Individual Mobility Prediction in Mass Transit Systems Using Smart Card Data: An Interpretable Activity-Based Hidden Markov Approach

Baichuan Mo, Zhan Zhao, Haris N. Koutsopoulos, and Jinhua Zhao

***Abstract*— Individual mobility is driven by demand for activities with diverse spatiotemporal patterns, but existing methods for mobility prediction often overlook the underlying activity patterns. Knowledge of activity patterns can improve the performance and interpretability of existing individual mobility models, leading to more informed policy design and better user experience in intelligent transportation systems. This study develops an activity-based modeling framework for individual mobility prediction in mass transit systems. Specifically, an input-output hidden Markov model (IOHMM) approach is proposed to simultaneously predict the (continuous) time and (discrete) location of an individual's next trip using transit smart card data. The prediction task can be transformed into predicting the hidden activity duration and end location. Based on a case study of Hong Kong's metro system, we show that the proposed model can achieve similar prediction performance as the state-of-the-art long short-term memory (LSTM) model. Unlike LSTM, the proposed IOHMM approach can also be used to analyze hidden activity patterns, which provides meaningful behavioral interpretation for why an individual makes a certain trip. Therefore, the activity-based prediction framework offers a way to preserve the predictive power of advanced machine learning methods while enhancing our ability to generate insightful behavioral explanations, which is useful for user-centric policy design and intelligent transportation applications such as personalized traveler information.**

***Index Terms*— Individual mobility, next trip prediction, hidden Markov model, smart card data, public transit.**

I. Introduction

INDIVIDUAL mobility prediction describes the prediction of human movements over space and time at the individual level. It has important smart city and smart transportation applications, including personalized traveler information, targeted demand management, etc. Despite the emergence of extensive urban data, it is a challenging problem to accurately predict individual mobility. Travel behavior concerns multiple dimensions (most notably the temporal and spatial dimensions), exhibits longitudinal variability for an individual, and varies across individuals [1], making the mobility prediction problem difficult to tackle. This is especially challenging for public transit systems because they can only observe part of individual mobility as transit trips, typically through smart card records. The same issue is also relevant to the new app-based mobility systems, as their apps only collect partial mobility information when a user consumes their services. Such data directly generated by usage of various mobility systems are referred to as *intrinsic mobility data* [2].

Individual mobility prediction is complex and multi-dimensional. While the literature mostly focuses on the problem of next location prediction [3]–[6], relatively less attention was given to the problem of next trip prediction, especially using intrinsic mobility data. In a prior related study, Zhao et al. [2] defined several sub-problems related to the next trip prediction problem in the context of mass rail transit systems based on smart card data. It is found that, while it is easier to predict whether an individual travels or not, it is much harder to predict when and where they go next. This is not surprising because of the large number of possible combinations of people's spatiotemporal choices. It is generally challenging to deal with high-dimensional problems, especially when the data is relatively sparse at the individual level. Besides, the existing methods are limited in that the time of travel is often treated as a categorical variable. The arbitrary discretization of time does not represent people's temporal choices adequately and may exacerbate the data sparsity issue. Furthermore, while spatial and temporal choices of travel are typically made simultaneously, existing methods often simplify the problem to a sequential prediction task [2]. This study aims to address these limitations with an activity-based approach.

Inspired by activity-based models commonly used in travel behavior research, the main objective of the paper is to develop a methodology to simultaneously predict the time and location of an individual's next trip through their latent activity patterns. Instead of directly predicting travel behavior, we propose an input-output hidden Markov model (IOHMM) approach to analyze the underlying activity behavior. Individual mobility is driven by demand for activities with diverse spatiotemporal patterns, and thus uncovering the latter can help us predict the former. For example, the prediction of the activity duration is equivalent to that of the start time of the next trip. The specific problem formulation and model design are based on smart card data in mass transit systems, but they can be adapted to other mobility systems, e.g., ride-hailing. Transit smart

Manuscript received January 1, 2021; revised July 1, 2021; accepted August 27, 2021. The Associate Editor for this article was R. Langari. *(Corresponding author: Zhan Zhao.)*

Baichuan Mo is with the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Zhan Zhao is with the Department of Urban Planning and Design, The University of Hong Kong, Hong Kong (e-mail: zhanzhao@hku.hk).

Haris N. Koutsopoulos is with the Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115 USA.

Jinhua Zhao is with the Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Data is available on-line at https://github.com/mbc96325/IOHMM-for-individual-mobility-prediction.

Digital Object Identifier 10.1109/TITS.2021.3109428

card data from Hong Kong's Mass Transit Railway (MTR) system are used to illustrate the applicability of the proposed methodology.

The contributions of this study are summarized as follows:

- Existing individual mobility prediction models often lack natural behavioral interpretation, limiting their applicability for intelligent policy design. This paper introduces an activity-based modeling framework that captures the underlying generative mechanism of travel behavior and uncovers people's travel purposes not directly observable in the data. To the best of our knowledge, this is the first study that adopts an activity-based model for individual mobility prediction using transit smart card data.
- We adopt an IOHMM approach for activity-based modeling, and extend it for simultaneous prediction of the discrete location and continuous time of travel. It has been shown that the temporal aspect of individual mobility is least predictable [2], partly because of the arbitrary discretization of time. In our extended IOHMM, the trip start time is represented as a continuous emission variable, and can be predicted jointly with a discrete location variable through inference of latent activity types.
- We associate the hidden states in the IOHMM with individual's hidden activities, and propose a Gibbs sampling method to extract and visualize hidden activity patterns with semantic explanations. We also show how the estimated model parameters and transition matrix can be used for the model's interpretability.
- The proposed methodology is demonstrated using transit smart card data from Hong Kong MTR. Compared to state-of-the-art deep learning models, the results show that our activity-based model can achieve competitive predictive performance, while offering significantly more interpretability into the underlying activity patterns and travel purposes. The combination of performance and interpretability makes our approach more versatile and actionable.

II. Literature Review

Demand prediction in public transit systems can be categorized into aggregate and individual levels. The aggregate demand prediction [7]–[11] has been extensively studied with the recent development of deep learning methods. A summary of previous works can be found in Fang et al. [12]. However, the individual-level prediction gains less attention compared to the aggregated ones.

The literature on individual mobility prediction mostly focuses on the problem of next location prediction, rather than next trip prediction. Most existing methods for next location prediction are based on mining sequential patterns of individual location histories. Simple Markov chain (MC) models have shown to be able to achieve good prediction performance [4], [13]. Asahara et al [14] proposed a mixed Markov chain model (MMM) for next location prediction by identifying the group a particular individual belongs to and applying a specific MC model for that group. Asahara et al. [15] presented a hybrid method of clustering location histories according to their characteristics before training a hidden Markov model (HMM) for each cluster. Recently, due to the rapid advance of deep learning, variants on Recurrent Neural Network (RNN) models have been adopted for next location prediction, and showed improved prediction performance over MC models [16]–[18]. Some of the most competitive models today are based on the Long Short-Term Memory (LSTM) architecture [19]–[21]. Similar methods have also been proven successful for vehicle trajectory prediction problems [22], [23]. However, none of these methods explicitly consider the temporal behavior of individual mobility, e.g., when to start the next trip. This is important for any mobility service because travel demand is dynamic and time-sensitive. In addition, despite the superior predictive performance, the deep learning models are generally black-box and difficult to interpret, making them unfit as supportive tools for policy design.

For the next trip prediction, we have to model both the spatial and temporal choices of individual trip-making. Only a few prior studies have dealt with this issue. Gidófalvi and Dong [24] developed a continuous-time Markov model to predict when an individual will leave their current location and where will they go next. Hsieh et al. [25] introduced a time-aware language model, *T-gram*, to predict when an individual leaves a location by extracting location-specific time distributions from social media check-in data. Focusing on mass transit systems, Zhao et al. [2] explicitly formulated the spatiotemporal choices of individual passengers as a sequence of decisions, and proposed a *mobility N-gram* model to predict the choices associated with the next trip—the trip start time, origin, and destination. It is found that the start time is the least predictable aspect of the next trip. While the low predictability for trip start time is to some extent rooted in people's inherent behavioral variability, the discrete representation of time in most existing models is likely to limit our ability to predict temporal behavior. The key challenge is to capture the complex interaction of continuous time choices and discrete location choices. One way to do this is through latent variables representing hidden activities between trips [26].

In the activity-based analysis of travel behavior, travel is treated as being derived from the need to pursue activities distributed in space and time [27]–[29]. With a more realistic representation of travel behavior, activity-based models are intuitive and interpretable, and have been widely used in transportation planning, though they usually depend on detailed manual survey data. Recent years have seen growing interests in automatically detecting activity patterns from large-scale human mobility traces [30], [31], including transit smart card data. Han and Sohn [32] developed a Continuous Hidden Markov Model (CHMM) to impute the sequence of activities for each trip chain. Zhao et al. [26] proposed a spatiotemporal topic model to discover latent activity patterns from smart card data. While these methods can enrich mobility data with behavioral semantics, none of them are suitable for predicting future trips. An activity-based approach to individual mobility prediction is needed to capture spatiotemporal correlation and enhance behavioral interpretability.

