Transportation Research Part B 91 (2016) 511–524

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Understanding urban mobility patterns with a probabilistic tensor factorization framework

Lijun Sun [a,b,*], Kay W. Axhausen [b,c]

[a] The Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[b] Future Cities Laboratory, Singapore-ETH Centre, Singapore 138602, Singapore
[c] Institute for Transport Planning and Systems (IVT), ETH Zürich, Zürich CH-8093, Switzerland

ARTICLE INFO

Article history:
Available online 29 June 2016

Keywords:
Human mobility
Urban computing
Smart card data
Tensor decomposition
Data-driven

ABSTRACT

The rapid developments of ubiquitous mobile computing provide planners and researchers with new opportunities to understand and build smart cities by mining the massive spatial-temporal mobility data. However, given the increasing complexity and volume of the emerging mobility datasets, it also becomes challenging to build novel analytical framework that is capable of understanding the structural properties and critical features. In this paper, we introduce an analytical framework to deal with high-dimensional human mobility data. To this end, we formulate mobility data in a probabilistic setting and consider each record a multivariate observation sampled from an underlying distribution. In order to characterize this distribution, we use a multi-way probabilistic factorization model based on the concept of tensor decomposition and probabilistic latent semantic analysis (PLSA). The model provides us with a flexible approach to understand multi-way mobility involving higher-order interactions—which are difficult to characterize with conventional approaches—using simple latent structures. The model can be efficiently estimated using the expectation maximization (EM) algorithm. As a numerical example, this model is applied on a four-way dataset recording 14 million public transport journeys extracted from smart card transactions in Singapore. This framework can shed light on the modeling of urban structure by understanding mobility flows in both spatial and temporal dimensions.

1. Introduction

Urban transportation systems are the backbone of cities, allowing people to have diverse types of interactions through social activities such as work, school, shopping and leisure. To provide people-centric transportation and to enhance the efficiency of spatial interaction become the primary goals of transportation planning. As the first step towards an efficient transportation system, understanding and modeling urban transportation demand is crucial to fleet management, infrastructure design, epidemic control, urban planning and policy making. In the last decades, the developments of transportation demand modeling have created a variety of landmark works, including the classic four-step model and the recent activity-based model (Axhausen and Gärling, 1992; Bhat and Koppelman, 2003). These sophisticated approaches normally involve the estimation of a parametric spatial interaction model using available data and then apply the estimated model to predict

* Corresponding author.
E-mail addresses: sunlijun@mit.edu (L. Sun), axhausen@ivt.baug.ethz.ch (K.W. Axhausen).

http://dx.doi.org/10.1016/j.trb.2016.06.011
0191-2615/

demand at a population scale. To bring these models into practice, collecting credible data recording individual's transportation activities from travel diaries and household surveys becomes the first critical step. However, the developments of novel transportation demand models still suffer from the small sampling shares, high cost, infrequent periodicity and limited accuracy of the survey data.

With the fast development of information and communication technologies (ICT) and ubiquitous mobile computing, large quantities of digital traces that register individual human activities at both spatial and temporal scales have become available. These datasets not only help planners and researchers understand cities as complex systems, but also allow practitioners to understand cities through data-centric technologies (Batty et al., 2012). In terms of urban transportation, such data set allows us to further reveal the spatial-temporal structure of cities inherited from human mobility (González et al., 2008; Roth et al., 2011; Sun et al., 2015).

The emergence of such mobility data indeed brings us new opportunities to integrate more information into decision making. However, the complexity of the data also increases with the dimensions of its contents, which imply complex dependence and higher-order interactions among space, time and individual social-demographic attributes. Given the increasing volumes and complexity of datasets, retrieving important information and critical features from them becomes challenging. It is crucial to developing advanced data-driven approaches and models to enrich our understanding about urban transportation systems and human mobility patterns from the large amount of data.

This paper is dedicated to providing a data-driven approach to characterize the collective mobility patterns from high-dimensional structured datasets. Considering the multivariate nature of urban mobility data, we focus on depicting the complex dependence and interactions with a probabilistic framework, which provides us with a natural way to represent the structural properties and uncertainty in the data. To this end, we concentrate on a multi-way categorical setting, which allows us to summarize the high-dimensional data into contingency tables and tensors (Agresti, 2002; Kolda and Bader, 2009). In a probabilistic setting, collective mobility data can be modeled in the sense that each mobility record is a sample generated from a universal multivariate distribution. The main contribution of this paper is to introduce an analytical framework to deal with multi-dimensional transportation/mobility data. In so doing, we adopt statistical techniques to better interpret human behavior and urban dynamics. To better understand the structure of urban mobility, we apply a factorization model to decompose high-dimensional mobility data into important patterns, from which we can extract key information by reasoning about the semantics of regions and activities. Based on this information, we can further reveal the dynamics of cities. Using a real large-scale dataset from public transport systems in Singapore, we find both travel behavior and spatial configuration of the city are well structured.

The remainder of this paper is organized as follows. In Section 2, we review relevant papers on revealing urban structure using various spatial-temporal datasets and in particular the developments of applying tensor decomposition on high-dimensional urban data analytics. Section 3 introduces the framework of probabilistic factorization on multivariate transportation data based on the concept of Tucker decomposition and multi-way probabilistic latent semantic analysis (PLSA). The expectation-maximization (EM) algorithm is applied to efficiently infer the model. In Section 4, we present a case study of applying the model on a four-way dataset (time, passenger type, origin zone and destination zone) extracted from public transport smart card transactions in Singapore. We also discuss the implications and insights of the experiment. Finally, Section 5 summarizes our key findings and discusses the potential application of such a data-driven approach in helping us better understand urban dynamics.

2. Literature review

With the recent advances in ICT, large quantities of individualized data—such as call detail record (CDR), geo-referenced social media data, Global Positioning System (GPS) trajectories and transit smart card transactions—are generated through various social activities. Thanks to the rich spatial-temporal information in such datasets, analyzing and modeling the spatial-temporal structure of cities from a variety of proxies on human activities have become an emerging topic in geography, urban and transportation research (e.g., Calabrese et al., 2011; Jiang et al., 2012; Sun et al., 2012; Yuan et al., 2012; Coffey and Pozdnoukhov, 2013; Allahviranloo and Recker, 2013; Wang et al., 2014; Sun et al., 2015; Han and Sohn, 2016). In the meanwhile, the dimensions of information also become richer and richer. For example, smart card transactions contain not only spatial-temporal information of trips, but also some basic social-demographic profiles of cardholders. In most cases, such data is well-structured with pre-defined fields and each record can be considered a draw of multivariate variables.

A tensor representation allows us to summarize multivariate categorical data into a multi-dimensional array (Kolda and Bader, 2009). The goal of tensor decomposition is to efficiently reproduce the complex dependence and higher-order interactions between different modes in multivariate data by using simple structures with relatively few parameters. As a natural choice in dealing with multi-dimensional data, tensor decomposition becomes increasingly important in interpreting the underlying structure of complex datasets. It has been successfully applied in a variety of fields, such as signal processing, computer vision, online recommendation, web data mining, psychometrics and survey analysis (Kolda and Bader, 2009; Sun et al., 2006).

Given its strength in retrieving information from large datasets, tensor decomposition also attracts more and more attentions in the field of transportation data analysis. For example, Tan et al. (2013) integrated Tucker decomposition and EM algorithm to efficiently impute the missing values in a four-way tensor of traffic data (by link, day, hour and a five-minute domain). The numerical experiments suggest that the tensor-based method shows superior performance compared with

other state-of-the-art approaches. Han and Moutarde (2014) proposed to use non-negative tensor decomposition to identify spatial-temporal patterns of traffic states. The authors applied non-negative CANDECOMP/PARAFAC (CP) decomposition on a three-way (link × sampling step × time period) tensor constructed from simulated traffic state data.

With the emergence of individual-based spatial-temporal datasets, tensor decomposition also plays an increasingly important role in urban computing and travel behavior research. For instance, mobility data obtained from taxis trajectories, smart phones and transit smart cards offer researchers new opportunities to study collective mobility patterns (Jiang et al., 2012; Ma et al., 2013; Sun et al., 2013). In a two-dimensional setting, Reades et al. (2009) applied principle component analysis (PCA) on a zone by time matrix to identify eigenplaces from mobile phone data. Peng et al. (2012) transformed taxi trips into a zone × time matrix and applied a non-negative matrix factorization (NTF) model (Lee and Seung, 1999) to identify basic spatial and temporal patterns of human mobility. The factorization suggests that taxi trips in Shanghai are essentially characterized by three purposes based on their temporal profiles: home-to-work, work-to-work, and others such as leisure activities. Integrating taxi trajectory data with other urban sensing data, Zhang et al. (2013) modeled urban refueling events as a three-way gas-station × hour × day tensor and adapted Tucker decomposition to take additional contextual features of gas stations into account. Using a large taxi trajectory dataset in Beijing, Wang et al. (2014) modeled collective mobility as a three-way (origin × destination × time, with a large size of 651 × 651 × 24) tensor, with each cell (i, j, k) corresponding to traffic volume from zone i to zone j at time domain k. To better fit the urban analytics framework and interpret the results, the authors applied a regularized non-negative Tucker decomposition approach to reveal the principle patterns on each mode. Fan et al. (2014) utilized non-negative CP decomposition on three-way (region × day × time) tensor constructed from a large mobile phone GPS log dataset. The authors performed regularization on the region mode and the tensor is filled up by measuring the proportion of GPS points across all regions given time and day. The tool was used to quantify the fluctuation of urban mobility before and after the Great East Japan Earthquake in 2011.

