Inferring Spatiotemporal Mobility Patterns from Multidimensional Trip Data

Jeongyun Kim*, Andrea Conti†, and Moe Z. Win*
*Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
†Department of Engineering and CNIT, University of Ferrara, 44122 Ferrara, Italy
Email: kjy@mit.edu, a.conti@ieee.org, moewin@mit.edu

***Abstract*—The massive amount of data related to spatiotemporal mobility offers new opportunities to understand human behaviors. However, with the increase of volume and complexity of mobility data, it has become challenging to retrieve important information and critical features of spatiotemporal mobility. In particular, predicting large-scale travel demands is challenging and requires a high computational load. This paper introduces a data-driven approach for estimating high-dimensional travel demands. We propose a method to identify mobility patterns using a probabilistic tensor decomposition approach for interpreting the complexity and uncertainty of mobility data. Expectation-maximization (EM) algorithm is applied for inferring mobility patterns. A case study is presented, where the proposed model is applied to New York city taxi data. The results show the model performance according to the number of origin and destination patterns and the number of trip data used. The probabilistic modeling results provide a deeper understanding of large-scale mobility data in the spatiotemporal dimension.**

***Index Terms*—Human mobility, travel demand modeling, probabilistic mobility pattern, tensor decomposition, data-driven estimation**

I. Introduction

Mobility data are essential to understand human activities and evaluate mobility systems. Over the past decade, information and communications technologies have helped to obtain large quantities of mobility data [1] such as smart phone data [2], [3], taxi trip data [4], and smart card transit data [5], [6].

Travel demand models have been determined in the last decades using mobility data including the four-step models, activity-based models, and statistical models [7], [8]. As the knowledge of travel demand in terms of origin and destination became important to plan efficient routes for decision making in transport service design, route planning, and fleet management [9], [10], there has been active discussion of origin-destination (OD) demand estimation in previous literature. Autoregressive integrated moving average (ARIMA) [11], [12], Poisson model [12], least-square modeling [13], and Kalman-filter [14] are well-known approaches for estimating time-series travel demand using historical data.

With the increased number of attributes and the spatial and temporal resolution of data, it becomes challenging to reveal major mobility patterns [15]. As a part of an effort to find the low-order dynamics that makes spatiotemporal patterns, the application of decomposition approaches to traffic data has recently received attentions. The goal of tensor decomposition is to capture the multi-dimensional structural dependencies by clustering data attributes and forcing representations of large datasets in terms of a small number of substructures [16]. Given its strength in retrieving information from large datasets, tensor decomposition has played an important role in discovering urban travel patterns [17]. Regardless of the number of attributes, identification of spatial and temporal travel patterns has been the major goal of the travel pattern identification research using tensor decomposition [18]–[21].

Most of the OD demand estimation studies over the past decades have focused on the quantity of travel demand. On the other hand, as reliability and resilience become issues in many mobility services, a more comprehensive understanding of demand stochasticity is required for transportation systems [22]. Few previous studies have identified spatial and temporal mobility patterns using a multi-way probabilistic factorization [15]. To expand the usability of the mobility patterns to predict future demand, it needs identification of patterns that are fixed or changeable over time, which helps to reduce the computational cost, especially when the data dimension is large. In addition, the effect of the number of patterns to the demand estimation accuracy needs to be further examined [15], [17], [23], [24].

The goal of this paper is to identify spatiotemporal mobility patterns from high-dimensional datasets using probabilistic tensor decomposition. We design an estimation algorithm to infer probabilistic mobility patterns, which allows to interpret the complexity and uncertainty inherent in the mobility data. We advocate the importance of temporal interactions among OD patterns to improve efficiency and accuracy of the demand estimation. The key contributions of this paper are as follows:

- we provide a probabilistic characterization of mobility patterns via a probabilistic tensor decomposition approach;
- we investigate how the number of possible spatiotemporal mobility patterns affects the demand estimation accuracy; and
- we apply the proposed model to New York city taxi data to reveal key trip information.

The remaining sections are organized as follows: Section II presents an approach for modeling and estimating trip demand. Section III provides modeling and estimation results using New York taxi data. Finally, Section IV gives our conclusions.

Notations: Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\mathbf{x}$ and $\boldsymbol{x}$; a random matrix and its realization are denoted by $\mathbf{X}$ and $\boldsymbol{X}$, respectively. Sets and random sets are denoted by upright sans serif and calligraphic font, respectively. For example, a random set and its realization are denoted by X and $\mathcal{X}$, respectively.

II. TRAVEL DEMAND MODEL ESTIMATION

This section presents the framework on OD demand modeling in a probabilistic way. The goal of this section is to characterize the demand probability distribution by discovering basis mobility patterns and temporal interactions among patterns.