This study presents the first activity-based approach to individual mobility prediction. It is based on the Input-Output

TABLE I
SUMMARY OF INDIVIDUAL MOBILITY PREDICTION STUDIES

Study	Problem	Data	Model
Asahara et al. [14]; Gambs et al. [13]	Location Prediction	GPS Data	Markov Chain
Mathew et al. [15]	Location Prediction	GPS Data	HMM
Al-Molegi et al. [17]	Location Prediction	GPS Data	RNN
Liu et al. [16]; Kong and Wu [19]	Location Prediction	LBS Data	RNN / LSTM
Feng et al. [18, 21]	Location Prediction	Mobile Phone Data	RNN / LSTM
Gidófalvi and Dong [24]	Location+Time Prediction	GPS Data	Continuous-Time Markov Chain
Zhao et al. [2]	Location+Time Prediction	Transit Smart Card Data	Bayesian N-Gram
This Study*	Location+Time Prediction	Transit Smart Card Data	IOHMM

*Only this study adopts an activity-based approach

Hidden Markov Model (IOHMM), which is an extension of standard HMMs [33]. The standard HMMs assume homogeneous transition and emission probabilities, in which the contextual information cannot be captured. To overcome this limitation, the IOHMM was proposed to incorporate additional information. Specifically, transition probabilities in IOHMMs are conditional on the input and thus depend on time. IOHMM was designed for sequence data processing, and has been applied for diverse problems including grammar inference [34], gesture recognition [35], audio processing [36], electricity price forecasting [37], and urban activity generation [38]. As we will demonstrate in Section III-C, the IOHMM architecture can be adapted to (1) capture the dynamics of individual travel-activity histories, (2) incorporate rich contextual information for improved prediction performance, and (3) jointly predict both discrete (location) and continuous (time) attributes of trips/activities. Table I summarizes the key difference of our model compared to some of the existing ones for individual mobility prediction.

III. METHODOLOGY

A. Problem Description

In this section, we illustrate the methodology based on transit smart card data for consistency with the case study. However, as mentioned in Section I, this framework applies to all intrinsic mobility data. More discussions on how to extend the model to other intrinsic mobility data are shown in Section V.

Transit smart card data is one of the most important intrinsic mobility data. It includes passengers' tap-in and tap-out[1] transaction records, which can provide the chronological public transit (PT) trip histories of each individual. The trip structure

[1]This study focuses on closed public transit systems with both tap-in and tap-out records.

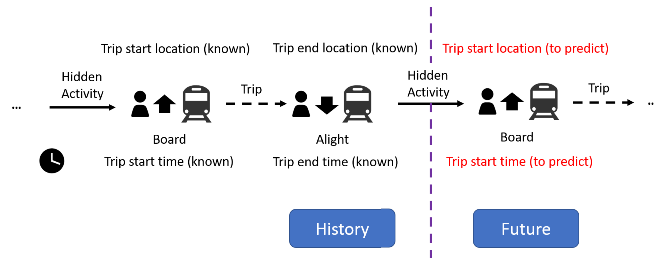


Fig. 1. Public transit trip structure.

is shown in Figure 1. Each trip starts with boarding at an origin station and ends with alighting at a destination station. The boarding (and alighting) times and locations are known from the transit smart card data. The unique ID of each smart card allows us to track the trip histories of each anonymous individual. Between two consecutive trips, a passenger may have some activities such as working, staying at home, etc. In this study, the latent behavior of an individual between two adjacent trips is referred to as a *hidden activity*. Different from a typical definition of an activity where passengers stay in a place, the hidden activity in this study may include unobserved trips such as taking a taxi to another place. Due to data limitations, we cannot identify people's trips outside the PT system. Thus, we assume that no matter what people have done between two adjacent transit trips, this process is treated as a single hidden activity. The alighting station of the last trip and the boarding station of the next trip are referred to as activity start and end locations, respectively. Our goal is to predict when and where the next trip will start given a sequence of recorded trip histories.

Since the alighting time of the last trip is known, predicting the next trip start time is equivalent to estimating the duration of the current hidden activity. Similarly, predicting the next trip start location is equivalent to predicting the end location of the current hidden activity. In this way, we can transform the next trip prediction problem into an activity duration and location prediction problem. The new perspective has more relevant behavioral implications, as activities are usually what drive people to travel.

B. Activity-Based Modeling Framework

The duration of people's hidden activities can vary greatly, anywhere from one hour (e.g. shopping) to several days (e.g. vacation). The wide range of activity duration makes it challenging to predict. In this study, we set the basic prediction interval as one day. A sequence of consecutive activities is extracted from the smart card data on a specific day and each individual may have multiple sequences of activities. The choice of prediction interval of one day not only reduces the scope of the prediction problem (from infinity to 24 hours) but also represents the basic period of regularity for human mobility and activity patterns [2], [39], [40]. Specifically, each day spans from 4:00 AM to 4:00 AM of the next calendar day, which better matches people's daily activity schedules and the operating time of transit.

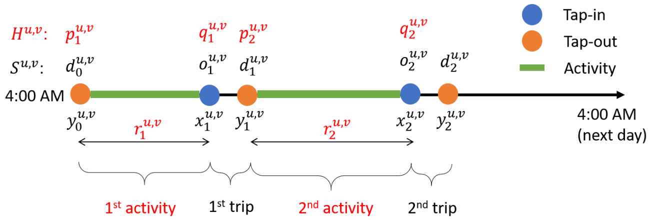


Fig. 2. Relationship between $S^{u,v}$ and $H^{u,v}$. There are two trips (thus two activities) in the day for this example.

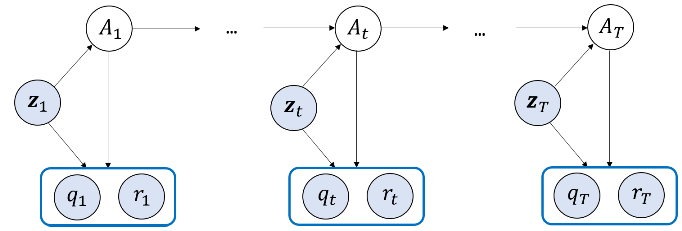


Fig. 3. Structure of IOHMM. The solid nodes represent observed information, while the transparent (white) nodes represent latent random variables.

For a user u, the recorded public transit trips in day v are represented as

$$S^{u,v} = \{(o_1^{u,v}, d_1^{u,v}, x_1^{u,v}, y_1^{u,v}), \ldots, (o_{T^{u,v}}^{u,v}, d_{T^{u,v}}^{u,v}, x_{T^{u,v}}^{u,v}, y_{T^{u,v}}^{u,v})\} \tag{1}$$

where $o_t^{u,v}, d_t^{u,v}, x_t^{u,v}, y_t^{u,v}$ are the origin, destination, start time, and end time of t-th trip for user u in day v, respectively. $T^{u,v}$ is the total number of trips. The corresponding hidden activity sequence is defined as

$$H^{u,v} = \{(p_1^{u,v}, q_1^{u,v}, r_1^{u,v}), \ldots, (p_{T^{u,v}}^{u,v}, q_{T^{u,v}}^{u,v}, r_{T^{u,v}}^{u,v})\} \tag{2}$$

where $p_t^{u,v}, q_t^{u,v}, r_t^{u,v}$ are the start location, end location, and duration of t-th activity for user u in day v, respectively. Particularly, for $t = 1, \ldots, T^{u,v}$, we have

$$p_t^{u,v} = d_{t-1}^{u,v} \tag{3}$$
$$q_t^{u,v} = o_t^{u,v} \tag{4}$$
$$r_t^{u,v} = x_t^{u,v} - y_{t-1}^{u,v} \tag{5}$$

For the first activity, we explicitly define $d_0^{u,v}$ = "null" and $y_0^{u,v}$ = 4:00 AM. An example to illustrate the relationship between $S^{u,v}$ and $H^{u,v}$ is shown in Figure 2. After a trip ends, $p_t^{u,v}$ is directly observed from the transit smart card records. Therefore, our goal is to predict $q_t^{u,v}$ and $r_t^{u,v}$ given historical trajectories and other information (e.g., weather).

It is worth noting that we do not consider the time period from $y_{T^{u,v}}^{u,v}$ to 4:00 AM next day as the last activity interval because its duration is deterministic ($y_{T^{u,v}}^{u,v}$ is known). This study focuses on predicting the next trip's time and location, but there is no corresponding next trip for this activity. Therefore, there is no need to predict the last activity, and it is excluded from further analysis.

C. IOHMM for Activity Prediction

IOHMM is proposed to capture exogenous contextual information over time, which allows the modeling of heterogeneous transition and emission probabilities. The structure of IOHMM for individual activity modeling is shown in Figure 3. A_t is the t-th hidden activity (a latent random variable) and $\boldsymbol{z_t}$ is a vector of observed input variables containing contextual information (e.g., weather, day of week, p_t, etc.). The superscript (u, v) is ignored for simplicity. Since each hidden activity can be encoded as a latent state in IOHMM, the IOHMM architecture matches well with the activity-based modeling framework.

The model consists of three key components: 1) initial state probability $\pi_i = \mathbb{P}(A_1 = i \mid z_1;\ \boldsymbol{\theta_{in}})$, where $i \in \mathcal{A}$ and $\mathcal{A}$ is the state space; It quantifies the distribution of the first activity's type. 2) transition probability: $\varphi_{ij,t} = \mathbb{P}(A_t = j \mid A_{t-1} = i, z_t;\ \boldsymbol{\theta_{tr}})$, which quantifies the probability that next activity is j given this activity is i, and 3) emission probability: $\delta_{i,t} = \mathbb{P}(q_t, r_t \mid A_t = i, z_t;\ \boldsymbol{\theta_{em}})$, which quantifies the distributions of activity duration and end location. $\boldsymbol{\theta_{in}}$, $\boldsymbol{\theta_{tr}}$, and $\boldsymbol{\theta_{em}}$ are parameters of initial, transition, and emission probability functions, respectively. The likelihood of a data sequence under this model is given by:

$$L(\boldsymbol{\theta}) = \sum_{A_1,\ldots,A_T} \mathbb{P}(A_1 \mid \boldsymbol{z_1};\ \boldsymbol{\theta_{in}}) \cdot \prod_{t=2}^{T} \mathbb{P}(A_t \mid A_{t-1}, \boldsymbol{z_t};\ \boldsymbol{\theta_{tr}}) \cdot \prod_{t=1}^{T} \mathbb{P}(q_t, r_t \mid A_t, \boldsymbol{z_t};\ \boldsymbol{\theta_{em}}) \tag{6}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta_{in}}, \boldsymbol{\theta_{tr}}, \boldsymbol{\theta_{em}}]$.

