, \tau)$.

Through the above three steps, we use external POI information to transform low-level geo-location data into a high-level *ROI* sequences. These *ROI* sequences exist in semantic space, in which user's behavior and intention is easier to explain. For example, given that a person usually approaches $tower_x$ between 12:00p.m. and 1:00p.m., along with the fact that $tower_x$ is associated with ROI label [catering service:20, service for life:1, leisure and entertainment:1, … hotel:1], conclusion can be drawn that the user is most likely having lunch in this place. The term "location trajectory" used in following sections refers to this high level *ROI* sequence, instead of raw check-in location sequence.

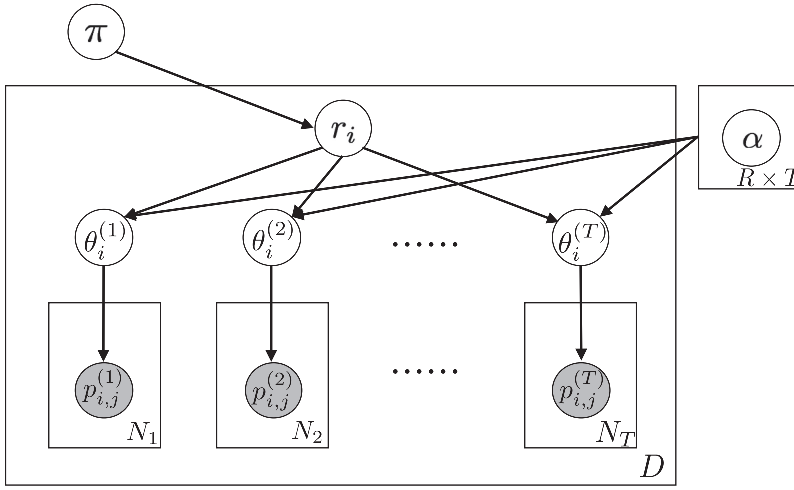


Fig. 2. Plate notation of the STRMM model.

4 MODEL SPECIFICATION

In the last section, we defined three key concepts related to human travel pattern: Location Trajectory, One-day Pattern, and Routine Pattern. These three concepts together form a hierarchical view of human travel behaviors. From top to down, first we have Routine Pattern, the underlying "theme" of the day's commute pattern. For example, Routine Pattern can be "stay at home all day," "commute and work from nine to five," or "go to parents house for dinner." Then there is One-day Pattern, the variation of the "theme." Three days may follow the same Routine Pattern "go to parents house for dinner," but people may choose from three different roads when they drive depending on the traffic of the day. Therefore, on each of the three days people may follow a slightly different travel pattern, corresponding to three possible One-day Patterns. Each One-day Pattern is a realization of the underlying Routine Pattern. Routine Pattern defines what the most typical One-day Pattern is likely to be and how One-day Patterns may vary from day to day. Last, we have Location Trajectory, which is the real location trajectory that user follows. Location trajectories are recorded as mobile phone users move across cellular towers. After preprocessing Location Trajectory becomes a sequence of *ROI*, each of which represents the semantic meaning of the tower.

Based on the five defined concepts, we propose Spatio-Tempotal Routine Mining Model (STRMM). See Figure 2 for its plate notation.

Here, r_i is a categorical variable denoting the day's routine pattern. θ is one-day pattern and p is location trajectory. We assume the number of routine types R is fixed (in experiments R is pre-selected by the measurement of perplexity). T Dirichlet parameters are used to describe the T timespans in each of the R routine pattern. Parameter π is the probabilistic distribution of R routine patterns. For the ith day of user U, we sample this day's routine category $r_i \sim Multinomial(\pi)$. Then, for each of the T timespans of day i, we sample its one-day pattern $\theta_i^{(k)}$ from the corresponding Dirichlet parameter $\theta_i^{(k)} \sim Dirichlet(\boldsymbol{\alpha}_{r_i}^{(k)}),\ k = 1, 2, \ldots, K$. Last, we sample the location in each time slice according to Multinomial distribution $p_{i,j}^{(k)} \sim Multinomial(\theta_i^{(k)}),\ j = 1, 2, \ldots, N_k$. Table 1 summarizes the notations used in our model. The generative process of STRMM is described as follows.

Table 1. Notations Used in the STRMM Model

Notation	Description
π	Probabilistic distribution of R types of routines
$\alpha_r^{(k)}$	Dirichlet parameter of the kth timespan of the rth routine
r_i	Type of routine that the ith day is following
$\theta_i^{(k)}$	Multinomial parameter of the kth timespan of the ith day
$p_{i,j}^{(k)}$	ROI of the jth time slice of the kth timespan of the ith day
D	D days of location trajectories in total
R	R types of routine patterns in total
C	C regions of interest (ROI)
N_k	Number of time slices in the kth timespan
T	A day is divided into T timespans
τ	A day is divided into τ time slices

Generative Process of the Model

Set parameter π
Set parameter α
for $i = 1 \rightarrow D$ **do**
  Draw $r_i \sim Multinomial(\pi)$
  for $k = 1 \rightarrow T$ **do**
    Draw $\theta_i^{(k)} \sim Dirichlet(\alpha_{r_i}^{(k)})$
    for $j = 1 \rightarrow N_k$ **do**
      Draw $p_{i,j}^{(k)} \sim Multinomial(\theta_i^{(k)})$
    end for
  end for
end for

Parameters to be learnt is routine distribution π and routine pattern $\alpha_r^{(k)},\ k = 1, 2, \ldots T,\ r = 1, 2, \ldots, R$. π is easy to understand, as π_r denoting the probability of the rth routine's appearance in this user's daily life. The rth routine pattern $\alpha_r^{\cdot}$ can be understood from the following two perspectives:

(1) It characterizes the typical one-day pattern in the k timespans of the rth routine. Specifically, if we extend the Dirichlet parameter $\alpha_r^{(k)}$ into its vector form $\alpha_r^{(k)} = (\alpha_{r,1}^{(k)},\ \alpha_{r,2}^{(k)}, \ldots,\ \alpha_{r,C}^{(k)})$, then we have

$$E\left[\theta_r^{(k)}\right] = \left(\frac{\alpha_{r,1}^{(k)}}{\alpha_{r,\cdot}^{(k)}}, \frac{\alpha_{r,2}^{(k)}}{\alpha_{r,\cdot}^{(k)}}, \ldots, \frac{\alpha_{r,C}^{(k)}}{\alpha_{r,\cdot}^{(k)}}\right), \tag{1}$$

where $\alpha_{r,\cdot}^{(k)} = \sum_{y=1}^{C} \alpha_{r,y}^{(k)}$. Here, $\theta_r^k \sim Dirichlet(\alpha_r^{(k)})$ is the possible one-day pattern in the kth timespan generated from routine pattern $\alpha_r^{(k)}$ and $E[\theta_r^{(k)}]$ represents the "typical" one-day pattern. The rth routine's typical one-day pattern is the set $\{E[\theta_r^{(k)}] \mid k = 1, 2, \ldots, T\}$.

(2) It characterizes the uncertainty of one-day patterns of the rth routine from day to day. The measurement of uncertainty can be done with entropy

$$H\left(\theta_r^{(k)}\right) = logB\left(\alpha_r^{(k)}\right) - \left(C - \alpha_{r,\cdot}^{(k)}\right)\psi\left(\alpha_{r,\cdot}^{(k)}\right) - \sum_{y=1}^{C}\left(\alpha_{r,y}^{(k)} - 1\right)\psi\left(\alpha_{r,y}^{(k)}\right), \tag{2}$$

where $B(\cdot)$ is Beta function and $\psi(\cdot)$ is Digamma function. Normally, the smaller the entropy is, the closer one-day pattern $\theta_r^{(k)}$ is distributed near $E[\theta_r^{(k)}]$. This means traveling behaviors are relatively stable and of less uncertainty across days. As the Dirichlet density function is relatively steep when entropy is low, minor difference on $\theta_i^{(k)}$ and $\theta_{i'}^{(k)}$ leads to large difference in the overall model likelihood and thus would suggest day i and day i' are subject to different routines. On the contrary as entropy grows larger, Dirichlet density function flattens and one-day pattern $\theta_r^{(k)}$ is allowed to distributed farther from $E[\theta_r^{(k)}]$. In this case, minor difference on $\theta_i^{(k)}$ and $\theta_{i'}^{(k)}$ might not change the overall model likelihood function very much. This property of Dirichlet distribution creates an effect that when deciding the routine type of one day, more emphasis is placed on the timespan with lower uncertainty.

