

## Research Article

# Simultaneous Incomplete Traffic Data Imputation and Similarity Pattern Discovery with Bayesian Nonparametric Tensor Decomposition

Yaxiong Han and Zhaocheng He 

Guangdong Provincial Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China

Correspondence should be addressed to Zhaocheng He; hezhch@mail.sysu.edu.cn

Received 27 May 2020; Revised 2 July 2020; Accepted 27 July 2020; Published 12 August 2020

Academic Editor: Bin Yu

Copyright © 2020 Yaxiong Han and Zhaocheng He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A crucial task in traffic data analysis is similarity pattern discovery, which is of great importance to urban mobility understanding and traffic management. Recently, a wide range of methods for similarities discovery have been proposed and the basic assumption of them is that traffic data is complete. However, missing data problem is inevitable in traffic data collection process due to a variety of reasons. In this paper, we propose the Bayesian nonparametric tensor decomposition (BNPTD) to achieve incomplete traffic data imputation and similarity pattern discovery simultaneously. BNPTD is a hierarchical probabilistic model, which is comprised of Bayesian tensor decomposition and Dirichlet process mixture model. Furthermore, we develop an efficient variational inference algorithm to learn the model. Extensive experiments were conducted on a smart card dataset collected in Guangzhou, China, demonstrating the effectiveness of our methods. It should be noted that the proposed BNPTD is universal and can also be applied to other spatiotemporal traffic data.

## 1. Introduction

Recent advances in data acquisition technologies and mobile computing lead to a collection of large quantities of urban traffic data from various sources, such as loop detectors data, GPS data, and smart card data. These datasets can capture rich spatial-temporal information of the whole transportation system and enable some traffic analysis. A crucial task in a data-driven transportation system is similarity pattern discovery. For example, as shown in Figure 1(a), two vehicles have different trajectories but share the same trip origin and trip destination. Figure 1(b) shows that the passenger inflow time series of subway station A and subway station B are similar in a week, which both have peaks in the morning and the peak values of weekends are relatively small to weekdays.

These similarities are beneficial for urban mobility pattern understanding and the authorities' policy-making.

For example, for aggregate-level, the classification management can be adopted in metro systems and the managers should pay more attention to station A and station B to prevent congestion during the morning peak. For individual-level, the travelers that have similar trip rules can be found, that is, familiar stranger [1] so that the authorities can perform more precise demand control and provide personalized travel service. In addition, the similarities can be used for anomaly detection and improving the traffic prediction as a prior knowledge [2–4].

In general, the similarity pattern could be extracted by clustering methods, such as K-means algorithm and the density-based spatial clustering of applications with noise (DBSCAN) algorithm [5]. However, the conventional clustering algorithms always require the complete data, while the missing data problem is inevitable in traffic data collection process. For example, some observations will not be recorded if the sensors were broken. Traditionally, the

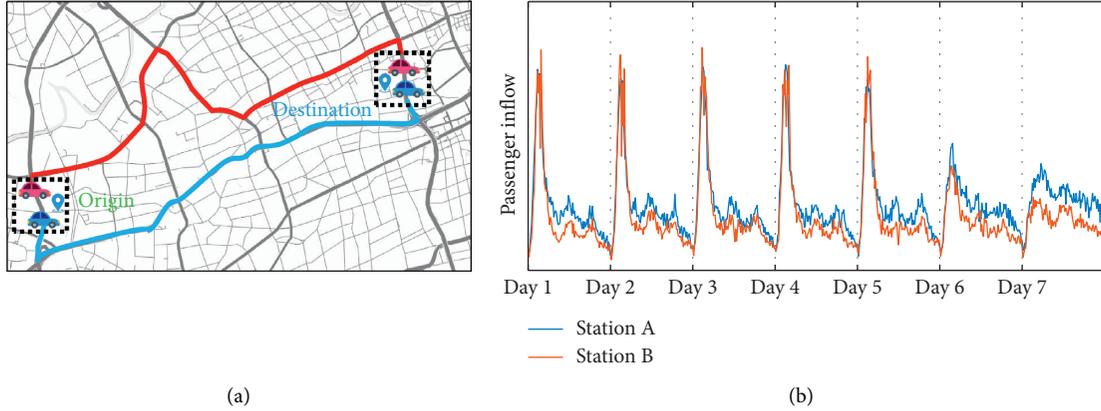


FIGURE 1: The graph description of similar patterns in the transportation system. (a) The same trip origin and trip destination. (b) The similar passenger flow time series.

missing data should be filled firstly; then, a clustering approach is implemented for the imputed data [6]. The missing data imputation and clustering are conducted separately, which is complex and time-consuming. Besides, the frequently used clustering methods such as K-means algorithm must set the number of clusters in advance, and the tedious sensitivity analysis is necessary.

Aiming at the missing data imputation, a variety of methods have been proposed such as multioutput Gaussian processes [7], deep generative models [8], and Bayesian tensor decomposition [9], among which Bayesian tensor decomposition is proved to be more effective and efficient than the other methods. Tensor (i.e., multiway arrays) is a generalization of matrix, which is widely used for modeling multidimensional traffic data and can capture the multidimensional structural dependencies; for example, the passenger flow of subway stations can be represented with a three-order tensor (station  $\times$  day  $\times$  time of the day). Subsequently, the tensor decomposition can be employed to project station, day, and time into a latent space, also known as factor matrices. Finally, the missing data can be filled by the interaction of factor matrices. Recently, many studies have solved tensor decomposition in the Bayesian formula [9,10], which can provide more robust estimation and uncertainty measures for those missing data.