A. Probabilistic trip demand model

A set of observed trip data $\mathcal{X} \triangleq \{\boldsymbol{x}_i, \forall i = 1, 2, \ldots, n\}$ consists of realizations of the independent and identically distributed (i.i.d.) random trip demand $\mathbf{x}_i \triangleq [\mathsf{o}_i, \mathsf{d}_i, \mathsf{t}_i]^{\mathrm{T}}$ with the common probability mass function (PMF) $p_{\mathbf{x}}(\boldsymbol{x})$. The random variables $\mathsf{o}_i, \mathsf{d}_i$, and t_i represent the indices of the origin, destination, and departure time of the trip demand such that $\mathsf{o}_i \in \{1, 2, \ldots, W_{\mathrm{o}}\}, \mathsf{d}_i \in \{1, 2, \ldots, W_{\mathrm{d}}\}, \mathsf{t}_i \in \{1, 2, \ldots, W_{\mathrm{t}}\}$. Similarly, the i-th trip data is expressed as

$$\boldsymbol{x}_i = \left[o_i, d_i, t_i\right]^{\mathrm{T}}, \forall i = 1, 2, \ldots, n. \tag{1}$$

The spatial and temporal probability distribution of trip demand can be expressed as a $W_{\mathrm{o}} \times W_{\mathrm{d}} \times W_{\mathrm{t}}$ tensor $\boldsymbol{V}$. Each element of $\boldsymbol{V}$ is defined as

$$[\boldsymbol{V}]_{c_{\mathrm{o}}, c_{\mathrm{d}}, c_{\mathrm{t}}} = p_{\mathbf{x}}([c_{\mathrm{o}}, c_{\mathrm{d}}, c_{\mathrm{t}}]^{\mathrm{T}}) \tag{2}$$

where $c_{\mathrm{o}}, c_{\mathrm{d}}$, and c_{t} are the cell indices of the origin, destination, and time axes of $\boldsymbol{V}$, respectively.

We aim to reveal the origin pattern and destination pattern inherent in the trip data $\mathcal{X}$ and their temporal interactions based on a probabilistic factorization approach. While an arbitrary tensor can be broken down using traditional tensor decomposition techniques, the probabilistic trip demand tensor can only be decomposed if certain conditions are met: i) the sum of the modeled tensor is 1 and ii) each pattern as well as temporal interactions are probabilistic. Given these constraints, we employ a probabilistic factorization approach to achieve the desired results. Specifically, the time domain is deliberately left undecomposed to retain the temporal interactions between origin and destination patterns. From this approach, we provide insights on the correlation between the patterns over time.

Latent class models are applied in this paper to infer the distribution $p_{\mathbf{x}}(\boldsymbol{x})$ in order to establish a connection between observed multivariate categorical data and a set of latent classes. This connection is necessary to capture the interactions between origin and destination pattern. The distribution

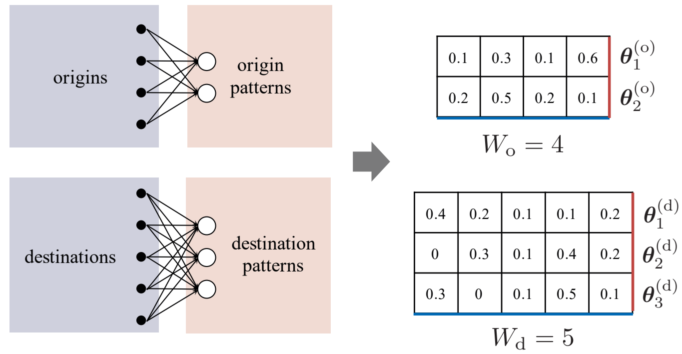


Fig. 1. Description on $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$.

$p_{\mathbf{x}}(\boldsymbol{x})$ is assumed to follow the categorical distribution with a parameter $\boldsymbol{\Theta}$, i.e., $q_{\mathbf{x}}(\boldsymbol{x}; \boldsymbol{\Theta})$ with

$$\boldsymbol{\Theta} \triangleq [\boldsymbol{\Theta}^{(\mathrm{o})}, \boldsymbol{\Theta}^{(\mathrm{d})}] \tag{3}$$

where $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ denote the origin pattern and destination pattern, respectively. When the numbers of patterns in origins and destinations are set to K_{o} and K_{d}, respectively, the dimensions of $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ are given by $K_{\mathrm{o}} \times W_{\mathrm{o}}$ and $K_{\mathrm{d}} \times W_{\mathrm{d}}$, respectively. For example, let us assume that we have trip data composed of four origins and five destinations, where $W_{\mathrm{o}} = 4$ and $W_{\mathrm{d}} = 5$. Since we want to find macroscopic movement patterns based on the trip data, the number of origin and destination patterns is set as smaller than that of the origins and destinations, i.e, $K_{\mathrm{o}} = 2$ and $K_{\mathrm{d}} = 3$. In this case, $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ can be described as Figure 1.