To sum up, routine pattern α carries two aspects of information: the typical one-day pattern and the measurement of uncertainty. This allow us to mine more meaningful and intuitive routines from data, according to experiment results in latter section.

5 INFERENCE AND LEARNING

Inference and learning of the model is done by maximizing the marginal likelihood function of the observed data. We first derive the simplified marginal likelihood leveraging the conjugacy between Dirichlet distribution and Multinomial distribution, then we adopt Stochastic Expectation Maximization algorithm to approximate the true value of parameters (Bishop 2009, 2009; Hastie et al. 2009).

We denotes model parameters $\Theta = \{\pi, \alpha\}$, latent variables $\boldsymbol{Z} = \{\theta, r\}$ and observed data $\boldsymbol{V} = \{p\}$. Given the ith day's location trajectory $\boldsymbol{V}_i = p_{i,j}^{(k)},\ j = 1, 2, \ldots, N_k,\ k = 1, 2, \ldots, T$, the complete likelihood function of the ith day is

$$\begin{aligned} P(\boldsymbol{V}_i,\ \boldsymbol{Z}_i \mid \Theta) &= P\left(p_{i,\cdot}^{(\cdot)},\ r_i,\ \theta_i^{(\cdot)} \mid \pi,\ \alpha\right) \\ &= \boldsymbol{Mult}(r_i \mid \pi) \prod_{k=1}^{T}\left(\boldsymbol{Dir}\left(\theta_i^{(k)} \mid \alpha_{r_i}^{(k)}\right)\prod_{j=1}^{N_k}\boldsymbol{Mult}\left(p_{i,j}^{(k)} \mid \theta_i^{(k)}\right)\right). \end{aligned} \tag{3}$$

After integrating out latent variables θ and r, we have the marginal likelihood function as follows:

$$\begin{aligned} P(\boldsymbol{V}_i \mid \Theta) = \sum_{r_i}\Bigg(&\boldsymbol{Mult}(r_i \mid \pi) \\ &\times \prod_{k=1}^{T}\int_{\theta_i^{(k)}}\boldsymbol{Dir}\left(\theta_i^{(k)} \mid \alpha_{r_i}^{(k)}\right)\prod_{j=1}^{N_k}\boldsymbol{Mult}\left(p_{i,j}^{(k)} \mid \theta_i^{(k)}\right)d\theta_i^{(k)}\Bigg). \end{aligned} \tag{4}$$

Given that Dirichlet distribution is the conjugate prior to the Multinomial distribution, the integral in the above equation can be simplified as

$$\int_{\theta_i^{(k)}} \boldsymbol{Dir}\left(\theta_i^{(k)} \mid \alpha_{r_i}^{(k)}\right) \prod_{j=1}^{N_k} \boldsymbol{Mult}\left(p_{i,j}^{(k)} \mid \theta_i^{(k)}\right) d\theta_i^{(k)} = \frac{\Gamma(\alpha_{r_i,\cdot}^{(k)})}{\Gamma(n_{i,\cdot}^{(k)} + \alpha_{r_i,\cdot}^{(k)})} \prod_{y=1}^{C} \frac{\Gamma(n_{i,y}^{(k)} + \alpha_{r_i,y}^{(k)})}{\Gamma(\alpha_{r_i,y}^{(k)})}, \tag{5}$$

where $\alpha_{r_i,\cdot}^{(k)} = \sum_{y=1}^{C} \alpha_{r_i,y}^{(k)}$. $n_{i,y}^{(k)}$ denotes the number of times that the yth ROI appears in the location trajectory of the kth timespan of the ith day. $n_{i,\cdot}^{(k)}$ denotes the total number of ROI that appears in the location trajectory of the kth timespan of the ith day, $n_{i,\cdot}^{(k)} = \sum_{y=1}^{C} n_{i,y}^{(k)} = N_k$.

By substituting Equation (5) into Equation (4), we derive the simplified version of the marginal likelihood function of the ith day:

$$\begin{aligned} P(\boldsymbol{V}_i \mid \boldsymbol{\Theta}) &= \sum_{r_i} \pi_{r_i} \prod_{k=1}^{T} \left(\frac{\Gamma(\alpha_{r_i,\cdot}^{(k)})}{\Gamma(n_{i,\cdot}^{(k)} + \alpha_{r_i,\cdot}^{(k)})} \prod_{y=1}^{C} \frac{\Gamma(n_{i,y}^{(k)} + \alpha_{r_i,y}^{(k)})}{\Gamma(\alpha_{r_i,y}^{(k)})} \right) \\ &\triangleq \sum_{r_i} \pi_{r_i} f\left(\boldsymbol{V}_i \mid \alpha_{r_i}^{(\cdot)}\right), \end{aligned} \tag{6}$$

where π_{r_i} is the probability of routine r_i and $f(\boldsymbol{V}_i \mid \alpha_{r_i}^{(\cdot)})$ is the multiplication of T probability mass function of Dirichlet compound multinomial distribution (DCM).

Our aim is to find the best parameter $\boldsymbol{\Theta}$ satisfying $\boldsymbol{\Theta} = \text{argmax}_{\boldsymbol{\Theta}} P(\boldsymbol{V}|\boldsymbol{\Theta}) = \text{argmax}_{\boldsymbol{\Theta}} \prod_{i=1}^{D} P(\boldsymbol{V}_i|\boldsymbol{\Theta})$. Here, we adopt Stochastic Expectation Maximization algorithm (SEM) (Celeux et al. 1996). On the S step, we sample latent variable $\boldsymbol{r}$ from its posterior distribution

$$\begin{aligned} P(r_i \mid \boldsymbol{V_i}, \boldsymbol{\Theta}) &\propto P(r_i \mid \boldsymbol{\Theta}) P(\boldsymbol{V_i} \mid r_i, \boldsymbol{\Theta}) \\ &= \pi_{r_i} f\left(\boldsymbol{V}_i \mid \alpha_{r_i}^{(\cdot)}\right) \\ &= \pi_{r_i} \prod_{k=1}^{T} \left(\frac{\Gamma(\alpha_{r_i,\cdot}^{(k)})}{\Gamma(n_{i,\cdot}^{(k)} + \alpha_{r_i,\cdot}^{(k)})} \prod_{y=1}^{C} \frac{\Gamma(n_{i,y}^{(k)} + \alpha_{r_i,y}^{(k)})}{\Gamma(\alpha_{r_i,y}^{(k)})} \right). \end{aligned} \tag{7}$$

On the M step, we estimate model parameter $\boldsymbol{\Theta}$ by Maximum Likelihood Estimation (MLE) based on observed data $\boldsymbol{V}$ and the sampled latent variable $\boldsymbol{r}$. The MLE of π can be done straightforwardly from the histogram of $\boldsymbol{r}$. The MLE of Dirichlet parameter α can be solved using the iterative methods in Minka (2003).

The learning algorithm iterates between sampling $\boldsymbol{r}$ and estimating $\boldsymbol{\Theta}$. After the Markov Chain reaches its steady state, we average multiple samples to get the final output value of π and α.

6 EXPERIMENT

In this section, we conduct routine mining experiments on one synthetic dataset and two real datasets. In synthetic datasets, we aim to evaluate the effectiveness of model learning process. In Reality Mining dataset, we aim to mine shared routines (i.e., common behaviors) of multiple mobile subscribers. In China Mobile Operator dataset, we aim to mine personalized routines for individuals. Furthermore, we apply the mined routine patterns to routine type prediction task and trajectory prediction task and make comparison with state-of-the-art methods.