The model is estimated by the Expectation-Maximization (EM) algorithm.

1) E-Step: Denote the estimated parameters at iteration $k-1$ of M-step as $\boldsymbol{\theta}^{(k-1)}$ (if $k = 1$, use the initial values of the parameters). From $\boldsymbol{\theta}^{(k-1)}$ we can obtain the three probabilities as $\pi_i^{(k-1)}, \delta_{i,t}^{(k-1)}, \varphi_{ij,t}^{(k-1)}$. Then, the forward and backward variables (denoted as $\alpha_{i,t}^{(k)}$ and $\beta_{i,t}^{(k)}$, respectively) are calculated as

$$\alpha_{i,t}^{(k)} = \mathbb{P}(q_{1:t}, r_{1:t}, A_t = i \mid \boldsymbol{z_{1:t}}) = \delta_{i,t}^{(k-1)} \sum_{l \in \mathcal{A}} \varphi_{li,t}^{(k-1)} \cdot \alpha_{l,t-1}^{(k)} \tag{7}$$

$$\begin{aligned} \beta_{i,t}^{(k)} &= \mathbb{P}(q_{t+1:T}, r_{t+1:T} \mid A_t = i, \boldsymbol{z_{t:T}}) \\ &= \sum_{l \in \mathcal{A}} \varphi_{il,t}^{(k-1)} \cdot \beta_{l,t+1}^{(k)} \cdot \delta_{l,t+1}^{(k-1)} \end{aligned} \tag{8}$$

where $\alpha_{i,1}^{(k)} = \pi_i^{(k-1)} \delta_{i,1}^{(k-1)}$ and $\beta_{i,T} = 1$. The subscripts $1:t$ indicates a list of the corresponding variable with subscript from 1 to t. Then, we calculate the posterior state probability and posterior transition probability as:

$$\gamma_{i,t}^{(k)} = \mathbb{P}(A_t = i \mid q_{1:T}, r_{1:T}, z_{1:T}) = \alpha_{i,t}^{(k)} \cdot \beta_{i,t}^{(k)} / L_c^{(k)} \tag{9}$$

$$\begin{aligned} \xi_{ij,t}^{(k)} &= \mathbb{P}(A_t = j, A_{t-1} = i \mid q_{1:T}, r_{1:T}, z_{t:T}) \\ &= \varphi_{ij,t}^{(k-1)} \cdot \alpha_{i,t-1}^{(k)} \cdot \beta_{j,t}^{(k)} \cdot \delta_{j,t}^{(k-1)} / L_c^{(k)} \end{aligned} \tag{10}$$

where $L_c^{(k)}$ is the complete data likelihood at iteration k, defined as $L_c = \sum_{i \in \mathcal{A}} \alpha_{i,T}^{(k)}$. Obtaining $\alpha_{i,t}^{(k)}$, $\beta_{i,t}^{(k)}$, $\gamma_{i,t}^{(k)}$, and $\xi_{ij,t}^{(k)}$ for all $i, j \in \mathcal{A}$ and $t = 1, \ldots, T$ finishes the E-step.

2) M-Step: The probability parameters in iteration k is updated by maximizing the expected data log likelihood:

$$\begin{aligned}Q(\boldsymbol{\theta};\ \boldsymbol{\theta}^{(k-1)}) &= \sum_{i\in\mathcal{A}} \gamma_{i,1}^{(k)} \cdot \log \mathbb{P}(A_1 = i \mid \boldsymbol{z_1};\ \boldsymbol{\theta_{in}}) \\ &+ \sum_{t=2}^{T}\sum_{i,j\in\mathcal{A}} \xi_{ij,t}^{(k)} \cdot \log \mathbb{P}(A_t = j \mid A_{t-1} = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{tr}}) \\ &+ \sum_{t=1}^{T}\sum_{i\in\mathcal{A}} \gamma_{i,t}^{(k)} \cdot \log \mathbb{P}(q_t, r_t \mid A_{t-1} = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{em}}) \end{aligned} \tag{11}$$

We have $\boldsymbol{\theta}^{(k)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta};\ \boldsymbol{\theta}^{(k-1)})$. The M-step can be implemented by any supervised learning model that supports gradient ascent on the log probability. With proper specification of three probability functions, the optimization problem can be convex and easily solved. It is also worth noting that since $Q(\boldsymbol{\theta};\ \boldsymbol{\theta}^{(k-1)})$ consists of three components with independent parameters, maximizing $Q(\boldsymbol{\theta};\ \boldsymbol{\theta}^{(k-1)})$ is equivalent to maximizing the three components separately, which allows for more flexibility in optimization [38].

D. Model Specification

1) Contextual Information: In terms of the contextual information $\boldsymbol{z_t}$, five different dimensions are considered: weather, day of the week, holidays, last trip information, and historical travel statistics. Denote the total number of contextual features as Z. The specific variables used can be found in Appendix A.

2) State Space: The state space $\mathcal{A}$ is specified for each individual (i.e., $\mathcal{A}^u$ for user u). Since an activity label is a latent categorical variable during the modeling process, we only need to define the cardinality N^u, which follows that $\mathcal{A}^u = \{1, \ldots, N^u\}$. N^u indicates how many hidden activities we considered for user u. A semantic label can be associated to each element in $\mathcal{A}^u$ with an in-depth analysis in Section IV. Generally, the value of N^u can be determined using the validation data set [41] or optimization approaches [42]. In this study, we select N^u by maximizing the *silhouette coefficient* [43]. We assume hidden activities can be characterized by $\boldsymbol{z_t}^{u,v}$. For user u, we can cluster $\boldsymbol{z_t}^{u,v}$ into m clusters, representing m possible hidden activities. The silhouette coefficient of the m-clustering (denoted as $SC^u(m)$ for user u) is a measure of how similar each object is to its own cluster compared to other clusters (i.e., the quality of clustering). It is defined as

$$SC^u(m) = \text{mean}\{\frac{b(i) - a(i)}{\max\{a(i), b(i)\}}\} \tag{12}$$

where $a(i)$ and $b(i)$ are the intra-cluster distance and nearest-cluster (that i does not belong to) distance of data point i, respectively. "mean{}" indicates taking the average over all samples. $SC^u(m)$ ranges from -1 to $+1$, where a high value indicates that the samples are well matched to their own cluster and poorly matched to neighboring clusters. Hence, N^u is obtained by

$$N^u = \arg\max_{m\in\mathcal{M}} SC^u(m) \tag{13}$$

where $\mathcal{M}$ is the set of possible numbers of hidden activities. In this study, $\mathcal{M} = \{3, 4, \ldots, 7\}$ is used. It worth noting that we also tested other cluster quality metrics, such as Akaike information criterion (AIC) and Bayesian information criterion (BIC). Numerical results show that the silhouette coefficient works best for determining the number of hidden activities.

3) Three Probability Functions: The multinomial logistic regression is used to model the initial probability and transition probability. Specifically, we have:

$$\mathbb{P}(A_1 = i \mid \boldsymbol{z_1};\ \boldsymbol{\theta_{in}}) = \frac{\exp(\boldsymbol{\theta_{in,i}} \cdot \boldsymbol{z_1})}{\sum_{j\in\mathcal{A}} \exp(\boldsymbol{\theta_{in,j}} \cdot \boldsymbol{z_1})} \tag{14}$$

where $\boldsymbol{\theta_{in,i}}$ are the coefficients of the initial state probability function at state i.

$$\mathbb{P}(A_t = j \mid A_{t-1} = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{tr}}) = \frac{\exp(\boldsymbol{\theta_{tr,ij}} \cdot \boldsymbol{z_t})}{\sum_{j\in\mathcal{A}} \exp(\boldsymbol{\theta_{tr,ij}} \cdot \boldsymbol{z_t})} \tag{15}$$

where $\boldsymbol{\theta_{tr,ij}}$ are the coefficients of the state transition probability function when the next state is j given the current state is i.

Note that we use the multinomial logit (MNL) model to characterize the probabilities, instead of directly estimating the individual values (like the typical HMM), because these probabilities are also determined by $\boldsymbol{z_t}$. And the linear-in-parameters structure of the MNL model facilitates the model's interpretability. The same probability modeling methods can also be found in [38].

In terms of the emission probability, it is worth noting that q_t is a discrete random variable while r_t continuous. Given a hidden activity, we assume conditional independence between q_t and r_t, that is, we assume that the correlation between q_t and r_t within the same activity type is negligible, which simplifies the model estimation.

$$\begin{aligned}&\mathbb{P}(q_t, r_t \mid A_t = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{em}}) \\ &\quad = \mathbb{P}(q_t \mid A_t = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{emq}}) \cdot \mathbb{P}(r_t \mid A_t = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{emr}})\end{aligned} \tag{16}$$

For the activity end location distribution, a similar multinomial logistic regression model is used, where

$$\mathbb{P}(q_t = l \mid A_t = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{emq}}) = \frac{\exp(\boldsymbol{\theta_{emq,i,l}} \cdot \boldsymbol{z_t})}{\sum_{l\in\mathcal{L}} \exp(\boldsymbol{\theta_{emq,i,l}} \cdot \boldsymbol{z_t})} \tag{17}$$

where $\boldsymbol{\theta_{emq,i,l}}$ are the coefficients for emission probability of activity location where the location is l given the current state is i. $\mathcal{L}$ is the set of location candidates. For user u, $\mathcal{L}^u$ is defined as all stations that he/she has visited in the smart card data records.

