Understanding Urban Dynamics via State-Sharing Hidden Markov Model

Tong Xia, Yong Li, *Senior Member, IEEE*, Yue Yu, Fengli Xu, Qingmin Liao, *Member, IEEE*, and Depeng Jin, *Member, IEEE*

Abstract—With the ever-increasing urbanization process, systematically modeling people's activities in the urban space is being recognized as a crucial socioeconomic task. It is extremely challenging due to the lack of reliable data and suitable methods, yet the emergence of population-scale urban mobility data sheds new light on it. However, recent works on discovering activity patterns from urban mobility data are still limited in terms of concisely and specifically modeling the temporal dynamics of people's urban activities. To bridge the gap, we present a State-sharing Hidden Markov Model (SSHMM), a novel time-series modeling method that uncovers urban dynamics with massive urban mobility data. SSHMM models the urban dynamics from two aspects. First, it extracts the urban states from the whole city, which captures the volume of population flows as well as the frequency of each type of Point of Interests (PoIs) visited. Second, it characterizes the urban dynamics of each urban region as the state transition on the shared-states, which reveals distinct daily rhythms of urban activities. We evaluate our method via large-scale real-life mobility dataset. The results demonstrate that SSHMM learns semantics-rich urban dynamics, which are highly correlated with the functions of the region. Besides, it recovers the urban dynamics in different time slots with RMSE of 0.0793 when only learn limited states for the whole city, which outperforms the general HMM by 54.2 percent.

Index Terms—Urban computing, time-series analysis, urban dynamics, mobility, hidden markov model

1 Introduction

The rapid urbanization process has been nurturing large and complex urban systems worldwide. It is estimated that over the next thirty years, more than two-thirds of the people will dwell in modern cities [1]. Therefore, understanding urban dynamics, namely, the temporal patterns of urban activities, is fundamental for tackling the increasingly prominent urban challenges, e.g., excessive energy consumption, air pollution, and traffic congestion [37]. However, citizen's activities in the urban space are extremely complex and highly volatile, which poses challenges to model temporal urban dynamics precisely and systematically. Traditional approaches rely on expensive manual surveys, yet the understanding is still coarse-grained and limited in geographical scope [2].

Fortunately, the advent of the ubiquitous mobile Internet and Location-Based Social Networks (LBSNs) makes it possible to collect population-scale urban mobility data, which sheds new light on this open problem. These datasets contain semantic-rich human mobility information, which includes the timestamps, location coordinates as well as the visited Points of Interest (PoIs). Previous works have demonstrated that the daily movements of citizens can be utilized to infer the functions of urban regions [30], [33], and the patterns of urban activities (e.g., working, resting, commuting, etc.) are closely correlated with urban mobility patterns [29], [34], which indicates the feasibility of leveraging urban mobility data to model urban dynamics.

In this paper, we aim to harness the power of massive urban mobility data to deepen the understanding of urban dynamics. The research problem is non-trivial mainly for three reasons: (1) Urban mobility behaviour is a *noisy* representation of urban activities. Similar urban activities may correspond to slightly different mobility patterns, e.g., central business districts may experience different population flow during the working hours of different days. Therefore, it is difficult to robustly and accurately infer the underlying urban activities from the empirical observation of urban mobility behaviour. (2) The semantic-rich mobility data, i.e., check-in data on LBSNs, is sparse in urban space, especially in sparsely populated areas. The sparse and unevenly distributed mobility data poses significant challenges to extract reliable patterns of urban dynamics for different regions in urban systems. (3) Modern cities are complicated socioeconomic systems, where each urban region possesses different urban dynamics due to its unique activities. Therefore, it is challenging to interpret the identified urban dynamics and reveal the underlying mechanisms.

Motivated by these challenges, we propose a novel State-sharing Hidden Markov Model (SSHMM) to reveal the urban dynamics. The key insight of SSHMM is that the mobility behavior of an urban region can be viewed as a

- *T. Xia, Y. Li, Y. Yu, F. Xu, and D. Jin are with the Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. E-mail: {xia-t17, yue-yu15, xfl15}@mails.tsinghua.edu.cn, {liyong07, jindp}@tsinghua.edu.cn.*
- *Q. Liao is with the Graduate School at Shenzhen, Tsinghua University, Beijing 100084, China. E-mail: liaoqm@sz.tsinghua.edu.cn.*

Manuscript received 15 Mar. 2019; revised 23 Dec. 2019; accepted 31 Dec. 2019. Date of publication 22 Jan. 2020; date of current version 10 Sept. 2021. (Corresponding author: Yong Li.)
Recommended for acceptance by M. A. Cheema.
Digital Object Identifier no. 10.1109/TKDE.2020.2968432

1041-4347 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

probabilistic observation of the underlying urban activity, and similar activities have the similar probability distribution of the observations across different urban regions since they are likely conducted by similar population. Specifically, the model learns the urban dynamics of a region as transitions between hidden states, where each state maps to a certain urban activity. The corresponding mobility behavior is generated from the hidden state through an emission probability function, which allows the same urban activities to be mapped to slightly different mobility behavior and effectively addresses the problem of noisy mobility data. More importantly, SSHMM facilitates different urban regions to share the same set of hidden states because the similar urban activities in different regions correspond to similar mobility behavior patterns. Therefore, it addresses the challenge of data sparsity by allowing different regions to share parameters, which fully exploits the correlation between different regions. Finally, as a generative model, apart from predicting the mobility behavior of urban regions, SSHMM can also characterize the urban dynamics as hidden state sequences. Based on the identified state sequences, we further design an unsupervised clustering analysis technique to reveal their correlation with urban functions (i.e., land use) and provides a meaningful interpretation of the urban dynamics.

The contributions of our research are three-fold:

1) We propose a novel urban dynamics revealing model SSHMM. It can robustly and accurately infer the underlying urban activities from noisy and sparse urban mobility data by sharing model parameters across different regions. In addition, it achieves qualitative representations of urban dynamics as the transition patterns between urban activities. Compared with previous works, it can model urban dynamic in a concise and probabilistic way.
2) We propose an effective and efficient algorithm to infer the parameters of our model. By splitting the long observations into shorter ones and updating the parameters in parallel, we reduce the training time of learning R groups of parameters to that of only one group, where R is the number of regions.
3) We evaluate our method using a population-scale mobility dataset, which demonstrates that our SSHMM model learns meaningful and explainable urban dynamics. Besides, the activity regularities can be recovered with an error of 0.0793 by 100 states, outperforming the baseline by 54.2 percent. We also achieve RMSE for population flow prediction of 0.195 and *Top3-accuracy* for PoI popularity prediction of 41.4 percent, outperforming the baseline method by 16 and 8 percent respectively.

This paper is an extension of our previous conference paper [28]. In this paper, we further offer the following new contributions. *First*, we design a system framework, which consists of data collecting and preparing module, feature extracting module, model learning module as well as application module. The pipeline of the complete framework makes our work more practical and systematic. *Second*, we provide the detailed theoretical derivation of the model parameter inference to bring more valuable takeaways for the readers who plan to apply the model in their research problems. *Last but not least*, we propose an effective parameter updating algorithm to continuously learn the model when new data available, which shows a significant performance improvement in terms of both activity recovery and prediction.

2 Related Work

2.1 Urban Dynamics Modeling

The development of the city has witnessed a great series of studies on urban dynamic problems. Urban dynamics, generally defined as how sociological indicators (e.g., the population, the land use) change over time [10], can be divided into two aspects. One is to investigate the urbanization and sustained economic growth [19] via the dataset with a long period, while another focuses on characterizing human daily activity rhythms in the city with more fine-grained time granularity [2], [21], which is more relevant to our research. Yuan *et al.* [33] proposed an LDA model to detect different functional regions in a city through the GPS trajectory and PoI. Zhang *et al.* [37] used the geo-tagged social data to model urban activities with more attention paid to temporal dimension, which demonstrated that the activity volume of an area is not uniformly distributed across time and different areas have different activity volume temporal distributions. Sohiance *et al.* [2] built activity time series for different cities and different neighborhoods within the same city to identify different dynamic patterns via the geo-tagged data from Twitter. By clustering the activity time series, they found that close neighborhoods tend to share similar rhythms. Mireanda *et al.* [21] captured the spatio-temporal activity in a city across multiple temporal resolutions and visualized different activity levels in different time. Song *et al.* [25] clustered time slots into different urban states by their population moving volume.

In this problem, we regard urban activities as time series and aim to reveal the daily regularity hidden in them. Different from the existing works based on statistical analysis [2] and data visualization [21], we propose a specific model achieving understanding and prediction at the same time. With the similar data, i.e., human mobility and PoIs, and specific LDA model, our goal to reveal the dynamic regularity is different from Yuan *et al.*'s work to find urban functions [33]. On top of the application of urban function inference, our SSHMM can also applied in human activity prediction.

2.2 Hidden Markov Model and its Application

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e., hidden) states [23], which has been widely used in time series analysis and prediction in the past years. In HMM, model mixture and parameter sharing are very helpful to deal with increasingly complex tasks [13]. One of the well-known mixture methods is the Gaussian mixture model-based HMM (GMM-HMM). GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities, while GMM-HMM assumes that a set of Gaussian components can represent a distribution based on the spectral envelope [11]. The second example is shared-distribution HMM, where clustering is carried out at the distribution level for parameters sharing and output distributions are shared with each other if they exhibit acoustic similarity

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

[15]. Another model tied-mixture HMM uses both mixture and parameter sharing, which belongs to semi-continuous HMM. It is recognized as a useful complexity reduction method, because of its ability to maintain modeling accuracy of large-mixture probability density functions (PDF) by enforcing PDF sharing [4], [14], [18]. One emission-sharing HMM named HTMM also has been proposed recently [12] for document topic modeling. It assumes that the topic matrix is shared among all documents, meaning all the given HMMs share the same emission probability. Yet, each document has a specific topic distribution, whereas each document has its own topic transition probability.