6.1 Synthesized Data

We generate synthetic data according to STRMM's Generative Process mentioned in Section 4 with pre-defined parameters $\boldsymbol{\Theta} = \{\pi,\ \alpha\}$. Using the method discussed in Section 5, we in turn

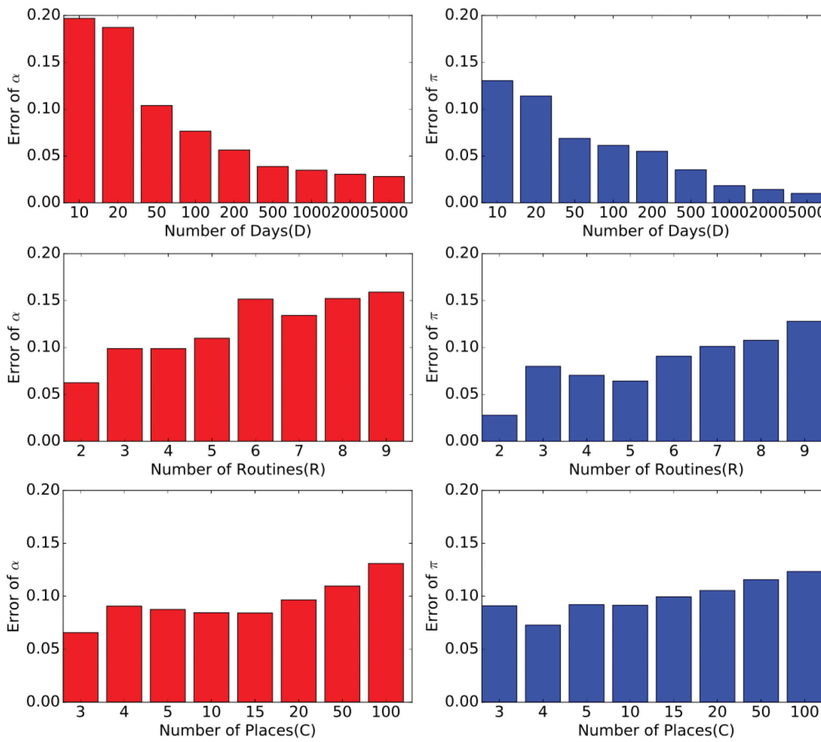


Fig. 3. Estimation error of the pre-defined parameters.

learn parameters $\Theta = \{\pi, \alpha\}$ given synthetic data and compare the learned parameters with the ground-truth parameters. The base setting is to use 100 days' location trajectories ($D = 100$), the number of Places (ROI) is 6 ($C = 6$) and the number of Routine Pattern is 3 ($R = 3$). We varied D, R, and C, respectively, in different experiments to evaluate the accuracy of our model under different settings. Result is shown in Figure 3. $Error\ of\ \pi$ and $Error\ of\ \alpha$ are both calculated by Euler distance. The larger the D is, the better our model learns correct parameters. However, as number of routines (R) and number of *ROIs* (C) increases, learning becomes more difficult. This justifies our approach in grouping raw cellular towers into semantic *ROIs*. Typically there are thousands of cellular towers, which would severely deteriorate model's learning ability. Learning on top of extracted *ROIs* becomes much easier as dimension is usually reduced to less than a hundred.

6.2 Reality Mining Dataset

The Reality Mining project (Eagle et al. 2009) was conducted from 2004 to 2005 at the MIT Media Laboratory. The dataset consists of 96 subjects whose behaviors are automatically logged by the ContextLog application installed on their mobile phones. These behaviors include location (both cell tower ids and semantic labels), other proximate subjects (from Bluetooth device discovery scans at 5-minute intervals), phones activities (voice calls and text messages), active applications (such as the calendar or games), and the phone's charging status. The part used in this article is the semantic location trajectories of each subjects.

We construct the table of 6,887 day's location trajectories as in Figure 1. Each day is divided to $\tau = 24$ time slices and $T = 8$ timespans: 0:00–7:00, 7:00–9:00, 9:00–11:00, 11:00–14:00, 14:00–17:00,

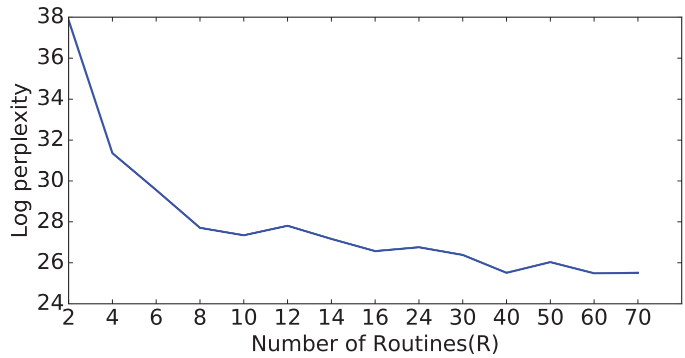


Fig. 4. Perplexity calculated for the 20% test data.

17:00–19:00, 19:00–21:00, 21:00–24:00. We set the number of *ROIs* to be $C = 5$. In each time slice, one of the five location labels are recorded: 1-home, 2-work, 3-elsewhere, 0-no signal, NaN-phone is off. These five location labels have the same functionality of *ROIs* defined in Section 3.

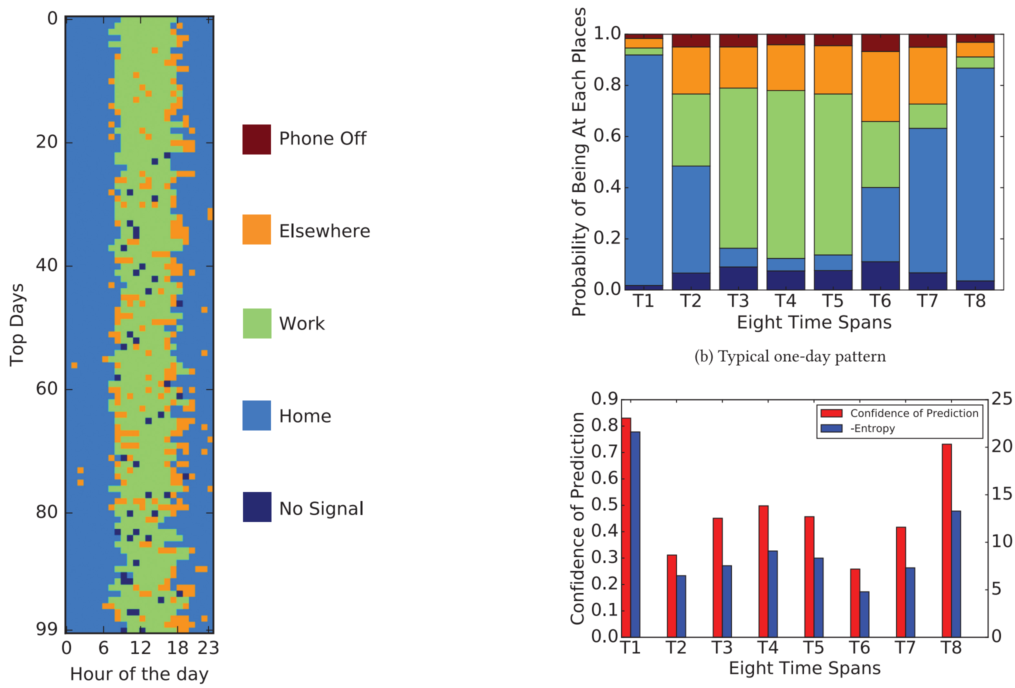
Perplexity is leveraged to find the best routine number R. Perplexity is a measurement of how well a probability distribution or probability model predicts a sample. We randomly select 80% of the dataset to be training data and calculate the trained model's perplexity on the rest 20% dataset. Result is shown in Figure 4.

As the number of routine pattern R grows larger, the model predicts better of unknown data. However, perplexity changes little after 40, so we set $R = 40$ in the following experiments. Due to the limitation of space, we only present two of the mined routine patterns as examples of how the STRMM model extracts meaning routines from raw location trajectories. More are presented in the appendix.

Figure 5(a) plots the most typical 100 days in the dataset subject the 31th routine pattern. From bottom to top the likelihood of data increases. Each line is the location trajectory of a day. Five colors stand for five location labels respectively. Figure 5(b) is the stacked column diagram plot of the typical one-day pattern of the 31th routine, calculated according to Equation (1). Figure 5(c) measures the degree of uncertainty of the one-day pattern on different timespan of a day. First, we compute the confidence of prediction (red column) by calculating the probability that location trajectories is generated from the typical one-day pattern for each timespan. Then, we compute the entropy of Dirichelt distribution under each timespan according to Equation (2). The blue column is the negative entropy.