Inspired by the recent work of Bayesian tensor decomposition, we propose a novel framework named Bayesian nonparametric tensor decomposition (BNPTD) to achieve incomplete traffic data imputation and similarity pattern discovery simultaneously. The BNPTD consists of two components: (1) Bayesian tensor decomposition and (2) Dirichlet process mixture model (DPMM). DPMM is a nonparametric Bayesian mixture model that has shown great promise for data clustering and allows for the automatic determination of an appropriate number of clusters. These two components are combined with a hierarchical probabilistic model and a variational inference algorithm is presented to derive the posterior distributions of all the model parameters and hyperparameters. It should be noted that the combination not only finds groups of similar objects in the case of missing data but also offers adaptive prior to

Bayesian tensor decomposition model which can further improve better imputation performance.

In summary, our contributions are summarized as follows:

- (i) We proposed a Bayesian nonparametric tensor decomposition model to achieve incomplete traffic data imputation and similarity pattern discovery simultaneously.
- (ii) We presented a variational inference algorithm to learn the BNPTD model. Variational inference tends to be faster and easier to scale to large-scale datasets than classical methods, such as Markov chain Monte Carlo sampling.
- (iii) Extensive experiments were conducted on a smart card dataset from Guangzhou, China. The results demonstrate that our approach successfully finds interpretable similarity pattern and recovers the missing values well in the case of incomplete traffic data.

The rest of this paper is structured as follows. Section 2 presents a literature review. Section 3 describes the materials and methods. The results and discussion are presented in Section 4. Finally, Section 5 summarizes the main conclusions of this work.

## 2. Literature Review

In this section, we review the previous approaches regarding missing data imputation and similarity pattern discovery in the traffic data analysis.

*2.1. Missing Data Imputation.* Missing data problem is inevitable in traffic data collection process, and the reasons are manifold. For example, some observations will be lost if the sensors were broken. Besides, when we use mobile sensors such as floating car, some locations will not be covered at some times. So missing data imputation is always a hot topic in academic research and predecessors have solved this problem using a wide range of methods. The traditional

approach is to use time series method such as autoregressive integrated moving average (ARIMA) and its variants [11]. Recently, machine learning methods that considering spatial-temporal correlations have achieved the state-of-the-art performance. For example, Rodrigues et al. [7] proposed the use of multioutput Gaussian processes to model the complex spatial and temporal patterns for crowdsourced traffic data imputation. Yoon et al. [12] adapted the well-known Generative Adversarial Nets (GAN) framework and proposed Generative Adversarial Imputation Nets.

The collaborative filtering methods such as tensor decomposition have been proved more effective and efficient than other methods. Tensor (i.e., multiway arrays) is a generalization of matrix. The two most popular tensor decomposition frameworks are Tucker and CANDECOMP/PARAFAC (CP), which both could capture the underlying multilinear factors. Kolda et al. [13] gave a comprehensive overview of tensor decomposition and its applications. Tan et al. [14] introduced the tensor to preserve the multiway nature of traffic data and developed a tensor decomposition based imputation method for missing traffic data completion. Zhao et al. [10] formulated CP decomposition using a hierarchical probabilistic model and incorporated a sparsity-inducing prior over factor matrices, resulting in automatic rank determination. Chen et al. [9] extended the Bayesian probabilistic matrix factorization to higher-order tensor to recover missing values. Zhang et al. [15] proposed an iterative tensor decomposition approach and proved its effectiveness under different missing cases.

**2.2. Similarity Pattern Discovery.** A crucial task in traffic data analysis is similarity pattern discovery and the similarities can help to demand control, personalized travel service, and anomaly detection. The similarity pattern can be extracted by clustering methods. Zhao et al. [2] proposed an unsupervised clustering-based approach to classify passengers in terms of the similarity of their travel patterns and find out the abnormal passengers in metro systems. Gan et al. [16] used the k-means algorithm for clustering metro stations and identified urban mobility patterns from a spatiotemporal perspective.

Besides, the nonparametric clustering such as Dirichlet process mixture model (DPMM) is widely used for similarity pattern discovery. DPMM is a nonparametric Bayesian mixture model which allows for the automatic determination of an appropriate number of clusters. The most commonly used description of DPMM is Chinese Restaurant Process that can be solved by MCMC sampling [17]. Blei et al. [18] presented an alternative stick-breaking representation and adopted a variational inference algorithm for DPMM. It is universally acknowledged that variational inference is more efficient than MCMC sampling and adapted to large-scale datasets. In traffic data analysis, Ngan et al. [19] and Ngan et al. [20] employed DPMM to outlier detection and validated its effectiveness via numerical experiments. However, DPMM only can apply to complete data while the real-world traffic data is incomplete inevitably.

In this paper, we propose Bayesian nonparametric tensor decomposition that is comprised of Bayesian tensor decomposition and Dirichlet process mixture model via a hierarchical probabilistic model to achieve incomplete traffic data imputation and similarity pattern discovery simultaneously. Compared to the previous studies, DPMM is employed to low-rank feature extracted from Bayesian tensor decomposition rather than raw data, which can solve similarity pattern discovery in the case of incomplete traffic data effectively. Moreover, DPMM is beneficial for missing data imputation as an adaptive prior distribution.

### 3. Materials and Methods

The proposed BNPTD consists of two components: Bayesian tensor decomposition model and Dirichlet process mixture model. In the following subsections, we first give detailed descriptions of these two components. Secondly, we provide an overview of our Bayesian nonparametric tensor decomposition model. Finally, we derive variational approximations to learn the proposed BNPTD. The description below is taken from smart card data as an example and can be applied to other spatiotemporal traffic data.

**3.1. Bayesian Tensor Decomposition Model.** A smart card data record usually has multidimensional attributes such as origin station, destination station, day, and time of day. We organize the smart card data into a three-order tensor  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ , where  $I$ ,  $J$ , and  $K$  represent the size of station dimension, day dimension, and time dimension, respectively. Each entry in the tensor  $\mathcal{X}$  is denoted as  $x_{ijk}$  to represent the inflow (or outflow) of  $i$ th station in  $k$ th time of  $j$ th day.