SSHMM is also a kind of parameters sharing HMM. While different from previous works, we design it to automatically learn a set of states for the continuous observations and no following clustering is conducted to force the parameters shared. Both based on the Gaussian emission function, we use multi-dimensional Gaussian instead of GMM. With a similar idea of state-sharing while transition-independent, our model differs from HTMM [12], because HTMM can only deal with discrete distribution but our observations are continuous. Therefore, we derive a novel EM algorithm to infer model parameter. In addition, though HMM has been successfully applied to the topic of mobility modeling, most of the work concentrated on individual mobility prediction [20], [35], [39], and we are the first to apply it in urban dynamics modeling.

3 Overview

3.1 Problem Description

HMM is a generative model, which assumes that the observation sequence is generated by a hidden state sequence. To apply this model to urban dynamics problem, we regard human aggregated activities in different regions as time series. More specifically, we extract mobility behaviours in different time slots as observation sequences, and we aim to reveal urban dynamics in terms of human daily life rhythms by the corresponding hidden state sequence. Before formally define our investigated problem, we give the definition of mobility behaviour observation as follows,

Definition 1 (Mobility Behaviour Observation). *The mobility behaviour sequences of region r is a time-ordered sequence $O_r = [O_{r,1}, O_{r,2}, \ldots, O_{r,N}]$, where $O_{r,n}$ is a tuple of length L, standing for the observation in n-th time slot. It contains two parts: (1) $\{o_{r,n,1}, o_{r,n,2}, o_{r,n,3}\}$ denotes the number of people who arrive at, leave from and stay in this region in this time slot. (2) $\{o_{r,n,4}, o_{r,n,5}, \ldots, o_{r,n,L}\}$ denotes the check-in frequency of different categories of PoIs.*

The motivation for choosing these two types of features as the observation is very intuitive. On the one hand, how many people leave from, arrive at and stay in each region, represents the daily commute pattern and activity level, is crucial for urban dynamics modeling [6], [32]. On the other hand, since dynamic pattern has close relation with the land use, while the PoIs are static, therefore, we utilize the check-ins, which reflects people's dynamic demand for a region's function [38]. As a consequence, these two aspects can give deep insights into urban activities.

Now, we formulate the urban dynamics understanding problem. Given the mobility behaviour observations of a group of regions in the city, we aim to (1) discover hidden states represented by the volume of population and the visiting frequency of different PoIs to understand the intrinsic life modes in the city; (2) reveal urban dynamics represented by hidden state sequences to understand the daily life rhythms and their correlation with urban functions; (3) characterize dynamic regularity by learning state transition probability to achieve dynamic prediction.

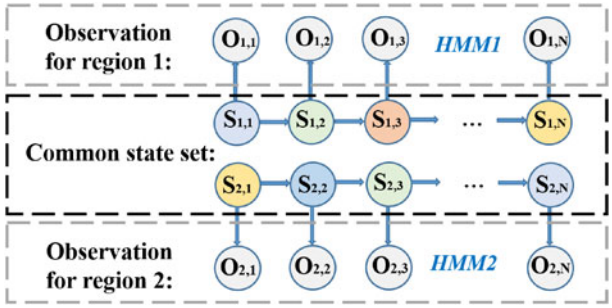


Fig. 1. An illustration of the SSHMM, where two kinds of dynamics for region 1 and 2 are generated from the same set of states and each state is presented by a unique color.

3.2 Model Description

Our problem can be simply abstracted into that when neither the hidden state sequence nor the underlying probabilities are known—we only have access to a sequence of observations, and our job is to reveal the HMM structure. As such, the HMM we need to infer is specified as:

Definition 2.1 (SSHMM). *HMM is parameterized by two parts, one is the state transition parameter characterizing how the states transit and the other is the state emission parameter characterizing how the observation generated by the state. SSHMM contains R groups of HMMs for R groups of observation sequences. However, these HMMs share state emission parameter, which means all the observations are generated by the same set of hidden states.*

Definition 2.2 (Hidden state set). *We define the hidden state set including K hidden states as $S = \{s_1, s_2, \ldots, s_K\}$, where each state describes L dimensional features.*

Our model is based on the general assumptions of HMM, where each observation $O_{r,n}$ is generated from a hidden state s_n, and the n-th hidden state s_n merely depend on the previous state s_{n-1}. In our problem, we use one HMM denoted by θ_r to model the dynamics of r-th region, therefore we need to learn R groups of transition parameters for R regions in the city. Fig. 1 gives the illustration of our model. There are two observation sequences for region 1 and 2, so we learn two HMMs for them respectively. These observations are generated by the same set of states, where each state is presented by a unique color. These states appear in different time slots, which reveals the different dynamics of these two regions. Inspired by previous works indicating that different regions present similar states in different time slots [9], [21], [27], we intuitively share the states for the whole city. However, as shown in [33], although topics in human mobility and PoI are common, each region has its own topic distribution, which means we should learn unique state transition for each region. That is why we only share state on emission models instead of sharing transition probability.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

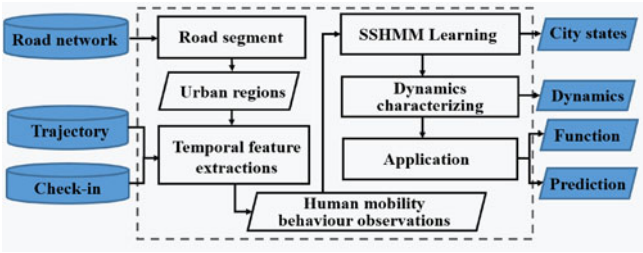


Fig. 2. The framework of our system.

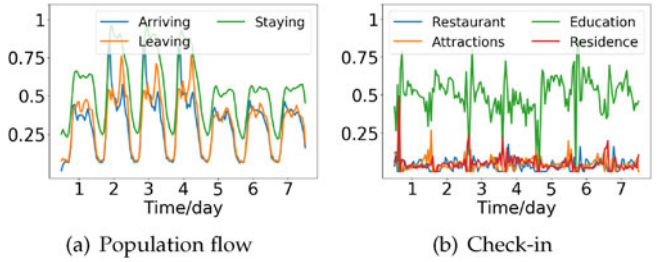


Fig. 3. Normalized mobility behaviour observations for Tsinghua University, Beijing from April 1st to 7th, 2018.

Consequently, the key insights of sharing states are two aspects. On the one hand, we investigate the dynamics of the city, where life modes (i.e., sleeping, working, etc.) is limited and similarities in different regions exist objectively. Therefore, it is reasonable to use a common state set and it also makes the following dynamics analysis through state sequences easier. One the other hand, massive mobility data especially the semantic check-in data is sparse and mobility behaviour is a noisy representation of urban activities, which indicates that by sharing states we can achieve more robust and accurate models.

3.3 System Overview

To learn and apply our propose SSHMM in the real world, we design a system that consists of data collecting and preparing module, feature extracting module, model learning module as well as application module. Along this pipeline, our model is more practical. Fig. 2 summarizes this system framework. We first apply a map segment method [33] to obtain the geographical boundaries of each region formed by the road network. Then we extract the temporal feature, i.e., the population flow and check-in frequency in each time slot for each region and normalize them as mobility behaviour observations. After that, we learn SSHMM to discover the city states and reveal the dynamics. Besides understanding the behind dynamic regularity, SSHMM can be further applied for urban function identification and human activity prediction.

4 Methodology

In this section, we introduce the implement of our system according to the pipeline as shown in Fig. 2.

4.1 Data Preparing

We first adopt the map segment method [33] to obtain the geographical boundaries of each region formed by the road network. In view of the urban mobility patterns and lifestyles, those regions are used as a unit to reveal the dynamics instead of simply dividing the city into grids. Then, we process individual trajectory and check-in dataset into aggregated mobility behaviour of each urban region.

To share the states in the city, we normalize mobility behaviour observations to eliminate the problem regarding the difference in population between regions. For mobility, i.e., the number of arriving, leaving and staying of citizens, we directly conduct min-max normalization for each region over different time slots respectively. For check-ins, i.e., the visit frequency of PoIs, we first compute the TF-IDF weights based on the region-PoI matrix in each time slot [22]. Then, for each region, we conduct the min-max normalization on the TF-IDF weights over different time slots. After normalization, each dimension of the mobility behaviour observations is rescaled to $0-1$. Fig. 3 gives an example of the normalized mobility behaviour observations. From 3(a) we can observe that both the volume of population flow and the check-in frequency change from morning to night every day.

After pre-processing, we regard the aggregated activities in the city as normalized time series. As defined in Definition 1, we present the mobility behaviour observations for R regions in the city as $O = \{O_1, O_2, \ldots, O_r, \ldots, O_R\}$ with O_r denoting the observations for r-th region.