As can be seen from Figure 5(a) and (b), the 31th routine pattern probably depicts those who "work from nine to five." In timespan T1 and T8 (21:00–7:00) people have high probability of staying home (blue color), while in timespan T3, T4, T5 (from 9:00 to 17:00) people have high probability of staying at working place (green). The probability of staying elsewhere (orange) remains around 0.2 in daytime and reaches its peak value in T6 and T7 (17:00–21:00), which suggest that people sometimes "go our for leisure after work."

In Figure 5(c), we observe the proportional relation between confidence of prediction and negative entropy. The higher confidence of prediction is (as in T1 and T8), the less uncertain this timespan is. At the same time, entropy is also relatively low in T1 and T8. By contrast, confidence of prediction is low in T2–T7 because human behaviors are much more complex and uncertain during daytime, and entropy are also relatively large in these timespans. The apparent proportional relation between confidence of prediction and negative entropy demonstrates that apart from



(a) Most typical 100 days following this routine

(b) Typical one-day pattern

(c) Degree of uncertainty

Fig. 5. The 31th routine pattern.

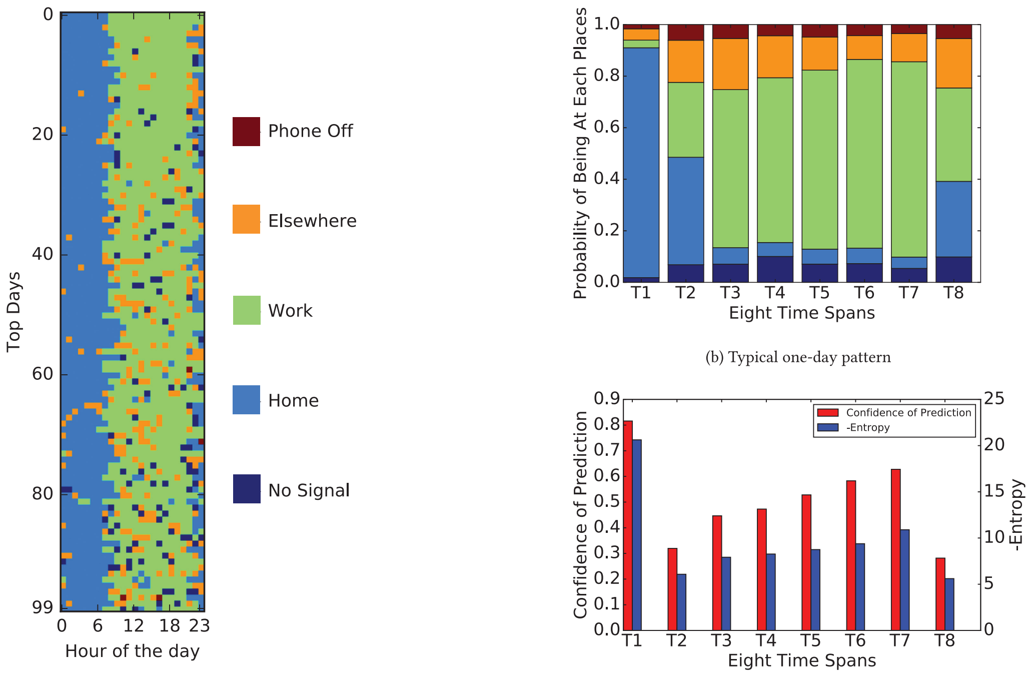
typical one-day pattern, our model is also able to take into account different degree of uncertainty of one-day pattern in each timespan.

Different from the 31th routine, the 29th routine (Figure 6) likely depicts those who "work late into the night." Instead of coming home after 17:00, under this routine users tend to stay late at working place. In Figure 6(b), the probability of staying at working place remains high until T8. Also, in Figure 6(c), we observe the similar proportional relation between confidence of prediction and negative entropy.

6.3 A China Mobile Operator Dataset

The mobile dataset used in this section is provided by a China Mobile Operator, involving a population of 300,000 mobile subscribers in a medium-size city of China from 2010 to 2011. It contains the time and location of several types of events, such as Call, SMS, Phone Power On/Off, Tower Switch (phone connecting to a different cellular tower). Call/SMS log helps locate user when he/she is actively using the phone. Tower Switch log helps locate user when he/she is moving across cells, without the need of user's action. These two parts of the data are complement to each other in order to construct a more precise location trajectory. The dataset spans across 297 cellular towers, with known longitude and latitude. In the occurrence of events mentioned above, user can be coarsely located at one of these cellular towers. Data used in the following analysis, includes mobile logs from September to October 2010. To protect user privacy, all phone numbers are encrypted and there is no way to identify individual users in the dataset.

In this section, we aim to mine personalized routines for individuals. Different from Reality Mining Dataset, we do not have semantic location labels (home, work, etc.) now. What we have



(a) Most typical 100 days following this routine

(b) Typical one-day pattern

(c) Degree of uncertainty

Fig. 6. The 29th routine pattern.

is the original geo-location log recorded when mobile phone transfers data with nearby cellular tower. To construct the table of trajectories (*ROIs*) as model input, we need the following two additional preprocessing steps:

(1) Construct daily trajectories. The original log data is event-driven, that is, there is a record in the database only when users call, SMS or have transition between towers. We need to interpolate location trajectories (described in Section 3.1) from this non-uniform-in-time data. Specifically, in this experiment we set $\tau = 48$. Every half an hour, we select the cellular tower which user's phone interact most with and set its coordinate as the user's current location. The reason for this approach is to alleviate tower oscillation. It is fairly common to see user's location oscillates between multiple cellular towers. This happens when signal from one tower is intermittently blocked by surroundings. Picking the most frequent tower within a time period can smooth out oscillation noise to some extend. In case there are no data records within a half-hour period, we assume the user stays where he/she was in the previous half-hour.
(2) Denoise location trajectory and label cellular towers with real-world meaningful POIs, and finally group cellular towers by their semantic meaning according to the three steps proposed in Section 3. After the preprocessing, cellular tower ids are replaced by cluster labels (ROI).

Figure 7 plots the raw trajectories of the 22,148th user. We group all cellular towers into five *ROIs* ($C = 5$), which are represented by five different colors. Setting the appropriate C is case dependent and involves an inherent bias variance tradeoff. C determines the number of *ROI* used in the model.

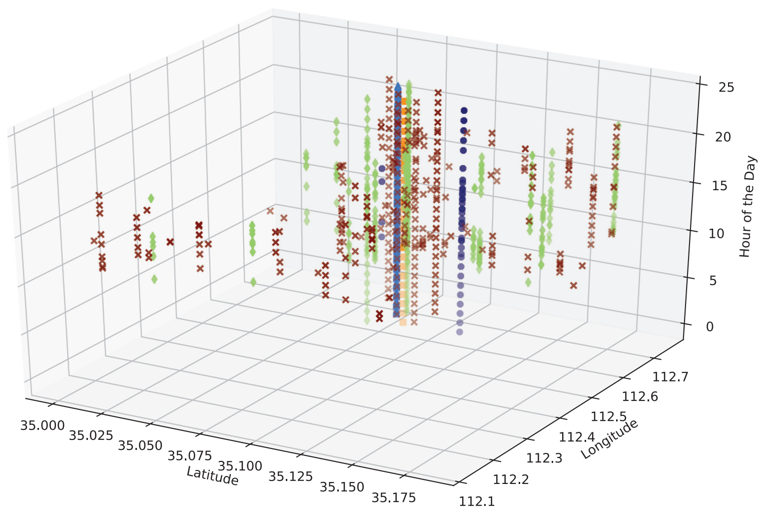


Fig. 7. Three-dimensional plot of the 22,148th user's daily trajectories.