In detail, the k-th row of $\boldsymbol{\Theta}^{(\mathrm{o})}$ is the k-th origin pattern and denoted as $\boldsymbol{\theta}_{k_{\mathrm{o}}}^{(\mathrm{o})}, k_{\mathrm{o}} = 1, 2, \ldots, K_{\mathrm{o}}$. Each element of $\boldsymbol{\theta}_{k_{\mathrm{o}}}^{(\mathrm{o})}$, which is denoted as $\theta_{c_{\mathrm{o}} k_{\mathrm{o}}}^{(\mathrm{o})}$, represents the composition ratio of a cell c_{o} in the k-th origin. In a similar way, $\theta_{c_{\mathrm{d}} k_{\mathrm{d}}}^{(\mathrm{d})}$ is the element of $\boldsymbol{\theta}_{k_{\mathrm{d}}}^{(\mathrm{d})}$. These distributions should satisfy the following constraints:

$$\sum_{c_{\mathrm{o}}=1}^{W_{\mathrm{o}}} \theta_{c_{\mathrm{o}} k}^{(\mathrm{o})} = \sum_{c_{\mathrm{d}}=1}^{W_{\mathrm{d}}} \theta_{c_{\mathrm{d}} k}^{(\mathrm{d})} = 1. \tag{4}$$

As a source of modeling occurrence probability distribution of the trip demand, a core tensor $\boldsymbol{\Pi}$ is defined as temporal interaction information among $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$. Since there are K_{o} origin basis patterns and K_{d} destination basis patterns, there are $K_{\mathrm{o}} K_{\mathrm{d}}$ interactions for each time. Since the length of time is W_{t}, the dimension of $\boldsymbol{\Pi}$ is $(K_{\mathrm{o}}, K_{\mathrm{d}}, W_{\mathrm{t}})$. For each time index c_{t}, the probability that a trip belongs to the k_{o}-th origin pattern and the k_{d}-th destination pattern is defined as $\pi_{k_{\mathrm{o}} k_{\mathrm{d}}}^{(c_{\mathrm{t}})}$, which satisfies

$$\sum_{k_{\mathrm{o}}, k_{\mathrm{d}}} \pi_{k_{\mathrm{o}} k_{\mathrm{d}}}^{(c_{\mathrm{t}})} = 1. \tag{5}$$

Using $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, the occurrence probability of $\mathbf{x}_i = \boldsymbol{x}_i$ is modeled by a parameterized function as $p_{\mathbf{x}}(\boldsymbol{x}_i) \approx q_{\mathbf{x}}(\boldsymbol{x}_i; \boldsymbol{\Theta})$, which is given by

$$q_{\mathsf{x}}(\boldsymbol{x}_i;\boldsymbol{\Theta}) = \mathbb{P}\{\mathsf{x}_i = \boldsymbol{x}_i;\boldsymbol{\Theta}\} = \sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}} \pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)}\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}. \quad (6)$$

In (6), the probability that o_i is o_i and belongs to the k_o pattern is given by $\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}$, and the probability that d_i is d_i and belongs to the k_d pattern is given by $\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}$.

B. Model inference

For tractable maximum likelihood estimation of $\boldsymbol{\Theta}$ using the expectation-maximization (EM) algorithm, we introduce latent variables $\mathsf{z}_i \in \left\{[k_\mathrm{o},k_\mathrm{d}]^\mathrm{T} \mid k_\mathrm{o}=1,2,\ldots,K_\mathrm{o}, k_\mathrm{d}=1,2,\ldots,K_\mathrm{d}\right\}$ on the joint membership across all combinations of the OD patterns. From (6), the joint probability that $\mathsf{x}_i = x_i$ and $\mathsf{z}_i = z_i, \forall i$ is given by

$$p_{\mathsf{X},\mathsf{Z}}(\mathcal{X},\mathcal{Z};\boldsymbol{\Theta}) = \prod_{i=1}^{n}\prod_{k_\mathrm{o}=1}^{K_\mathrm{o}}\prod_{k_\mathrm{d}=1}^{K_\mathrm{d}}\left[\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)}\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}\right]^{\mathbb{I}(\boldsymbol{z}_i=[k_\mathrm{o},k_\mathrm{d}])} \quad (7)$$

where $\mathbb{I}$ is an indicator function with $\mathbb{I}=1$ if e is true and 0 otherwise. The log likelihood of $\boldsymbol{\Theta}$ can be written as,

$$\log\mathcal{L}(\boldsymbol{\Theta};\mathcal{X},\mathcal{Z}) = \log p_{\mathsf{X},\mathsf{Z}}(\mathcal{X},\mathcal{Z};\boldsymbol{\Theta}) \quad (8)$$

$$= \sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\mathbb{I}(\boldsymbol{z}_i=[k_\mathrm{o},k_\mathrm{d}]) \times\left(\log\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)}+\log\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}+\log\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}\right). \quad (9)$$

The EM algorithm first initializes $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$ with random values satisfying (4) and (5). Then, the following iterative procedures successively approximate $\boldsymbol{\Theta}$ until certain convergence criterion is met. Parameters updated at t-th iteration are denoted as $\boldsymbol{\Theta}^{[t]}$ and $\boldsymbol{\Pi}^{[t]}$.