In terms of the duration distribution, we assume a Gaussian distribution with the mean expressed as a linear function of explanatory variables. This formulation enables the evaluation of the impact of contextual variables on activity duration and can be estimated efficiently (like a linear regression).

$$\mathbb{P}(r_t \mid A_t = i, \boldsymbol{z_t};\ \boldsymbol{\theta_{emr}}) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(r_t - \boldsymbol{\theta_{emr,i}} \cdot \boldsymbol{z_t})^2}{2\sigma_i^2}} \tag{18}$$

where $\boldsymbol{\theta_{emr,i}}$ and σ_i denote the coefficients and the standard deviation of the model when the hidden state is i. Ideally, σ_i can also be a function of $\boldsymbol{z_t}$. However, this will significantly increase the model's estimation difficulty. Hence, we assume that under a specific activity type, the variance of duration is a constant, and different variance patterns can be captured by different activity types.

It is worth noting that the conditional independence assumption allows us to model the joint probability of r_t and q_t through the hidden activity. This makes it feasible to estimate the two variables simultaneously. Moreover, the conditional independence between q_t and r_t under the same activity type can be further justified by adding, if needed, more activity types as a simple extension of the current framework. For example, one may argue that the duration of a lunch activity depends on where they have the meal. This can be captured by separating the lunch activity into more sub-activities, where each sub-activity is associated with an identical restaurant (location). In this way, the current conditional independence framework still works.

E. Prediction Formulation

The IOHMM supports predicting the next activity duration and location given today's trajectories after training. Here we only give the formulation for predicting the duration, and that for location prediction can be derived in the same way. Observe that predicting the duration of the next activity is equivalent to obtaining $\mathbb{P}(r_{t+1} \mid q_{1:t}, r_{1:t}, \mathbf{z_{1:t+1}})$. This is because $q_{1:t}, r_{1:t}, \mathbf{z_{1:t+1}}$ are all observed information. By the conditional independence, we have $\mathbb{P}(r_{t+1} \mid q_{1:t}, r_{1:t}, \mathbf{z_{1:t+1}}) = \mathbb{P}(r_{t+1} \mid r_{1:t}, \mathbf{z_{1:t+1}})$. By the law of total probability:

$$\begin{aligned}&\mathbb{P}(r_{t+1} \mid r_{1:t}, \mathbf{z_{1:t+1}}) \\ &\quad= \sum_{i \in \mathcal{A}} \mathbb{P}(r_{t+1} \mid A_{t+1} = i, \mathbf{z_t}) \cdot \mathbb{P}(A_{t+1} = i \mid r_{1:t}, \mathbf{z_{1:t+1}})\end{aligned} \tag{19}$$

The first term in the right hand side (RHS) of Eq. 19 is the emission probability. And the second term can be expanded as:

$$\mathbb{P}(A_{t+1} = i \mid r_{1:t}, \mathbf{z_{1:t+1}}) = \frac{\mathbb{P}(A_{t+1} = i, r_{1:t} \mid \mathbf{z_{1:t+1}})}{\sum_{j \in \mathcal{A}} \mathbb{P}(A_{t+1} = j, r_{1:t} \mid \mathbf{z_{1:t+1}})} \tag{20}$$

where

$$\begin{aligned}&\mathbb{P}(A_{t+1} = i, r_{1:t} \mid \mathbf{z_{1:t+1}}) \\ &\quad= \sum_{i \in \mathcal{A}} \mathbb{P}(A_{t+1} = j \mid A_t = i, \mathbf{z_{t+1}}) \cdot \mathbb{P}(A_t = i, r_{1:t} \mid \mathbf{z_{1:t}})\end{aligned} \tag{21}$$

The first term in the RHS of Eq. 21 is the state transition probability. And the second term is essentially the forward variable (Eq. 7) when only incorporating the emission probability of duration. Therefore, based on the forward variable, state transition probability, and emission probability, one can output the distribution of r_{t+1} given $r_{1:t}, \mathbf{z_{1:t+1}}$. We adopt the value with the highest probability density as the predicted duration.

For the location distribution, we can derive $\mathbb{P}(q_{t+1} \mid q_{1:t}, \mathbf{z_{1:t+1}})$ using the same method. The location with the highest probability is selected as the prediction.

F. Model Interpretability

The IOHMM allows us to explore the mobility patterns of an individual. In the following discussion, the subscript t is ignored as we focus on deriving the general pattern over history.

1) Activity Pattern Identification: To identify the latent activity, four distributions conditioning on a specific activity label are calculated: 1) duration distribution $\mathbb{P}(r \mid A = i)$, 2) end location distribution $\mathbb{P}(q \mid A = i)$, 3) start time (i.e., last trip end time) $\mathbb{P}(y \mid A = i)$, and 4) start location distribution $\mathbb{P}(p \mid A = i)$. Since the transition matrix describes how passengers moving from one activity to another, we are also curious about $\mathbb{P}(A_t = j \mid A_{t-1} = i)$ for all $i, j \in \mathcal{A}$. Based on these distributions, we can assign a semantic label (e.g., home, work) to each hidden activity manually.

These distributions/parameters are calculated based on Gibbs sampling as illustrated in Algorithm 1. It is worth noting that we assume $\mathbb{P}(\mathbf{z})$ is the same as the distribution of $\mathbf{z}$ in historical trajectories. Hence, instead of sampling $\mathbf{z} \sim \mathbb{P}(\mathbf{z})$, we can generate samples by going through all histories (i.e. $t = 1, \ldots, T^{u,v}$, for $v = 1, \ldots, V^u$, where V^u is the total number of travel days for user u. This can be seen as bootstrapping). The sampling process is repeated N times. And the intended distributions can be directly obtained from the generated sequences (e.g., for a discrete variable, we can directly count the conditional frequency in the generated sequences).

Algorithm 1 Activity Pattern Identification for a User u Using Gibbs Sampling

Input: Trained IOHMM; Trip history of user u
Output: Intended probability distribution of user u

Initialize the number of sampling N.
for $n = 1$ to N **do**
  for $v = 1$ to V^u **do**
    Sample $A_1^{u,v} \sim \mathbb{P}(A_1^{u,v} \mid \mathbf{z_1^{u,v}})$
    Sample $r_1^{u,v} \sim \mathbb{P}(r_1^{u,v} \mid A_1^{u,v}, \mathbf{z_1^{u,v}})$ and $q_1^{u,v} \sim \mathbb{P}(q_1^{u,v} \mid A_1^{u,v}, \mathbf{z_1^{u,v}})$.
    for $t = 2$ to $T^{u,v}$ **do**
      Sample $A_t^{u,v} \sim \mathbb{P}(A_t^{u,v} \mid A_{t-1}^{u,v}, \mathbf{z_t^{u,v}})$
      Sample $r_t^{u,v} \sim \mathbb{P}(r_t^{u,v} \mid A_t^{u,v}, \mathbf{z_t^{u,v}})$ and $q_t^{u,v} \sim \mathbb{P}(q_t^{u,v} \mid A_t^{u,v}, \mathbf{z_t^{u,v}})$.
    end for
  end for
  Save the generated activity sequences and corresponding contextual information in iteration n as H_n^u and z_n^u, respectively.
end for
Obtain the intended distribution described above for user u based on $[(H_1^u, z_1^u), \ldots, (H_N^u, z_N^u)]$.

2) Probability Coefficient Explanation: After training the model, we can obtain $\boldsymbol{\theta}$ for each probability function. Since all functions adopt a linear relationship between $\boldsymbol{\theta}$ and z, the value of $\boldsymbol{\theta}$ enables interpretability and validation of the training results. For example, we may expect rain to have a positive effect on the duration of all activities.

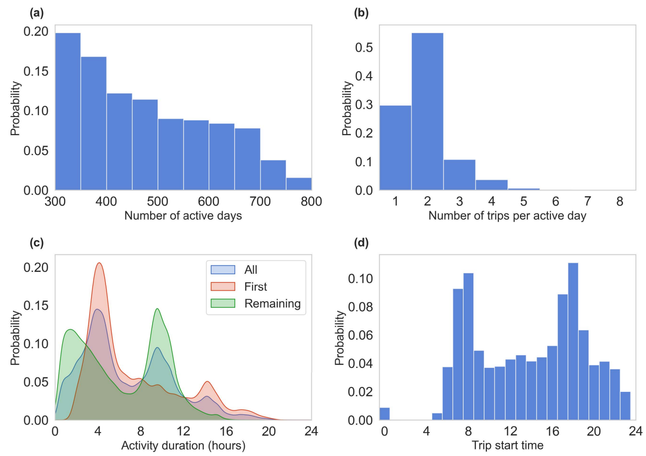


Fig. 4. Distribution of individual travel characteristics. In (c), "first" means the duration distribution of the first activity in a day. "Remaining" means all activities in a day excluding the first one. "All" means all activities.

IV. Case Study

A. Data

The dataset used for the case study contains transit smart card records from 500 anonymous users between July 2014 and March 2017 in the Hong Kong Mass Transit Railway (MTR) system. These users are selected randomly from all individuals with at least 300 active days of transit usage during the study period, which excludes occasional users and short-term visitors such as tourists. This is because a minimum amount of personal travel history is required to achieve reasonable prediction performance. The mobility prediction for infrequent users and short-term visitors requires future research. It is worth noting that though an individual may hold more than one smart card and a smart card data may represent multiple users, we assume that each card ID corresponds to only one user [2].

We partition the personal daily activity sequences of each user into training and test sets. The test set consists of the sequences from 20% randomly selected active days. The remaining sequences form the training set. The proposed model is specified for each user based on their own training data.