4.2 SSHMM Learning

To reveal urban dynamics by the hidden states sequence, the first step is to learn the model parameters by maximizing the probability of the observation sequences given the model. Since SSHMM is a novel model proposed by us, how to learn the model parameters reasonably and effectively, and how to apply it in the real-world large-scale dataset are two difficulties we face. To clear introduce our solution, we first give the model design. After that, we present the parameter inference process from theoretical derivation to programming implementation. Last but not least, considering that new data can be continuously collected, we introduce how to apply and update our model online.

4.2.1 Model Design

In this section, we give the basic assumptions and the formula definition. When generating $o_{r,n}$ from s_n, we assume that the emission probability is Gaussian distribution, i.e., $p(o_{r,n,l}|s_n) = N(o_{r,n,l}|\mu_n, \sigma_n)$. The reason to chose the Gaussian is intuitive: μ_n describes the fundamental feature of s_n, while the corresponding observations generated by s_n is slightly different from the mean μ_n with the difference controlled by the variance σ_n.To make the following parts more readable, we consistently use the symbols for description as shown in Table 1.

In summary, we build the SSHMM parameterized by R groups of parameters $\theta = \{\theta_1, \theta_2, \ldots, \theta_r, \ldots, \theta_R\}$ with $\theta_r = \{\pi_r, A_r, \mu, \sigma\}$ denoted for r-th region, where

1) $\pi_r \in \mathbb{R}^{K\times 1}$ denotes the initial distribution over K hidden states, i.e., $\pi_{r,k} = p(s_1 = k)(1 \le k \le K)$;
2) $A_r \in \mathbb{R}^{K\times K}$ denotes the transition probabilities among the K hidden states. If $(n-1)$-th state is $s_{n-1} = j$, then the probability for n-th state s_n to be k is given by $A_{r,j,k}$, i.e.,$p(s_n = k|s_{n-1} = j) = A_{r,j,k}$;

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
Notation and Description

Symbol	Description
$_r, R$	r-th region,number of regions
$_n, N$	n-th time slot, number of time slots
$_l, L$	l-th feature, dimension of features
o, O	Observation in one time slot, observation sequence
s, S	Hidden state in one time slot, hidden state set
θ	HMM parameter set, including π, A, μ, σ
π	Hidden state initial probability
A	Hidden state transition probability
μ, σ	Mean and variance of the Gaussian distribution

3) $\mu, \sigma \in \mathbb{R}^{K\times L}$ denotes the mean and variance of observation probability, i.e., $p(O_{r,n}|s_n = k) = \prod_{l=1}^{L} \frac{1}{\sqrt{2\pi\sigma_{k,l}}} \exp(-\frac{(o_{r,n,l}-\mu_{k,l})^2}{2\sigma_{k,l}})$.

It is worth noting that we do not use the subscript r to distinguish μ, σ of different HMMs because in SSHMM they are determined by the same common state set.

4.2.2 Parameter Inference

To infer the parameters, we use Expectation-Maximization method (EM) as the solution. Since SSHMM shares the same set of hidden states, the existing famous Baum-Welch algorithm cannot be applied directly. To address this, we give a new parameter derivation process, and in order to show it more clearly, we first give some definitions and theorem as follows.

Definition 3.1 (Log likelihood $L(\theta)$). *Let the log likelihood for r-th region $L(\theta_r) = \ln p(O_r|\theta_r) = \sum_S \ln p(O_r|S, \theta_r)$ $p(O_r|\theta_r)$. Therefore, the log likelihood for all observation sequences $L(\theta)$ can be derived as follows,*

$$\begin{aligned} L(\theta) &= \ln p(O|\theta) = \sum_S \ln p(O|S,\theta)p(O|\theta) \\ &= \sum_S \ln \prod_{r=1}^{R} p(O_r|S,\theta_r)p(O_r|\theta_r) \\ &= \sum_{r=1}^{R}\sum_S \ln p(O_r|S,\theta_r)p(O_r|\theta_r) \\ &= \sum_{r=1}^{R} L(\theta_r). \end{aligned} \tag{1}$$

Definition 3.2 (Q-function $Q(\theta, \theta^t)$). *Let the Q-function for r-th region $Q(\theta_r, \theta_r^t) = \sum_S p(S|O_r; \theta_r^t) \ln p(O_r, S|\theta_r)$, where θ_r means the old parameter. Therefore, the total Q-function for regions $Q(\theta, \theta^t)$ can be derived as follows,*

$$\begin{aligned} Q(\theta,\theta^t) &= \sum_S \ln p(O|S,\theta)p(S|O,\theta^t) \\ &= \sum_S \ln \prod_{r=1}^{R} p(O_r|S,\theta_r)p(S|O_r,\theta_r^t) \\ &= \sum_{r=1}^{R}\sum_S \ln p(O_r|S,\theta_r)p(S|O_r,\theta_r^t) \\ &= \sum_{r=1}^{R} Q(\theta_r,\theta_r^t). \end{aligned} \tag{2}$$

Theorem 1. *To achieve the maximized log likelihood, parameters can be updated by maximizing $Q(\theta, \theta^t)$ step by step.*

Proof. For each HMM parameterized by θ_r, Baum and his colleagues have proven that maximization of $Q(\theta_r, \theta_r^t)$ leads to increased likelihood, i.e,

$$\max_{\theta_r^t} Q(\theta_r, \theta_r^t) \Rightarrow p(O_r|\theta_r^t) \geq p(O_r|\theta_r). \tag{3}$$

Eventually the likelihood converges to a critical point. Since as derived in (2) and (1),

$$Q(\theta,\theta^t) = \sum_{r=1}^{R} Q(\theta_r,\theta_r^t), L(\theta) = \sum_{r=1}^{R} L(\theta_r). \tag{4}$$

Take the summary of the Q-function for all regions, the following result from (3) could be obtained,

$$\sum_{r=1}^{R} \max_{\theta_r^t} Q(\theta_r, \theta_r^t) \Rightarrow \sum_{r=1}^{R} p(O_r|\theta_r^t) \geq p(O_r|\theta_r). \tag{5}$$

Combine (4) with (5), we can derive the relationship between Q-function and log likelihood as follows,

$$\max_{\theta^t} Q(\theta, \theta^t) \Rightarrow p(O|\theta^t) \geq p(O|\theta). \tag{6}$$

This is to say, by maximizing $Q(\theta, \theta^t)$, the likelihood can converge to its maximum. □

Since $p(O_r|S, \theta_r^t)$ can be calculated as follows,

$$\begin{aligned} p(O_r|S,\theta_r^t) = \pi_{r,k}p(O_{r,1}|s_{r,1}) \cdot p(s_{r,2}|s_{r,1})p(O_{r,2}|s_{r,2}) \\ \cdots (s_{r,N}|s_{r,N-1})p(O_{r,N}|s_{r,N}). \end{aligned} \tag{7}$$

Replace (7) into the (2), the Q-function can be further unfolded as follows,

$$\begin{aligned} Q(\theta,\theta^t) = &\sum_{r=1}^{R}\sum_S p(S|O_r,\theta_r^t)\ln \pi_{r,k} \\ &+ \sum_{r=1}^{R}\sum_S\sum_{n=1}^{N-1} p(S|O_r,\theta_r^t)\ln p(s_{n+1}|s_n) \\ &+ \sum_{r=1}^{R}\sum_S\sum_{n=1}^{N} p(S|O_r,\theta_r^t)\ln p(O_{r,n}|\mu,\sigma), \end{aligned} \tag{8}$$

From (8), we can observe that the parameters π, A (i.e., $p(s_{n+1}|s_n)$), and $\{\mu, \sigma\}$ to be optimized appear separately in three items in the upper formula, so it is only necessary to maximize each item separately. In addition, the above three terms satisfy the equality constraints, so the results can be deduced by Lagrange multiplier method.

We give the detailed parameter derivation as follows. Based on Baum-Welch algorithm, the forward distribution $\alpha(s_{r,n})$ and backward distribution $\beta(s_{r,n})$ are first defined as follows,

$$\begin{aligned} \alpha(s_{r,n}) &= p(O_{r,n}|s_{r,n}) \sum_{s_{r,n-1}} \alpha(s_{r,n-1})p(s_{r,n}|s_{r,n-1}), \\ \beta(s_{r,n}) &= \sum_{s_{r,n+1}} \beta(s_{r,n+1})p(O_{r,n+1}|s_{r,n+1})p(s_{r,n+1}|s_{r,n}), \end{aligned} \tag{9}$$

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

where $\alpha(s_{r,1}) = \pi_{r,k} p(O_{r,1}|s_{r,1} = k)$ and $\beta(s_{r,N} = k) = 1$. Then, two probability can be derived from α and β as follow,

$$\begin{aligned}\gamma(s_{r,n}) &= p(s_{r,n}|O_r) = \alpha(s_{r,n})\beta(s_{r,n})/p(O_r),\\ \xi(s_{r,n}, s_{r,n+1}) &= p(s_{r,n_1}, s_{r,n}|O_r)\\ &= \alpha(s_{r,n-1}) p(s_{r,n}|s_{r,n-1}) P(O_{r,n}|s_{r,n}) \beta(s_{r,n})/p(O_r),\end{aligned} \tag{10}$$

where $p(O_r) = \sum_{s_{r,N}} \alpha(s_{r,n})$.

Since the three terms in (8) are independent, it is easy to optimize the parameters π, A and $\{\mu, \sigma\}$ separately. Due to the limited space, we omit the detailed procedure here.