$\mathcal{X}$  could be an incomplete tensor when the traffic data is missing. We assume the element  $x_{ijk}$  is observed if  $(i, j, k) \in \Omega$ , where  $\Omega$  denotes a set of indices. Furthermore, we define a binary tensor  $\mathcal{O}$  as an indicator of the observed entries, which has the same size as  $\mathcal{X}$ .

Then, CP decomposition factorizes the tensor into a sum of rank-one tensors; that is,

$$X \approx U, V, W \equiv \sum_{r=1}^R u_r \circ v_r \circ w_r, \quad (1)$$

where  $U, V$ , and  $W$  represent the factor matrices of each dimension, respectively, and  $u_r, v_r$ , and  $w_r$  are the  $r$ th column vectors of factor matrices. The symbol  $\circ$  denotes the outer product of the vectors and  $R$  is the tensor rank. Figure 2 is a visualization of the CP decomposition of a third-order tensor.

Elementwise, (1) is written as

$$x_{ijk} \approx \hat{x}_{ijk} = \sum_{r=1}^R u_{ir} v_{jr} w_{kr}. \quad (2)$$

CP decomposition can be solved by the alternating least squares method [13]. In order to obtain a more robust solution and avoid overfitting, we resort to Bayesian formula and assume Gaussian observation as follows:

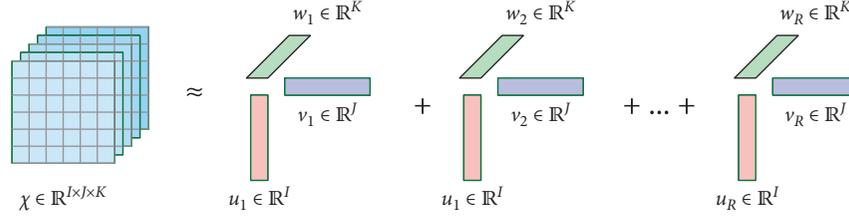


FIGURE 2: CP decomposition of a third-order tensor.

$$x_{ijk} \sim N(\hat{x}_{ijk}, \tau^{-1}), \quad (3)$$

where  $\mathcal{N}(\cdot)$  denotes the Gaussian distribution and  $\tau$  is precision.

The maximum likelihood estimation of (2) is equivalent to the minimum of square loss function  $\mathcal{X} - \hat{\mathcal{X}}_{\Omega}^2$ . To avoid overfitting, we place conjugate prior distributions over factor matrices and hyperparameters, which correspond to regularization terms:

$$\mathbf{u}_i \sim N(\boldsymbol{\mu}_u, [\Lambda_u]^{-1}) \quad (\boldsymbol{\mu}_u, \Lambda_u) \sim \text{Gaussain - Wishart}(\boldsymbol{\mu}_0, \beta_0, \mathbf{W}_0, \nu_0), \quad (4)$$

where  $u_i$  is the  $i$ th row vector of factor matrix  $U$  and  $V, W$  have similar prior distributions.

We also place a conjugate prior distribution over the precision; that is,

$$\tau \sim \text{Gamma}(a_0, b_0). \quad (5)$$

The graphical model of Bayesian tensor decomposition is illustrated in Figure 3.

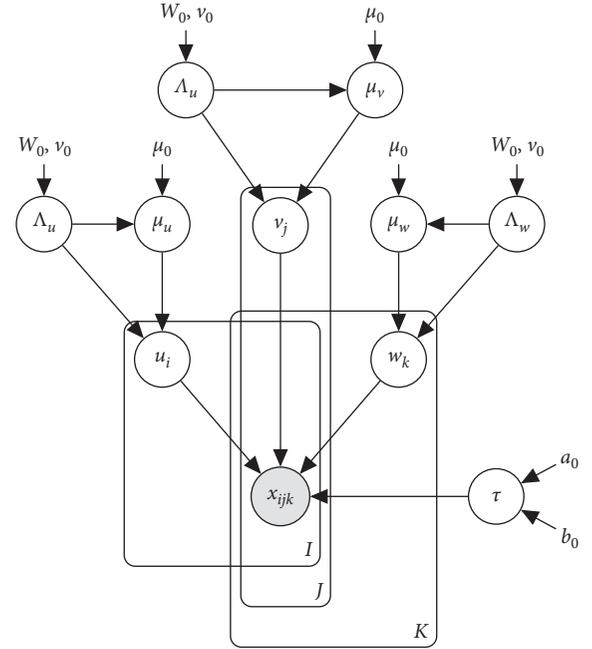


FIGURE 3: The graphical model of Bayesian tensor decomposition.

**3.2. Dirichlet Process Mixture Model.** The smart card data can be represented with a three-order tensor (station  $\times$  day  $\times$  time of the day); then Bayesian tensor decomposition factorizes the tensor into factor matrices  $U$  and  $V, W$ . The factor matrices also denote the features of each attribute; for example,  $u_i$  reflects passenger departure (or arrival) demand patterns of  $i$ th station in the specific time interval. In order to find groups of stations that have similar demand pattern, we place a Dirichlet process mixture prior over the factor matrix  $U$ .

Dirichlet process mixture model is the Bayesian non-parametric technique which allows for the automatic determination of an appropriate number of mixture components. In this paper, we adopt the stick-breaking construction of DPMM [18]. Specifically, let  $z_i$  be a categorical variable that indicates demand pattern category of  $i$ th station.  $T$  is truncation level that could be chosen large relative to the number of stations. The generative process of DPMM is given as follows:

- (1) For each mode  $t \in \{1, 2, \dots, T\}$ 
  - (a) Draw  $\beta_t \sim \text{Beta}(1, \alpha)$
  - (b) Compute  $\pi_t(\beta) = \beta_t \prod_{m=1}^{t-1} (1 - \beta_m)$
  - (c) Draw  $\mu_t \sim G_0$

- (2) For each station  $i \in \{1, 2, \dots, I\}$ 
  - (a) Draw  $z_i \sim \text{Categorical}(\pi(\beta))$
  - (b) Draw  $\mathbf{u}_i \sim \mathcal{N}(\mu_{z_i}, \mathbf{I})$

where  $\alpha$  is hyperparameter that could be set in advance.  $\pi_t(\beta)$  is the probability value of mode  $t$ .  $G_0$  is base distribution, which could be chosen Gaussian distribution  $\mathcal{N}(0, \lambda \mathbf{I})$  in this work.  $\mathbf{I}$  is identity matrix and  $\lambda$  is hyperparameter. The graphical model of DPMM is illustrated in Figure 4.