Despite applying tensor decomposition directly, the analysis of multi-way categorical data also demonstrates an increasing interest in a probabilistic context in statistical modeling and machine learning. For instance, Hofmann (1999) introduced probabilistic latent semantic analysis (PLSA), which in principle is equivalent to a non-negative matrix factorization model when the Kullback-Leibler divergence is minimized (Ding et al., 2008). The model can be also extended to a multi-way scenario, which is often referred to as multi-way probabilistic latent semantic analysis (Peng and Li, 2011; Zafeiriou and Petrou, 2011). In order to better model large sparse tabular data, Dunson and Xing (2012) and Bhattacharya and Dunson (2012) proposed non-parametric Bayesian factorizations using Markov chain Monte Carlo (MCMC) algorithm. Essentially, the two models are in line with the concepts of non-negative CP and Tucker decompositions, respectively. Owing to the advantages of the Bayesian paradigm, the models are capable of handling sparse tensors/contingency tables by specifying proper priors and adopting efficient samplers. Although these approaches are proposed with different purposes and estimation algorithms, they all model particular type of data in a probabilistic framework.

In this paper, we consider individual-based mobility record a multivariate observation sampled from a universal distribution and apply a tucker-like PLSA framework to identify principle patterns on each mode from a large-scale transit smart card dataset. Although performing tensor decomposition in a probabilistic context is not new by itself, this article focuses on its application on a high-dimensional mobility dataset, revealing valuable insights in spatial-temporal-mobility interactions.

3. Modeling framework

In this section we present the framework on modeling multi-way urban mobility data. We assume that trips are multivariate variables sampled from a universal distribution. In order to characterize this universal distribution, we apply a multi-way probabilistic factorization model that is able to capture the principle patterns on different modes and their joint interactions using few latent factors in Section 3.3. To better understand its interconnection with tensor decomposition techniques, we first introduce the basic concepts of CP and Tucker decompositions in Section 3.2.

3.1. Notation

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^\top$ denote a trip observation, where m represents the dimension of information (i.e., number of attributes in a single trips record), and $d = 1, \ldots, m$ denotes the indices of attributes. In characterizing transit trips, an element x_{i*} may represent the type of passenger (e.g., adult, senior and children), time of trip, origin zone, destination zone, travel mode (e.g., bus/metro), etc. We use $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ to represent all the trip observations we have, with $i = 1, \ldots, n$ denoting the index of each record.

To facilitate the factorization model, we first map all variables in trip records into the categories. In doing so, we define $x_{id} \in \{1, \ldots, w_d\}$ as discrete values starting from one for attribute d. Taking passenger type as an example, we may use $x_{id} = 1, \ldots, 3$ to represent adults, senior citizens and children correspondingly. For those continuous variables, we map the values on a discrete domain. For example, we may redefine trip start time to hours by letting $x_{id} \in \{1, \ldots, 24\}$, with values 1, 2,..., 24 corresponding to the hour of timestamp.

With the notations above, all the trip records $\mathbf{x}$ can be summarized into a m-way tensor/contingency table with dimension $D = w_1 \times \cdots \times w_m$, and each cell $(c_1, \ldots, c_m)$ representing the count of observations $\sum_{i=1}^{n} \mathbb{I}(x_{i1} = c_1, \ldots, x_{im} = c_m)$. In this paper, we use $\mathbb{I}(e)$ as an indicator function with $\mathbb{I}(e) = 1$ if e is true and 0 otherwise. In other words, each cell in

the tensor represents a particular combination of trip information and its value indicates number of trips belonging to this combination.

Similarly, in a probabilistic setting we can also model collective mobility using a probabilistic tensor, with each cell representing the probability of a trip belonging to a particular cell $p(x_{i1} = c_1, \ldots, x_{im} = c_m)$. To facilitate the modeling and interpretation, in this paper we adopt the probabilistic representation.

3.2. Tensor decomposition

In this section we briefly introduce two types of tensor decomposition techniques: CP decomposition and Tucker decomposition. Both CP and Tucker decompositions can be considered a high-order generalization of matrix singular value decomposition (SVD) and principle component analysis (PCA). We refer interested readers to Kolda and Bader (2009) for a comprehensive review about tensor decomposition and its application.

3.2.1. CANDECOMP/PARAFAC (CP) decomposition

A N-way tensor is rank-one if it can be written as the outer product of N vectors. The CP decomposition factorizes a tensor $\mathcal{X}$ into the sum of a series of rank-one tensors

$$\mathcal{X} = \sum_{k=1}^{h} \lambda_k \mathbf{u}_k^{(1)} \otimes \cdots \otimes \mathbf{u}_k^{(m)}, \tag{1}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_h > 0$ represent weight parameters, $\mathbf{u}_k^{(d)} = (u_{1k}^{(d)}, u_{2k}^{(d)}, \ldots, u_{w_d k}^{(d)})^\top$ is the kth vector of dimension d and '$\otimes$' denotes vector outer product. Therefore, each cell in the decomposed tensor is

$$x_{c_1 \ldots c_m} = \sum_{k=1}^{h} \lambda_k \prod_{d=1}^{m} u_{c_d k}^{(d)}. \tag{2}$$

3.2.2. Tucker decomposition

Tucker decomposition is also referred to as higher-order singular value decomposition (HOSVD) in the literature. In general, Tucker decomposition factorizes a tensor $\mathcal{X}$ into a core tensor multiplied by a series of factor matrices along all modes

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{u}^{(1)} \times_2 \cdots \times_m \mathbf{u}^{(m)} = \sum_{k_1=1}^{h_1} \cdots \sum_{k_m=1}^{h_m} g_{k_1 \ldots k_m} \mathbf{u}_{k_1}^{(1)} \otimes \cdots \otimes \mathbf{u}_{k_m}^{(m)}, \tag{3}$$

where $\times_d$ represents the tensor-matrix multiplication on mode d, $\mathcal{G} \in \mathbb{R}^{h_1 \times \cdots \times h_m}$ is the core tensor ($g_{k_1 \ldots k_m}$ represents value in each cell), $\mathbf{u}^{(d)} \in \mathbb{R}^{w_d \times h_d}$ is a factor matrix of mode d and h_d is the number of patterns (vectors) in the factor matrix $\mathbf{u}^{(d)}$. The core tensor $\mathcal{G}$ characterizes the degree of interaction between different modes. Elementwise, we have

$$x_{c_1 \ldots c_m} = \sum_{k_1=1}^{h_1} \cdots \sum_{k_m=1}^{h_m} g_{k_1 \ldots k_m} u_{c_1 k_1}^{(1)} \times \cdots \times u_{c_m k_m}^{(m)}. \tag{4}$$

Since Tucker decomposition utilizes all combinations of vectors in factor matrices ($\prod_{d=1}^{m} h_d$ joint patterns from $\sum_{d=1}^{m} h_d$ patterns in factor matrices), it is able to capture complex dependence with few components compared with CP decomposition.

3.3. Probabilistic factorization model

Given the large number of parameters, working on the probabilistic tensor directly is infeasible. A common approach to simplify the model and reduce number of parameters is to characterize each attribute as a multinomial distribution (categorical distribution), so that the joint distribution of $\mathbf{x}_i$ can be modeled as using a product-multinomial. However, by doing so the interactions among different attributes are essentially ignored and thus it becomes difficult to capture crucial dependence in the data.

An efficient alternative to capture those interactions is to apply latent class models, which assume that each observation is generated from a mixture of underlying classes and each class is associated with a unique probability distribution. In standard latent class models on categorical variables, a general assumption is that the observed x_{i*} are mutually independent within each class. This assumption essentially simplifies the joint distribution as a mixture of product-multinomial and the probability of observing a trip $\mathbf{x}_i$ (i.e., the probability mass function) can be written as

$$p(\mathbf{x}_i|\boldsymbol{\theta}) = \sum_{k=1}^{h} \pi_k p_k(\mathbf{x}_i|\boldsymbol{\theta}_k) = \sum_{k=1}^{h} \pi_k \prod_{d=1}^{m} p_k(x_{id}|\boldsymbol{\theta}_k), \tag{5}$$

where h is the number of classes and π_k denotes the probability that an observation belongs to class k. Each class k has a unique set of parameters denoted by $\boldsymbol{\theta}_k$. Therefore, an element in the probability tensor is given by

$$p(x_{i1} = c_1, \ldots, x_{im} = c_m|\boldsymbol{\theta}) = \sum_{k=1}^{h} \pi_k \theta^{(1)}_{c_1 k} \times \cdots \times \theta^{(m)}_{c_m k}, \tag{6}$$

where $\boldsymbol{\theta}^{(d)}_k = (\theta^{(d)}_{1k}, \ldots, \theta^{(d)}_{w_d k})^\top$ is a $w_d \times 1$ probability vector with $\theta^{(d)}_{c_d k} = p_k(x_{id} = c_d)$ and $\sum_{c_d=1}^{w_d} \theta^{(d)}_{c_d k} = 1$.