- **E-step** compute the expected log likelihood of $\boldsymbol{\Theta}$ as follows,

$$\mathbb{E}_{\mathsf{Z}|\mathsf{X};\boldsymbol{\Theta}^{[t-1]}}\left\{\log p_{\mathsf{X},\mathsf{Z}}(\mathcal{X},\mathcal{Z};\boldsymbol{\Theta})|\mathcal{X};\boldsymbol{\Theta}^{[t-1]}\right\} = \sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\gamma_i^{k_\mathrm{o}k_\mathrm{d}}\left[\log\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)}+\log\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}+\log\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}\right] \quad (10)$$

where

$$\gamma_i^{k_\mathrm{o}k_\mathrm{d}} \triangleq \mathbb{E}_{\mathsf{z}_i|\mathsf{x}_i;\boldsymbol{\Theta}^{[t-1]}}\left\{\mathbb{I}(\boldsymbol{z}_i=[k_\mathrm{o},k_\mathrm{d}])|\boldsymbol{x}_i;\boldsymbol{\Theta}^{[t-1]}\right\} = \mathbb{P}\left\{\boldsymbol{z}_i=[k_\mathrm{o},k_\mathrm{d}]|\boldsymbol{x}_i;\boldsymbol{\Theta}^{[t-1]}\right\}. \quad (11)$$

- **M-step** Update $\boldsymbol{\Theta}$ with

$$\boldsymbol{\Theta}^{[t]} = \underset{\boldsymbol{\Theta}}{\mathrm{argmax}}\,\mathbb{E}_{\mathsf{Z}|\mathsf{X};\boldsymbol{\Theta}^{[t-1]}}\left\{\log p_{\mathsf{X},\mathsf{Z}}(\mathcal{X},\mathcal{Z};\boldsymbol{\Theta})|\mathcal{X};\boldsymbol{\Theta}^{[t-1]}\right\}. \quad (12)$$

In (10), $\gamma_i^{k_\mathrm{o}k_\mathrm{d}}$ can be updated by applying the Bayes' theorem as follows,

$$\gamma_i^{k_\mathrm{o}k_\mathrm{d}} = \frac{\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)[t-1]}\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}}{\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)[t-1]}\theta_{o_ik_\mathrm{o}}^{(\mathrm{o})}\theta_{d_ik_\mathrm{d}}^{(\mathrm{d})}} \quad (13)$$

where $\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)[t]}$ denotes the t-th updated value of $\pi_{k_\mathrm{o}k_\mathrm{d}}^{(t_i)}$. The optimal solution for the problem in (12) can be derived analytically, and the new value of $\boldsymbol{\Theta}^{[t]}$ is updated given by

$$\theta_{c_\mathrm{o}k'_\mathrm{o}}^{(\mathrm{o})[t]} = \frac{\sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\mathbb{I}(o_i=c_\mathrm{o},k_\mathrm{o}=k'_\mathrm{o})\gamma_i^{k_\mathrm{o}k_\mathrm{d}}}{\sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\mathbb{I}(k_\mathrm{o}=k'_\mathrm{o})\gamma_i^{k_\mathrm{o}k_\mathrm{d}}} \quad (14)$$

$$\theta_{c_\mathrm{d}k'_\mathrm{d}}^{(\mathrm{d})[t]} = \frac{\sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\mathbb{I}(d_i=c_\mathrm{d},k_\mathrm{d}=k'_\mathrm{d})\gamma_i^{k_\mathrm{o}k_\mathrm{d}}}{\sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\mathbb{I}(k_\mathrm{d}=k'_\mathrm{d})\gamma_i^{k_\mathrm{o}k_\mathrm{d}}} \quad (15)$$

where $\theta_{c_\mathrm{o}k'_\mathrm{o}}^{(\mathrm{o})[t]}$ and $\theta_{c_\mathrm{d}k'_\mathrm{d}}^{(\mathrm{d})[t]}$ denote the t-th updated value of $\theta_{c_\mathrm{o}k'_\mathrm{o}}^{(\mathrm{o})}$ and $\theta_{c_\mathrm{d}k'_\mathrm{d}}^{(\mathrm{d})}$, respectively. Since $\boldsymbol{\Pi}$ denotes the probability that a trip belongs to the k_o-th origin pattern and the k_d-th destination pattern, based on (11), $\boldsymbol{\Pi}^{[t]}$ is updated as

$$\pi_{k_\mathrm{o}k_\mathrm{d}}^{(c_\mathrm{t})[t]} = \frac{1}{n_{c_\mathrm{t}}}\sum_{i_{c_\mathrm{t}}=1}^{n_{c_\mathrm{t}}}\gamma_i^{k_\mathrm{o}k_\mathrm{d}} \quad (16)$$

where i_{c_t} denotes an index of the observed trip data that have the departure time t_i belonging to c_t, while n_{c_t} is the number of the observed trip data that fall into under this category.