Finally, the parameters can be inferred as follows,

$$\begin{aligned}\pi_{r,k}^{(t+1)} &= \gamma(s_{r,1}^k),\\ A_{r,j,k}^{(t+1)} &= \frac{1}{\Xi_j} \sum_{n=2}^{N} \xi(s_{r,n-1}^j, s_{r,n}^k),\\ \mu_{k,l}^{(t+1)} &= \frac{1}{\Gamma_K} \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma(s_{r,n}^k) o_{r,n,l},\\ \sigma_{k,l}^{(t+1)} &= \frac{1}{\Gamma_K} \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma(s_{r,n}^k)(o_{r,n,l} - \mu_{k,l}^{(t+1)})^2,\end{aligned} \tag{11}$$

where $\Gamma_K = \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma(s_{r,n}^k)$, $\Xi_j = \sum_{n=2}^{N} \sum_{i=1}^{K} \xi(s_{r,n-1}^j, s_{r,n}^i)$.

Algorithm 1. SSHMM Parameter Learning

Input: Observations $O = \{O_1, O_2, \ldots, O_R\}$, Maximum Iterations $MaxIter$;
Output: $\theta_r = \{\pi_r, A_r, \mu, \sigma\}$ $\forall 1 \le r \le R$;
Procedure:
  Initialization: $t = 0$, initial $\pi_{r,k}^{(0)} = 1/K$, $A_{r,j,k}^{(0)} = 1/K$, $\mu_{k,l}^{(0)} = random(0,1)$, $\sigma_{k,l}^{(0)} = random(0, 0.1)$, $\forall 1 \le j, k \le K, 1 \le l \le L$.
  while $t < MaxIter$ **do**
    for $w = 1, 2, \ldots, W$ **do**
      E-step: $\forall 1 \le r \le R$, calculate $\alpha(s_{r,n})^{(t+1)}$, $\beta(s_{r,n})^{(t+1)}$, $\gamma(s_{r,n})^{(t+1)}$, $\xi(s_{r,n})^{(t+1)}$ **in parallel** based on old parameters $\theta_r^{(t)}$ utilizing the w-th subsequence of O_r.
      M-step: $\forall 1 \le r \le R$, update $\pi_{r,k}^{(t+1)}$ and $A_{r,j,k}^{(t+1)}$ **in parallel**.
      Update $\mu_{k,l}^{(t+1)}$, $\sigma_{k,l}^{(t+1)}$ utilizing $\gamma(s_{r,n})^{(t+1)}$, $\xi(s_{r,n})^{(t+1)}$ $\forall 1 \le r \le R$.
    end
    Update t**:** $t = t + 1$
  end

4.2.3 Model Implement

The derivation of the parameter has been achieved, however, the process could be inefficient and inefficient when applying it to the large-scale mobility dataset due to two aspects. First is about observation sequence length N. A large dataset means long-time observation with a large N. However, with the increase of N, the computation of α and β becomes more complicated until the result exceeds the max value of the floating-point numbers that the computer can store. The second is about the total number of states K. The computational complexity of Baum-Welch algorithm is $O(RNLK^2)$ as we need R groups of parameter, where R is the number of regions, N is the length of observation sequence, L is the dimension of each observation and K is the number of states. The time complexity is quadratic in K, rendering it inefficient for large K. However, if K is not large enough, the model is unable to capture all the dynamic patterns in the city, thus reducing the representational ability of the model.

To overcome these problems, we first split the long observation sequences into several shorter subsequences of length N. In each round, we use only one subsequence. By doing this, we can not only utilize all the data for parameter learning but also avoid float-point number exceeding. Second, from (11), we can observe that the updating of $\pi_{r,k}$ and $A_{r,j,k}$ for region r in independent. Thus, we conduct their updating in parallel, which reduces the training time to that when only learning one group of parameters.

The detailed procedure is shown in Algorithm 1. In each iteration, we run EM-steps for W rounds, and in each round, we fed the subsequences of length N into the model. After calculating the $\alpha(s_{r,n})^{(t+1)}$, $\beta(s_{r,n})^{(t+1)}$, $\gamma(s_{r,n})^{(t+1)}$, $\xi(s_{r,n})^{(t+1)}$, $\pi_{r,k}^{(t+1)}$ and $A_{r,j,k}^{(t+1)}$ *in parallel* for all the regions in each round, we update the state parameters $\mu_{k,l}^{(t+1)}$, $\sigma_{k,l}^{(t+1)}$.

4.2.4 Continuous Learning

In most real-world applications, data incrementally available over time or data come from non-stationary distributions. In our problem, on the one hand, mobility data are continuously collected. On the other hand, existing PoIs have their life and new PoIs will appear, especially the commercial and food PoIs. As such, how to utilize the new data to efficiently and effectively improve model performance is also our concern.

To avoid learning from scratch, SSHMM is initialed by maximizing the log-likelihood of the old data. Then, when new data is available, we fine-tune the model to maximize the log-likelihood on the new data. We denote the new data as O^f, and then the likelihood can be written as $L(\theta^f) = \ln p(O^f|\theta)$. We update the model when each subsequence with length N is prepared. The updating role is the same in (11). We present the algorithm as follows.

Algorithm 2. SSHMM Parameter Updating

Input: New observations O^f, Old model θ;
Output: New model $\theta_r^f = \{\pi_r^f, A_r^f, \mu^f, \sigma^f\}$;
Procedure:
  Initialization: $t = 0$, initial $\theta_r^f = \theta$.
  while $t < MaxIter$ **do**
    %only one new subsequence;
    E-step: $\forall 1 \le r \le R$, calculate $\alpha(s_{r,n})^{(t+1)}$, $\beta(s_{r,n})^{(t+1)}$, $\gamma(s_{r,n})^{(t+1)}$, $\xi(s_{r,n})^{(t+1)}$ based on $\theta_r^{(t),f}$ utilizing O_r^f.
    M-step: $\forall 1 \le r \le R$, update $\pi_{r,k}^{(t+1),f}$ and $A_{r,j,k}^{(t+1),f}$, $\mu_{k,l}^{(t+1),f}$, $\sigma_{k,l}^{(t+1),f}$ $\forall 1 \le r \le R$.
    Update t**:** $t = t + 1$
  end

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

4.3 Dynamics Characterizing

Based on the model $\theta_r = \{\pi_r, A_r, \mu, \sigma\}$ obtained from the above algorithm, the state sequences can be decoded by Viterbi algorithm [26]. As a result, the dynamics can be revealed by the state sequences. For a long sequence, the Viterbi algorithm also suffers from similar problems arose in model learning. To this end, we decode the hidden state sequence for each subsequence one by one. In addition, this process is independent for each region, and thus we present the decoding algorithm for one region as an example, which is illustrated in Algorithm 3. For each subsequence, it is a dynamic programming algorithm for finding the most likely sequence of hidden states. Consequently, the hidden state sequences characterize urban dynamics, and by aligning the state subsequence of each region, we can reveal its dynamic regularity.

Algorithm 3. Viterbi Algorithm for State Sequence Decoding

Input: Observation dataset $O_r = \{O_r^1, O_r^2, \ldots, O_r^W\}$, Model θ;
Output: Hidden state subsequence for region r $S_r^w = (S_{r,1}^w, S_{r,2}^w, \ldots, S_{r,N}^w)\ \forall 1 \le w \le W$;
Procedure:
  for $w = 1, 2, \ldots, W$ **do**
    (1) Initialization: $n = 1$, initial $\delta_{r,1}(k) = \pi_p(O_{r,1}^w | s_{r,1}(k))$, $\psi_{r,1}(k) = 0, \forall 1 \le k \le K$
    (2) Iteration:
    for $n = 2, 3, \ldots, N$ **do**
      $\delta_{r,n}(k) = \max[\delta_{r,n-1}(j) A_{r,j,k}] p(O_{r,n}^w | s_{r,n}(k))$, $\psi_{r,n}(k) = \arg\max_j [\delta_{r,n-1}(j) A_{r,j,k}], \forall 1 \le k \le K$
    end
    (3)Last state: $s_{r,N}^w = \arg\max_j [\delta_{r,N}(j)]$
    (4)Backforward:
    for $n = N-1, N-2, \ldots, 1$ **do**
      $s_{r,n}^w = \psi_{r,n+1}(s_{r,n+1}^w)$
    end
  end

4.4 Application

4.4.1 Urban Function Identification

Considering that regions with similar dynamic patterns could have similar functions, we match the dynamics with the functions by clustering the state sequence of different regions via k-mediods algorithm [16]. In this clustering algorithm, we define the distance of two sequences as the sum of the euclidean distance of the mean of the two corresponding states. The specific calculation is as follows,

Definition 4.1 (The distance between corresponding states $d(s_{i,n}, s_{j,n})$). *$d(s_{i,n}, s_{j,n})$ is the distance between the state of i-th and j-th region in n-th time slot. It is the euclidean distance between the mean value vector of Gaussian distribution of these two states,*

$$d(s_{i,n}, s_{j,n}) = \sqrt{\sum_{l=1}^{L} (\mu_{i,n,l} - \mu_{j,n,l})^2}, \tag{12}$$

where $\mu_{i,n,l}, \mu_{j,n,l}$ is the mean of l-th dimension of $s_{i,n}, s_{j,n}$, respectively.