**3.3. The Proposed Bayesian Nonparametric Tensor Decomposition Model.** The BNPTD is comprised of Bayesian tensor decomposition and DPMM, and these components are coupled via a hierarchical probabilistic model. The graphical model of BNPTD is illustrated in Figure 5.

The advantages of the combination are as follows: (1) missing data imputation and similarity pattern discovery can be achieved simultaneously, which is more efficient and avoids errors accumulation, and (2) DPMM can offer adaptive prior distribution to Bayesian tensor decomposition, corresponding to adaptive regularization term and further better imputation performance.

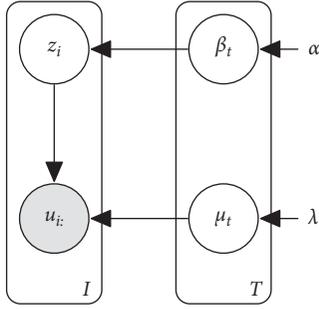


FIGURE 4: The graphical model of DPMM.

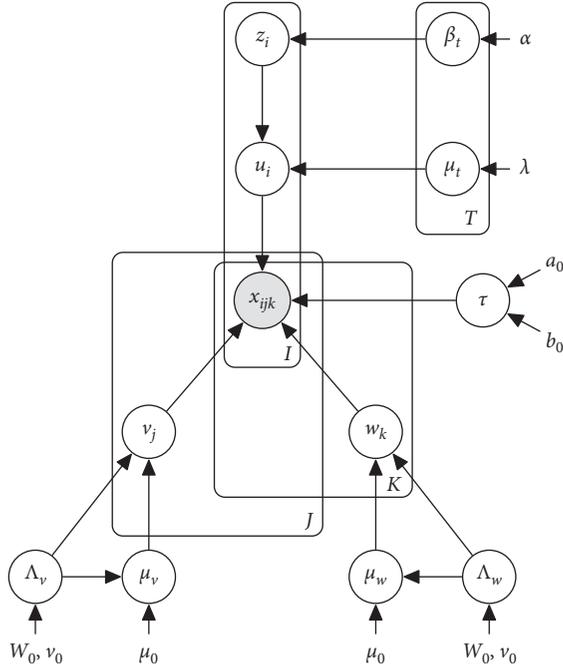


FIGURE 5: The graphical model of BNPTD.

**3.4. Model Learning via Variational Inference.** The goal of BNPTD model learning is to derive the posterior distributions of parameters and hyperparameters. Variational inference is the method that approximates the posterior distributions through optimization [21]. Specifically, for a group of random variables  $\Theta$ , variational inference seeks a variational distribution  $q(\Theta)$  to approximate the true posterior distribution  $p(\Theta|\mathcal{X}_\Omega)$ . The approximation is measured by Kullback–Leibler divergence, and a Kullback–Leibler divergence of 0 indicates that the two distributions are identical:

$$\begin{aligned} KL(q(\Theta) \parallel p(\Theta|\mathcal{X}_\Omega)) &= \int q(\Theta) \ln \left\{ \frac{q(\Theta)}{p(\Theta|\mathcal{X}_\Omega)} \right\} d\Theta \\ &= E[\ln q(\Theta)] - E[\ln p(\Theta|\mathcal{X}_\Omega)], \end{aligned} \quad (6)$$

where all expectations are taken with respect to  $q(\Theta)$ . Because the KL divergence is difficult to compute, variational

inference optimizes an alternative objective that the minimization of KL is equivalent to the maximization of the evidence lower bound (i.e., ELBO).

$$\begin{aligned} KL(q(\Theta) \parallel p(\Theta|\mathcal{X}_\Omega)) &= E[\ln q(\Theta)] - E[\ln p(\Theta, \mathcal{X}_\Omega)] \\ &\quad + \ln p(\mathcal{X}_\Omega) = \text{ELBO}(q(\Theta)) + \ln p(\mathcal{X}_\Omega). \end{aligned} \quad (7)$$

We further focus on the mean-field variational family that the variational distribution  $q(\Theta)$  can be factorized with regard to each variable  $\Theta_j$ ; that is,

$$q(\Theta) = \prod_{j=1}^m q_j(\Theta_j). \quad (8)$$

On the basis of the mean-field variational family, the optimal variational posterior distribution of each  $\Theta_j$  can be derived in turn via maximizing the ELBO, and the optimal form is given by

$$\ln q_j(\Theta_j) = \mathbb{E}_{q(\Theta \setminus \Theta_j)}[\ln p(\mathcal{X}_\Omega, \Theta)] + \text{const.} \quad (9)$$

where notation  $\mathbb{E}_{q(\Theta \setminus \Theta_j)}[\cdot]$  denotes an expectation with respect to distributions  $q(\Theta)$  over all variables except for  $\Theta_j$ . The deduction process of optimal variational posterior distributions with regard to parameters and hyperparameters can be found in Supplementary File. Variational inference updates the model parameters and hyperparameters alternatively. Algorithm 1 outlines the training process of BNPTD.

