Transportation Research Part C 86 (2018) 59–77

Contents lists available at ScienceDirect

Transportation Research Part C

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

Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition

Xinyu Chen, Zhaocheng He*, Jiawei Wang

Guangdong Provincial Key Laboratory of Intelligent Transportation Systems, Research Center of Intelligent Transportation System, Sun Yat-Sen University, Guangzhou 510006, China

ARTICLE INFO

Keywords:
Intelligent transportation systems
Tensor decomposition
Traffic speed data
Traffic patterns discovery
Incomplete data recovery
Data quality enhancement

ABSTRACT

Missing data is an inevitable and ubiquitous problem in data-driven intelligent transportation systems. While there are several studies on the missing traffic data recovery in the last decade, it is still an open issue of making full use of spatial-temporal traffic patterns to improve recovery performance. In this paper, due to the multi-dimensional nature of traffic speed data, we treat missing data recovery as the problem of tensor completion, a three-procedure framework based on Tucker decomposition is proposed to accomplish the recovery task by discovering spatial-temporal patterns and underlying structure from incomplete data. Specifically, in the missing data initialization, intrinsic multi-mode biases based traffic pattern is extracted to perform a robust recovery. Thereby, the truncated singular value decomposition (SVD) is introduced to capture main latent features along each dimension. Finally, applying these latent features, the missing data is eventually estimated by the SVD-combined tensor decomposition (STD). Empirically, relying on the large-scale traffic speed data collected from 214 road segments within two months at 10-min interval, our experiment covers two missing scenarios – element-like random missing and fiber-like random missing. The impacts of different initialization strategies for tensor decomposition are evaluated. From numerical analysis, a sensitivity-driven rank selection can not only choose an appropriate core tensor size but also determine how much features we actually need. By comparison with two baseline tensor decomposition models, our method is shown to successfully recover missing data with the highest accuracy as the missing rate ranges from 20% to 80% under two missing scenarios. Moreover, the results have also indicated that an optimal initialization for tensor decomposition could suggest a better performance.

1. Introduction

1.1. Motivation

With advancements in multiple surveillance equipment and sensor technologies, data-driven intelligent transportation systems can acquire traffic data with high spatial and temporal resolution (Asif et al., 2013; Li et al., 2013). However, when traffic data is collected from real transportation systems, the problem of missing data caused by communication malfunctions and transmission distortions is inevitable. On the one hand, without any improvement, it may further cause the problem of poor data quality and great holes in the database (Soriguera, 2016). Thus, data quality enhancement especially for the missing data is crucial to support the intelligent transportation systems. On the other hand, traffic parameters such as speed, volume and travel time usually exhibit

* Corresponding author.
E-mail address: hezhch@mail.sysu.edu.cn (Z. He).

https://doi.org/10.1016/j.trc.2017.10.023
Received 29 June 2017; Received in revised form 24 October 2017; Accepted 27 October 2017
Available online 10 November 2017
0968-090X/

intrinsic spatial-temporal patterns and are in essence multi-dimensional. It is still an open issue of making full use of spatial-temporal traffic patterns to improve recovery performance.

The goal of this study is to develop and implement a tensor decomposition method for discovering traffic patterns from partially observed data and further use them to estimate the missing data accurately. The framework is designed with following considerations. Firstly, the real-world multi-relational traffic data involves both spatial and temporal resolution. In relation to the matrix models, the tensor describes a richer algebraic structure and can thereby encode more information (Anandkumar et al., 2014), so that it is important to structure the original traffic data into a tensor. Secondly, the tensor decomposition can unravel hidden and structured patterns without external supervision. It has been proved that this kind of method requires only a small amount of samples to work well (Anandkumar et al., 2014; Wang et al., 2014). Thus, for missing traffic data, especially, the common scenario where some road segments have lost their speed observations during a couple of days, the tensor decomposition is a suitable method.

1.2. Related works

Tensor decomposition can be originated with the contribution of Hitchcock (1927a,b), and it was not until the work of Tucker (1966) that the tensor methods became more practical for data analysis. Nowadays, there are mainly two classical formulations of tensor decomposition, the first is Tucker decomposition (Tucker, 1966) which decomposes a given tensor into a core tensor and factor matrices in a sequence. The second is CP (CANDECOMP/PARAFAC) decomposition (Carroll and Chang, 1970). Both of two formulations can be regarded as higher-order generalizations of the singular value decomposition (SVD) (Kolda and Bader, 2009). Intuitively, the tensor decomposition is a multi-linear structure compared to the matrix decomposition, and the multiple facets of the data are taken into account (Schifanella et al., 2014).