In summary, such latent class models impose conditional independence on each class. By summing up all classes the model is able to capture complex interactions and dependence. In principle, this model is equivalent to a non-negative CP (NNCP) decomposition model as in Eq. (1), in which all elements are non-negative values (Dunson and Xing, 2012; Peng and Li, 2011). However, the independence assumption within each class still seems to be restrictive in practice. Even simple dependence structure in the data requires us to use a large number of classes and parameters to capture, in particular when number of attributes of interest is large. In this case, an natural solution is to reformulate the latent class models in a similar way as Tucker decomposition

$$p\left(\mathbf{x}_i|\boldsymbol{\theta}\right) = \sum_{k_1=1}^{h_1} \cdots \sum_{k_m=1}^{h_m} \pi_{k_1 \ldots k_m} p_{k_1 \ldots k_m}(\mathbf{x}_i|\boldsymbol{\theta}), \tag{7}$$

where h_d is the number of principle patterns in mode d (thus the total number of pattern combinations is $h_1 \times \cdots \times h_m$), and $\pi_{k_1 \ldots k_m}$ denotes the probability that a trip is characterized by combination $(k_1, \ldots, k_m)$. A common approach in modeling the joint probability is to assume local independence

$$p_{k_1 \ldots k_m}\left(\mathbf{x}_i|\boldsymbol{\theta}\right) = \prod_{d=1}^{m} p^{(d)}_{k_d}\left(x_{id}|\boldsymbol{\theta}^{(d)}_{k_d}\right) = \prod_{d=1}^{m} \theta^{(d)}_{x_{id} k_d}, \tag{8}$$

where $\boldsymbol{\theta}^{(d)}_{k_d} = (\theta^{(d)}_{1k_d}, \ldots, \theta^{(d)}_{w_d k_d})^\top$ is a $w_d \times 1$ probability vector with $\sum_{c_d=1}^{w_d} \theta^{(d)}_{c_d k_d} = 1$. Similar to $\mathbf{u}^{(d)}$ in Eq. (3), $\boldsymbol{\theta}^{(d)}$ can be considered a probabilistic factor matrix on mode d. Thus, each element in the probability tensor can be calculated as the sum across all pattern combinations

$$p\left(x_{i1} = c_1, \ldots, x_{im} = c_m|\boldsymbol{\theta}\right) = \sum_{k_1=1}^{h_1} \cdots \sum_{k_m=1}^{h_m} \pi_{k_1 \ldots k_m} \theta^{(1)}_{c_1 k_1} \times \cdots \times \theta^{(m)}_{c_m k_m}. \tag{9}$$

This model is essentially equivalent to a non-negative Tucker (NNT) decomposition as in Eq. (3) (Peng and Li, 2011), in which all vectors $\mathbf{u}^{(d)}_k \in \mathbb{R}^{w_d}$ in factor matrices are replaced with a probability vector $\boldsymbol{\theta}^{(d)}_k$ and the core tensor $g_{k_1 \ldots k_m}$ is replaced with $\pi_{k_1 \ldots k_m}$. Eq. (7) can be also regarded as a multi-way extension to the standard two-dimensional PLSA in topic model (Hofmann, 1999). For parameters interpretation, $\boldsymbol{\theta}^{(d)}$ characterize the principle patterns on each mode d and the core tensor $\boldsymbol{\pi}$ captures the interactions across different modes. One advantage of this model is that all combinations of model vectors are used to form the tensor and it may require less parameters than the latent class model (NNCP decomposition in other words). Therefore, instead of capturing local dependence using large number of classes, it is able to identify universal patterns along each mode and use a core tensor to depict their interactions. The Tucker alike representation in Eq. (7) is our focus in this paper.

3.4. Model inference

The most common approach to estimate the PLSA model is to introduce latent variables on the joint membership and then apply the EM algorithm (Dempster et al., 1977), which is a powerful tool for mixture model and missing data problems, to infer its parameters. For this probabilistic factorization problem, we first introduce latent variables $\mathbf{z}$ on membership across all combinations, with $z_i \in \{(k_1, \ldots, k_m)|k_1 = 1 \ldots h_1, \cdots, k_m = 1 \ldots h_m\}$ representing a combination that an observations $\mathbf{x}_i$ comes from. The joint distribution of $\mathbf{x}$ and $\mathbf{z}$ in the complete model can be formulated as

$$p\left(\mathbf{x}, \mathbf{z}|\boldsymbol{\theta}\right) = \prod_{i=1}^{n} \prod_{k_1=1}^{h_1} \cdots \prod_{k_m=1}^{h_m} \left[\pi_{k_1 \ldots k_m} \prod_{d=1}^{m} \theta^{(d)}_{x_{id} k_d}\right]^{\mathbb{I}(z_i = (k_1, \ldots, k_m))}, \tag{10}$$

and the complete data log likelihood can be written as

$$\log \mathcal{L}\left(\boldsymbol{\theta}|\mathbf{x}, \mathbf{z}\right) = \sum_{i=1}^{n} \sum_{k_1=1}^{h_1} \cdots \sum_{k_m=1}^{h_m} \left[\mathbb{I}(z_i = (k_1, \ldots, k_m)) \times \left[\log \pi_{k_1 \ldots k_m} + \sum_{d=1}^{m} \log \theta^{(d)}_{x_{id} k_d}\right]\right]. \tag{11}$$

By introducing latent variable $\mathbf{z}$, we are able to push the log function inside the sum and thus all parameters can be efficiently estimated using the EM algorithm. Applying the EM algorithm begins with random initial values of $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. Then, the EM algorithm applies the following two steps iteratively until certain convergence criterion is met: (1) compute

the expectation of complete data log likelihood, which is often denoted as an auxiliary function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[t-1]})$, and then (2) updates new parameter by maximizing

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[t-1]}) &= \mathbb{E}_{z|\mathbf{x},\boldsymbol{\theta}^{[t-1]}}\left[\log \mathcal{L}\left(\boldsymbol{\theta}|\mathbf{x},\mathbf{z}\right)\right] \\ &= \sum_{i=1}^{n}\sum_{k_1=1}^{h_1}\cdots\sum_{k_m=1}^{h_m}\gamma_i^{k_1\ldots k_m}\left[\log \pi_{k_1\ldots k_m} + \sum_{d=1}^{m}\log \theta^{(d)}_{x_{id}k_d}\right], \end{aligned} \tag{12}$$

where $\gamma_i^{k_1\ldots k_m} = \mathbb{E}_{z|\mathbf{x},\boldsymbol{\theta}^{[t-1]}}[\mathbb{I}(z_i = (k_1,\ldots,k_m))] \triangleq p(z_i = (k_1,\ldots,k_m)|\mathbf{x}_i,\boldsymbol{\theta}^{[t-1]})$ (also called the responsibility that class $(k_1,\ldots,k_m)$ takes for trip record $\mathbf{x}_i$). The EM procedure can be summarized as follows:

E-step Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[t-1]};\boldsymbol{\pi}^{[t-1]})$ in Eq. (12). The responsibilities γ_i^k can be calculated by applying the Bayes' theorem

$$\gamma_i^{k_1\ldots k_m} = \frac{\pi^{[t-1]}_{k_1\ldots k_m}\, p_{k_1\ldots k_m}(\mathbf{x}_i|\boldsymbol{\theta}^{[t-1]})}{\sum_{k_1=1}^{h_1}\cdots\sum_{k_m=1}^{h_m}\pi^{[t-1]}_{k_1\ldots k_m}\, p_{k_1\ldots k_m}(\mathbf{x}_i|\boldsymbol{\theta}^{[t-1]})}. \tag{13}$$

M-step Update parameters $\boldsymbol{\theta}_k$ for each class k with

$$\boldsymbol{\theta}^{[t]} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[t-1]};\boldsymbol{\pi}^{[t-1]}). \tag{14}$$

As can be seen, Eq. (12) only involves sum of logs for each multinomial distribution and thus the maximization problem in Eq. (14) can be decomposed to into $k \times d$ subproblems. The optimal solution for each subproblem can be derived analytically. For each element in $\boldsymbol{\theta}$ ($\forall c_d = 1,\ldots,w_d$; $\forall d = 1,\ldots,m$; $\forall k = 1,\ldots,h_d$), the new value is updated given by

$$\theta^{(d)[t]}_{c_d k} = \frac{\sum_{i=1}^{n}\sum_{k_1=1}^{h_1}\cdots\sum_{k_m=1}^{h_m}[\mathbb{I}(x_{id} = c_d, k_d = k)\times\gamma_i^{k_1\ldots k_m}]}{\sum_{i=1}^{n}\sum_{k_1=1}^{h_1}\cdots\sum_{k_m=1}^{h_m}[\mathbb{I}(k_d = k)\times\gamma_i^{k_1\ldots k_m}]}. \tag{15}$$

For each class $\forall k_1 = 1,\ldots,h_1;\ \cdots;\ \forall k_m = 1,\ldots,h_m$, update new weight parameter $\pi_{k_1\ldots k_m}$ with

$$\pi^{[t]}_{k_1\ldots k_m} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i^{k_1\ldots k_m}. \tag{16}$$

A well-known drawback of the EM algorithm is that it may only find a local optimal solution rather than a global one, depending on the initial values of parameters in the first iteration. In order to avoid local optimal and be more certain in getting a global solution, the estimation process should be repeated for multiple times with different initial parameters.