Being a level of abstraction over raw geographical location, *ROIs* exist in the semantic space in which user's behaviors and intention are easier to explain (refer to Section 3.2). A sensible value for C lies between zero and number of cellular towers. A larger C allows the model to learn more fine-grained Routine Pattern, while inevitably introducing higher variance to the learnt Routine Pattern. In the extreme case, when C equals the number of cellular towers, *ROI* sequence reduces to raw location trajectory and loses all of its semantic meaning. In this case, the model would be highly sensitive to even a slight change in actual location trajectory (e.g., user picking a different road to go to work) and the learnt Routine Patterns cannot be expected to be generalizable. On the other hand, if C is too small, Routine Patterns would be vague and have little explanatory power.

Using the distribution of *POIs* associated each of these five *ROIs*, we interpret their semantic meanings as follows:

(1) comprehensive area (which includes lots of accommodation, life service, Recreation & Entertainment, education and diet);
(2) office area (which includes lots of companies and some traffic facilities);
(3) entertainment area (which includes lots of recreation and some traffic facilities);
(4) common area (which is similar to comprehensive area, with smaller number of POI);
(5) other area.

Figure 8 plots the 2D version of daily trajectories. Each line represents a day, and five colors corresponds to five ROI labels obtained from clustering.

We compare three different methods of routine mining in their ability to group daily data by underlying routine category: the clustering method in Lv et al. (2013), LDA method in Farrahi and Gatica-Perez (2010b), and the STRMM model. We set $R = 3$, $C = 5$, $\tau = 48$. Partition of timespan remains the same as in last section. We use the mined routines to classify original daily trajectories into three categories, as shown in Figures 9–11. Classification methods are as follows: for STRMM model, we compute the posterior probabilities of each day's trajectory subject to the three candidate routines and choose the highest one; for clustering method, we compute the cosine distance between each day's trajectory and the centers of cluster and choose the one with shortest distance; for LDA, we take the most dominant (the one that has highest probability in topic

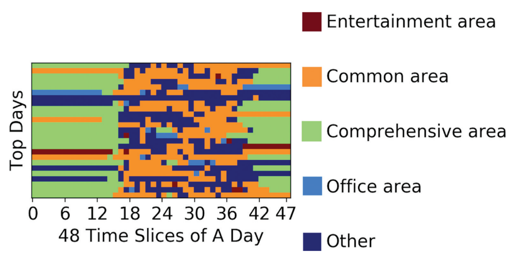
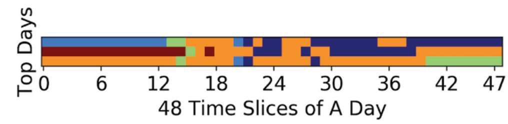
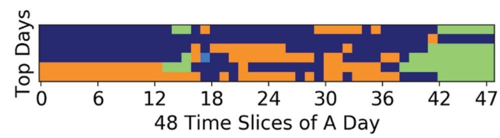


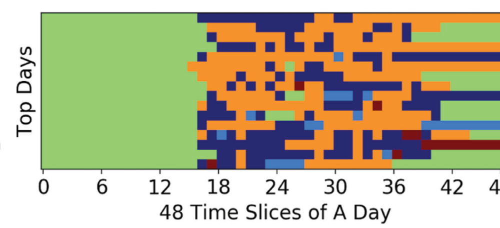
Fig. 8. Location trajectories for a given user.



(a) Days subject to the 1^{st} routine

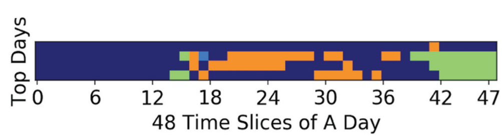


(b) Days subject to the 2^{nd} routine

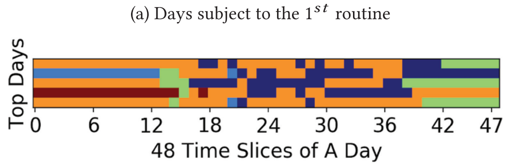


(c) Days subject to the 3^{rd} routine

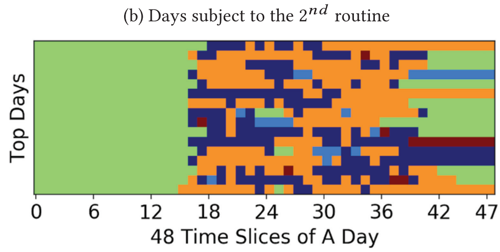
Fig. 9. Classification result by the STRMM model.



(a) Days subject to the 1^{st} routine

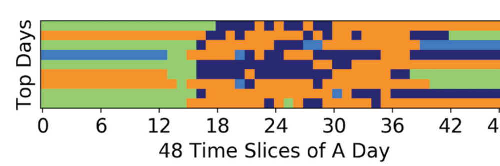


(b) Days subject to the 2^{nd} routine

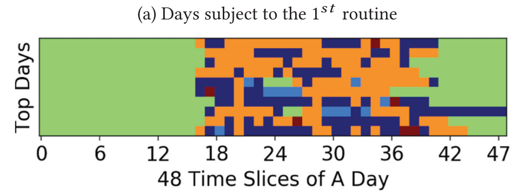


(c) Days subject to the 3^{rd} routine

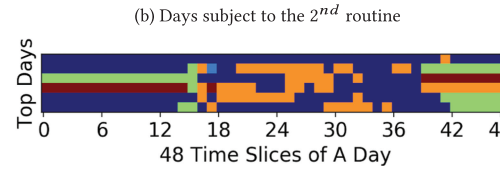
Fig. 10. Classification result by clustering.



(a) Days subject to the 1^{st} routine



(b) Days subject to the 2^{nd} routine

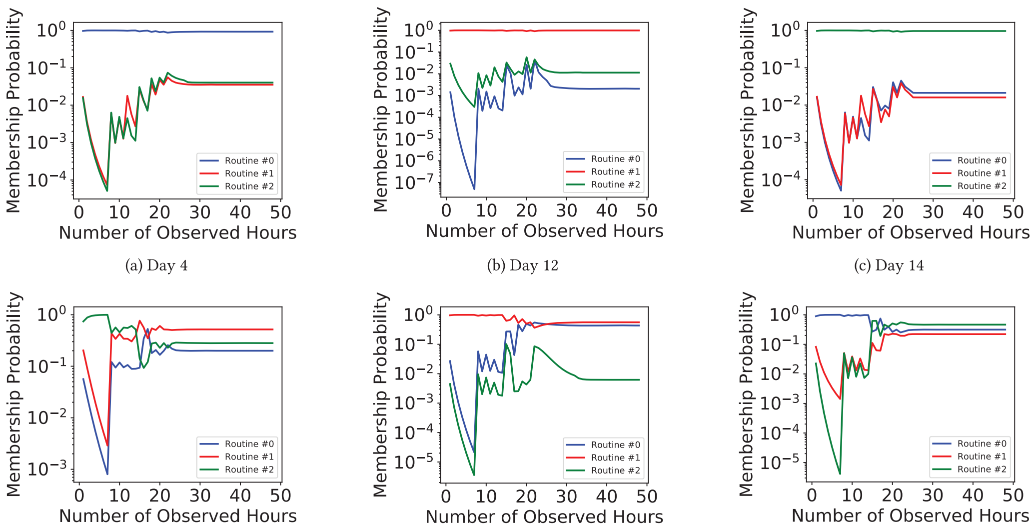


(c) Days subject to the 3^{rd} routine

Fig. 11. Classification result by LDA.

distribution) latent topic in each day's document to represent that day's routine, similar to Farrahi and Gatica-Perez (2008b).

Routines mined by the three methods have different properties. In Figure 11, LDA method fails to distinguish some trajectories in Figure 11(a) and (b), for example, the 1th line and 5th



(a) Day 4 (b) Day 12 (c) Day 14

(d) Day 6 (e) Day 8 (f) Day 11

Fig. 12. Routine category prediction.