Using the result of the EM algorithm, i.e., $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, $\boldsymbol{V}$ in (2) is obtained by as follows,

$$\boldsymbol{V} \approx \left(\boldsymbol{\Theta}^{(\mathrm{o})}\right)^\mathrm{T}\boldsymbol{\Pi}\boldsymbol{\Theta}^{(\mathrm{d})}. \quad (17)$$

III. Case Study

A. Data description

We use yellow taxi trip record data in New York city. Taxi and Limousine Commission opens the taxi trip data from 2009. The taxi trip data have individual trip information including pick-up and drop-off location, pick-up and drop-off date/time, trip distances, fares, payment types, and number of passengers etc. Pick-up and drop-off locations are indicated by 260 taxi zone IDs, of which 66 are taxi zones in Manhattan. The data have been frequently used in various works for human trip behavior analysis and demand estimation [25], [26].

In particular, we use the records of pick-up and drop-off location, date/time, and number of passengers of individual trips that occurred during December 2018. In December 2018, the number of trips by yellow taxis is 24,648,499, and the average daily number of trips is about 48,600. A single trip is represented by a multivariate tuple with three attributes of origin zone ID, destination zone ID, and departure time as in (1). The trip information fits with the tensor representation by aggregating departure time to 30-minute intervals since the spatial dimension (origin and destination) has aggregated values at the zone IDs. Then, the number of passengers of each combination can be counted correspondingly.

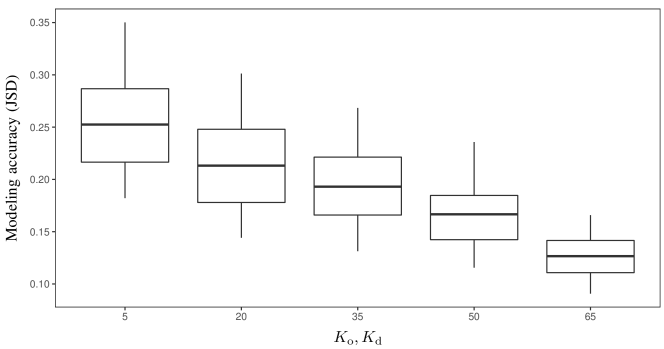


Fig. 2. Effect of K_o and K_d on the modeling accuracy with $K_\mathrm{o} = K_\mathrm{d}$.

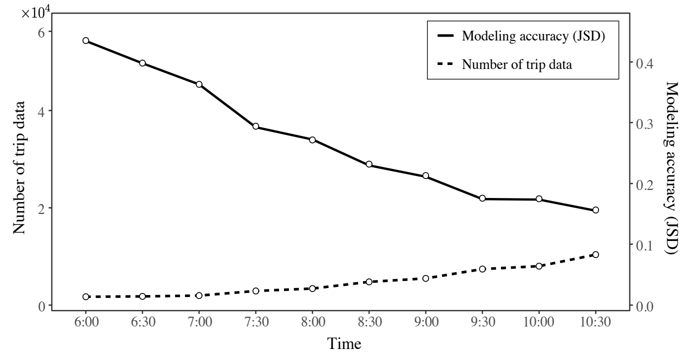


Fig. 3. Temporal changes in the modeling accuracy according to the number of trip data when $K_\mathrm{o} = K_\mathrm{d} = 5$.

B. Modeling performance

The New York taxi data is analyzed using the proposed method, revealing fundamental trip patterns and temporal interactions between patterns, and estimating the trip occurrence probability. The temporal range of trip pattern modeling is set to 6:30 – 10:30 AM. The total number of possible combinations is $265 \times 265 \times 8 = 540,800$ and the number of unique combinations observed is 15,175 with a minimum count of 1 and a maximum count of 175. The number of basis trip patterns determines the capacity of the model. More basis patterns can describe the trip patterns in more detail, which may help to fit better to the data, but in the meanwhile, it makes difficult to interpret the modeling result and requires higher computation loads. The convergence criteria is set to be the average of relative changes in basis patterns being less then 10^{-5}, and maximum iteration is set to 10,000.

The accuracy of the proposed model in (17) is evaluated by comparing it to the probability distribution given by the sample mean of the New York taxi data such that

$$\boldsymbol{V}_{c_\mathrm{o} c_\mathrm{d} c_\mathrm{t}} \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\big(o_i = c_\mathrm{o}, d_i = c_\mathrm{d}, t_i = c_\mathrm{t}\big). \qquad (18)$$

Specifically, the Jensen-Shannon divergence (JSD) between two probability distributions (17) and (18) is used as the model accuracy. The JSD is non-negative and symmetric, which becomes zero if only if two distributions are identical.

Figure 2 describes how the accuracy of the proposed model changes according to K_o and K_d. The modeling accuracy of the y axis in the figure is measured by JSD. Each box plot has modeling results for all time periods. As K_o and K_d increase, the difference between two distributions gets smaller, which means the modeling accuracy increases. These results indicate that setting K_o and K_d values large help the model to fit the data with the high capacity to describe a variety of phenomenon. Larger K_o and K_d, the better ability to express the detailed highs and lows of values.