B. Travel Patterns

The travel patterns of selected sample individuals are shown in Figure 4. Figure 4(a) shows the distribution of the number of active days (i.e., days with at least one trip). We observe most of the samples have less than 400 active days during the 2.5 years of the analysis period. Figure 4(b) shows the distribution of the number of trips per active day. An individual typically makes two trips in a day, likely as a result of commuting to and from work. Note that only rail-based trips are considered in the case study. Thus, this distribution is an underrepresentation of the true travel intensity of users. The distribution of activity duration is shown in Figure 4(c). For the first activity (i.e., the first activity in a day), we observe a prominent peak of around 4 hours. This may represent the weekday "staying at home" activity because the start of a day is set as 4:00 AM and people usually leave home for work at around 8:00 AM. There is a sub-peak at around 14 hours for the first activity, which may correspond to the holiday "staying at home" activity where people stay at home until 18:00 and then leave home for leisure. For the remaining activities (i.e., all activities in a day excluding the first one), a major peak at around 10 hours is observed, which may indicate the work activity (start at 8:00 and end at 18:00). Another peak for remaining activity duration is around 2 hours, which may represent short-term dining/entertainment activities. The trip start time distribution is shown in Figure 4(d), as expected, a morning peak at 8:00 AM and an evening peak at 18:00 are observed.

C. Evaluation Metrics and Benchmark Models

Recall that the proposed IOHMM can output the predicted activity duration ($r_t^{u,v}$) and end location ($q_t^{u,v}$). $q_t^{u,v}$ is a categorical variable and the prediction accuracy is used for the performance evaluation. $r_t^{u,v}$ is a continuous variable. The predicted R^2 (i.e. R^2 in the test data set) is used as the main performance metric because it typically ranges between 0 and 1, which is consistent with the range of prediction accuracy.

To properly evaluate the proposed IOHMM, we compare it against two types of models for benchmarking. The first group includes simple and straightforward models and can be seen as a "lower bound" of the prediction performance. The second group is based on more advanced machine learning methods that are commonly used for sequential prediction. It can be seen as providing an approximate "upper bound" of the prediction performance for existing approaches. Specifically, linear regression (LR) and the first-order Markov Chain (MC) model are used as the first type benchmark models for predicting $r_t^{u,v}$ and $q_t^{u,v}$, respectively. LR is used because it is the most commonly used model for continuous variable prediction. The MC model is used because it was shown in [4] that the first-order MC can approach the limit of predictability for the next location prediction and it was previously used in [2] as the baseline model for location prediction. Moreover, we also include the mobility n-gram (NG) model proposed in [2] as a benchmark for the location prediction.

The LR model for user u is formulated as

$$r_t^{u,v} = \beta_0^u + \beta^u \cdot \boldsymbol{z_t^{u,v}} + \epsilon^u \quad \forall v, t \tag{22}$$

where β_0^u is the intercept and β^u is the vector of parameters to estimate. ϵ^u is the error term.

In terms of the MC model, the distribution of the activity end location (i.e. next trip origin) is formulated as

$$\mathbb{P}(q_1^{u,v}) = \mathbb{P}(o_1^{u,v}) = \frac{C(o_1^{u,v}) + \alpha/|\mathcal{L}^u|}{V^u + \alpha} \tag{23}$$

$$\begin{aligned}\mathbb{P}(q_t^{u,v} \mid q_{t-1}^{u,v}) &= \mathbb{P}(o_t^{u,v} \mid o_{t-1}^{u,v}) \\ &= \frac{C(o_{t-1}^{u,v}, o_t^{u,v}) + \alpha/|\mathcal{L}^u|}{\sum_{\tilde{o}_t^{u,v} \in \mathcal{L}^u} C(o_{t-1}^{u,v}, \tilde{o}_t^{u,v}) + \alpha} \quad \forall t \geq 2\end{aligned} \tag{24}$$

where $C(o_1^{u,v})$ is a counting function that returns the number of times that the first activity of the day ends at $o_1^{u,v}$. Similarly, $C(o_{t-1}^{u,v}, o_t^{u,v})$ returns the number of times that an activity

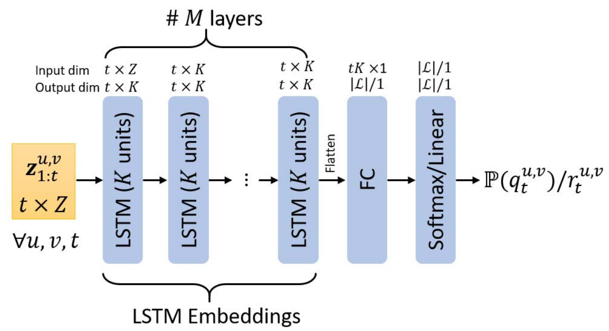


Fig. 5. LSTM network structure. M is the number of LSTM layers. K is the number of hidden units. "FC" means fully connected neural network layer.

ending at $o_{t-1}^{u,v}$ is followed (in the same day) by another activity ending at $o_t^{u,v}$. $\mathcal{L}^u$ is the set of candidates location for user u. The parameter α is used for smoothing so that a non-zero probability is generated for any possible value.

Long short-term memory (LSTM) [44] is selected to represent the second type of benchmark model. It is well-suited to classify, process, and predict time series given time lags of unknown duration as it has the advantage of memorizing long-range dependencies in the data. LSTM is often considered one of the state-of-the-art methods for time series prediction tasks. Since LSTM is not suitable for predicting continuous ($r_t^{u,v}$) and discrete ($q_t^{u,v}$) variables simultaneously, we train for each individual two separate LSTM models to predict $r_t^{u,v}$ and $q_t^{u,v}$, respectively. The structure of LSTM is shown in Figure 5. Input variables are fed into M LSTM layers and then aggregated by a fully connected (FC) neural network layer. For duration (resp. location) prediction, the linear (resp. softmax) activation layer is used for the outputs. The hyper-parameters (e.g., number of LSTM layers M, number of hidden units K, regularization strength, drop out rate) are tuned based on a searching process over a predetermined hyper-parameter space (see Appendix B). A validation data set (20% of the training data) is used for the hyper-parameters selection. The hyper-parameters with the highest R^2 and prediction accuracy in the validation data are used as the final models.

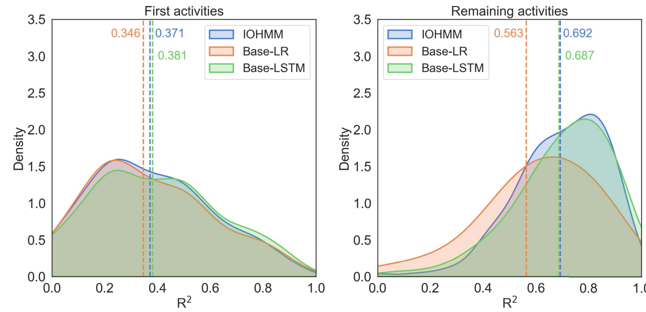
The key idea of the NG model for the next tap-in location prediction is to construct the following conditional probability based on individual's travel histories.

$$\mathbb{P}(q_t^{u,v} \mid x_{t-1}^{u,v}, o_{t-1}^{u,v}, d_{t-1}^{u,v}). \tag{25}$$

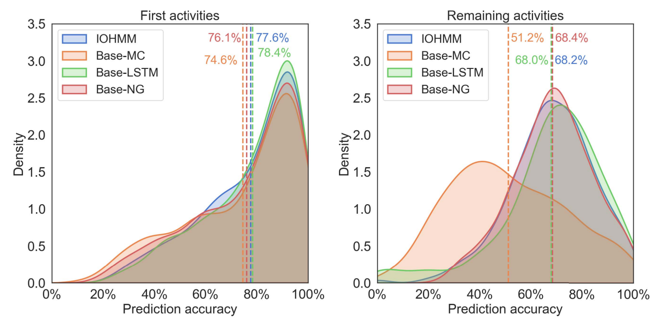
Details of the NG model can be found in Zhao et al. [2].

D. Prediction Performance

As each model outputs R^2 and prediction accuracy for each individual, we can plot the distribution of R^2 and prediction accuracy for overall performance evaluation. Figure 6 shows the prediction performance for the activity duration and location. Since the first activities are predicted by the initial probability and the remaining activities are predicted by the transition probability, we plot the performance distribution for two types of activities separately. What stands out in the figure is a high degree of individual heterogeneity in terms



(a) R^2 distribution for activity duration prediction



(b) Accuracy distribution for activity end location prediction

Fig. 6. Prediction performance. "First activities" means the first activity in a day. "Remaining activities" are all other activities in a day excluding the first one. The dash lines represent the mean value.

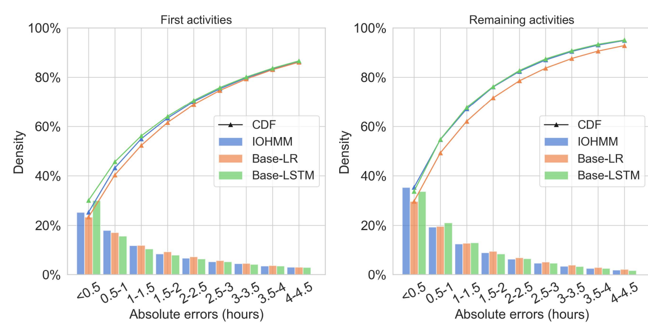


Fig. 7. Distribution of prediction errors for activity duration. The errors are aggregated in 0.5 hour intervals for better visualization. The solid lines are cumulative density functions (CDF) with colors corresponding to different models.

of predictability. Overall, the IOHMM shows very similar performance as the LSTM model in all prediction tasks. And both IOHMM and LSTM can outperform the first type of baseline models (i.e. LR and MC). This implies that the proposed IOHMM not only has the same predictive capacity as the advanced machine learning model but also has the potential to identify latent activities with model interpretability (details illustrated in Section IV-F).