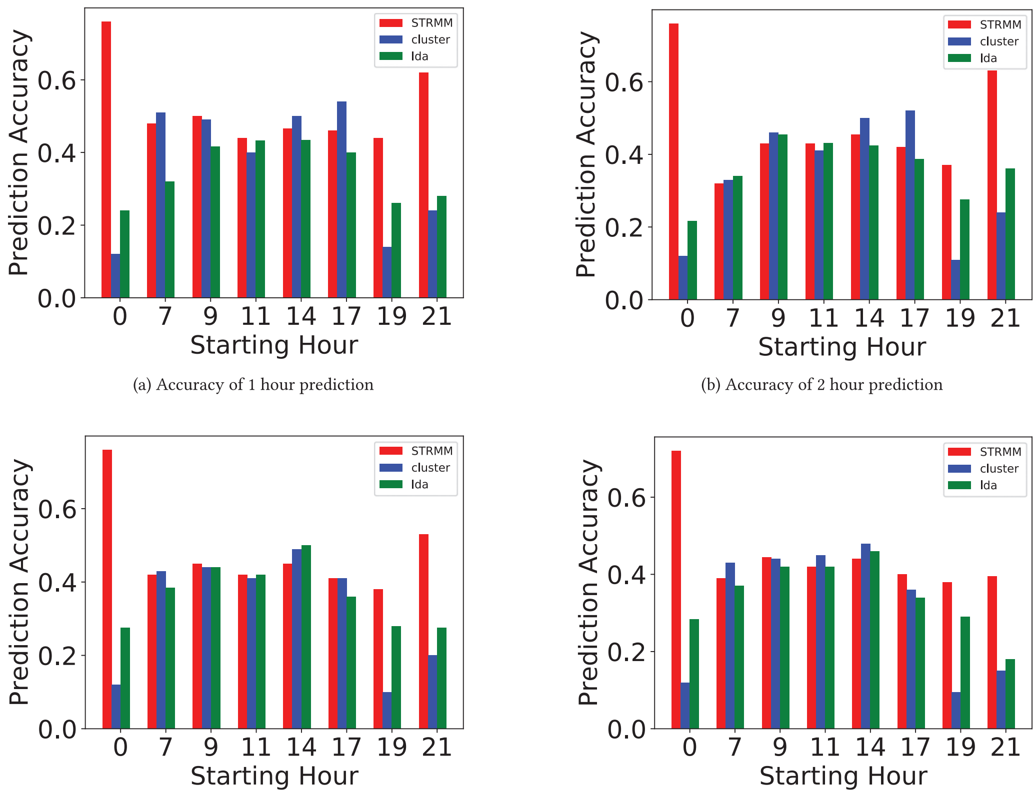
line in Figure 11(b), which is likely the consequence of "partial pattern" discussed in Section 2. Clustering method treats every timespan of the trajectory equally. Therefore, as can be seen from Figure 10(b) and (c), classifying of daily trajectories are based largely on morning trajectories in this case and daytime in some other cases. While in Figure 9, classifying of daily trajectories by the STRRM always focuses more on trajectories at night when user's behaviors are less uncertain. This is because the STRRM model is able to automatically learn the degree of uncertainty in each timespan. During daytime when user's behaviors are more complex and uncertain, the corresponding Dirichlet distribution of this timespan would have a higher entropy and allows more flexible trajectories from day to day. Thus, the routine mined from STRRM model would "assign a higher weight" to the timespans of less uncertainty, rather than treat each timespan equally. This property makes the mined routines more intuitive and meaningful.

Next, we apply routine mining technique to location prediction. Being able to accurately predicate user's next move is a way to prove the validity of the mined routine and the effectiveness of the routine mining model. Here, two kinds of prediction are made. Based on the observed trajectory of the first several hours of a day, we could first predict which routine pattern this day is following (routine category prediction). Based on the predicted routine category, we then predict user's location in the next few hours (trajectory prediction).

To predict the routine category of the ith day, we train the STRRM model with the partially observed trajectories and learn parameter $\Theta' = \{\alpha,'' \pi'\}$. Given the ith day's partial observed trajectory $\boldsymbol{V}'_{\boldsymbol{i}}$, we have

$$\begin{aligned} P(r_i \mid \boldsymbol{V}'_{\boldsymbol{i}}, \Theta') &\propto P(r_i \mid \Theta')P(\boldsymbol{V}'_{\boldsymbol{i}} \mid r_i, \Theta') \\ &= {\pi'}_{r_i} f\left(\boldsymbol{V}'_{\boldsymbol{i}} \mid {\alpha'}^{(\cdot)}_{r_i}\right). \end{aligned} \tag{8}$$

In Figure 12, we plot the prediction result of six days' routine category. X-axis denotes (half) hours of observed trajectory. Y-axis denotes the membership probability of this day subject to each of the three routine categories. Figure 12(a)–(c) are the three days with highest likelihood, while Figure 12(d)–(f) are the three days with lowest likelihood. Apparently the first three days



(a) Accuracy of 1 hour prediction

(b) Accuracy of 2 hour prediction

(c) Accuracy of 3 hour prediction

(d) Accuracy of 4 hour prediction

Fig. 13. Location prediction accuracy.

(Figure 12(a)–(c)) are easier to predict, with one line always higher than the other two. However, the three days with lowest likelihood (Figure 12(d)–(f)) are much harder to predict, and three lines twist together all the way. The results suggest that STRMM model quickly recognizes typical traveling behaviors if it is subject to one of the learnt routine patterns.

Last, we conduct trajectory prediction task and compare three methods (Cluster (Lv et al. 2013), LDA (Farrahi and Gatica-Perez 2010b), the STRRM model) on their prediction accuracy. Let $\boldsymbol{V}_i'$ denotes the partially observed trajectory, $\boldsymbol{V}_i^*$ denotes the trajectories in the next few hours to be predicted. Let $\boldsymbol{V}_i'' = \{\boldsymbol{V}_i', \boldsymbol{V}_i^*\}$.

For the STRMM model, we fill $\boldsymbol{V}_i^*$ with trajectories on the same time of other days $\boldsymbol{V}_{-i}^*$, respectively, and get $D-1$ possible complete trajectories $\boldsymbol{V}_j'', j \neq i$. Then, we compute the membership probability of these $D-1$ complete trajectories according to Equation (8) and get $(D-1) \times R$ probability value. We select the maximum value, and output the corresponding $\boldsymbol{V}_j^*$ as the predicted location. For clustering methods, the procedure is similar except that the calculation the membership probability is replaced by distance. For LDA methods, we adopt the trajectory prediction algorithm proposed in Farrahi and Gatica-Perez (2010b).

Prediction accuracy of these three methods is shown in Figure 13. The x-axis is the starting time from which we predict, and y-axis is the prediction accuracy. Subgraphs (a)–(d) show the accuracy of 1 hour, 2 hour, 3 hour, and 4 hour prediction, respectively.

From Figure 13, we can see that the STRRM model outruns other two methods in most of the time. The STRMM model (red) has higher prediction accuracy at night than at daytime because there is less degree of uncertainty at night. Yet this phenomenon is not observed on clustering method (blue) and the LDA model (green). Further, prediction accuracy of clustering methods in 0:00–7:00 is below 20% in all four subgraphs, which suggests that clustering method fails to take into account the different degree of uncertainty in different timespan of a day and may have emphasized too much on those timespans with high level of uncertainty (daytime). As for LDA (green), its limitation lies in that the mined routines are only partial patterns of the day. In most cases, partial pattern is not enough to represent the whole day's pattern, and thus yields lower prediction accuracy.

7 CONCLUSION

In this article, we study how to mine routine patterns from raw location trajectories of mobile subscribers. We model people's commute behaviors as three levels: location trajectory, one-day pattern, and routine pattern. We further propose to use Multinomial distribution and Dirichelt distribution to describe their generative process. Stochastic Expectation Maximization algorithm is adopted for learning and inference of the model. The output of the model includes the probability of routine categories π and routine pattern α. Two characteristic of routine pattern α are discussed. First, it unveils typical one-day pattern subject to this routine. Second, it shows the uncertainty of one-day patterns in each timespan.