4. Case study

In this section we apply the framework and estimation procedure on a large-scale public transport smart card transactions dataset collected in Singapore, which include both bus and metro modes. Section 4.1 briefly introduces the smart card data used in this study and the way we transform it into multivariate categorical data $\mathbf{x}$. The results of applying the factorization model are reported in Section 4.2.

4.1. Smart card data

We use smart card transactions as a proxy to collective mobility. The public transport systems in Singapore, including both bus and metro, are equipped with smart card based automated fare collection (AFC) systems. Since public transport in Singapore adopts a distance-based fare scheme, passengers not only tap-in but also have to tap-out for each journey. As a result, a full transaction contains spatial information (stop/station ID) and temporal information (tapping timestamp) for both boarding and alighting activities (Sun et al., 2014). The smart card data used in this study was recorded from April 11 to 14 (Monday–Thursday) in 2011.

We represent a single transit trip by a multivariate tuple (card ID, passenger type, time, boarding stop/station, alighting stop/station). Considering that a single transit journey may involve multiple stages with transfers, we first combine those transactions with intervals (from tapping-off to the next tapping-in) being less than 30 min. By doing so we are able to infer the real origin-destination information in a journey level instead of using transactions from all stages. In order to fit trip information with the tensor representation, we apply the transformation listed in Table 1 to map different variables into categorical values. In the spatial dimension, boarding/alighting stops/stations are aggregated at the planning zone level.

By applying this procedure, we transform the raw transit trip data into $\mathbf{x}$, with each observation containing four variables. The total number of journey observations n is about 14 million. We aggregate those trips belonging to the same

Table 1
Transformation of smart card transactions.

Attribute	No. of categories	Description
Time of day	20	(1) 4:30-5:30, (2) 5:30-6:30,..., (20) 23:30-0:30
Passenger type	3	(1) Adults, (2) Senior citizens, (3) Children/students
Origin zone	51	One to one mapping to zone ID
Desination zone	51	One to one mapping to zone ID

Table 2
Core tensor $\boldsymbol{\pi}$ summarized in origin zone mode and destination zone mode.

	T1			T2			T3			T4			
Sum	0.3647			0.2513			0.2175			0.1665			1.000
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3	
Sum	0.3157	0.0339	0.0151	0.2190	0.0294	0.0029	0.1060	0.1004	0.0111	0.0631	0.0169	0.0866	1.000
O1	0.1706	0.0148	0.0043	0.0034	0.0000	0.0000	0.0474	0.0259	0.0009	0.0006	0.0013	0.0006	0.2699
O2	0.0306	0.0058	0.0030	0.0668	0.0106	0.0007	0.0157	0.0269	0.0026	0.0103	0.0040	0.0249	0.2020
O3	0.0350	0.0056	0.0026	0.0485	0.0065	0.0008	0.0151	0.0198	0.0027	0.0124	0.0034	0.0194	0.1718
O4	0.0455	0.0039	0.0016	0.0357	0.0055	0.0002	0.0105	0.0130	0.0015	0.0147	0.0051	0.0141	0.1514
O5	0.0215	0.0021	0.0018	0.0356	0.0032	0.0009	0.0103	0.0074	0.0012	0.0136	0.0016	0.0122	0.1115
O6	0.0125	0.0016	0.0017	0.0290	0.0036	0.0003	0.0070	0.0074	0.0021	0.0115	0.0013	0.0154	0.0934
D1	0.0409	0.0016	0.0005	0.1289	0.0145	0.0001	0.0562	0.0261	0.0011	0.0035	0.0018	0.0006	0.2758
D2	0.0817	0.0122	0.0043	0.0144	0.0039	0.0005	0.0119	0.0259	0.0024	0.0087	0.0032	0.0252	0.1942
D3	0.0644	0.0081	0.0035	0.0227	0.0038	0.0012	0.0127	0.0201	0.0027	0.0119	0.0038	0.0193	0.1740
D4	0.0494	0.0056	0.0019	0.0344	0.0041	0.0000	0.0085	0.0127	0.0015	0.0211	0.0062	0.0147	0.1601
D5	0.0444	0.0031	0.0026	0.0130	0.0018	0.0003	0.0093	0.0077	0.0013	0.0108	0.0013	0.0119	0.1077
D6	0.0350	0.0033	0.0024	0.0055	0.0013	0.0008	0.0073	0.0080	0.0021	0.0071	0.0005	0.0148	0.0882

combination [time of day, passenger type, origin zone, destination zone] and count the number of occurrences of each combination correspondingly. In doing so, we convert the raw data into a combination-frequency table and use this table in the estimation procedure. In total, we observe 105,688 unique combinations (the total number of possible combinations is $20 \times 3 \times 51 \times 51 = 156{,}060$), with a minimum count of 1 and a maximum count of 11,845.

4.2. Factorization of collective transit journeys

We apply the EM procedure introduced in previous section on the trip record **x**. The size of the core tensor determines the capacity and generalization of the model. The values in the core tensor shows the degrees of interactions between different modes. A larger core tensor may better fit the data, but in the meanwhile it involves more parameters that may increase the risk of overfilling. On the other hand, a small core tensor may be not capable of fully capturing complex interdependence in the data. In a probabilistic framework we can compute the likelihood of observing data **x**, and thus one may conduct model selection using Akaike information criterion (AIC) or Bayesian information criterion (BIC) in determining the size of the core tensor. In this study, we use a small 4 (time of day, T) $\times$ 3 (passenger, P) $\times$ 6 (origin zone, O) $\times$ 6 (destination zone, D) core tensor $\boldsymbol{\pi}$ for efficient estimation. In practice, a smaller core tensor could facilitate the interpretation of results. We chose this size of core tensor based on subjective analysis and similar analysis such as in Wang et al. (2014). To test the representativeness of our results, we also apply the procedure on different sizes of core tensors ([$4 \times 3 \times 5 \times 5$], [$4 \times 3 \times 8 \times 8$], [$5 \times 3 \times 5 \times 5$], [$5 \times 3 \times 6 \times 6$] and [$5 \times 3 \times 8 \times 8$]). We find that the results are basically consistent across different sizes of core tensors. In general, we find that larger core tensors do provide better fit given the large amount of data. However, a larger core also makes the results more difficult to interpret. In the following of this paper, we report the main result based on [$4 \times 3 \times 6 \times 6$] and use [$5 \times 3 \times 8 \times 8$] as a comparison. The convergence criteria is set to be relative change being less than 10^{-5} and maximum iteration is set to 10,000. In order to find global optimal, we ran the EM procedure 20 times with different random initial values. However, we find that the estimation always converges to the same solution. Table 2 provides the values summed over the origin zone mode and the destination zone mode in the core tensor, respectively. A cell $\pi_{k_1 \ldots k_4}$ characterizes how much the four components in a particular combination interact with each other and shows the probability that a trip is characterized by temporal pattern T(k_1), passenger pattern C(k_2), origin pattern O(k_3) and destination pattern D(k_4).

The decomposition allows us to reason about each pattern given its distribution profile. Fig. 1(a) depicts the profiles of the four temporal patterns (four columns in the time factor matrix). We obtain a similar temporal composition profile as in Peng et al. (2012) and Wang et al. (2014). The largest temporal component—T1, with peaks around 6 p.m—covers about 36.5% of all journeys. The second largest component T2 concentrates at the morning peak hours from 7 a.m. to 9 a.m. T3 is mainly distributed in between T1 and T2. The smallest temporal component—T4, looks like a bimodal distribution with a morning peak at 6–7 a.m., which is before the peak of T2, and an afternoon/evening peak that is flatter than the peak in the

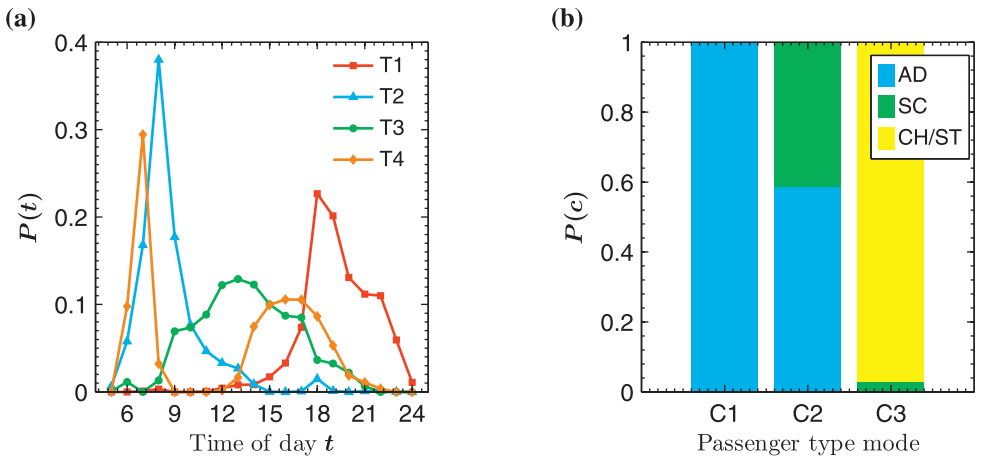

Fig. 1. Principle patterns in time and passenger modes. (a) temporal distribution for each pattern (column) in the time factor matrix, and (b) passenger composition for each pattern (column) in passenger factor matrix.