Definition 4.2 (The distance between state sequences $D(S_i, S_j)$). *$D(S_i, S_j)$ is the distance between the state sequence of i-th and j-th region. It is the average distance of all the corresponding states,*

$$D(S_i, S_j) = \frac{\sum_{n=1}^{N} d(s_{i,n}, s_{j,n})}{N}, \tag{13}$$

where $\mu_{i,n,l}, \mu_{j,n,l}$ is the mean of l-th dimension of $s_{i,n}, s_{j,n}$, respectively.

We adopt *Davies-Bouldin index* (DBI) [8] to determine the number of clusters, which reflects the ratio between inter-cluster distance and inter-cluster distance. A smaller DBI usually indicates a more effective division. Finally, similar dynamics in each cluster would present the same kind of function in the city. To annotate the function after clustering, we consider the following two aspects: 1) The PoI distribution of each region. We compute the density value of each POI category for each region and rank POI categories by their density as prior knowledge. 2) The human-labeled regions. People may know the functions of a few well-known regions, e.g., the region contains the Forbidden City is an area of historic interests. After clustering, the human-labeled regions will help us understand other regions in a cluster. Refer to the experiments for the detailed results and analysis [33].

4.4.2 Activity Prediction

SSHMM also enables the ability of prediction. Through the Viterbi algorithm, we can identify the last state of the region based on the observation. With the advantage of a probabilistic model, we can infer the state that the region will be in the next time slot. More specifically, we first obtain the current state of the region according to the current observation, then predict the next state according to the latest state by maximizing the transition probability.

Formally, given the current observation sequence of r-th region denoted by $O_{r,1}O_{r,2}, \ldots, O_{r,n}$, we decode its corresponding hidden state sequence as $s_{r,1}s_{r,2}, \ldots, s_{r,n}$. If the latest state $s_{r,n}$ is i-th state in the state set, then the state in the next time slot can be predicted as

$$s_{r,n+1} = \arg\max_{1 \le j \le K} A_{r,i,j}. \tag{14}$$

This is to say, we can achieve the prediction for the volume of population flows and the percentage of PoIs visited in the next time slot by utilizing the mean value of the identified next state.

5 Experiments

In this section, we empirically evaluate the performance of SSHMM in a large-scale real-life dataset. All experiments are implemented by Python, and codes can be accessed in https://github.com/XTxiatong/SSHMM.git.

5.1 Data

The mobility dataset was collected by collaborating with *Tencent*[1], one of the largest Internet integrated service

1. https://heat.qq.com/

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

Fig. 4. Road network that divides the downtown area of Beijing into 665 regions.

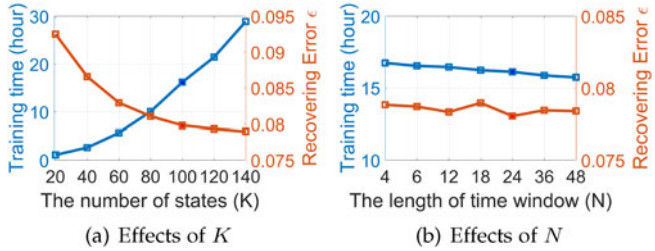


Fig. 6. Effects of parameters.

providers in China. Whenever the user sends a location request by all of the services provided by *Tencent*, his GPS location, and request timestamp as well as check-in PoI could be recorded.

The collected dataset covers near 2 million users in Beijing, China with a duration of one month from April 1st to 30th, 2018. We divide the check-in PoIs into nine categories: *Company*, *Agency*, *Shopping*, *Service*, *Entertainment*, *Attractions*, *Education* and *Residence*. We also crawl the road network from Map service and divide the downtown area in Beijing into 665 non-overlapping regions. It is worth noting that to protect user privacy, all data is anonymous and stored in Tencent offline servers. We pre-process the data under their overseeing and only take the aggregated results for further analysis. The regions we select are as shown in Fig. 4. We have counted the number of users as well as check-ins in each region and each half an hour, the Cumulative Distribution Functions (CDF) of which are as shown in Fig. 5. From it, we can observe that our sample has covered a large amount of the population as they are more than 600 users for more than 50 percent of the cases, while the semantic check-ins are rather sparsity as only 20 percent of the cases with more than 8 records.

5.2 System Settings

5.2.1 Data Usage

In the experiments, we divide the dataset into two parts. We utilize 21 days of data to generate the mobility behaviour observations for model learning and use the rest for prediction evaluation. We set the length of the time slot to 1 hour. The observation in each time slot is 12-dimensional, including 3-dimensional population flow volume and 9-dimensional check-in frequency.

5.2.2 Model Evaluation

To evaluate the effectiveness of our model and the inferred parameters, we utilize the obtained states to recover the observations by concatenating the mean value of corresponding state in the hidden state sequence in chronological order. We adopt ε (see Definition 5), the error between all of the observations and the activities recovered by the mean value of the corresponding hidden state as an evaluation metric.

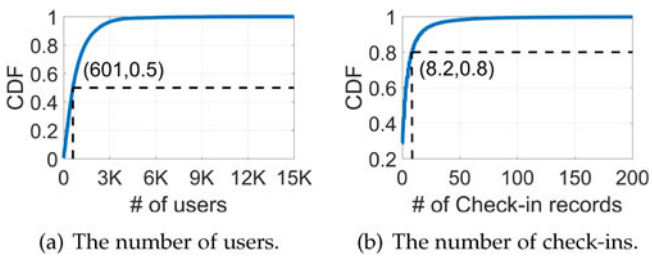


Fig. 5. The statics of dataset.

Definition 5 (Recovering Error ϵ). *Given the observation sequence of r-th region $[O_{r,1}, O_{r,2}, \ldots, O_{r,N}]$ and the corresponding state sequence $[s_{r,1}, s_{r,2}, \ldots, s_{r,N}]$, then the raw observation and the recovered observation in n-th time slot is $\{o_{r,n,1}, o_{r,n,2}, \ldots, o_{r,n,L}\}$ and $\{\mu_{n,1}, \mu_{n,2}, \ldots, \mu_{n,L}\}$, respectively. Therefore, it is defined as*

$$\epsilon = \sqrt{\frac{\sum_{r=1}^{R}\sum_{n=1}^{N}\sum_{l=1}^{L}(\mu_{n,l} - O_{r,n,l})^2}{RNL}}. \quad (15)$$

To determine the number of hidden states and the length of time slots for training in each round, we try different values for comparison. As shown in Fig. 6a, we can observe that when the number of state K increases,ϵ decreases while the training time increases superlinearly. Thus, we set $K = 100$ as a trade-off between model complexity and accuracy. As Fig. 6b show, we find the length of time window has almost no impact on the training time and ϵ, which is in line with our expectations. We set is as 24, i.e., training the model day by day.

In order to demonstrate that the dynamics are highly related to urban functions, as our model can be used to infer functions, we have manually labeled of dozens of regions for dynamics validation. They have different functions shown in Table 2.

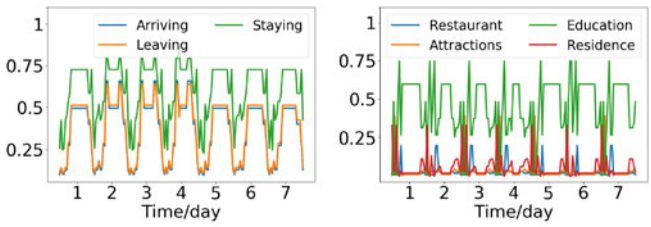
5.2.3 Prediction Performance

In terms of prediction, we utilize the rest 9 days' data for evaluation. We compare the mean value of the next state predicted with the ground-truth observations. For mobility prediction, we adopt *RMSE* as the metric, which is defined as average root-mean-square error between the normalized the population flow of predicted and the ground truth [36].

TABLE 2
Typical Regions With Different Functions

Region Name	Functions
Tsinghua University	Education
Peking University	Education
Tian'anmen	Tourist attraction
Xidan	Shopping & Business
Wangjing	Residence
Zhongguancun Software Park	Company

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.



(a) Mobility (b) Check-in

Fig. 7. Recovered activities for Tsinghua University.

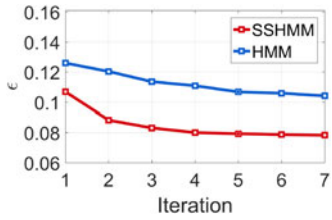


Fig. 8. Model Effectiveness comparison.

For check-ins, we adopt *TopN-accuracy* that reflects the average accuracy on topN frequently-visited PoI prediction of all regions [31]. A lower *RMSE* or a higher *TopM-accuracy* indicates better prediction performance. These two metrics are calculated as follows.

Definition 5.1 (RMSE and TopM-accuracy). *Given the ground-truth observation sequence of r-th region* $[O_{r,n}, O_{r,n+1}, \ldots, O_{r,N}]$ *and the corresponding prediction* $[P_{r,n}, P_{r,n+1}, \ldots, P_{r,N}]$*, then the ground truth and the prediction in n-th time slot both contains two parts as Definition 1: (1)* $\{o_{r,n,1}, o_{r,n,2}, o_{r,n,3}\}$ *and* $\{p_{r,n,1}, p_{r,n,2}, p_{r,n,3}\}$ *denotes the number of arriving, leaving and staying; (2)* $\{o_{r,n,4}, \ldots, o_{r,n,L}\}$ *and* $\{p_{r,n,4}, \ldots, p_{r,n,L}\}$ *denotes the check-in frequency of different categories of PoIs. By sorting the PoIs with their frequency in a descending order and keep the first M PoIs, we obtain the raw TopM list and predicted TopM list denoted by* $V^g_{r,n}$ *and* $V^p_{r,n}$ *respectively. Above all, the metrics are defined as follows,*

$$RMSE = \sqrt{\frac{\sum_{r=1}^{R}\sum_{n=1}^{N}(p_{r,n,l} - o_{r,n,l})^2}{R \times N}},$$
$$TopM - accuracy = \frac{\sum_{r=1}^{R}\sum_{n=1}^{N}(|V^g_{r,n} \bigcap V^p_{r,n}|/M)}{R \times N}. \tag{16}$$

5.3 Results

5.3.1 Model Effectiveness

To show the representative ability for human activities of our model, we first provide a case study of *Tsinghua University* and then give the overall evaluations.