Since tensor models can preserve the multi-dimensional nature of the data, we can thus conveniently extract the latent factors along each dimension and exploit complex underlying interactions among different dimensions of a higher-order array (Acar et al., 2011). The objective of tensor decomposition for incomplete data is to capture the underlying multi-linear factors from only partially observed entries, which can in turn estimate the missing entries (Zhao et al., 2015). As experimented in the recent literature, growing evidence shows that tensor decomposition techniques can significantly contribute to recovering the incomplete traffic data. Asif et al. (2016) and Goulart et al. (2017) evaluated various matrix and tensor based methods to estimate missing traffic speed data. Tan et al. (2013) proposed a Tucker decomposition based imputation algorithm (TDI) to impute missing traffic volume data, and low-rank tensor completion methods have also been tested in their studies (Tan et al., 2013, 2014; Ran et al., 2016). To estimate the travel time of any path in real time with problem of data sparsity, Wang et al. (2014) modeled different drivers' travel times on different road segments in different time slots with a third-order tensor and utilized a context-aware tensor decomposition approach to estimate the tensor's missing values.

Considering the high-dimensional spatial and temporal traffic characteristics, tensor decomposition techniques are also useful for solving plenty of problems in the field of traffic data mining, such as traffic prediction (Han and Moutarde, 2016; Tan et al., 2016; Li et al., 2016), data compression (Asif et al., 2013, 2015), and even human mobility analysis (Sun and Axhausen, 2016). Han and Moutarde (2016) applied non-negative tensor factorization model to achieve long-term prediction of the large-scale traffic evolution by simulated traffic network data. Based on dynamic tensor completion (DTC), Tan et al. (2016) designed a short-time traffic flow prediction approach by converting future traffic prediction into a tensor completion problem, and thereby, the DTC made an effective use of multi-mode periodicity (such as daily and weekly periodicity) and spatial information. Asif et al. (2013, 2015) applied the tensor decomposition technique to find low-dimensional structures of large-scale traffic networks and harness them to compress original data. Dealing with high-dimensional human mobility data, Sun and Axhausen (2016) utilized a probabilistic tensor decomposition framework to identify human mobility patterns and reveal the underlying spatial-temporal structure from city-wide daily trips.

Other than the tensor decomposition models, there also exist some methods such as PCA-based method (Qu et al., 2009; Li et al., 2013, 2014; Asif et al., 2016; Goulart et al., 2017), matrix decomposition (Asif et al., 2016), HaLRTC (Liu et al., 2013; Ran et al., 2016; Goulart et al., 2017) and deep learning (Duan et al., 2016) that have been used to handle the problem of missing traffic data. Specifically, Li et al. (2014) categorized earlier missing data imputation methods into prediction, interpolation and statistical learning, in which statistical learning methods have been illustrated to perform more effective than the other two kinds of imputation methods, and notably, probabilistic principal component analysis (PPCA, also examined by earlier Qu et al., 2009; Li et al., 2013) has achieved the best performance. However, expressing traffic data as a matrix will inevitably damage its multi-linear structure with only two factors considered. Furthermore, Asif et al. (2016) and Goulart et al. (2017) applied various matrix and tensor decomposition methods for incomplete traffic speed data recovery, and where the accuracy of tensor decomposition methods is considerably higher than matrix based methods.

1.3. Challenges and contributions

Though the incomplete data recovery is a hot topic especially for traffic data quality enhancement, the problem is still open and has not been solved yet given the following challenges. The first is what traffic patterns are hidden in the data, and how to leverage them to better missing data recovery. For instance, basic road similarity (Wang et al., 2014) and daily periodicity (Tan et al., 2013; Ran et al., 2016) are significant in urban road network, and the traffic condition of a road segment is highly related to the road segment and time period. Despite of the fact that the estimation of missing data is more important than the interpretability of models, the second is how to improve the interpretability of the proposed models such as implications of latent factors in tensor

decomposition models. While tensor decomposition techniques have been demonstrated to be effective in previous studies, the different initialization strategies for tensor decomposition which may lead to different performance have not been well detailed and tested.