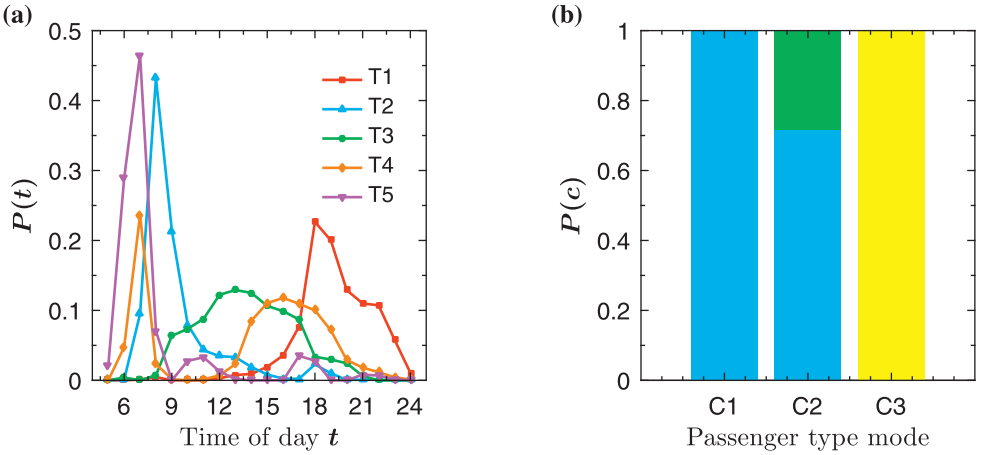

Fig. 2. Principle patterns in time and passenger modes with core tensor size being 5 × 3 × 8 × 8.

morning. The factorization also allows us to measure the corresponding interactions along other modes. Fig. 1(b) shows the composition of each passenger pattern C1, C2 and C3, which cover 70.4%, 18.0% and 11.6% of total journeys, respectively. We see in the figure that senior citizens (SC) and children/students (CH/ST) are clearly distinguished, indicating that different types of passengers have their own unique travel patterns. Compared with the results from a [5 × 3 × 8 × 8] core, we find that the passenger type mode patterns almost remain the same, with children and senior citizens being clearly identified, with C1, C2, and C3 covering 60.8%, 27.9% and 11.4%, respectively. In terms of temporal distribution, we can see that mode T1 almost remains identical and a new pattern T5 is generated by mainly taking the mass on 6–7 a.m. from T2 and T4 (see Fig. 2). The proportion of T1–T5 is 36.2%, 21.3%, 21.5%, 11.7% and 9.3%.

To further quantify the latent interactions, we can calculate the conditional probability across different modes in the core tensor using the Bayes' theorem. For example, the conditional distribution of temporal patterns (T) given passenger composition pattern (C) is

$$P(T|C) = \begin{bmatrix} & T1 & T2 & T3 & T4 \\ C1 & 0.4486 & 0.3112 & 0.1506 & 0.0896 \\ C2 & 0.1879 & 0.1626 & 0.5561 & 0.0934 \\ C3 & 0.1303 & 0.0250 & 0.0959 & 0.7487 \end{bmatrix}. \tag{17}$$

As can be seen, passenger pattern C1 (adults) is mainly associated with temporal profiles T1 and T2. Considering the temporal profiles, T1 and T2 can be principally interpreted as work-home and home-work trips. C2 is mainly (55.6%) covered by temporal pattern T3, which may denote work-work trips and some secondary activities such as going for lunch, leisure and shopping. Passenger pattern C3 is dominated by temporal distribution T4 (74.9%), corresponding to the bimodal nature of home-school/school-home journeys. On the other hand, from Table 2 we can also see that journeys of children and students (C3) cover more than 50% (0.0866/0.1665) of those trips belonging to T4. Similarly, one can further study the interaction

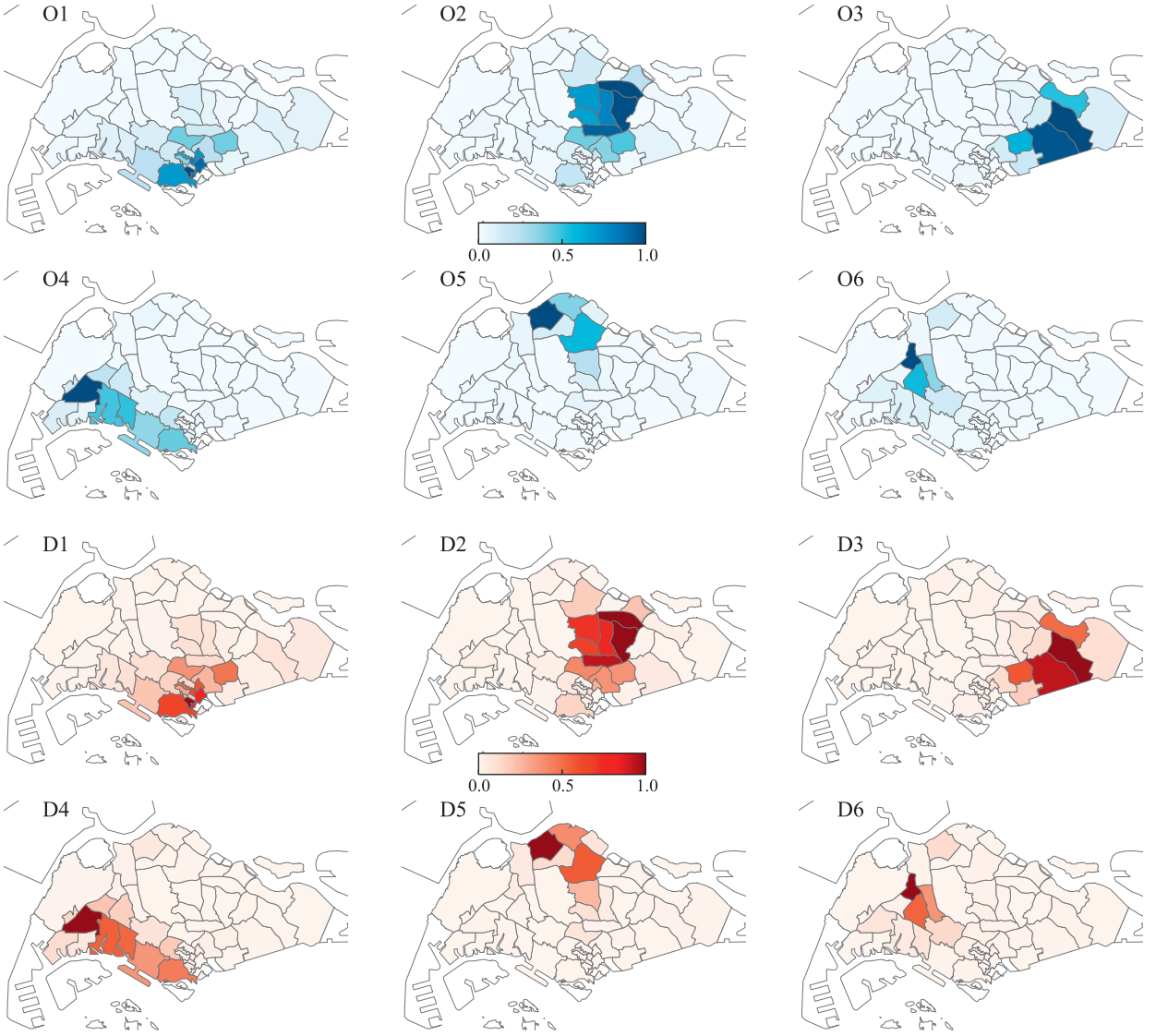

Fig. 3. Rescaled probability distribution of $\boldsymbol{\theta}'^{(O)}$ and $\boldsymbol{\theta}'^{(D)}$ of spatial modes.

by looking at $P(C|T)$. Essentially, by analyzing the factorization results we are able to interpret the latent interactions in the core tensor and each factor matrix.