We show the recovering results of *Tsinghua University* of one week as examples in Fig. 7, which exhibits a high similarity with the original observations compared with Fig. 3. The PoI visited frequencies in original observations are volatile due to the sparsity of check-ins whole the recovered results are more regular. Since the mobility is smooth, the recovered results can still capture the peak as well as valley and the difference between working day and non-working day. Fig. 8 shows the relationship between RMSE and the number of iterations. We compare it with the result of training independent HMM for each region with the same number of states. From Fig. 8, we can observe that as the number of iterations increases, the RMSE decreases first, then tends to remain unchanged. Furthermore, our representative ability for human activities by a limited number of states is better than HMM: When the model converges, our RMSE is 0.0793, which outperforms HMM by 54.2 percent.

5.3.2 States and Dynamics

In order to demonstrate the ability of our model in discovering hidden states and revealing urban dynamics, we give a series of special examples and detailed explanations. Fig. 9 visualizes the results for the regions in Table 2. We first plot the mean value of the states with frequent occurrence, then show the state transition process in working day and non-working day, respectively. In China, besides the normal weekends, April 5th (Thursday), 6th (Friday) and 7th (Saturday) are the Qingming Festival, which belongs to the non-working day, while April 8th (Sunday) is a working day.

First, we Look into the Discovered States Shown in Fig. 9. Each state has the semantics of two aspects: (1) the density of the population flow. For example, *state 32* presents a large volume of flows and high density of populations, *state 21* presents a small volume of flows while a high density of populations, and *state 17* present a small volume of flows and low density of populations. (2) the visit frequency of different PoIs. For example, *state 31* indicates the most frequently visited PoI is *Education*, while *state 21* indicates the most frequently visited PoI is *Attractions*. As shown in Fig. 9a, this region have the *state 79* during the day as *Tsinghua University* occupies most of this region. Similarly, *Peking University* has *state 99* and *Tian'anmen* has *state 81*. By combining these two semantics, we can infer the activity level and lifestyle of the region. For example, *state 79* shows active school status while *state 70* shows quiet school status. These high representative states, indicating different activity levels and lifestyles, are sufficient to demonstrate the ability of SSHMM to model urban activities.

Then we discuss *the dynamics represented by the state transition processes.* Take the dynamics of *Tsinghua University* as an example again. As shown in Fig. 9a, during the night, there are fewer people than the day as *state 70*, and *31* have a smaller mean value of staying than that of *state 79*. Besides, in working days, there is a sudden increase in crowd flow as *state 32* appear in 8:00-9:00 and 17:00-19:00. The transition from *state 70* to *32*, from *32* to *79* in working days reveals the dynamics that only students live in the region at night and many teachers go to school in the morning, causing the population denser than night. Compared with Figs. 9a and 9b, and 9c shows that both in working day and non-working day, the density of population is consistently high and the PoI visited most frequently is *Attractions*, as *Tian'anmen* is one of the most famous tourist spots in China. Another interesting finding is that for shopping and residential areas, as Figs. 9d and 9e shown, non-working days are more prosperous and lively than working days, but working areas are quiet and peaceful as Fig. 9f shown. It is worth noting that although the state index changes during the night in Figs. 9a, 9b, and 9e, the semantic expressed by the states remain consistent as their mean value μ is closed.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

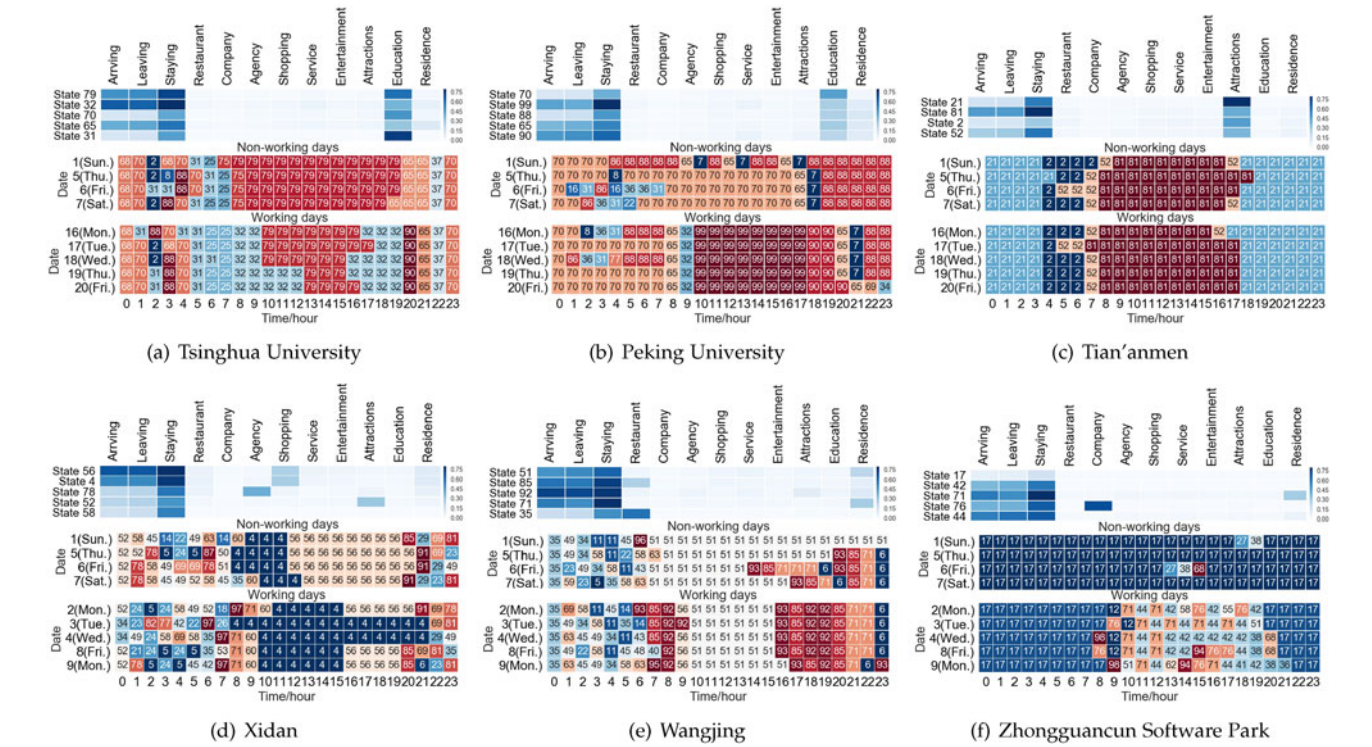


Fig. 9. Visualization for representative states and state sequences for different regions, where each row in the state sequences heatmap exhibits the state transition process of one day with the number indicating the corresponding state.

To conclude, Fig. 9 gives the insights as follows:

(1) The dynamic of each region takes one day as its cycle, as the states in the same time slot but with different dates are usually the same. (2) The dynamic patterns within working days or no-working days are very similar, while the difference across working and non-working days are determined by the function of the region. For some regions like tourist attractions and residence areas, they are similar like Figs. 9c and 9e, but for educational and working places as Fig. 9b and 9f express, the dynamics vary a lot. (3) Regions with similar functions are more likely to have similar dynamics, i.e., sharing more hidden states or having similar transition patterns like Figs. 9a and 9b, which are both universities.

Compared with deep learning model, one superiority of our SSHMM is its interpretability, i.e., each state has specific physical meaning. As discussed above, 100 states are the best when considering the overall performance while too many states weaken their interpretability because some of many only have little difference. To inspect it, we reduce the number of states to 20 and show the results in Fig. 10. It is clear that each state represents a unique function. For example, state 4 and 10 map to residence, state 7 maps to scenic spots, state 0, 2, and 16 map to business. Therefore, it is easy to infer the functions of region. For instance, Tsinghua University in Fig. 10b is a school with state 13 appearing most frequently. In summary, our model is highly explicable for understanding urban dynamics.

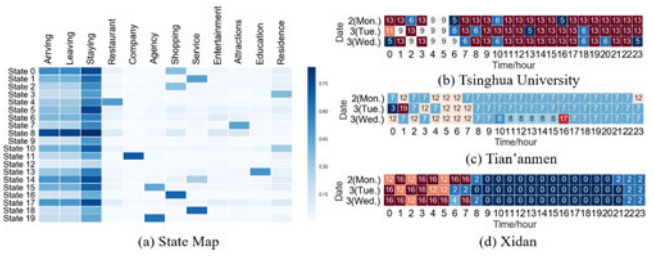


Fig. 10. Results when state number $K = 20$.