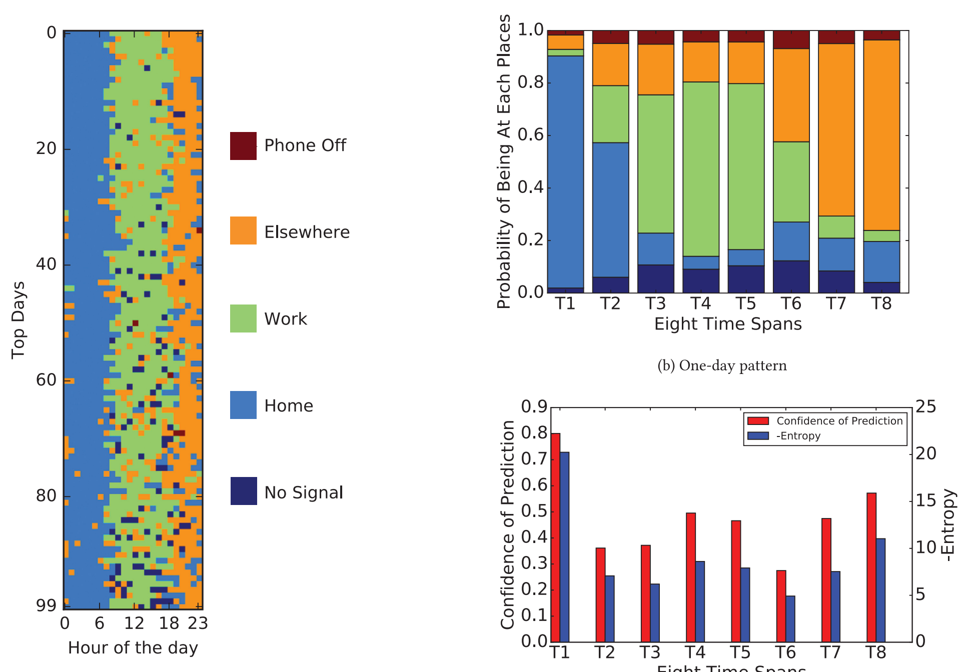
Comprehensive experiments are conducted on synthetic dataset and two real-world datasets. On Reality Mining dataset, we explore the use of perplexity to determine the best routine number R. Then, we graphically present two of the forty mined routines. On China Mobile Operator dataset, we compare three methods on their effectiveness of mining meaningful routines. We also conduct routine category prediction and trajectory prediction tasks. Empirical results show that STRMM model outruns the other methods on prediction accuracy.

Future works may further explore the Markov property of location trajectories. Also, *ROI* extraction and routine mining could be co-modeled, instead of being separate processes as in this article. We argue that they could be mutual beneficial: routine mining requires meaningful location labels from ROI extraction, while in turn ROI extraction could benefits from user's spatio-temporal routine semantics on each location. Apart from ROI extraction, other advanced trajectory preprocessing techniques such as segmentation, compression can also be used to obtain location trajectories with higher information density and thus lead to more meaningful routines.

APPENDIX

We visualize some of the mined routine patterns from Reality Mining Dataset. All users' data are included and here routine refers to common travel behaviors of members of this MIT community.

The 10th mined routine (Figure 14) depicts a common travel pattern that users stay at home before 7:00a.m., go to work from 7:00a.m. to 17:00p.m., and go out for recreation at night. This is one of the typical weekday patterns.



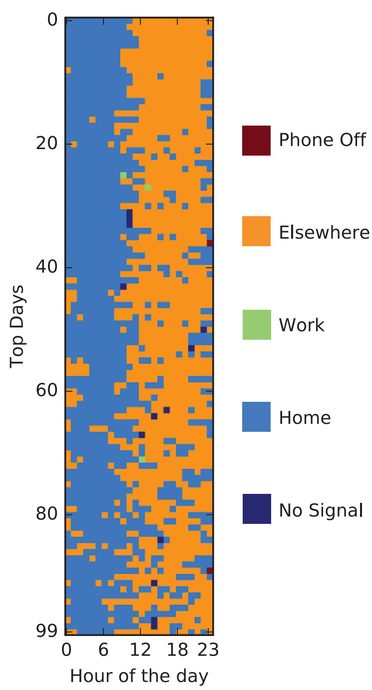
(a) Most typical 100 days following this routine

(b) One-day pattern

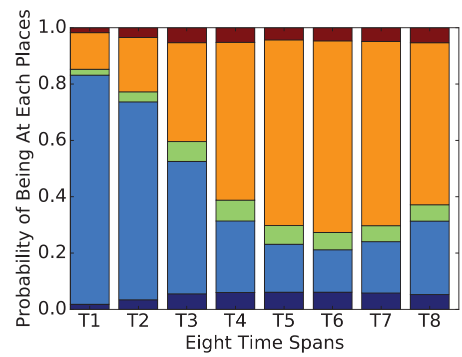
(c) Degree of uncertainty

Fig. 14. The 10th mined routine pattern. (a) plots the most typical 100 days in the dataset following to this routine pattern. From bottom to top the likelihood of data increases. Each line is the location trajectory of a day. Five colors stand for five location labels, respectively. (b) is the stacked column diagram plot of the one-day pattern of the this routine, calculated according to Equation (1). Figure (c) measures the degree of uncertainty of the one-day pattern on different timespan of a day. First, we compute the confidence of prediction (red column) by calculating the probability that location trajectories is generated from the typical one-day pattern for each timespan. Then, we compute the entropy of Dirichelt distribution under each timespan according to Equation (2). The blue column is the negative entropy.

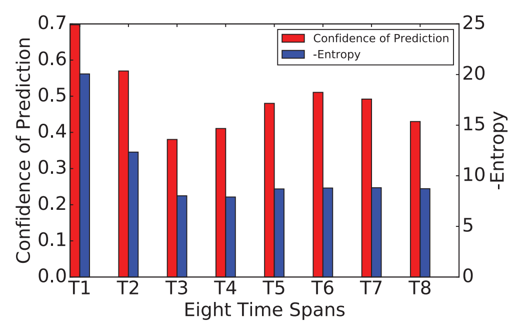
The 20th mined routine (Figure 15) depicts a common travel pattern that users stay at home before 7:00am, and go outside until midnight. This routine probably represent activities such as “go out camping” or “on business trip.”



(a) Most typical 100 days following this routine



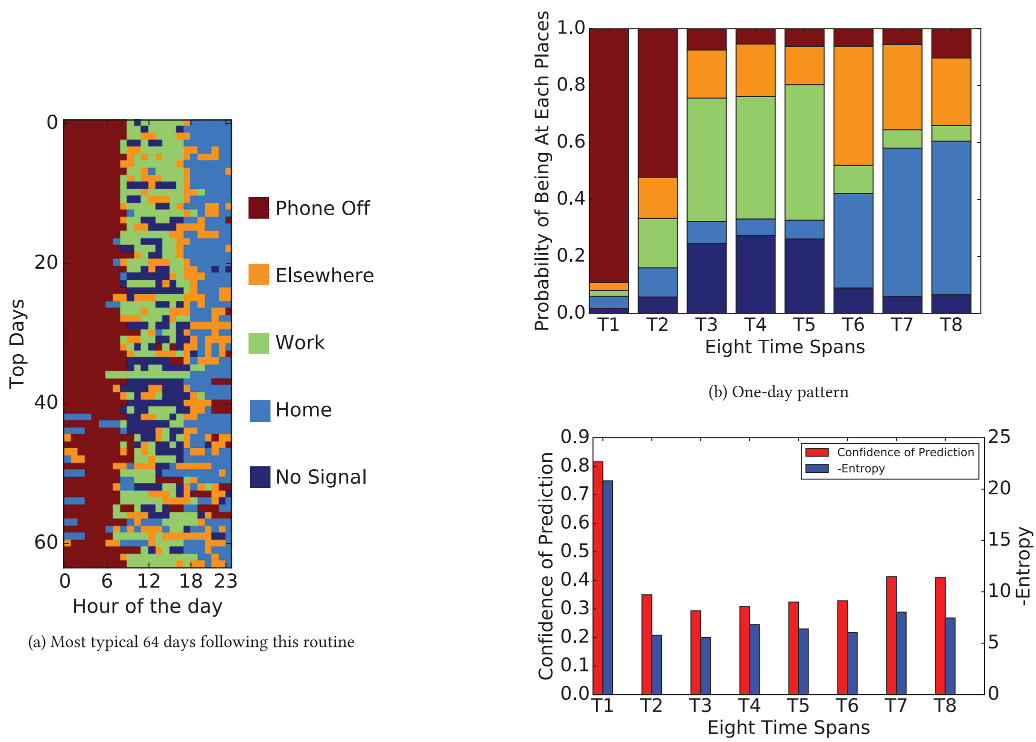
(b) One-day pattern



(c) Degree of uncertainty

Fig. 15. The 20th mined routine pattern. (a) plots the most typical 100 days in the dataset following this routine pattern. From bottom to top the likelihood of data increases. Each line is the location trajectory of a day. Five colors stand for five location labels, respectively. (b) is the stacked column diagram plot of the one-day pattern of the this routine, calculated according to Equation (1). Figure (c) measures the degree of uncertainty of the one-day pattern on different timespan of a day. First, we compute the confidence of prediction (red column) by calculating the probability that location trajectories is generated from the typical one-day pattern for each timespan. Then, we compute the entropy of Dirichelt distribution under each timespan according to Equation (2). The blue column is the negative entropy.