Fig. 3 displays the column factor of the spatial factor matrix $\boldsymbol{\theta}^{(O)}$ and $\boldsymbol{\theta}^{(D)}$. To better visualize the spatial configuration, the values in each column are rescaled to 0–1 by letting $\theta'^{(d)}_{c_d k} = \theta^{(d)}_{c_d k} / \max(\boldsymbol{\theta}^{(d)}_k)$. Interestingly, we find that the spatial components across the origin zone mode and the destination zone mode are almost identical with minor differences. This indicates that the public transit journeys demonstrate strong bi-directional homogeneity due to those regular home-based trips, such as home-work/work-home and home-school/school-home. The results also suggest that we may further reduce the number of parameters in the factorization model by letting $\boldsymbol{\theta}^{(O)} = \boldsymbol{\theta}^{(D)}$. From now on we discuss the spatial patterns in origin and destination mode jointly.

To compare the inferred spatial patterns with the regions in Singapore's master plan, we show in Fig. 4 the urban regional systems defined by Urban Redevelopment Authority of Singapore. As can be seen, the identified spatial patterns match the master very well except that divided the western region into two patterns. The largest component (O1/D1) is mainly distributed to the central business district (CBD) of Singapore. The rest components are locally distributed in certain regions across the whole island. Even though the estimation process does not take any spatial location information into account, we still observe that each component forms a strong spatial cluster due to the homogeneity of collective travel patterns inside it.

Table 3 shows the joint distribution of origin mode and destination mode in the core tensor. The distribution is obtained by summing up the values in the core tensor along the time and passenger modes. As can be seen, the distribution shows

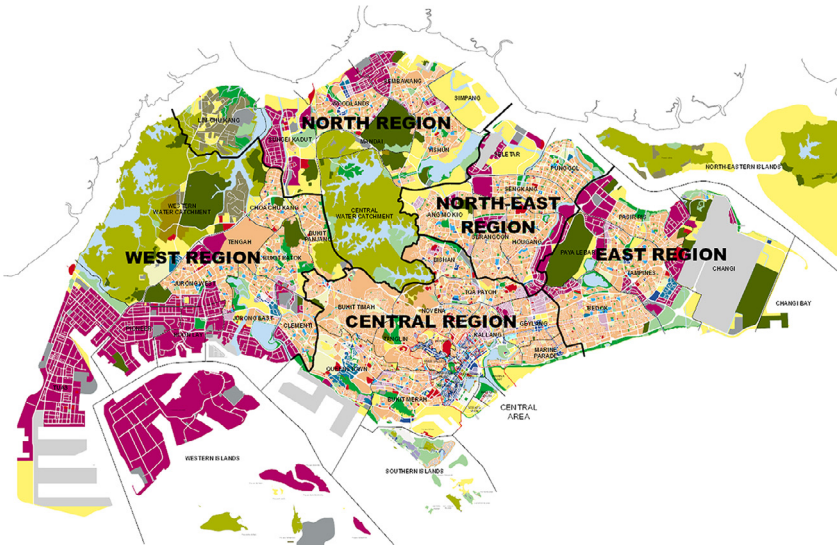

Fig. 4. Planning region of Singapore Master Plan 2014 (source: https://www.ura.gov.sg/maps/).

Table 3
Core tensor summed over origin × destination.

		Destination						
		D1	D2	D3	D4	D5	D6	Sum
Origin	O1	0.1167	0.0584	0.0370	0.0179	0.0224	0.0175	0.2699
	O2	0.0629	0.1252	0.0061	0.0021	0.0045	0.0012	0.2020
	O3	0.0370	0.0038	0.1255	0.0029	0.0019	0.0007	0.1718
	O4	0.0190	0.0006	0.0020	0.1154	0.0034	0.0108	0.1514
	O5	0.0225	0.0053	0.0025	0.0061	0.0705	0.0046	0.1115
	O6	0.0178	0.0010	0.0008	0.0155	0.0050	0.0534	0.0934
	Sum	0.2758	0.1942	0.1740	0.1601	0.1077	0.0882	1.000

very high values on the diagonal of the matrix, indicating that the overall spatial interaction is dominated by inner-pattern interactions. The interactions between pattern 2 to pattern 6 (O2-O6/D2-D6) is found to be very weak, while the CBD pattern (O1/D1) demonstrates strong interactions with all the rest patterns. Considering that CBD often plays the role of primary work location, it generates and attracts large volumes of traffic across the whole city, showing strong interactions with all the other regions. On the contrary, the local components (2–6) seem to be well self-sustained, with little communication between each other.

To further explore the patterns of spatial interactions, we show the disaggregated distribution along each temporal component in Fig. 5. As can be seen, T3 and T4 are almost locally distributed without much interactions with other spatial patterns. The essential difference is the value of cell (O1,D1) in T3 is substantially larger than that in T4, indicating that T3 has strong interactions within spatial pattern 1 (O1/D1), while T4 is barely revealed in CBD. This can be explained by integrating information from the temporal and passenger type dimension. As mentioned before, T4 is mainly composed of trips from children and students and it is mainly covered by inner region home-school/school-home journeys that are not related to CBD. In comparing the spatial interaction of T1 and T2, we find that they are almost opposite to each other due to the balanced work-home and home-work journeys. However, despite the difference in strength, we observe a considerable amount of trips within the CBD pattern in evening (T1), while such trips rarely appear in the morning (T2). In fact, the CBD of Singapore contains only a few residential buildings so it does not generate large flows in the morning. The considerable inner CBD trips in T1 suggest that those people working in CBD tend to have some secondary activities—such as leisure and shopping—around their work location before going home. This also explains why the interaction strength on the diagonal from pattern 2 to 6 in T1 is stronger than T2 and why temporal component T1 covers a larger proportion than T2. Similarly, we can also analyze the interactions along the passenger mode. In principle, the numerical example shows that the core tensor $\boldsymbol{\pi}$ characterizes the interactions between different modes in a very efficient and informative manner. The factorization allows us to explain the complex dependence and higher-order interactions in the joint distribution $p(\mathbf{x})$ by using relatively few latent factors.

Based on the factor matrices on the two spatial modes, we can further explore the community structure of a city (Sun et al., 2015; Wang et al., 2014). In fact, by working on the factor matrices, we can compute the conditional probabil-

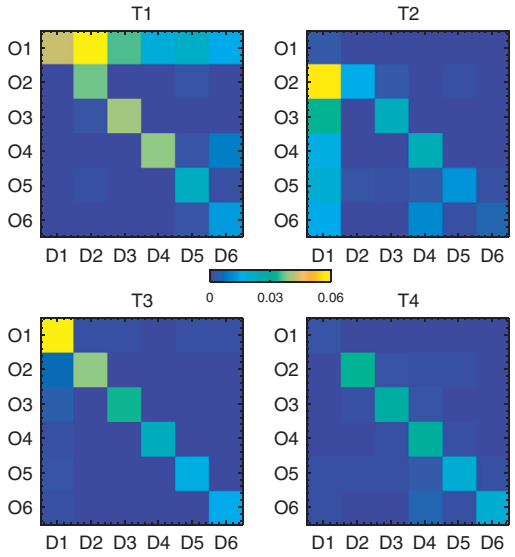

Fig. 5. Spatial interaction (origin × destination) for different temporal patterns.

ity that a zone belongs to a particular pattern (corresponding to community), and thus we can infer not only the most probable structure, but also the degree of overlapping (Yang and Leskovec, 2013). In defining communities, we denote the community that zone c_d belongs to as the pattern index k ($k = 1, \ldots, h_d$) with highest $\theta^{(d)}_{c_d k}$. In other words, we let $\text{community}(c_d) = \arg\max_k \theta^{(d)}_{c_d k}$. For the case study, both origin zones and destination zones are categorized into six communities. Panel (a,c) and (b,d) in Fig. 6 show the community structure of origin zone mode and destination zone mode, respectively. The opacity in panel (a) and (b) indicates the absolute value of $\theta^{(d)}_{c_d,k}$ of the corresponding community in the origin/destination projection matrices, respectively. While in panel (c) and (d), the opacity denotes the conditional probability that an zone c_d belongs to community k. As can be seen, the community structures for origin zones and destination zones are almost identical. The uncertainty of membership mainly comes from two factors. First, zones with little traffic might be assigned to different communities, since the variation in traffic will affect the conditional probability significantly. On the other hand, the uncertainty results from zones have strong interactions with different communities. In this case we may consider them as overlapping structures of different communities.

We conduct the same exercise on the [5 × 3 × 8 × 8] core tensor and show the results of spatial-temporal interactions and community structure in Fig. 7 and 8. Comparing Fig. 5 with Fig. 7, we find that the spatial interaction with 8 modes is characterized by a similar structure as the one with 6 modes. Similar interpretations for T1-T4 from 6 modes could be applied on the results from 8 modes. The last pattern T5, although only covers a tiny proportion of all trips, does exhibit a complex interaction across different regions. Given its temporal pattern shown in Fig. 2, this could be explained as those long-distance commuters, who departed very early in the morning and traveled across the whole city to reach their work place in time. Similar to Fig. 6, we show the community structure with 8 spatial modes in Fig. 8. Essentially, we find that the community colored in yellow in Fig. 6 is divided into two (yellow and brown) in Fig. 8. And the blue community in Fig. 6 is divided into blue and black in Fig. 8.