5.3.3 *Application of Urban Function Identification*

Intuitively, human activity is significantly influenced by the function of each region, so the inferred city state can be applied to identify region functionality. To evaluate the identified function, considering the real fine-grained function distribution is hard to obtain and the land use plan published by the government is generally out of date, which means ground truth is not available, we compare our results with the state-of-the-art data-driven function detection method, which is an LDA model by utilizing the static PoIs and mobility together [33].

To achieve this goal, we cluster the state sequences by the method introduced in Section 4.4. We have empirically tried different values for the number of clusters as Fig. 11 shown, which indicates and found that dividing dynamic patterns into 8 types is most suitable as the minimum DBI is achieved in this condition. After clustering, we label their functions by the revealed semantic dynamics, including attractions, residence, compound (suburb & business), business, education, compound (residence & business) as well as company. Finally, the geographical distribution of the regions with their function types is shown in Fig. 12, where different colors present different functions. We manually check some regions including the regions shown in Fig. 9 on the map to verify their functions, which shows most of the regions with the same functions are divided into one

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

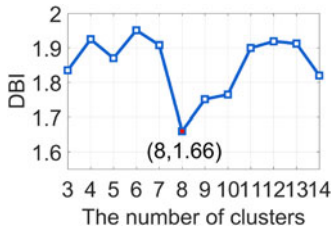


Fig. 11. Cluster number identification.

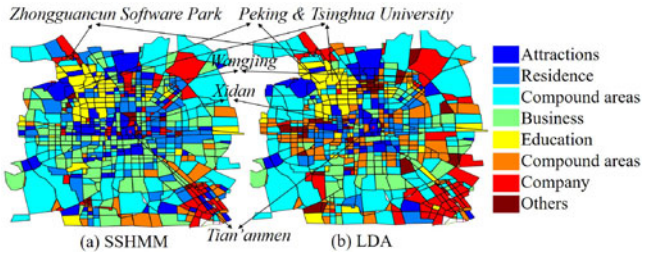


Fig. 12. Visualization of the distribution of the regions with similar dynamics and functions.

cluster. Besides, we present the confusion matrix of the two results in Table 3. The clusters obtained by our model and LDA model have with the Normalized Mutual Information (NMI) of 0.25, which measures the similarity of two divisions with the range from -0.5 to 1 [17].

While the disagreement mainly reflects in two aspects: 1) near half of Business areas from SSHMM are Compound region from LDA. 2) one or two regions in different functions from SSHMM are identified as Others from LDA. To verify these results, we have checked them in the Map one by one. We found that 30 percent regions of the Business areas from SSHMM is not pure Business but also have some apartments for people to live, thus LDA is correct to identify them as Compound region. But the rest regions of the Business areas from SSHMM is sure Business and thus LDA is wrong. However, about 80 percent of Others from LDA has different POI distribution compared with the other seven functions, which is to say LDA achieves higher accuracy the SSHMM. In fact, because we derive functions by clustering, Attraction, Residence, Education, and Company these pure and distinctive functions are easy to be distinguished (SSHMM and LDA are highly consistent on these functions), while Business, Compound and other are complex and more difficult to identify and thus disagreement appears. In summary, SSHMM is good at finding pure and common compound functions, while LDA is more suitable to find the functions that is a quite special minor sort.

In summary, our goal is to model citizens' daily activities and reveal the behind regularity. Because human activity is significantly correlated with the function of each region, the inferred city state can be applied to identify region functionality with high accuracy.

5.3.4 Application of Activity Prediction

In order to evaluate the prediction accuracy of our model, we first show the prediction of the number of staying for 9 days in *Tsinghua University* in Fig. 13a as an example. Compared with the generate HMM, the recovered observations of our model are closer to the ground-truth in different time slots. The results for all metrics are shown in Fig. 13b, where the average *RMSE* of population flow prediction is 0.195 and the *Top3-accuracy* for PoI popularity prediction is 41.4 percent, outperforming the HMM by 16 and 8 percent respectively. These results demonstrate our SSHMM outperforms HMM in urban dynamics prediction problem.

5.3.5 Impact on Continuous Learning

In the above experiments, we divide our mouth data into the first 21 and the following 9 days two parts. We learn SSHMM parameter on the first part, uncovering states and revealing dynamics. When we evaluate the performance of prediction, we use the model learning from old data but test it on the new data. Once the parameter is determined, SSHMM can decode the hidden state sequence and predict the next state on both old and new observations. Now, to evaluate the continuous learning ability of our model, we compare the observation recovery and activity prediction performance when updating and not updating the parameter. The overall results are presented in Table 4. Both the recovery and prediction error after continuous updating have been dropped by $2\% \sim 5\%$. This demonstrates learning SSHMM online, if possible, can achieve better performance. We also believe that the performance can be further improved when the data collected for several mouth and years, with more PoIs disappearing and arising.

6 Discussion

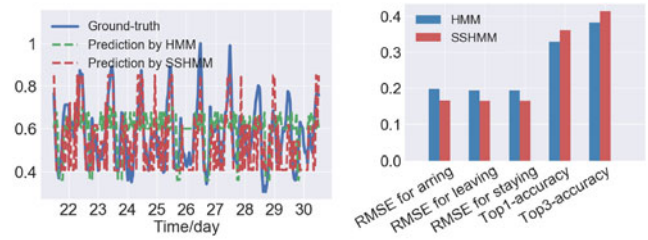
6.1 Implications

In this paper, we propose a model to reveal urban dynamics. As the time is an important context and observations of

TABLE 3
Confusion Matrix for the Clusters by SSHMM and LDA

Number		SSHMM							
		Attractions	Residence	Compound	Business	Education	Compound	Company	Others
LDA	Attractions	45	1	0	6	0	1	4	2
	Residence	10	33	4	10	1	2	9	0
	Compound	9	25	54	29	1	8	10	0
	Business	7	3	12	35	3	4	5	8
	Education	6	2	0	2	52	2	0	0
	Compound	13	15	13	54	0	42	12	1
	Company	6	10	11	0	0	7	28	0
	Others	10	0	1	22	0	1	14	0

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.



(a) Prediction for the number of staying of *Tsinghua University* (b) Evaluation metrics

Fig. 13. Prediction performance compared with HMM.

adjacent time slots are continuous as shown in Fig. 3, we regard human activities as time series and apply a sequence model to detect dynamic patterns. To deal with the first challenge of noisy representation, we adopt Hidden Markov Model, where mobility data as observations are generated by the hidden states. In this manner, noise disruption among different days' collected data can be significantly reduced. Yet previous works dealt with the noise by discretizing continuous observations into states through a manually-set threshold [7], [21], or by clustering [25]. In contrast, we learn the hidden state from the data distribution automatically, which can minimize the impact of manual factors. For the second challenge of sparse semantics, we share the states for all regions, which means all the observations are utilized to learn the emission function to mitigate the data bias of each region. It is ignored by the prior works [2], [21], [25] as they only reveal dynamic in terms of intensity while semantics from check-ins cannot be captured. Nevertheless, although we share states to handle data sparsity, considering the unique character of each region, we learn independent state transition for each region to model their dynamic regularities. Consequently, we can model the aggregated activities in a concise and probabilistic way, which means dynamics prediction can be achieved at the same time.

6.2 Applications

Our results facilitate urban planning in various aspects. Foremost, our findings provide a more comprehensive understanding of urban dynamics and urban functions. For architects, planners, and urban designers, neighborhood activity patterns from intensive ethnographic surveys that take years to conduct can be outdated quickly given the rapid development of the society. Moreover, the features provided by our work helps to examine the intensity and diversity of human activity within a given neighborhood, thus offering novel insights into the functioning of the entire neighborhood and enabling the government to make better plan on land use. Last but not the least, understanding the regularity of the population flows is essential for traffic dispatching, transportation infrastructure construction, etc., while predicting the popular PoIs in different regions and different time slots will benefit business siting and precision marking.

On the other hand, our proposed SSHMM, which learns a group HMMs tied by common state set, outperform independent HMMs in the prediction task as it overcomes the data sparsity and uses the correlation between sequences.

TABLE 4
Performance Comparison

Error ϵ	Learnt by previous data	Continuous updating
Recovering previous observation	0.080	0.078
Recovering future observation	0.085	0.083
Predicting future observation	0.096	0.091

Similar to the aggregated mobility data, there is much other time series with the multi-dimensional nature and sheer size (e.g., voice sequence, music sequence), which can utilize SSHMM for modeling and prediction.

6.3 Limitations and Future Work

In this paper, we propose a novel state-sharing but transition-independent HMM. On the one hand, we choose Gaussian function as the emission function as its simplicity and effectiveness [24]. We leave finding a complex but more effect emission function as our future work to extend the model to adapt to various scenarios. On the other hand, our motivation to share state is that previous related works show that different regions indeed have the same states in different time slots [9], [21], [27]. Yet there might be some common patterns for state evolving between regions. Therefore, finding some insightful motifs [3] in the revealed hidden state is also the next step of this work. Besides, inspired by IOHMM [5], we will also consider incorporating input variables, such as time of day and day of week, as context to improve the performance of our model.

7 Conclusion

In this paper, we study the problem of understanding urban dynamics. We propose a State-sharing Hidden Markov Model (SSHMM), where the states are shared by all the regions, but each region has its own transition regularity. To make it practical, we improve the Baum-Welch algorithm by splitting the long observation sequences into short ones and updating the parameters in parallel. We evaluate our method via a real-life data in Beijing. The results demonstrate that SSHMM learns a meaningful semantics-rich urban dynamics model, recovers different activity regularities by a limited number of states and incurs low training costs.