Figure 3 shows the impact of the number of trip data used for modeling to the model accuracy in $K_\mathrm{o} = K_\mathrm{d} = 5$ case. The number of trip data in New York city varies in different time periods. Starting from 1,719 at 6:00 AM, it rises to 10,377 by 10:30 AM. We can observe that the JSD changes in the opposite direction to the number of trip data. The model accuracy increases as more data are used in modeling. As the number of trip data increases by 503.66 %, the modeling accuracy is improved by 66.64 %. Based on Figures 2 and 3, it can be concluded that the number of basis trip patterns and the amount of trip data sufficient to describe complex trip behaviors are key factors for improving the modeling accuracy.

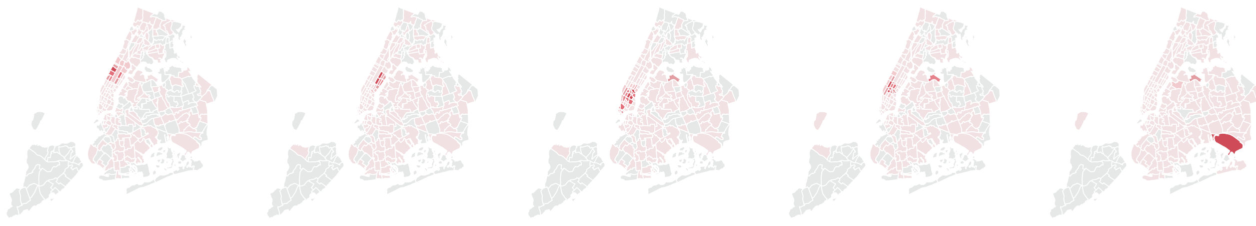
C. Basis trip patterns analysis

To analyze the modeling result in more detail, the origin and destination basis trip patterns are visualized. To make the result interpretable, K_o and K_d are set to 5. Figure 4 and 5 display the column factors of the basis trip patterns $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$, respectively. To better visualize the pattern configuration, the values in each column are rescaled to 0-1 so that $\theta'^{(\mathrm{o})}_{c_\mathrm{o} k_\mathrm{o}} = \theta^{(\mathrm{o})}_{c_\mathrm{o} k_\mathrm{o}} / \max(\boldsymbol{\theta}^{(\mathrm{o})}_{k_\mathrm{o}})$.

In Figures 4 and 5, it is noticeable that Manhattan areas show strong patterns in most of $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$. Manhattan consists of 69 taxi zones, which is larger number compared to other areas since Manhattan occupies the high portion of taxi demands. In both $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$, the 69 taxi zones in Manhattan are grouped into four patterns except $\boldsymbol{\theta}_5^{(\mathrm{o})}$ and $\boldsymbol{\theta}_5^{(\mathrm{d})}$, which contains trip patterns from and to John F. Kennedy International Airport and LaGuardia Airport. From $\boldsymbol{\theta}_1^{(\mathrm{o})}$ and $\boldsymbol{\theta}_1^{(\mathrm{d})}$ to $\boldsymbol{\theta}_4^{(\mathrm{o})}$ and $\boldsymbol{\theta}_4^{(\mathrm{d})}$, areas of uptown, upper-east side and midtown-east, downtown including financial district, and midtown are grouped to different origin and destination patterns.

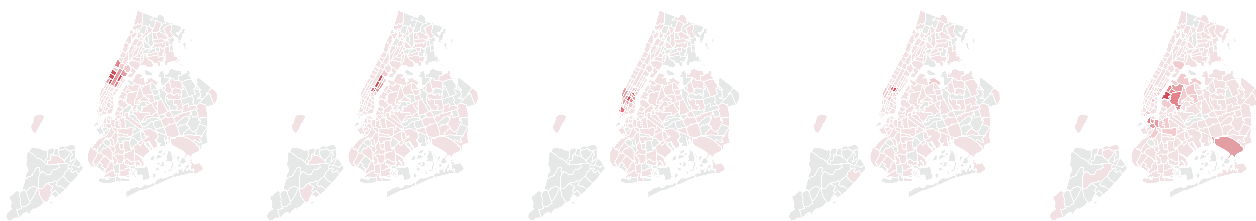
D. Temporal interactions analysis

Figure 6 shows the heat plots of $\boldsymbol{\Pi}$. It can be observed that $\boldsymbol{\Pi}$ does not change suddenly; rather, it changes slowly and steadily over several hours. Before morning peak hour in 6(a), trips from downtown areas described in $\boldsymbol{\theta}_3^{(\mathrm{o})}$ to outside of Manhattan in $\boldsymbol{\theta}_5^{(\mathrm{d})}$ are exclusively dominant demand. From morning peak hours, it can be seen that four types of trips remain popular: from downtown to outside of Manhattan, from airports to downtown, from uptown to midtown, and trips within the uptown area. For the trips from and to airports, the downtown area is the dominant origins and destinations, and demand is consistently high during the day.