5. Concluding remarks

In this paper, we apply a probabilistic factorization framework on multi-way transit trip records to reveal the spatial-temporal patterns of urban mobility. To this end, we consider transit journey a multivariate random variable and assume that each journey observation is a sample generated from a universal joint distribution. The joint distribution describes the complicated dependence and higher-order interactions among different dimensions. In order to characterize those dependence and interactions in such a high-dimensional setting, we use a probabilistic factorization model—which is consistent with probabilistic latent semantic analysis and non-negative Tucker decomposition—to reconstruct the joint distribution using a low-rank approximation and simple latent structures. The model can be efficiently estimated using the EM algorithm and the results can be better interpreted with the tensor representation.

The model focuses on providing a generalized data-driven framework to better utilize the increasing amount of individual-based spatial-temporal data. We apply the framework on 14 million transit journeys extracted from smart card transactions in Singapore. The original trip data is summarized with a four-way journey representation: time of day, passenger type, origin zone and destination zone. The factorization clearly identifies and classifies principle patterns and the result

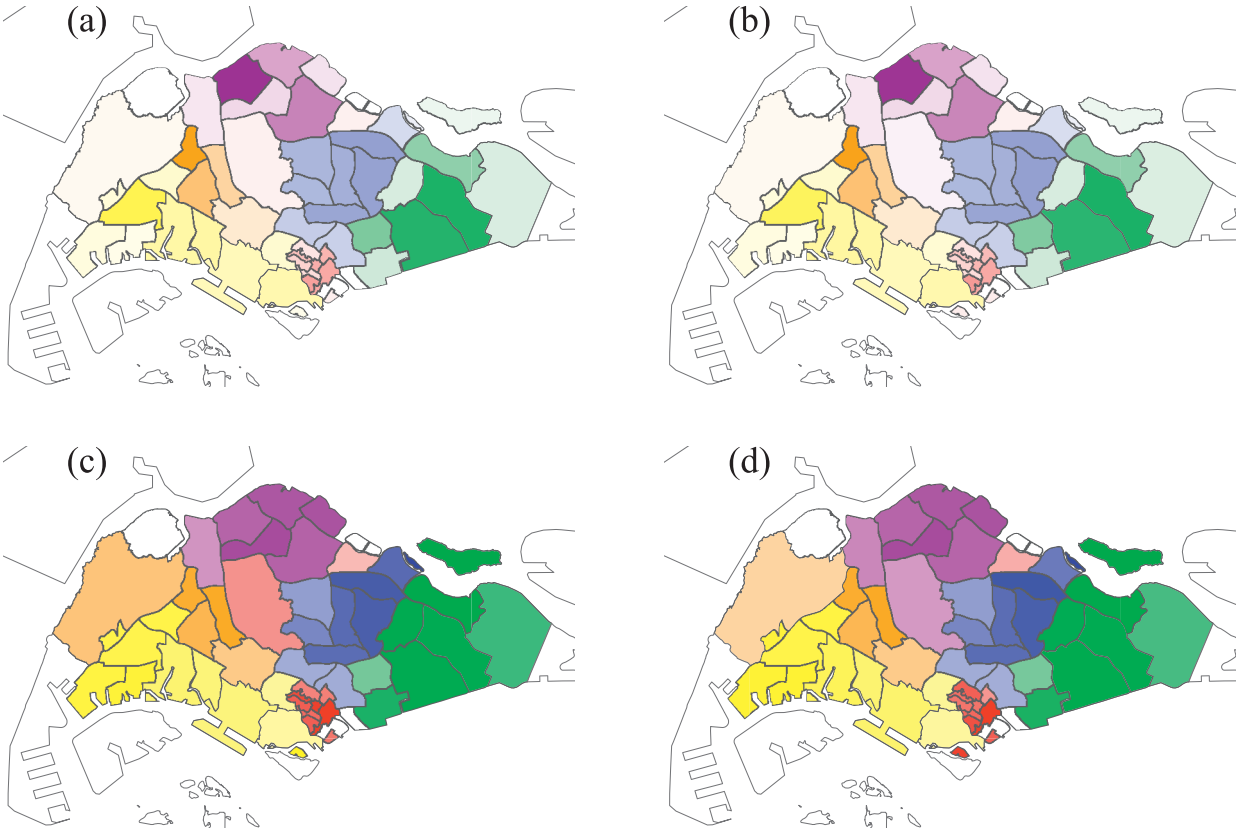

Fig. 6. Spatial structure of communities of origin mode (a,c) and destination mode (b,d). The six different colors indicate difference communities in all these panels. In [a,b], the opacity indicates the exact value of $\theta^{(d)}_{c_d,k}$ in factor matrices. In [c,d], the opacity represents the conditional probability that a zone belongs to a community.

is in agreement with previous studies on temporal patterns of urban mobility. Despite capturing patterns using factor matrices, the joint interactions are also characterized in the core tensor, which can be considered weight parameters in latent class models in the classification. As a result, the model also provides us a new clustering approach that takes information from different dimensions into account, helping us better reveal the underlying spatial-temporal structure of a city from human mobility data. In the numerical experiment, we demonstrate the patterns of transit journeys in terms of temporal distribution and passenger composition. By examining these patterns, we can classify groups of people with unique temporal activity distribution. Although spatial information is only used implicitly, we still observe strong spatial patterns in both the spatial factor matrices and the core tensor. By inspecting the results, it is possible to characterize all zones into separate communities, in which inner community interactions are stronger than those of inter community.

The questions about the dynamics and complexity of cities are essential to urban/transportation planning. Our paper enriches the information in mobility data by reasoning about critical mobility pattern in a multi-dimensional setting, providing a deep understanding of spatial-temporal urban dynamics through collective transit mobility. The probabilistic factorization is a general framework and it demonstrates great flexibility for classification from complex data sets. Several directions are worth noting for future research. For example, additional information such as trip purposes, trip chain and transport modes could be inferred and integrated to the model by adding new dimensions to the data (Han and Sohn, 2016; Nassir et al., 2015). Modeling individual journeys in a probabilistic setting also allows us to adopt a Bayesian framework to further support the estimation. In this case, we can use Dirichelet distribution as priors to the multinomial distribution in both the factor matrices and the core tensor. The model can also be adapted for different purposes, such as estimating origin-destination matrices. For example, focusing only on the destination dimension we could use it as a data-driven destination choice model that is sensitive to time and passenger types. This may help us to build a new transportation demand modeling framework and other agent-based models. On the other hand, the concept of tensor decomposition provides us a natural solution to deal with missing data under some assumptions and to reconstruct the full joint distribution from a small set of available data. With the increasing amount of available data nowadays, such approaches become advantageous for us to understand and interpret the latent interactions and complex dependence in the large datasets registering individual behaviors, offering new insights to urban and transportation planning by enhancing our understanding about collective human mobility.

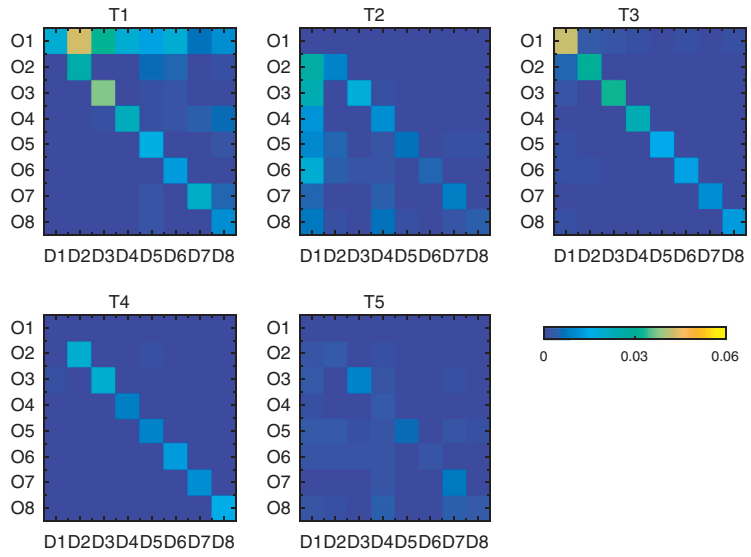

Fig. 7. Spatial interaction (origin × destination) for different temporal patterns.