In terms of the duration prediction (Figure 6a), we observe that IOHMM and LSTM are only slightly better than LR (with mean $R^2 = 0.371, 0.381$, and 0.346, respectively). This implies that people's first trip start time on a day has high randomness and is hard to predict as many uncaptured reasons can cause morning departure times to be adjusted.

However, for the remaining activities, the IOHMM and LSTM significantly outperform the LR model (with mean R^2 = 0.692, 0.687, and 0.563, respectively).

The results for location prediction (Figure 6b) are similar to those of duration prediction. IOHMM, NG, and LSTM models are slightly better than the MC model in the first activity end location prediction, but significantly better in the prediction of remaining activities. An interesting finding is that, though the duration of the first activity is relatively difficult to predict, the prediction accuracy for the first activity end location (i.e. first trip origin) is high (with a mean of 77.6%, 76.1%, 78.4%, and 74.6% for IOHMM, NG, LSTM, and MC, respectively). This implies that despite randomness in start time, the first trip origins are relatively stable for these frequent public transit users. The location prediction accuracy for the remaining activities is lower than that of the first activities (with a mean of 68.2%, 68.4%, 68.0%, and 51.2% for IOHMM, NG, LSTM, and MC, respectively). This may be attributed to the higher degree of behavioral randomness after leaving home. For the first activity, people are likely to use the nearest rail station around the home. But for remaining activities, people may have more choices that are not easy to capture.

In addition to R^2, it is also useful to examine the magnitude of duration prediction errors. The distribution of absolute errors for duration prediction is shown in Figure 7. Overall, errors within 30 minutes account for the highest fraction for all models. For the remaining activities, more than half of the activity duration can be predicted with errors within 1 hour for our IOHMM model. For the first activities, we observe that the LSTM model has a higher density in errors smaller than 30 minutes compared to IOHMM. However, for the remaining activities, IOHMM accounts for a higher density for prediction errors within 30 minutes. As for errors within 1.5 hours, the performance of IOHMM and LSTM models is similar, and both models outperform the LR model. This indicates that LSTM may have more advantages for the first activity duration prediction while IOHMM for remaining activities. This may be because the duration of the first activity (usually staying at home) is hard to predict given the complex interactions of different factors (such as weather, holidays, or even some unobserved factors such as users' moods). LSTM is more complicated than IOHMM in terms of model structure and the number of parameters. Hence, it may have more prediction power to capture the underlying factor interactions, thus outperforming the first activity duration prediction task.

Figure 8 shows the cumulative distribution of prediction rank for activity end location. The cumulative probability (on the y-axis) at rank k represents the probability that the true activity end location is among the top-k (on the x-axis) most likely outcomes predicted by the model. We observe that, for the first activity, there is more than 90% probability that one of the top 3 predictions in the proposed model is correct. But for the remaining activities, we need to include the top 10 predicted outcomes to achieve 90% probability. The results imply that the origins of the first trips (i.e. first activity end locations) are easier to predict with limited variations than those of following trips. Similarly, IOHMM, NG, and LSTM models achieve comparable (essentially the same) performance in the remaining activity prediction, and both consistently outperform the MC model. However, for the first activities, the NG model becomes worse when counting for more than the top 5 predictions.

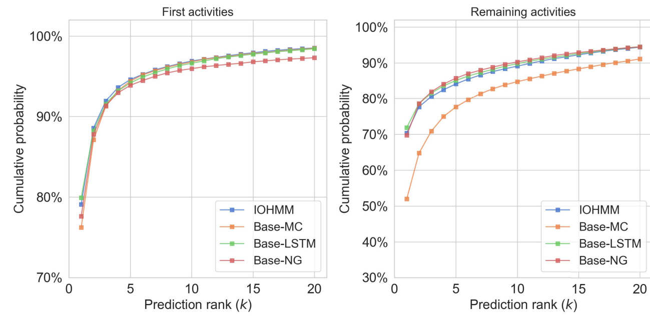


Fig. 8. Cumulative distribution of the prediction ranks for activity end location.

E. Factors Impacting Individual Mobility Predictability

As shown in Figure 6, the prediction performance varies greatly among passengers. Hence, it is worth evaluating which attributes affect the individual's predictability. We estimated two linear regression models with the R^2 (for duration prediction) and prediction accuracy (for location prediction) of IOHMM as dependent variables. Independent variables are factors related to a user's travel frequency, regularity, fare card type, number of estimated hidden activities, and inferred home location. To reveal how longitudinal behavior changes influence predictability, we introduce the "number of change points" for departure time and visited locations calculated from a Bayesian model in our previous study [45] as new independent variables. These two variables describe the number of substantial behavior pattern changes in terms of departure time and ODs, respectively.

Table II shows the results of estimated coefficients. We observe that the number of days with travel is significantly positive to location prediction, which implies that longer historical trips can increase location predictability. Similarly, the mean number of trips per day is significantly positive for both duration and location prediction. This may be because a high mean number of trips per day reflects longer daily travel sequences, which can potentially make it easier to uncover sequential dependencies and ultimately help with prediction performance. Variables that indicate high travel irregularity, such as the standard deviation (std.) of travel frequency and departure time, number of change points for departure time and locations, have significant negative effects on the prediction performance. We also find that senior passengers' activity duration is harder to predict. In addition, the activity locations for users living in New Territories (one of the three main regions of Hong Kong, alongside Hong Kong Island and the Kowloon Peninsula) are easier to predict. This may be because New Territories is further away from the commercial business center of Hong Kong, and its residents generally have less diverse socioeconomic activities outside commuting between home and work. The number of hidden activities has

TABLE II

FACTORS ON INDIVIDUAL MOBILITY PREDICTABILITY

Variables	Coefficients			
	Duration		End location	
Intercept	0.6423	**	0.8441	**
Total # of days with travel	3.24×10^{-6}		9.92×10^{-5}	**
Mean # of trips per day	0.1237	**	0.1074	**
Std. # of trips per day	-0.1488	**	-0.3058	**
Std. of departure time of the first trip in a day	-0.0009	**	-0.0005	**
Student	-0.0298		-0.0094	
Senior	-0.0786	**	0.0023	
[1]Living in Hong Kong Island	0.0030		0.0103	
[1]Living in New Territories	0.0068		0.0165	**
# change points (time)	-0.0132	**	N.A.	
# change points (location)	N.A.		-0.0164	**
# hidden activities (N^u)	-0.0014		-0.0033	

Number of observations: 500.
Duration R^2: 0.295; End location R^2: 0.484
**: p-value < 0.01; *: p-value < 0.05.
[1]: A user's home location is inferred as the most frequently used tap-in station for the first trip of a day.

no significant impact on either location or duration prediction accuracy.

F. Latent Activity Identification

Although IOHMM has a similar performance as the LSTM model, the advantage of IOHMM is its ability to identify latent activity patterns for each individual. Figure 9 presents the spatiotemporal distributions of three latent activities (see Section III-F) for a selected individual. Note that the activity labels are manually assigned based on their corresponding characteristics. Also, since the mean duration is a function of z_t (Eq. 18), the activity duration under a specific activity type can be multi-modal depending on the distribution of z_t.

The first activity (i.e. the first column) has a start time peak around 4:00 AM, a duration peak of around 4 hours, a dominant activity start location "Null" (i.e. the one with the highest probability and is much higher than others), and a dominant activity end location "CSW" (Cheung Sha Wan). This is obviously associated with "home" activity because people usually stay at home from the beginning of a day (4:00 AM) to the departure time for working (around 8:00 AM) with a duration of 4 hours. By definition, the first activity of a day has no activity start location (i.e. Null). And the dominant activity end location (i.e. CSW) should be the nearest station to his/her house.

The second activity (i.e. the second column) has a start time peak around 8:00 AM, a duration peak of around 10 hours, a dominant activity start and end location "MOK" (Mong Kok). We can easily associate it with the "work" activity, because the activity start time and duration match the typical work schedule, and the activity start and end locations are the same, which means during the activity the user does not move. MOK station is located in the CBD area in Hong Kong, which should be the nearest station to his/her working place.

The third activity (i.e. the third column) has a relatively dispersed start time and duration compared to the first two.

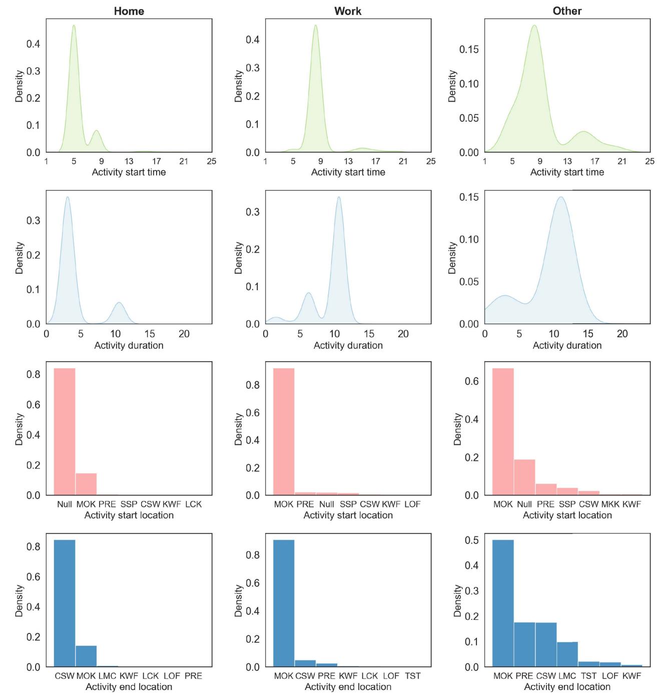


Fig. 9. Activity patterns of a selected individual with $N^u = 3$. The i-th column represents the conditional distributions $\mathbb{P}(\cdot \mid A = i)$. Each column (i.e. each latent activity) is associated with a semantic label (i.e. "home", "work", and "other") based on its distribution patterns.