## 4. Results and Discussion

In order to evaluate the performance of our methods, we conduct extensive experiments on a real-world smart card dataset collected from subway stations in Guangzhou, China. This dataset contains a large number of tap-in/tap-out records from 3/7/2017 to 16/7/2017. Because the subway agency does not provide services all day, we focus on the smart card data from 6 a.m. to 10 p.m., which contain the main trip time of the whole day. We construct the dataset into a three-order tensor (station  $\times$  day  $\times$  time of the day). Each entry of this tensor denotes the passenger inflow. The length of each time interval is set as 10 minutes, and the size of the tensor is  $148 \times 14 \times 96$ .

**4.1. The Missing Data Imputation Performance.** In order to show the effectiveness of the adaptive prior distribution, we compare BNPTD with some baselines including DA (daily average), BTD (Bayesian tensor decomposition), and GAIN (deep generative models [8]) in terms of missing data imputation. The BTD is also trained via variational inference algorithm and two methods have the same initialization. We implement experiments in the two missing scenarios, which are random missing scenario and continuous missing scenario. Besides, we set  $T = 12$  (the maximum clusters),  $R = 10, 20, 30$  (CP rank).

**Input:** incomplete tensor  $\mathcal{X}_\Omega \in \mathbb{R}^{I \times J \times K}$ , indicator tensor  $\mathcal{O} \in \mathbb{R}^{I \times J \times K}$ , CP rank  $R$ , maximum clusters  $T$ .  
**Output:** estimated tensor  $\hat{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ , the results of similarity pattern discovery. Initialize factor matrices and hyperparameters.

- (1) **repeat**
- (2) for  $t = 1$  to  $T$  do
- (3) Update the variational posterior distribution  $q(\mu_t), q(\beta_t)$ .
- (4) Update the variational posterior distribution  $q(\mu_v, \Lambda_v)$  and  $q(\mu_w, \Lambda_w)$ .
- (5) for  $i = 1$  to  $I$  do
- (6) Update the variational posterior distribution  $q(z_i), q(\mathbf{u}_i)$ .
- (7) for  $j = 1$  to  $J$  do
- (8) Update the variational posterior distribution  $q(\mathbf{v}_j)$ .
- (9) for  $k = 1$  to  $K$  do
- (10) Update the variational posterior distribution  $q(\mathbf{w}_k)$ .
- (11) Update the variational posterior distribution  $q(\tau)$ .
- (12) until maximum iterations exhausted

ALGORITHM 1: Bayesian nonparametric tensor decomposition (BNPTD).

In our experiments, we evaluate the imputation accuracy with two widely applied metrics, namely, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (10)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|. \quad (11)$$

The missing ratio ranged from 10% to 50%. Table 1 shows the performance of our proposed method as compared to other baselines in the random missing scenario, and the result of the continuous missing scenario is shown in Table 2. Our proposed BNPTD outperforms other baselines by achieving the lowest RMSE and MAE in most of the missing ratio cases. In some cases, the performance of BNPTD in Tables 1 and 2 is close to BTM, and the reason may be that the adaptive prior distribution is close to single prior distribution in numerical results. But the adaptive prior distribution is conducive to similarity pattern discovery compared to single prior distribution.

**4.2. The Similarity Pattern Discovery.** The CP rank  $R$  represents the size of low-rank feature, and large CP rank may lead to redundant feature which is bad for similarity pattern discovery. In this subsection, we set CP rank  $R = 8$  and the maximum clusters  $T = 12$  empirically. The random missing ratio of incomplete smart card data is 10%. The result of similarity pattern discovery is illustrated in Figure 6, and the passenger flow time series that has similar departure pattern is plotted in a subfigure. There are 7 clusters which represent different departure patterns. Moreover, Figure 7 gives a visualization of stations distribution in Google Maps. The detailed description of these patterns is given as follows:

(a) In the case of Cluster 1, the subway stations have large passenger volume every day, and there is no difference between a working day and nonworking

day. From Figure 7, we observe that these stations are adjacent to passenger transportation hub such as train station and high-speed railway station.

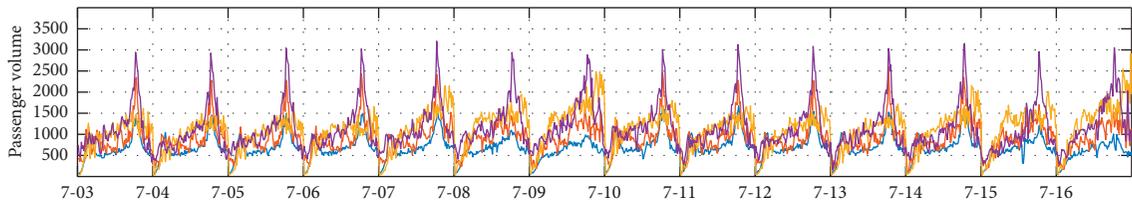
- (b) The second departure pattern is anomalous, and the passenger volume of nonworking days from 8/7/2017 to 9/7/2017 is larger than the working day. With our web search, the reason is that there is an exhibition happening near the subway stations. So, the proposed BNPTD also finds some special events effectively and can be used for anomaly detection.
- (c) Aimed at Cluster 3, the passenger volume of the stations is lower than 600 in 10 minutes. These stations are not usually crowded and the transportation potential could be further excavated to service more passengers.
- (d) Cluster 4 has a larger passenger volume compared to Cluster 3. The stations belonging to Cluster 4 have small peak values during the morning rush hour or evening rush hour. Besides, these stations are mainly situated in urban fringe area and far from the central business area from Figure 7.
- (e) Cluster 5 is a morning peak departure pattern, and the peak values of weekends are relatively small to weekdays. The stations are mainly in close proximity to the residential area in Figure 7. In the morning, people will leave home and go to work by subway.
- (f) The stations with the sixth departure pattern have larger passenger volume in the evening rush hour, and the peak values of weekends are relatively small to weekdays. The stations are mainly located in the work district, and many people working in this area usually take the subway and return home after going off work.
- (g) The seventh departure pattern has an extremely large passenger volume in the evening rush hour, and these stations are situated in the middle of the central business district (CBD) from Figure 7.