Acknowledgments

This work was supported in part by The National Key Research and Development Program of China under grant 2018YFB1800804, the National Nature Science Foundation of China under Grants U1936217, 61971267, 61972223, 61941117, and 61861136003, Beijing Natural Science Foundation under grant L182038, Beijing National Research Center for Information Science and Technology under Grant 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] World urbanization prospects: The 2018 revision. united nations department of economic and social affairs/population division, 2018. [Online]. Available: https://drive.google.com/file/d/1FcNuCu6mIi6l5ZkmpWXZElelwPQKca4V/view?usp=sharing

[2] S. Abbar, T. Zanouda, N. Al-Emadi, and R. Zegour, "City of the people, for the people: Sensing urban dynamics via social media interactions," in *Proc. Int. Conf. Social Inform.*, 2018, pp. 3–14.

[3] P. Argos, "A sequence motif in many polymerases," *Nucleic Acids Res.*, vol. 16, no. 21, pp. 9909–9916, 1988.

[4] J. R. Bellegarda and D. Nahamoo, "Tied mixture continuous parameter modeling for speech recognition," *IEEE Trans. Acoustics Speech Signal Process.*, vol. 38, no. 12, pp. 2033–2045, Dec. 1990.

[5] Y. Bengio and P. Frasconi, "An input output HMM architecture," in *Proc. Advances Neural Inf. Process. Syst.*, pp. 427–434, 1995.

[6] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1082–1090.

[7] M. G. Demissie, G. Correia, and C. Bento, "Analysis of the pattern and intensity of urban activities through aggregate cellphone usage," *Transportmetrica A: Transport Sci.*, vol. 11, no. 6, pp. 502–524, 2015.

[8] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

[9] M. G. Demissie, G. Correia, and C. Bento, "Analysis of the pattern and intensity of urban activities through aggregate cellphone usage," *Transportmetrica A: Transport Sci.*, vol. 11, no. 6, pp. 502–524, 2015.

[10] M. G. Docampo, "Theories of urban dynamics," *Int. J. Population Res.*, vol. 2014, pp. 1–11, 2014.

[11] F. F and N. G, "A comparative study of phoneme recognition using GMM-HMM and ann based acoustic modeling," *Int. J. Comput. Appl.*, vol. 98, no. 6, pp. 12–16, 2014.

[12] K. Huang, X. Fu, and N. Sidiropoulos, "Learning hidden markov models from pairwise co-occurrences with application to topic modeling," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2073–2082.

[13] X. Huang, F. Alleva, H.-W. Hon, M.-Y. Hwang, K.-F. Lee, and R. Rosenfeld, "The sphinx-ii speech recognition system: An overview," *Comput. Speech Lang.*, vol. 7, no. 2, pp. 137–148, 1993.

[14] X. D. Huang, "Phoneme classification using semicontinuous hidden Markov models," *IEEE Trans. Signal Process.*, vol. 40, no. 5, pp. 1062–1067, May 1992.

[15] M.-Y. Hwang and X. Huang, "Shared-distribution hidden markov models for speech recognition," vol. 1, pp. 414–420, 1993.

[16] X. Jin and J. Han, "K-medoids clustering," in *Encyclopedia of Machine Learning*. Berlin, Germany: Springer, 2011, pp. 564–565.

[17] Z. F. Knops, J. A. Maintz, M. A. Viergever, and J. P. Pluim, "Normalized mutual information based registration using k-means clustering and shading correction," *Med. Image Anal.*, vol. 10, no. 3, pp. 432–439, 2006.

[18] K. Rose, L. Gu, J. Nayak, "Discriminative training of tied-mixture HMM by deterministic annealing," in *Proc. Int. Conf. Spoken Lang. Process.*, 2000, vol. 4, pp. 183–186.

[19] M. Berliant and P. Wang, "Dynamic urban models: Agglomeration and growth," *Contributions Econ. Anal.*, vol. 266, pp. 531–581, 2004.

[20] W. Mathew, R. Raposo, and B. Martins, "Predicting future locations with hidden markov models," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 911–918.

[21] F. Miranda *et al.*, "Urban pulse: Capturing the rhythm of cities," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 791–800, Jan. 2017.

[22] J. H. Paik, "A novel TF-IDF weighting scheme for effective ranking," in *Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2013, pp. 343–352.

[23] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[24] H. Shi, C. Zhang, Q. Yao, Y. Li, F. Sun, and D. Jin, "State-sharing sparse hidden markov models for personalized sequences," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 1549–1559.

[25] S. Song, T. Xia, J. Feng, P. Hui, and Y. Li, "Urbanrhythm: Revealing daily urban dynamics hidden in mobility data," in *Proc. 8th SIGKDD Int. Workshop Urban Comput.*, 2019, pp. 32–39.

[26] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[27] T. Xia and Y. Li, "Revealing urban dynamics by learning online and offline behaviours together," *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 1, 2019, Art. no. 30.

[28] T. Xia *et al.*, "Understanding urban dynamics via state-sharing hidden markov model," in *Proc. World Wide Web Conf.*, 2019, pp. 3363–3369.

[29] F. Xu, T. Xia, H. Cao, Y. Li, F. Sun, and F. Meng, "Detecting popular temporal modes in population-scale unlabelled trajectory data," *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 46:1–46:25, Mar. 2018.

[30] Z. Yao, Y. Fu, B. Liu, W. Hu, and H. Xiong, "Representing urban functions through zone embedding with human mobility patterns," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 3919–3925.

[31] D. Yu, Y. Li, F. Xu, P. Zhang, and V. Kostakos, "Smartphone app usage prediction using points of interest," *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 174:1–174:21, Jan. 2018.

[32] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 186–194.

[33] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[34] C. Zhang *et al.*, "Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 361–370.

[35] C. Zhang, K. Zhang, Q. Yuan, L. Zhang, T. Hanratty, and J. Han, "GMove: Group-level mobility modeling using geo-tagged social media," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1305–1314.

[36] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.

[37] K. Zhang, Q. Jin, K. Pelechrinis, and T. Lappas, "On the importance of temporal dynamics in modeling urban activity," in *Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput.*, 2013, pp. 7:1–7:8.

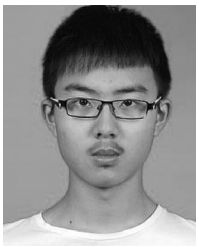
[38] X. Zhou, A. Noulas, C. Mascolo, and Z. Zhao, "Discovering latent patterns of urban cultural interactions in wechat for modern city planning," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery*, 2018, pp. 1069–1078.

[39] W. Zhu, C. Zhang, S. Yao, X. Gao, and J. Han, "A spherical hidden markov model for semantics-rich human mobility modeling," in *Proc. Assoc. Advancement Artif. Intell.*, pp. 1–8, 2018.

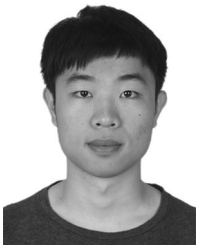
Tong Xia received the BS degree in electrical engineering from the School of Electrical Information, Wuhan University, Wuhan, China, in 2017. Currently, she is working towards the MS degree in big data from the Department of Electronic Engineering, Tsinghua University, Beijing, China. Her research interests include human mobility, mobile big data mining, user behavior modelling, and urban computing.

Yong Li (Senior Member, IEEE) received the BS degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, and the PhD degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. He is currently a faculty member of the Department of Electronic Engineering, Tsinghua University. He has served as general chair, TPC chair, SPC/TPC member for several international workshops and conferences, and he is on the editorial board of two IEEE journals. His papers have total citations more than 6900. Among them, ten are ESI highly cited papers in computer science, and four receive conference best paper (run-up) awards. He received IEEE 2016 ComSoc Asia-Pacific Outstanding Young Researchers, Young Talent Program of China association for Science and Technology, and the National Youth Talent Support Program.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.



Yue Yu is currently working toward the graduate degree in the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include spatio-temporal data mining and graph mining.



Fengli Xu received the BS degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2015, and he is currently working toward the PhD degree in the Electronic Engineering Department of Tsinghua University, Beijing, China. His research interests include human mobility, mobile big data mining, and user behavior modelling.

Qingmin Liao (Member, IEEE) received the BS degree in radio technology from the University of Electronic Science and Technology of China, Chengdu, China, in 1984, and the MS and PhD degrees in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 1990 and 1994, respectively. Since 1995, he has been with Tsinghua University, Beijing, China. He became a professor with the Department of Electronic Engineering, Tsinghua University, in 2002. From 2001 to 2003, he served as the invited professor with a tri-year contract with the University of Caen, France. Since 2010, he has been the director of the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. He has authored more than 90 papers internationally. His research interests include image/video processing, transmission and analysis, biometrics, and their applications to teledetection, medicine, industry, and sports.

Depeng Jin (Member, IEEE) received the BS and PhD degrees in electronics engineering from Tsinghua University, Beijing, China, in 1995 and 1999 respectively. Now he is an associate professor at Tsinghua University and vice chair of Department of Electronic Engineering. He was awarded National Scientific and Technological Innovation Prize (Second Class) in 2002. His research interests include telecommunications, high-speed networks, ASIC design, and future internet architecture.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on April 02,2022 at 02:12:50 UTC from IEEE Xplore. Restrictions apply.