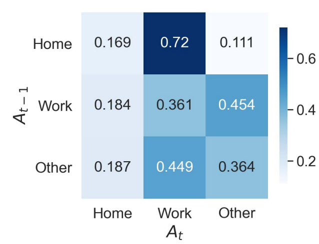


Fig. 10. Activity transition matrix ($\mathbb{P}(A_t \mid A_{t-1})$) of the selected individual.

And the activity locations are more diverse. So, we associated it with "other" activities such as entertainment and dining. The activity start and end locations with the highest probability are both MOK. It seems the user prefers to perform other activities around MOK as well, which makes sense as MOK is one of the busiest areas in Hong Kong with many shops and restaurants. It is also likely that some of the other activities are work-based, as they may be planned around the work location.

The activity transition matrix is shown in Figure 10. As expected, the transition from home to work shows the highest probability. It is worth noting that as the activity after the last trip in a day is omitted from analysis (see Section III-B), the typical work to home transition is not revealed in the model. The most likely activity following work is other, which may represent work-based shopping, dining, and entertainment activities (corresponding to results in Section IV-F). And the most likely activity following other is work. This also makes sense because the user usually conducts work-based other activities (such as dining) and after that he/she may need to return to work.

TABLE III
ESTIMATED PARAMETERS FOR DURATION PREDICTION

Activity	Estimated parameters ($\boldsymbol{\theta}_{emr,i}$)			
	Rainy	Monday	Sunday	Public holiday
Home	0.159	0.001	0.025	0.174
Work	0.336	0.173	-0.221	-0.065
Others	0.364	0.211	-0.001	-0.013

TABLE IV
SUMMARY OF CONTEXTUAL VARIABLES z_t

Variables	Type	Variables	Type
Rainy (Yes = 1)	D	National holidays (Yes = 1)	D
Heavy rain (Yes = 1)	D	Last activity duration	C
Sunny (Yes = 1)	D	Last trip travel time	C
Cloudy (Yes = 1)	D	# days with travel in past 20 days	I
Daily mean temperature	C	# consecutive days without travel	I
Monday - Sunday (Yes = 1)	D	# trips yesterday	I

Variable type: D: Dummy; C: Continuous; I: Integer.

Since the mean activity duration is specified with a linear model in IOHMM (see Section III-D), the estimated parameters in the linear model are useful for understanding other contextual factors affecting activity duration. Table III summarizes some estimated parameters with interpretability for the same selected individual. As expected, the duration for all activities is higher on rainy days (compared to those without rain). It is worth noting that the sum of three activity duration is not a fixed value, since the activity after the last trip is not considered. Thus, the rainy parameter can be positive for all three activities, which means users may delay their trip departure time when it rains outside. Monday has a positive impact on the work and other activities duration (the impact of Monday on home activity is negligible compared to the other two). Sunday has a negative impact on the work activity duration and a positive impact on home activity duration. On public holidays, the duration of home activities increases and decreases for the other two. Since "other" activities for this individual are usually work-based, all these effects are reasonable.

V. CONCLUSION AND DISCUSSION

This paper proposes an activity-based IOHMM framework to simultaneously predict the time and location of an individual's next trip using smart card data. The prediction task can be transformed into predicting the hidden activity duration and end location, which enables a natural behavioral representation. Based on a case study with data from Hong Kong's MTR system, we show that the proposed model has a similar prediction performance as the advanced LSTM model, and significantly outperforms the benchmark models. Unlike LSTM, the proposed activity-based model can also be used to analyze hidden activity patterns, which provides meaningful behavioral interpretation for why an individual makes a certain trip. Therefore, the activity-based prediction framework offers a way to combine the predictive power of machine learning methods and the behavioral interpretability of activity-based models. The estimated activity (or travel purpose) information can facilitate the development of situational awareness in intelligent transportation applications, such as personalized traveler information or targeted travel demand management in public transit systems [46]. Activity-based models have been used extensively in travel demand forecasting and simulation. As demonstrated in [38], IOHMM is well suited to simulating individual activity-travel behavior, which can help public transit agencies to better design policies and plan future services.

A natural extension of this study is to apply the proposed activity-based modeling framework to other data sources and mobility systems. Note that while the Hong Kong MTR system is used as a case study, our approach is agnostic to particular modes. The only information required is the longitudinal observations of individual travel history, including the start/end time, origin, destination, and individual identifier of each trip. Such travel information is generally available in most of the intrinsic mobility data sources. New mobility service providers, such as ride-hailing systems, bike-sharing programs as well as on-demand "pop-up" bus services, also collect individual-level travel records (typically through mobile apps) similar to the transit smart card data, though the predictability of individual mobility may vary by different systems. It is expected that the predictability is higher for public transit systems, because of a higher proportion of commuting trips. A further distinction can be made between stationed systems (e.g., subway, buses, docked bike-sharing) and stationless systems (e.g., taxis, ride-hailing, and dockless bike-sharing) [47]. For the latter, certain spatial aggregation is needed for the proposed method to work.

The proposed methodology is not without limitations. One limitation is that the model requires a long observation period of individual trip records, and does not work well with infrequent or new users with little to no travel histories. Future studies can leverage user clustering techniques to extract similar users' travel patterns as additional input [48], which can compensate for the sparsity of individual data. Another limitation lies in the assumption of stable travel patterns. However, individual travel patterns may change over time, leading to reduced predictive performance due to domain shift (or distributional shift) issues. To address this, we could potentially adopt a change detection module [45] to guide the update of individual mobility prediction model parameters. One way to do this is through dynamic weighting of data points based on behavior change patterns.

APPENDIX A
SUMMARY OF CONTEXTUAL VARIABLES

Table IV shows the summary of contextual variables z_t. Five different dimensions are considered: weather, day of the week, holidays, last trip information, and historical travel statistics.

APPENDIX B
HYPER-PARAMETER SPACE OF THE LSTM MODEL

The hyper-parameters of the LSTM model used in this study (for all individuals) are $M = 1$, $K = 50$, dropout rate $= 0.3$, l_1 regularization $= 0$, l_2 regularization $= 0$, batch size $= 30$.

TABLE V
HYPER-PARAMETER SPACE OF THE LSTM MODEL

Hyper-parameters	Value space
# of LSTM layers M	{1, 2, 3, 4, 5}
# of units in LSTM layer K	{30, 50, 100, 150, 200}
Dropout rate	{0.1, 0.3, 0.5, 0.7}
l_1 regularization	$\{0, 10^{-6}, 10^{-4}, 0.01, 0.1, 0.5\}$
l_2 regularization	$\{0, 10^{-6}, 10^{-4}, 0.01, 0.1, 0.5\}$
Batch size	{20, 30, 50, 70}

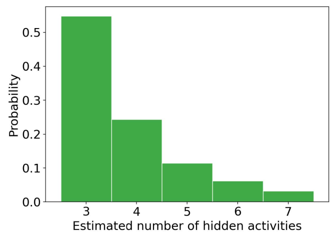


Fig. 11. Distribution of number of estimated hidden activities.

The model is trained using Adam optimizer (with the default learning rate) with 200 training epochs.

APPENDIX C
ANALYSIS ON NUMBER OF HIDDEN ACTIVITIES

Figure 11 shows the distribution of the number of estimated hidden activities for all 500 samples. Most of the users have three hidden activities. And with the increase in the number of activities, the proportion of users decreases. The high proportion of 3-activity users indicates that most of the frequent users (with at least 300 active days of transit usage during the study period) in the MTR system can be characterized by three major activity patterns: home, work, and other.

ACKNOWLEDGMENT

The authors would like to thank Hong Kong's Mass Transit Railway (MTR) for their support and data availability for this research.

REFERENCES

[1] G. Goulet-Langlois, H. N. Koutsopoulos, Z. Zhao, and J. Zhao, "Measuring regularity of individual travel patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1583–1592, May 2017.

[2] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Individual mobility prediction using transit smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 19–34, Apr. 2018.

[3] F. Calabrese, G. Di Lorenzo, and C. Ratti, "Human mobility prediction based on individual and collective geographical preferences," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2010, pp. 312–317.

[4] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Sci. Rep.*, vol. 3, no. 1, Oct. 2013, Art. no. 2923. [Online]. Available: http://www.nature.com/articles/srep02923

[5] B. Hawelka, I. Sitko, P. Kazakopoulos, and E. Beinat, "Collective prediction of individual mobility traces for users with short data history," *PLoS ONE*, vol. 12, no. 1, Jan. 2017, Art. no. e0170907. [Online]. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170907

[6] F. Alhasoun, M. Alhazzani, and M. C. Gonzalez, "City scale next place prediction from sparse data through similar strangers," presented in the 6th Int. Workshop Urban Comput. [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=StmgWSAAAAAJ&citation_for_view=StmgWSAAAAAJ:IjCSPb-OGe4C

[7] P. Noursalehi, H. N. Koutsopoulos, and J. Zhao, "Real time transit demand prediction capturing station interactions and impact of special events," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 277–300, Dec. 2018.

[8] P. Noursalehi, H. N. Koutsopoulos, and J. Zhao, "Dynamic origin-destination prediction in urban rail systems: A multi-resolution spatio-temporal deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, early access, Jan. 11, 2021, doi: 10.1109/TITS.2020.3047047.

[9] X. Yang, Q. Xue, M. Ding, J. Wu, and Z. Gao, "Short-term prediction of passenger volume for urban rail systems: A deep learning approach based on smart-card data," *Int. J. Prod. Econ.*, vol. 231, Jan. 2021, Art. no. 107920.