TABLE 1: Comparison of BNPTD with other baselines in the random missing scenario (the numbers 10, 20, 30 after the two methods denote CP rank and the best models are highlighted in bold).

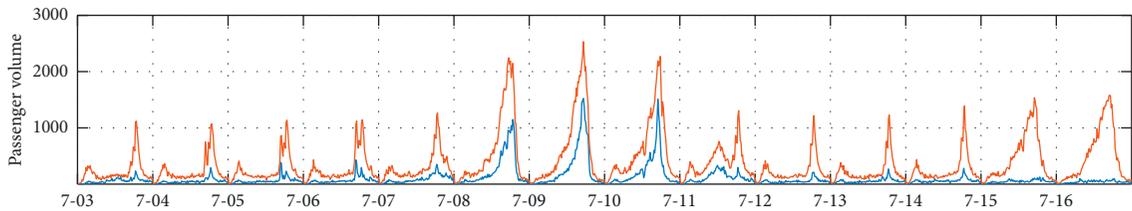
Method	10%		20%		30%		40%		50%	
	RMSE	MAE								
BTD (10)	57.37	34.04	56.58	33.61	62.35	36.20	65.23	37.64	65.51	37.83
BNPTD (10)	<b>56.35</b>	<b>33.51</b>	<b>56.25</b>	<b>33.45</b>	<b>58.97</b>	<b>34.11</b>	<b>62.94</b>	<b>36.54</b>	<b>65.41</b>	<b>37.79</b>
GAIN	56.91	34.77	56.29	34.46	62.22	37.29	65.24	38.91	65.66	39.27
BTD (20)	<b>46.52</b>	28.26	47.93	28.64	48.86	28.96	53.66	30.74	<b>50.01</b>	<b>29.42</b>
BNPTD (20)	46.61	<b>28.23</b>	<b>47.65</b>	<b>28.56</b>	<b>48.11</b>	<b>28.61</b>	<b>48.97</b>	<b>29.02</b>	50.13	29.43
GAIN	46.74	29.24	48.26	29.71	49.31	30.18	54.15	32.04	50.51	30.76
BTD (30)	43.79	26.59	44.96	26.90	45.79	27.21	46.80	27.66	49.21	28.77
BNPTD (30)	<b>43.46</b>	<b>26.5</b>	<b>44.7</b>	<b>26.86</b>	<b>45.27</b>	<b>27.09</b>	<b>45.96</b>	<b>27.48</b>	<b>48.43</b>	<b>28.25</b>
GAIN	44.19	27.66	45.46	28.02	46.39	28.44	47.41	28.92	49.92	30.20
DA	74.29	44.07	72.57	43.00	79.23	45.80	81.81	46.93	80.82	46.33

TABLE 2: Comparison of BNPTD with other baselines in the continuous missing scenario (the numbers 10, 20, 30 after the two methods denote CP rank and the best models are highlighted in bold).

Method	10%		20%		30%		40%		50%	
	RMSE	MAE								
BTD (10)	60.92	36.73	59.67	35.76	65.79	38.38	69.10	39.76	69.71	40.03
BNPTD (10)	60.19	<b>35.66</b>	59.10	<b>35.07</b>	65.01	<b>37.70</b>	<b>66.76</b>	<b>38.18</b>	<b>66.08</b>	<b>37.56</b>
GAIN	<b>59.29</b>	36.88	<b>58.33</b>	36.08	<b>64.59</b>	38.94	68.02	40.76	68.97	41.36
BTD (20)	49.85	30.70	51.29	30.74	52.22	30.97	57.62	32.74	53.77	31.35
BNPTD (20)	49.26	<b>29.81</b>	50.80	<b>30.14</b>	51.60	<b>30.41</b>	<b>55.67</b>	<b>31.44</b>	<b>50.97</b>	<b>29.42</b>
GAIN	<b>48.69</b>	31.01	<b>50.04</b>	31.11	<b>51.18</b>	31.52	56.46	33.56	53.06	32.39
BTD (30)	49.29	30.60	50.70	30.87	52.08	31.35	53.78	31.94	58.05	34.02
BNPTD (30)	<b>48.71</b>	<b>29.71</b>	<b>50.21</b>	<b>30.27</b>	<b>51.46</b>	<b>30.79</b>	<b>51.96</b>	<b>30.67</b>	<b>5.02</b>	<b>31.92</b>
GAIN	48.77	31.31	50.42	31.71	52.01	32.48	54.22	33.73	58.41	36.00
DA	73.36	43.89	71.81	42.61	78.36	45.15	80.93	46.12	80.67	45.95



(a)



(b)

FIGURE 6: Continued.