As most existing methods intend to bridge the gap between practical problems and tensor models, few studies described what latent traffic patterns have been discovered by tensor models. Furthermore, the core tensor size was always required to be specified manually in almost all the numerical studies, and it is not relevant to the total features. Hence, the problem of how much features we can discover and need should be discussed. In addition, within the previous studies, there were plenty of experiments concentrated on the random missing recovery (Tan et al., 2013, 2014; Asif et al., 2016; Ran et al., 2016), while others focused on the prediction problem (Tan et al., 2016) where missing scenarios differ from the completion issue. However, the structured missing (Acar et al., 2011) have not yet been comprehensively analyzed in the incomplete traffic data recovery.

To address such concerns, a three-procedure framework is proposed for dealing with incomplete traffic speed data. This method is able to discover the traffic patterns from partially observed data and further recover the missing values. Herein, the SVD-combined tensor decomposition (STD) is a mixture and improved model, and it has incorporated multi-mode biases based traffic pattern and latent features captured by the truncated SVD. In the framework, with the ratio threshold of singular values of tensor unfolding, the core tensor size is determined by a sensitivity test over final results. Through adopting different strategies for the tensor decomposition, their impacts on the final recovery performance are investigated. In order to simulate practical missing scenarios, the element-like and fiber-like random missing are also taken into account and implemented in the numerical studies with the missing raising from 20% to 80%.

This paper is organized as follows: The modeling framework is provided in Section 2. Empirical studies are conducted in Section 3. And the concluding remarks are given in Section 4.

2. Modeling framework

In this section, we present a three-procedure framework for dealing with missing data as shown in Fig. 1. The basic idea is to first exploit the multi-mode biases based traffic pattern and simply estimate the missing entries of the incomplete tensor; then capture the latent features, where these features are regarded as an optimal initialization for the tensor decomposition; and finally recover the missing data by training Tucker decomposition.

In the phase of missing data initialization, by constructing the space $\mathbb{R}^{n_1\times n_2\times n_3}$ and structuring the observed speed data into a tensor $\mathcal{X}$ at the outset, we can roughly estimate the missing entries of incomplete tensor through a multi-mode biases based approach in advance. Then, in the phase of truncated SVD, unfolding the tensor $\widetilde{\mathcal{X}}$, we tend to capture main features along each dimension by the leading left singular vectors of tensor unfolding. The truncated SVD could not only capture main features along three dimensions, but also remove noises hidden in the data. After that, with $U^{(SVD)}$,$V^{(SVD)}$,$W^{(SVD)}$,$\mathcal{G}^{(SVD)}$ and low rank-(r_1,r_2,r_3), the STD will provide a

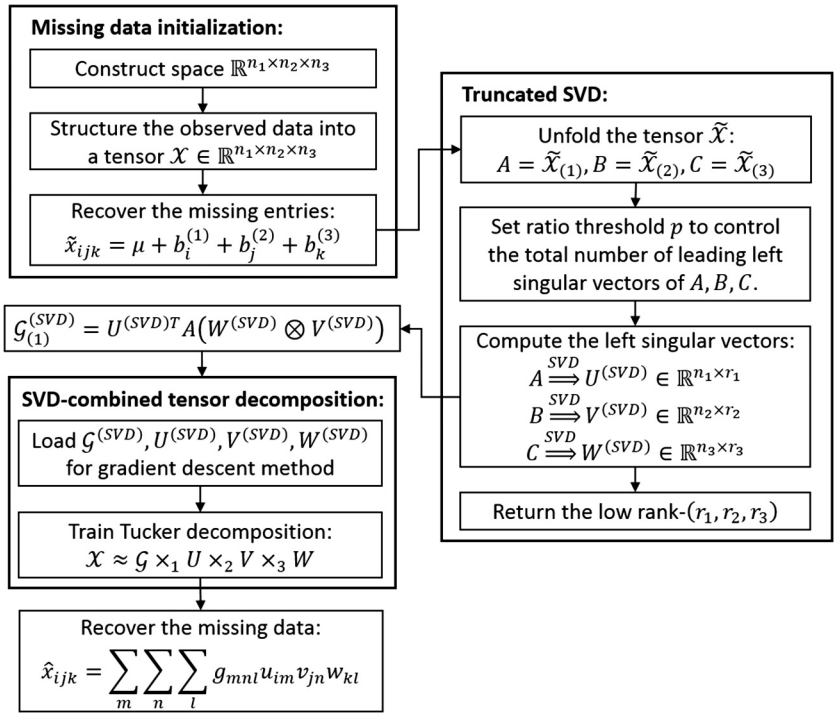


Fig. 1. Three-procedure (missing data initialization, truncated SVD and STD) framework for dealing with missing data.

promising solution for the missing data eventually.