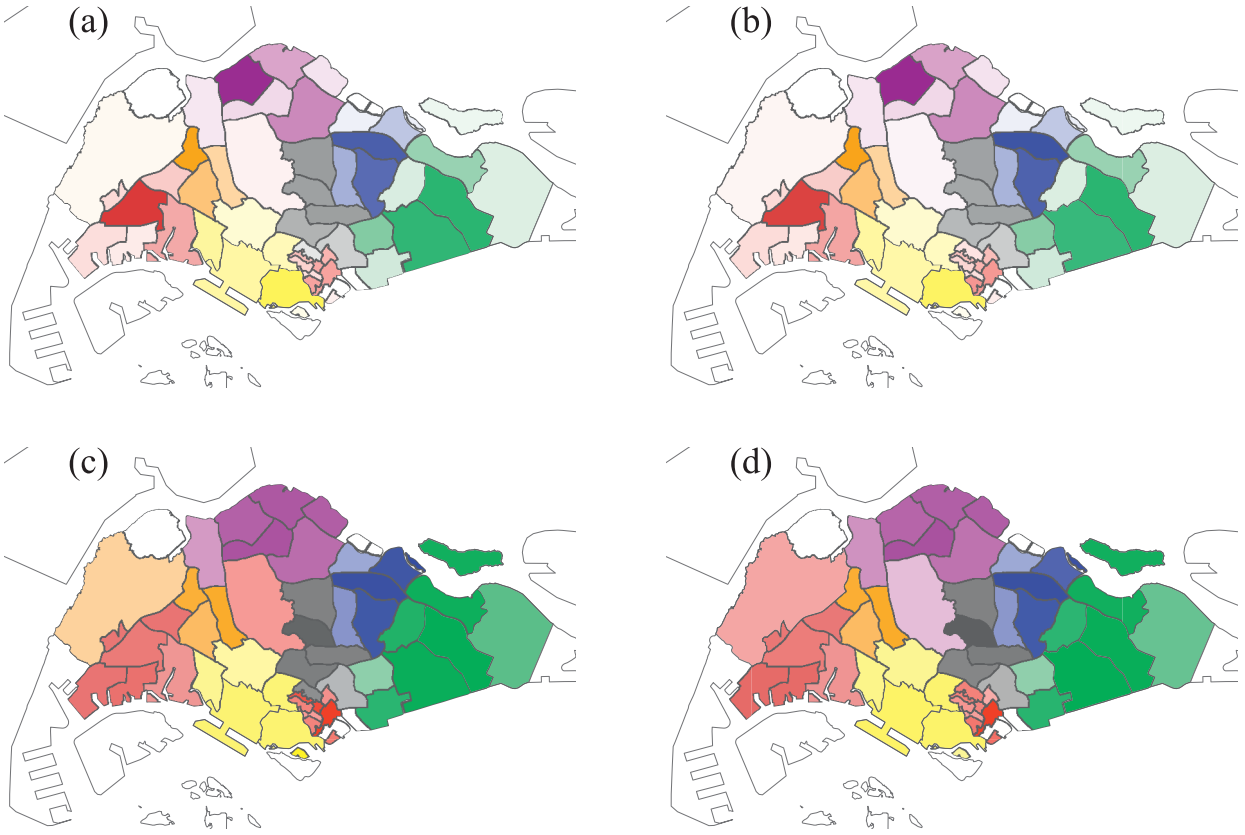

Fig. 8. Same as Fig. 6, but for 5 × 3 × 8 × 8 core tensor.

Acknowledgment

This research was partially supported by the Singapore National Research Foundation and ETH Zürich as a project at the Singapore-ETH Centre for Global Environmental Sustainability under the programme of the Future Cities Laboratory (NRF FI No. 370071002).

References

Agresti, A., 2002. Categorical Data Analysis, second ed. John Wiley & Sons, Hoboken, NJ.

Allahviranloo, M., Recker, W., 2013. Daily activity pattern recognition by using support vector machines with multiple classes. Transp. Res. Part B 58, 16–43.

Axhausen, K.W., Gärling, T., 1992. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. Transp. Rev. 12 (4), 323–341.

Batty, M., Axhausen, K., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., Portugali, Y., 2012. Smart cities of the future. Eur. Phys. J. Special Topics 214, 481–518.

Bhat, C.R., Koppelman, F.S., 2003. Activity-Based Modeling of Travel Demand, Vol. 3. Springer, US, pp. 39–65.

Bhattacharya, A., Dunson, D.B., 2012. Simplex factor models for multivariate unordered categorical data. J. Am. Stat. Assoc. 107 (497), 362–377.

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., Ratti, C., 2011. Real-time urban monitoring using cell phones: a case study in rome. Intell. Transp. Syst. IEEE Trans. on 12 (1), 141–151.

Coffey, C., Pozdnoukhov, A., 2013. Temporal decomposition and semantic enrichment of mobility flows. In: Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks. LBSN '13. ACM, Orlando, Florida, pp. 34–43.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. 39 (Series B), 1–38.

Ding, C., Li, T., Peng, W., 2008. On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. Comput. Stat. Data Anal. 52 (8), 3913–3927.

Dunson, D.B., Xing, C., 2012. Nonparametric bayes modeling of multivariate categorical data. J. Am. Stat. Assoc. 104 (487), 1042–1051.

Fan, Z., Song, X., Shibasaki, R., 2014. Cityspectrum: a non-negative tensor factorization approach. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, pp. 213–223.

González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. Nature 453, 779–782.

Han, G., Sohn, K., 2016. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model. Transp. Res. Part B 83, 121–135.

Han, Y., Moutarde, F., 2014. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. Int. J. Intell. Transp. Syst. Res. 14 (1), 36–49.

Hofmann, T., 1999. Probabilistic latent semantic indexing. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 50–57.

Jiang, S., FerreiraJr., J., González, M.C., 2012. Discovering urban spatial-temporal structure from human activity patterns. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing. ACM, Beijing, China, pp. 95–102.

Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 51 (3), 455–500.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401 (6755). 788–91

Ma, X., Wu, Y., Wang, Y.-J., Chen, F., Liu, J., 2013. Mining smart card data for transit riders' travel patterns. Transp. Res. Part C 36, 1–12.

Nassir, N., Hickman, M., Ma, Z.-L., 2015. Activity detection and transfer identification for public transit fare card data. Transportation 42, 483–705.

Peng, C., Jin, X., Wong, K.-C., Shi, M., Liò, P., 2012. Collective human mobility pattern from taxi trips in urban area. PloS One 7 (4), E34487.

Peng, W., Li, T., 2011. On the equivalence between nonnegative tensor factorization and tensorial probabilistic latent semantic analysis. Appl. Intell. 35 (2), 285–295.

Reades, J., Calabrese, F., Ratti, C., 2009. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. Env. Plann. B 36 (5), 824–836.

Roth, C., Kang, S.M., Batty, M., Barthélemy, M., 2011. Structure of urban movements: polycentric activity and entangled hierarchical flows. PloS One 6 (1), E15923.

Sun, J., Tao, D., Faloutsos, C., 2006. Beyond streams and graphs: dynamic tensor analysis. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '06. ACM, pp. 374–383.

Sun, L., Axhausen, K.W., Lee, D.H., Huang, X., 2013. Understanding metropolitan patterns of daily encounters. In: Proceedings of the National Academy of Sciences of the United States of America, 110, pp. 13774–13779.

Sun, L., Jin, J.G., Axhausen, K.W., Lee, D.H., Cebrian, M., 2015. Quantifying long-term evolution of intra-urban spatial interactions. J. R. Soc. Interface 12 (102). 20141089–20141089

Sun, L., Lee, D.-H., Erath, A., Huang, X., 2012. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of mrt system. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing. UrbComp '12. ACM, Beijing, China, pp. 142–148.

Sun, L., Tirachini, A., Axhausen, K.W., Erath, A., Lee, D.-H., 2014. Models of bus boarding and alighting dynamics. Transp. Res. Part A 69, 447–460.

Tan, H., Feng, G., Feng, J., Wang, W., Zhang, Y.-J., Li, F., 2013. A tensor-based method for missing traffic data completion. Transp. Res. Part C 28, 15–27.

Wang, J., Gao, F., Cui, P., Li, C., Xiong, Z., 2014. Discovering urban spatio-temporal structure from time-evolving traffic networks. In: Chen, L., Jia, Y., Sellis, T., Liu, G. (Eds.), Web Technologies and Applications. In: Lecture Notes in Computer Science, Vol. 8709. Springer International Publishing, pp. 93–104.

Yang, J., Leskovec, J., 2013. Overlapping community detection at scale: A nonnegative matrix factorization approach. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. WSDM '13. ACM, Rome, Italy, pp. 587–596.

Yuan, J., Zheng, Y., Xie, X., 2012. Discovering regions of different functions in a city using human mobility and pois. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '12. ACM, Beijing, China, pp. 186–194.

Zafeiriou, S., Petrou, M., 2011. Nonnegative tensor factorization as an alternative csiszar–tusnady procedure: algorithms, convergence, probabilistic interpretations and novel probabilistic tensor latent variable analysis algorithms. Data Mining Knowl. Discov. 22 (3), 419–466.

Zhang, F., Wilkie, D., Zheng, Y., Xie, X., 2013. Sensing the pulse of urban refueling behavior. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. UbiComp '13. ACM, Zurich, Switzerland, pp. 13–22.