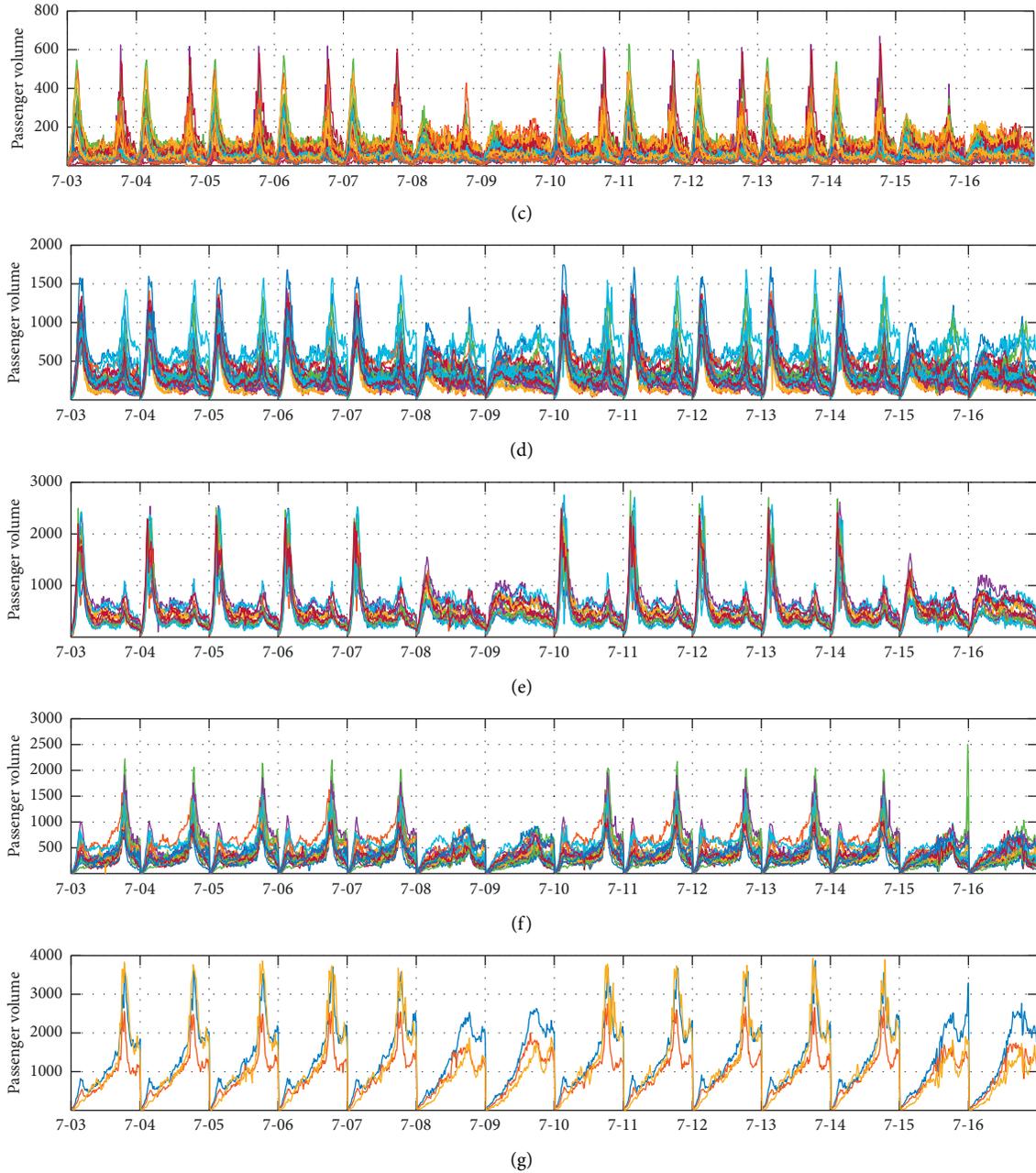


FIGURE 6: The result of similarity pattern discovery. (a) Cluster 1. (b) Cluster 2. (c) Cluster 3. (d) Cluster 4. (e) Cluster 5. (f) Cluster 6. (g) Cluster 7.

**4.3. The Clustering Stability with Different Missing Ratios.** In this subsection, an evaluation measure called purity is adopted to show the clustering stability of BNPTD with different missing ratios. Purity is an external index that is used to measure the similarity of the formed clusters to external clusters such as ground truth [22], considering  $G = \{G_1, G_2, \dots, G_m\}$  as external clusters and  $C = \{C_1, C_2, \dots, C_n\}$  are the clusters made by the BNPTD. To compute purity, each cluster of  $C$  is assigned to the class which is most frequent in the cluster, and the accuracy of this assignment is measured by counting the

number of correctly assigned stations and dividing by the number of stations  $I$ ; that is,

$$\text{purity}(G, C) = \frac{1}{I} \sum_{k=1}^n \max_m |C_k \cap G_m|. \quad (12)$$

It is easy to see that  $0 \leq \text{purity}(G, C) \leq 1$ , and there is a perfect clustering if  $\text{purity}(G, C)$  is equal to 1. We set the clustering result in missing ratio 10% as external clusters and implement experiments from missing ratios 15% to 50% with step size 5%. Figure 8 shows the purity with different