The 38th mined routine pattern (Figure 16) is also a typical weekday routine except that people tend to turn their phones off while sleeping.



(a) Most typical 64 days following this routine

(b) One-day pattern

(c) Degree of uncertainty

Fig. 16. The 38th mined routine pattern. (a) plots the most typical 64 days in the dataset following this routine pattern. From bottom to top the likelihood of data increases. Each line is the location trajectory of a day. Five colors stand for five location labels, respectively. (b) is the stacked column diagram plot of the one-day pattern of the this routine, calculated according to Equation (1). Figure (c) measures the degree of uncertainty of the one-day pattern on different timespan of a day. First, we compute the confidence of prediction (red column) by calculating the probability that location trajectories is generated from the typical one-day pattern for each timespan. Then, we compute the entropy of Dirichelt distribution under each timespan according to Equation (2). The blue column is the negative entropy.

REFERENCES

Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, and Anind K. Dey. 2016. Modeling and understanding human routine behavior. In *Proceedings of the 2016 ACM Annual Conference on Human Factors in Computing Systems.*

Jie Bao, Yu Zheng, David Wilkie, and Mohamed F. Mokbel. 2015. Recommendations in location-based social networks: A survey. *GeoInformatica* 19, 3 (2015), 525–565.

C. M. Bishop 2009. *Pattern Recognition and Machine Learning*. Springer.

Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. 2008. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* (2008).

Gilles Celeux, Didier Chauveau, and Jean Diebolt. 1996. Stochastic versions of the em algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation* (1996).

Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. 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, 1082–1090.

C. Y. Chow, J. Bao, and M. F. Mokbel. 2010. Towards location-based social networking services. In *Proceedings of the ACM SIGSPATIAL International Workshop on Location Based Social Networks.*

Ke Deng, Kexin Xie, Kevin Zheng, and Xiaofang Zhou. 2011. Trajectory indexing and retrieval. In *Computing with Spatial Trajectories*. Springer, New York, NY, 35–60.

Nathan Eagle, Alex Pentland, and David Lazer. 2009. Inferring social network structure using mobile phone data. *Proceedings of the National Academy of Sciences* 106, 36 (2009), 15274–15278.

Nathan Eagle and Alex (Sandy) Pentland. 2006. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing* 10 (2006), 255–268.

Nathan Eagle and Alex (Sandy) Pentland. 2010. Eigenbehaviors: Identifying structure in routine. In *Proceedings of Ubicomp.*

Katayoun Farrahi and Daniel Gatica-Perez. 2008a. Daily routine classification from mobile phone data. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, Berlin, Heidelberg, 173–184.

Katayoun Farrahi and Daniel Gatica-Perez. 2008b. What did you do today? Discovering daily routines from large-scale mobile data. In *Proceedings of ACM Multimedia Conference.*

Katayoun Farrahi and Daniel Gatica-Perez. 2010a. *Mining Human Location-Routines Using a Multi-Level Topic Model*. IDIAP Research Report.

Katayoun Farrahi and Daniel Gatica-Perez. 2010b. Probabilistic mining of socio-geographic routines from mobile phone data. *IEEE Journal of Seleccted Topics in Signal Processing* 4, 4 (2010), 746–755.

Katayoun Farrahi and Daniel Gatica-Perez. 2011. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 1 (2011), 3.

Katayoun Farrahi and Daniel Gatica-Perez. 2012. Extracting mobile behavioral patterns with the distant N-gram topic model. In *Proceedings of International Symposium on Wearable Computers.*

Marta C. Gonzalez, Cesar A. Hidalgo, and Albert Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779.

T. Hastie, R. Tibshirani, and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

Baoxing Huai, Enhong Chen, HengShu Zhu, Hui Xiong, Tengfei Bao, Qi Liu, and Jilei Tian. 2014. Toward personalized context recognition for mobile users: A semisupervised bayesian HMM approach. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9, 2 (2014), 10.

Shan Jiang, Joseph Ferreira, and Marta C. González. 2012. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery* 25, 3 (2012), 478–510.

Georgios Kellaris, Nikos Pelekis, and Yannis Theodoridis. 2009. *Trajectory Compression under Network Constraints.* Springer, Berlin, 392–398. http://dx.doi.org/10.1007/978-3-642-02982-0_27

Wang-Chien Lee and John Krumm. 2011. *Trajectory Preprocessing*. Springer, New York, New York, NY, 3–33. DOI: http://dx.doi.org/10.1007/978-1-4614-1629-6_1

Xiong Liang, Barnabás Póczos, and Jeff G. Schneider. 2011a. Group anomaly detection using flexible genre models. In *Advances in Neural Information Processing Systems.* 1071–1079.

Xiong Liang, Barnabás Póczos, Jeff Schneider, Andrew Connolly, and Jake VanderPlas. 2011b. Hierarchical probabilistic models for group anomaly detection. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics.* 789–797.

Siyuan Liu, Shuhui Wang, Kasthuri Jayarajah, Archan Misra, and Ramayya Krishnan. 2013. TODMIS: Mining communities from trajectories. In *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management (CIKM'13).* ACM, New York, NY, 2109–2118. DOI: http://dx.doi.org/10.1145/2505515.2505552

Mingqi Lv, Ling Chen, and Gencai Chen. 2013. Mining user similarity based on routine activities. *Information Sciences* 236 (2013), 17–32.

Gregory R. Madey, Gabor Szabo, and Albert-László Barabási. 2006. WIPER: The integrated wireless phone based emergency response system. In *Proceedings of Computational Science – ICCS.*

Thomas P. Minka. 2003. Estimating a dirichlet distribution. *Conference on Uncertainty in AI* (2003).

R. Montoliu and D. Gatica-Perez. 2010. Discovering human places of interest from multimodal mobile phone data. In *Proceedings of International Conference on Mobile and Ubiquitous Multimedia.*

Powers David M. W. 1998. Applications and explanations of Zipf's law. In *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning.* Association for Computational Linguistics, 151–160.

S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. 2011. Socio-spatial properties of online location-based social networks. In*Proceedings of the 5th International Conference on Weblogs and Social Media.*

Jiang Shan, Fiore Gaston A., Yang Yingxiang, Ferreira Joseph, Frazzoli Emilio, and González Marta C. 2013. A review of urban computing for mobile phone traces. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*.

X. Xiao, Y. Zheng, Q. Luo, and X. Xie. 2010. Finding similar users using category-based location history. In *Proceedings of ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems*.

Yang Ye, Yu Zheng, Yukun Chen, Jianhua Feng, and Xing Xie. 2009. Mining individual life pattern based on location history. In *Proceedings of the 10th International Conference on Mobile Data Management*.

Nicholas Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th SIGKDD Conference on Knowledge Discovery and Data Mining*.

Nicholas Jing Yuan, Yu Zheng, Xing Xie, Yingzi Wang, Kai Zheng, and Hui Xiong. 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering* 27, 3 (2015), 712–725.

Yu Zheng. 2015. Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 29.

Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: Concepts, methodologies, and applications. *ACM Transaction on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.

C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen. 2007. Discovering personally meaningful places: An interactive clustering approach. *ACM Transactions on Information Systems* 25 (2007), Article 12.

Received January 2017; revised March 2018; accepted March 2018