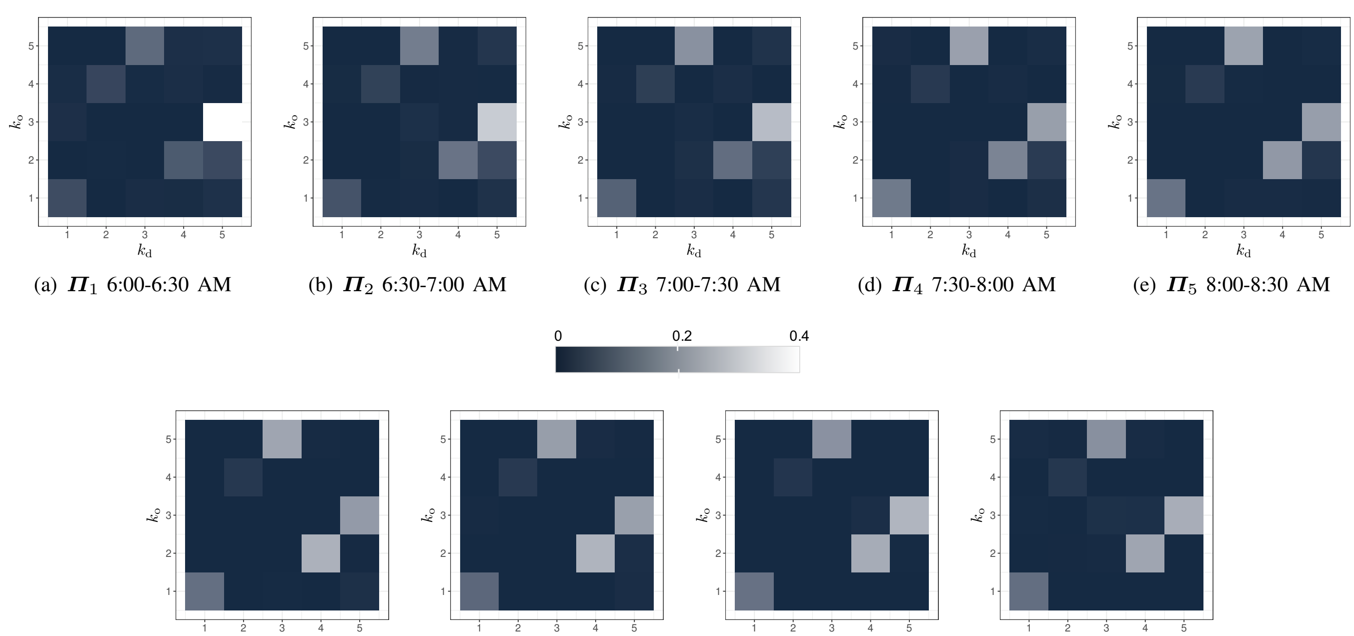
(a) $\boldsymbol{\theta}_1^{(\mathrm{o})}$ (b) $\boldsymbol{\theta}_2^{(\mathrm{o})}$ (c) $\boldsymbol{\theta}_3^{(\mathrm{o})}$ (d) $\boldsymbol{\theta}_4^{(\mathrm{o})}$ (e) $\boldsymbol{\theta}_5^{(\mathrm{o})}$

Fig. 4. Origin basis trip patterns $\boldsymbol{\Theta}^{(\mathrm{o})}$ modeled when $K_{\mathrm{o}} = 5$.



(a) $\boldsymbol{\theta}_1^{(\mathrm{d})}$ (b) $\boldsymbol{\theta}_2^{(\mathrm{d})}$ (c) $\boldsymbol{\theta}_3^{(\mathrm{d})}$ (d) $\boldsymbol{\theta}_4^{(\mathrm{d})}$ (e) $\boldsymbol{\theta}_5^{(\mathrm{d})}$

Fig. 5. Destination basis trip patterns $\boldsymbol{\Theta}^{(\mathrm{d})}$ modeled when $K_{\mathrm{d}} = 5$.



(a) $\boldsymbol{\Pi}_1$ 6:00-6:30 AM (b) $\boldsymbol{\Pi}_2$ 6:30-7:00 AM (c) $\boldsymbol{\Pi}_3$ 7:00-7:30 AM (d) $\boldsymbol{\Pi}_4$ 7:30-8:00 AM (e) $\boldsymbol{\Pi}_5$ 8:00-8:30 AM

(g) $\boldsymbol{\Pi}_6$ 8:30-9:00 AM (h) $\boldsymbol{\Pi}_7$ 9:00-9:30 AM (i) $\boldsymbol{\Pi}_8$ 9:30-10:00 AM (j) $\boldsymbol{\Pi}_9$ 10:00-10:30 AM

Fig. 6. Temporal interactions $\boldsymbol{\Pi}$ modeled when $K_{\mathrm{o}} = K_{\mathrm{d}} = 5$.

IV. Conclusion

This paper proposed a travel demand estimation model for identifying spatiotemporal mobility patterns using a tensor factorization. We utilized a probabilistic tensor factorization approach to reveal basis patterns and temporal interactions between patterns to model the probability distribution of trip demand. EM algorithm is applied for inferring mobility patterns efficiently. The model was applied to New York taxi data with 260 origins and destinations. The low-rank approximation results allowed us to analyze the basis patterns of the New York taxi trips. Results show the modeling performance with respect to the number of patterns and amount of trip data used. In addition, the result of the clustered mobility patterns reveal the underlying spatiotemporal structure of highly complex human mobility data. The proposed approach enriches the information in mobility data by providing intuitive low-dimensional trip patterns of multi-dimensional mobility.