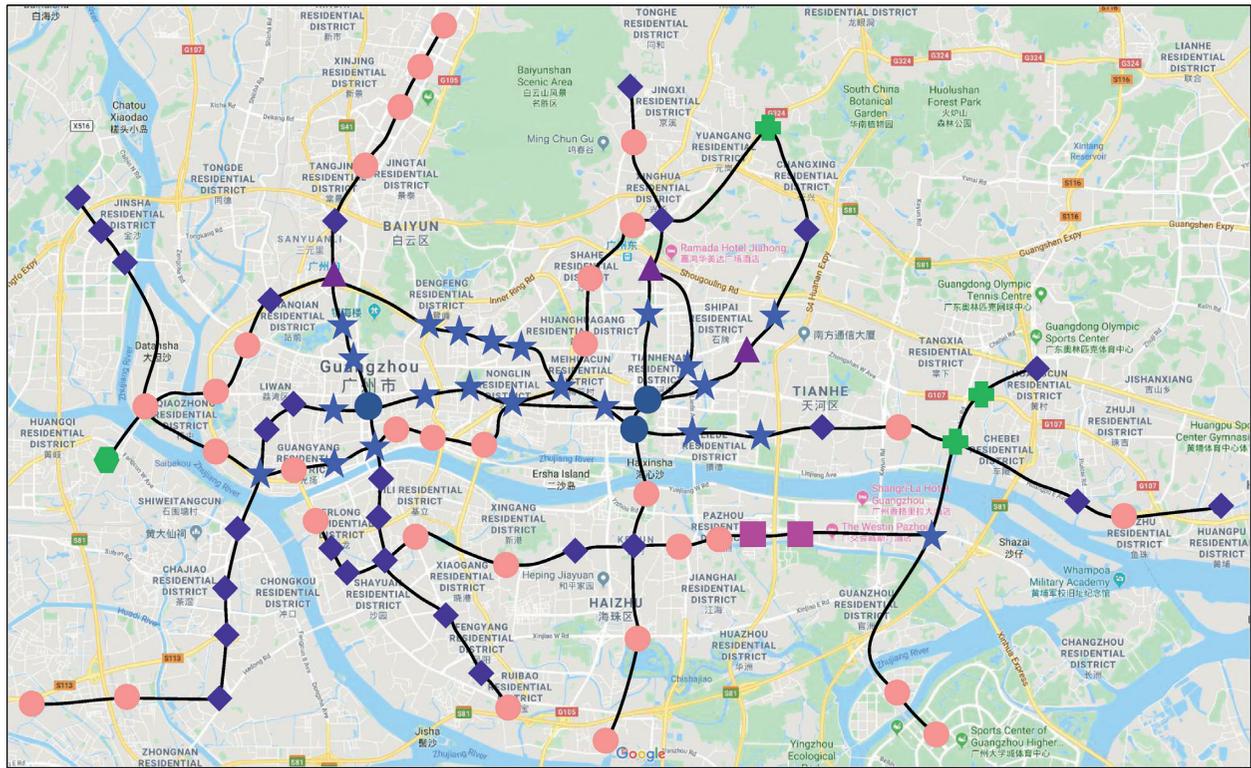


FIGURE 7: The subway stations distribution in Google Maps.

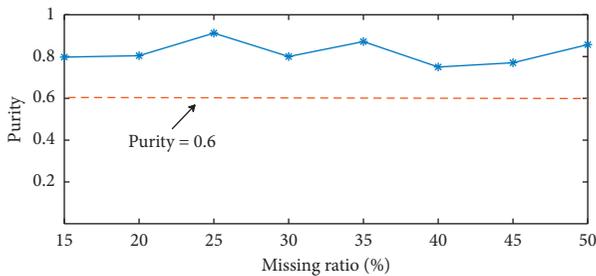


FIGURE 8: The purity with different missing ratios.

missing ratios. We observe that purity is greater than 0.6 in all cases, which indicates the clustering result is stability.

### 5. Conclusions

In this paper, a novel Bayesian nonparametric tensor decomposition model called BNPTD is proposed. The model combines the Bayesian tensor decomposition and Dirichlet process mixture model via a hierarchical probabilistic model, which can achieve incomplete traffic data imputation and similarity pattern discovery simultaneously. Moreover, we derive a variational inference algorithm to learn the

model efficiently. Experiments on a real-world smart card dataset show the effectiveness of the proposed model.

Actually, the similarity pattern extracted in this paper not only helps us understand the urban mobility patterns, but also can improve the traffic prediction as prior knowledge. For future work, we plan to develop multitask prediction based on the above similarities.

### Data Availability

The data used to support the results of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This research was supported by the project of National Natural Science Foundation of China (no. U1811463) and the Natural Science Foundation of Guangdong 292 Province, China (no. 20187616042030004).

## Supplementary Materials

The deduction process of optimal variational posterior distributions with regard to parameters and hyperparameters can be found in Supplementary File. (*Supplementary Materials*)

## References

- [1] 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. 13774–13779, 2013.
- [2] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3135–3146, 2017.
- [3] H. Yao et al., "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.
- [4] L. Bai et al., "Spatio-temporal graph convolutional and recurrent networks for citywide passenger demand prediction," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, November 2019.
- [5] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.
- [6] A. Ladino, A. Y. Kibangou, C. Canudas de Wit, and H. Fourati, "A real time forecasting tool for dynamic travel time from clustered time series," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 216–238, 2017.
- [7] F. Rodrigues, K. Henrickson, and F. C. Pereira, "Multi-output Gaussian processes for crowd sourced traffic data imputation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 594–603, 2018.
- [8] J. Yoon, J. James, and M. Van Der Schaar, "GAIN: missing data imputation using generative adversarial nets," 2018, <https://arxiv.org/abs/1806.02920>.
- [9] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 73–84, 2019.
- [10] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian CP factorization of incomplete tensors with automatic rank determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1751–1763, 2015.
- [11] Y. Li, L. Li, and Z. Li, "Missing traffic data: comparison of imputation methods," *IET Intelligent Transport Systems*, vol. 8, no. 1, pp. 51–57, 2014.
- [12] J. Yoon, J. James, and M. Van Der Schaar, "Missing data imputation using generative adversarial nets," 2018, <http://arxiv.org/abs/1806.02920>.
- [13] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [14] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013.
- [15] H. Zhang, P. Chen, J. Zheng et al., "Missing data detection and imputation for urban ANPR system using an iterative tensor decomposition approach," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 337–355, 2019.
- [16] Z. Gan, M. Yang, T. Feng, and H. Timmermans, "Understanding urban mobility patterns from a spatiotemporal perspective: daily ridership profiles of metro stations," *Transportation*, vol. 47, no. 1, pp. 315–336, 2018.
- [17] S. J. Gershman and D. M. Blei, "A tutorial on Bayesian nonparametric models," *Journal of Mathematical Psychology*, vol. 56, no. 1, pp. 1–12, 2012.
- [18] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Analysis*, vol. 1, no. 1, pp. 121–143, 2006.
- [19] H. Y. T. Ngan, H. C. Y. Nelson, and A. G. O. Yeh, "Modeling of traffic data characteristics by dirichlet process mixtures," in *Proceedings of the 2012 IEEE International Conference on Automation Science and Engineering (CASE)*, IEEE, Seoul, Korea/IEEE, Seoul, Korea, August 2012.
- [20] H. Y. T. Ngan, A. G. O. Yeh, and N. H. C. Yung, "Outlier detection in traffic data based on the Dirichlet process mixture model," *IET Intelligent Transport Systems*, vol. 9, no. 7, pp. 773–781, 2015.
- [21] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: a review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.
- [22] S. Aghabozorgi, S. Seyed, and W. Teh Ying, "Time-series clustering—a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.