2.1. Tensor construction and unfoldings

The fundamental traffic parameters such as speed can be observed and collected by various sensors. However, large-scale traffic data especially with missing values should be preprocessed before analysis. In this study, we focus on incomplete traffic speed data recovery task. For the reason that traffic speed involves road segment, day and time window, it makes the recovery task a high-dimensional problem. Thus, if $\mathscr{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes a tensor where n_1 represents the number of road segments, n_2 represents the number of days, n_3 represents the number of time windows and its entries represent speed, the ultimate goal is to discover spatial-temporal traffic patterns and further harness them to estimate the missing entries of the incomplete tensor accurately.

Tensor unfolding has an important role in the tensor computations. For instance, hidden structures within a tensor can sometimes be revealed by discovering patterns within its unfoldings. Given $\mathscr{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the mode-$q$ unfolding $\mathscr{X}_{(q)}, q = 1,2,3$ of the tensor is a combination of its slices along the qth mode (Lathauwe et al., 2000; Golub and Van Loan, 2013). Specifically, in block form, $\mathscr{X}_{(1)}$ is a $n_1 \times (n_2 n_3)$ matrix arranged by frontal slices $\mathscr{X}(:,:,1),\ldots,\mathscr{X}(:,:,n_3)$, $\mathscr{X}_{(2)}$ with size $n_2 \times (n_1 n_3)$ is arranged by the transpose of frontal slices, and $\mathscr{X}_{(3)}$ with size $n_3 \times (n_1 n_2)$ consists of the transpose of lateral slices $\mathscr{X}(:,1,:),\ldots,\mathscr{X}(:,n_2,:)$. Formally, these unfoldings are written as

$$\mathscr{X}_{(1)} = [\mathscr{X}(:,:,1),\cdots,\mathscr{X}(:,:,n_3)] \in \mathbb{R}^{n_1 \times (n_2 n_3)},$$

$$\mathscr{X}_{(2)} = [\mathscr{X}(:,:,1)^T,\cdots,\mathscr{X}(:,:,n_3)^T] \in \mathbb{R}^{n_2 \times (n_1 n_3)},$$

$$\mathscr{X}_{(3)} = [\mathscr{X}(:,1,:)^T,\cdots,\mathscr{X}(:,n_2,:)^T] \in \mathbb{R}^{n_3 \times (n_1 n_2)}.$$

2.2. Missing data initialization

In urban road network, it is a truth that the real-world traffic speed is highly related to the corresponding road segment (busy or not), day (a weekday or a weekend) and time window (at peak hours or off-peak hours). Traffic speed along three dimensions may have biases which are usually modeled in recommender systems (Paterek, 2007; Koren et al., 2009) over average speed. Regard bias of ith road segment as an entry $b_i^{(1)}$ of vector $\boldsymbol{b}^{(1)}$, jth day as an entry $b_j^{(2)}$ of vector $\boldsymbol{b}^{(2)}$, and kth time window as an entry $b_k^{(3)}$ of vector $\boldsymbol{b}^{(3)}$, mathematically, the biases $b_i^{(1)}, b_j^{(2)}, b_k^{(3)}$ associated with x_{ijk} are formulated as

$$\hat{x}_{ijk} = \mu + b_i^{(1)} + b_j^{(2)} + b_k^{(3)}, \quad \forall\, i,j,k \tag{1}$$

where $\mu = \frac{1}{|\Omega|}\sum_{(i,j,k)\in\Omega} x_{ijk}$ is the average speed. Consider a non-constraint optimization problem as below:

$$\min J = \frac{1}{2}\sum_{(i,j,k)\in\Omega} (x_{ijk} - \hat{x}_{ijk})^2 + \frac{\eta}{2}(||\boldsymbol{b}^{(1)}||_F^2 + ||\boldsymbol{b}^{(2)}||_F^2 + ||\boldsymbol{b}^{(3)}||_F^2)$$

where $||\cdot||_F$ denotes the Frobenius norm, for instance, $||\mathscr{X}||_F^2$ is equal to the sum of squared entries of $\mathscr{X}$. Using the gradient descent method, this optimization can be solved by Algorithm 1.