Acknowledgment

The fundamental research described in this paper was supported in part by the Office of Naval Research under Grant N62909-22-1-2009 and in part by the National Science Foundation under Grant CNS-2148251.

References

[1] B. Cheng, G. Solmaz, F. Cirillo, E. Kovacs, K. Terasawa, and A. Kitazawa, "Fogflow: Easy programming of iot services over cloud and edges for smart cities," *IEEE Internet of Things journal*, vol. 5, no. 2, pp. 696–707, 2017.

[2] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 162–177, 2015.

[3] S. Çolak, A. Lima, and M. C. González, "Understanding congested travel in urban areas," *Nature communications*, vol. 7, no. 1, pp. 1–8, 2016.

[4] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *Journal of Transport Geography*, vol. 43, pp. 78–90, 2015.

[5] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang, "Understanding metropolitan patterns of daily encounters," *Proceedings of the National Academy of Sciences*, vol. 110, no. 34, pp. 13 774–13 779, 2013.

[6] C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmitt, "Detecting the dynamics of urban structure through spatial network analysis," *International Journal of Geographical Information Science*, vol. 28, no. 11, pp. 2178–2199, 2014.

[7] K. W. Axhausen and T. Gärling, "Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems," *Transport reviews*, vol. 12, no. 4, pp. 323–341, 1992.

[8] C. R. Bhat and F. S. Koppelman, *Activity-based modeling of travel demand*. Springer, 1999.

[9] J. Kim and H. S. Mahmassani, "Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories," *Transportation Research Procedia*, vol. 9, pp. 164–184, 2015.

[10] M. Saberi, H. S. Mahmassani, T. Hou, and A. Zockaie, "Estimating network fundamental diagram using three-dimensional vehicle trajectories: extending edie's definitions of traffic flow variables to networks," *Transportation Research Record*, vol. 2422, no. 1, pp. 12–20, 2014.

[11] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers of Computer Science*, vol. 6, no. 1, pp. 111–121, 2012.

[12] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi–passenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, 2013.

[13] M. Bierlaire and F. Crittin, "An efficient algorithm for real-time estimation and prediction of dynamic od tables," *Operations Research*, vol. 52, no. 1, pp. 116–127, 2004.

[14] J. Barceló Bugeda, L. Montero Mercadé, M. Bullejos, O. Serch, and C. Carmona Bautista, "A kalman filter approach for the estimation of time dependent od matrices exploiting bluetooth traffic data collection," in *TRB 91st Annual Meeting Compendium of Papers DVD*, 2012, pp. 1–16.

[15] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transportation Research Part B: Methodological*, vol. 91, pp. 511–524, 2016.

[16] D. Skillicorn, *Understanding complex datasets: data mining with matrix decompositions*. Chapman and Hall/CRC, 2007.

[17] J. Ren and Q. Xie, "Efficient od trip matrix prediction based on tensor decomposition," in *2017 18th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2017, pp. 180–185.

[18] J. Reades, F. Calabrese, and C. Ratti, "Eigenplaces: analysing cities using the space–time structure of the mobile phone network," *Environment and Planning B: Planning and Design*, vol. 36, no. 5, pp. 824–836, 2009.

[19] C. Peng, X. Jin, K.-C. Wong, M. Shi, and P. Liò, "Collective human mobility pattern from taxi trips in urban area," *PloS one*, vol. 7, no. 4, p. e34487, 2012.

[20] F. Zhang, D. Wilkie, Y. Zheng, and X. Xie, "Sensing the pulse of urban refueling behavior," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, pp. 13–22.

[21] J. Wang, F. Gao, P. Cui, C. Li, and Z. Xiong, "Discovering urban spatio-temporal structure from time-evolving traffic networks," in *Asia-pacific web conference*. Springer, 2014, pp. 93–104.

[22] Y. Yang, Y. Fan, and R. J. Wets, "Stochastic travel demand estimation: Improving network identifiability using multi-day observation sets," *Transportation Research Part B: Methodological*, vol. 107, pp. 192–211, 2018.

[23] M. Bhanu, S. Priya, S. K. Dandapat, J. Chandra, and J. Mendes-Moreira, "Forecasting traffic flow in big cities using modified tucker decomposition," in *International Conference on Advanced Data Mining and Applications*. Springer, 2018, pp. 119–128.

[24] Z. Li, H. Yan, C. Zhang, and F. Tsung, "Long-short term spatiotemporal tensor prediction for passenger flow profile," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5010–5017, 2020.

[25] F. Rodrigues, I. Markou, and F. C. Pereira, "Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach," *Information Fusion*, vol. 49, pp. 120–129, 2019.

[26] B. Du, X. Hu, L. Sun, J. Liu, Y. Qiao, and W. Lv, "Traffic demand prediction based on dynamic transition convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1237–1247, 2020.