[10] Z. Cheng, M. Trepanier, and L. Sun, "Real-time forecasting of metro origin-destination matrices with high-order weighted dynamic mode decomposition," 2021, *arXiv:2101.00466*. [Online]. Available: http://arxiv.org/abs/2101.00466

[11] P. Noursalehi, H. N. Koutsopoulos, and J. Zhao, "Predictive decision support platform and its application in crowding prediction and passenger information generation," *Transp. Res. C, Emerg. Technol.*, vol. 129, Aug. 2021, Art. no. 103139.

[12] Z. Fang, Q. Cheng, R. Jia, and Z. Liu, "Urban rail transit demand analysis and prediction: A review of recent studies," in *Intelligent Interactive Multimedia Systems and Services* (Smart Innovation, Systems and Technologies), vol. 8, G. De Pietro, L. Gallo, R. Howlett, L. Jain, and L. Vlacic, Eds. Cham, Switzerland: Springer, 2018, pp. 300–309. [Online]. Available: https://link-springer-com.eproxy.lib.hku.hk/chapter/10.1007/978-3-319-92231-7_31, doi: 10.1007/978-3-319-92231-7_31.

[13] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, "Next place prediction using mobility Markov chains," in *Proc. 1st Workshop Meas., Privacy Mobility*. New York, NY, USA: ACM, 2012, pp. 1–6, doi: 10.1145/2181196.2181199.

[14] A. Asahara, K. Maruyama, A. Sato, and K. Seto, "Pedestrian-movement prediction based on mixed Markov-chain model," in *Proc. 19th Int. Conf. Adv. Geographic Inf. Syst.* New York, NY, USA: ACM, 2011, pp. 25–33, doi: 10.1145/2093973.2093979.

[15] W. Mathew, R. Raposo, and B. Martins, "Predicting future locations with hidden Markov models," in *Proc. ACM Conf. Ubiquitous Comput.*, New York, NY, USA, 2012, pp. 911–918.

[16] Q. Liu, S. Wu, L. Wang, and T. Tan, "Predicting the next location: A recurrent model with spatial and temporal contexts," in *Proc. Conf. Artif. Intell.* Phoenix, AZ, USA: AAAI Press, Feb. 2016, pp. 194–200.

[17] A. Al-Molegi, M. Jabreel, and B. Ghaleb, "STF-RNN: Space time features-based recurrent neural network for predicting people next location," in *Proc. IEEE Symp. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–7.

[18] J. Feng *et al.*, "DeepMove: Predicting human mobility with attentional recurrent networks," in *Proc. World Wide Web Conf.*, Apr. 2018, pp. 1459–1468, doi: 10.1145/3178876.3186058.

[19] D. Kong and F. Wu, "HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction," in *Proc. IJCAI*, 2018, pp. 2341–2347. [Online]. Available: https://www.ijcai.org/proceedings/2018/324

[20] F. Li *et al.*, "A hierarchical temporal attention-based LSTM encoder-decoder model for individual mobility prediction," *Neurocomputing*, vol. 403, pp. 153–166, Aug. 2020.

[21] J. Feng, C. Rong, F. Sun, D. Guo, and Y. Li, "PMF: A privacy-preserving human mobility prediction framework via federated learning," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, Mar. 2020, vol. 4, no. 1, pp. 1–21.

[22] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2018, pp. 1672–1678.

[23] Y. Liang and Z. Zhao, "Vehicle trajectory prediction in city-scale road networks using a direction-based sequence-to-sequence model with spatiotemporal attention mechanisms," 2021, *arXiv:2106.11175*. [Online]. Available: http://arxiv.org/abs/2106.11175

[24] G. Gidófalvi and F. Dong, "When and where next: Individual mobility prediction," in *Proc. 1st ACM SIGSPATIAL Int. Workshop Mobile Geographic Inf. Syst.* New York, NY, USA: ACM, 2012, pp. 57–64.

[25] H.-P. Hsieh, C.-T. Li, and X. Gao, "T-gram: A time-aware language model to predict human mobility," in *Proc. 9th Int. Conf. Web Social Media*, Apr. 2015, pp. 614–617. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10559

[26] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model," *Transp. Res. C, Emerg. Technol.*, vol. 116, Jul. 2020, Art. no. 102627.

[27] K. W. Axhausen and T. Gärling, "Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems," *Transp. Rev.*, vol. 12, no. 4, pp. 323–341, Oct. 1992.

[28] C. R. Bhat and F. S. Koppelman, "Activity-based modeling of travel demand," in *Handbook Transportation Science* (International Series in Operations Research Management Science). Boston, MA, USA: Springer, 1999, pp. 35–61.

[29] S. Rasouli and H. Timmermans, "Activity-based models of travel demand: Promises, progress and prospects," *Int. J. Urban Sci.*, vol. 18, no. 1, pp. 31–60, Jan. 2014.

[30] S. Hasan and S. V. Ukkusuri, "Urban activity pattern classification using topic models from online geo-location data," *Transp. Res. C, Emerg. Technol.*, vol. 44, pp. 363–381, Jul. 2014.

[31] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore," *IEEE Trans. Big Data*, vol. 3, no. 2, pp. 208–219, Jun. 2017.

[32] G. Han and K. Sohn, "Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model," *Transp. Res. B, Methodol.*, vol. 83, pp. 121–135, Jan. 2016.

[33] Y. Bengio and P. Frasconi, "An input output HMM architecture," in *Proc. Adv. Neural Inf. Process. Syst.*, G. Tesauro, D. S. Touretzky, and T. K. Leen, Eds. Cambridge, MA, USA: MIT Press, 1995, pp. 427–434.

[34] Y. Bengio and P. Frasconi, "Input-output HMMs for sequence processing," *IEEE Trans. Neural Netw.*, vol. 7, no. 5, pp. 1231–1249, Sep. 1996.

[35] S. Marcel, O. Bernier, J. E. Viallet, and D. Collobert, "Hand gesture recognition using input-output hidden Markov models," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Mar. 2000, pp. 456–461.

[36] Y. Li and H.-Y. Shum, "Learning dynamic audio-visual mapping with input-output Hidden Markov models," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 542–549, Jun. 2006.

[37] A. M. González, A. M. S. Roque, and J. García-González, "Modeling and forecasting electricity prices with input/output hidden Markov models," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 13–24, Feb. 2005.

[38] M. Yin, M. Sheehan, S. Feygin, J.-F. Paiement, and A. Pozdnoukhov, "A generative model of urban activities from cellular data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 6, pp. 1682–1696, Jun. 2018.

[39] N. Eagle and S. A. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, 2006.

[40] M. Kim and D. Kotz, "Periodic properties of user mobility and access-point popularity," *Pers. Ubiquitous Comput.*, vol. 11, no. 6, pp. 465–479, Aug. 2007.

[41] S. Chiappa and S. Bengio, "HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems," in *Proc. Eur. Symp. Artif. Neural Netw.*, Bruges, Belgium, Apr. 2004, pp. 199–204. [Online]. Available: https://csilviavr.github.io/assets/publications/silvia04hmm.pdf

[42] J. Geiger, J. Schenk, F. Wallhoff, and G. Rigoll, "Optimizing the number of states for HMM-based on-line handwritten whiteboard recognition," in *Proc. 12th Int. Conf. Frontiers Handw. Recognit.*, Nov. 2010, pp. 107–112.

[43] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987.

[44] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[45] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Detecting pattern changes in individual travel behavior: A Bayesian approach," *Transp. Res. B, Methodol.*, vol. 112, pp. 73–88, Jun. 2018.

[46] B. Mo, Z. Ma, H. N. Koutsopoulos, and J. Zhao, "Capacity-constrained network performance model for urban rail systems," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 5, pp. 59–69, May 2020.

[47] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Uncovering spatiotemporal structures from transit smart card data for individual mobility modeling," in *Demand for Emerging Transportation Systems*, C. Antoniou, D. Efthymiou, and E. Chaniotakis, Eds. Amsterdam, The Netherlands: Elsevier, Jan. 2020, pp. 123–149.

[48] F. Alhasoun, M. Alhazzani, F. Aleissa, R. Alnasser, and M. González, "City scale next place prediction from sparse data through similar strangers," in *Proc. ACM KDD Workshop*, 2017, pp. 191–196.

Baichuan Mo received the bachelor's degree in civil engineering from Tsinghua University and the dual master's degree in transportation and computer science from MIT, where he is currently pursuing the Ph.D. degree with the Department of Civil and Environmental Engineering. His research interests include data-driven transportation modeling, demand modeling, and applied machine learning, with a specific application in public transit systems.

Zhan Zhao received the bachelor's degree from Tongji University, the master's degree from The University of British Columbia, and the Ph.D. degree from Massachusetts Institute of Technology. Prior to joining The University of Hong Kong (HKU), he was a Senior Data Scientist at Via Transportation, Inc. He is currently an Assistant Professor with the Department of Urban Planning and Design, HKU. His research interests include human mobility, public transportation systems, and urban data science.

Haris N. Koutsopoulos is currently a Professor with the Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA, and a Guest Professor with the KTH Royal Institute of Technology, Stockholm. He founded the iMobility Laboratory, which uses information and communication technologies to address urban mobility problems. His current research focuses on the use of data from opportunistic and dedicated sensors to improve planning, operations, monitoring, and control of urban transportation systems, including public transportation. The laboratory received the IBM Smarter Planet Award in 2012.

Jinhua Zhao is currently the Edward H. and Joyce Linde Associate Professor of city and transportation planning at MIT. He brings behavioral science and transportation technology together to shape travel behavior, design mobility systems, and reform urban policies. He directs the MIT Urban Mobility Laboratory and Public Transit Laboratory.