Algorithm 1. Missing data initialization

1. **Input**: incomplete tensor $\mathscr{X}$
2. Initialize vectors $\boldsymbol{b}^{(1)} \in \mathbb{R}^{n_1}, \boldsymbol{b}^{(2)} \in \mathbb{R}^{n_2}, \boldsymbol{b}^{(3)} \in \mathbb{R}^{n_3}$ with uniformly distributed random numbers between 0 and 0.1
3. Set the learning rate θ and the regularization parameter η
4. Compute average speed $\mu = \frac{1}{|\Omega|}\sum_{(i,j,k)\in\Omega} x_{ijk}$ where $\Omega = \{(i,j,k) | x_{ijk} \text{ is observed}\}$
5. **while** not (convergence) **do**
6. **for** each $(i,j,k) \in \Omega$ **do**
7. **for** $i = \{1,\ldots,n_1\}$ **do**
8. $b_i^{(1)+} = (1-\theta\eta) b_i^{(1)} + \theta \sum_{j,k:(i,j,k)\in\Omega} (x_{ijk} - \hat{x}_{ijk})$
9. **for** $j = \{1,\ldots,n_2\}$ **do**
10. $b_j^{(2)+} = (1-\theta\eta) b_j^{(2)} + \theta \sum_{i,k:(i,j,k)\in\Omega} (x_{ijk} - \hat{x}_{ijk})$
11. **for** $k = \{1,\ldots,n_3\}$ **do**
12. $b_k^{(3)+} = (1-\theta\eta) b_k^{(3)} + \theta \sum_{i,j:(i,j,k)\in\Omega} (x_{ijk} - \hat{x}_{ijk})$
13. Update $b^{(1)} \Leftarrow b^{(1)+}$, $b^{(2)} \Leftarrow b^{(2)+}$, $b^{(3)} \Leftarrow b^{(3)+}$
14. **for** each $(i,j,k) \notin \Omega$ **do**
15. $\hat{x}_{ijk} = \mu + b_i^{(1)} + b_j^{(2)} + b_k^{(3)}$

So far, the complete tensor $\widetilde{\mathscr{X}}$ can be defined by

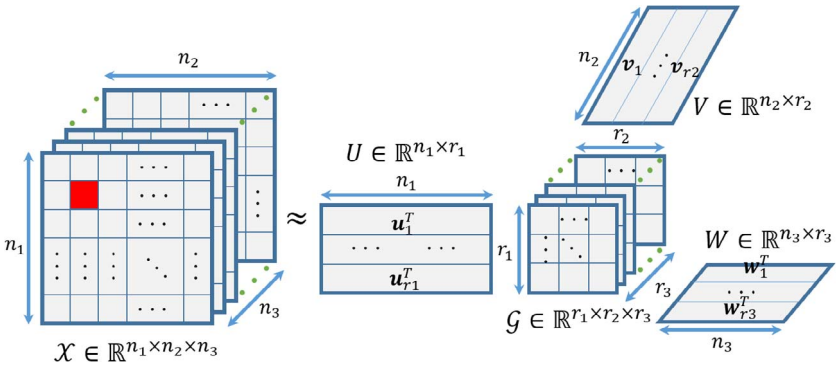


Fig. 2. Illustration of Tucker decomposition for third-order tensor $\mathcal{X}$ where the red square represents (2,2,1)th entry of $\mathcal{X}$ and can also be viewed as (2,2)th entry of slice $\mathcal{X}(:,:,1)$; $U = (\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{r1})$, $V = (\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{r2})$ and $W = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{r3})$ are factor matrices.

$$\widetilde{x}_{ijk} = \begin{cases} x_{ijk},(i,j,k) \in \Omega, \\ \hat{x}_{ijk},(i,j,k) \notin \Omega. \end{cases} \tag{2}$$

Although the training process of Eq. (1) may be efficient for dealing with missing data, the decision variables $b_i^{(1)},b_j^{(2)},b_k^{(3)}$ are independent from each other. Because Algorithm 1 is limited in describing complex interactions and dependence among different dimensions, we further discover traffic patterns towards the truncated SVD and tensor decomposition.

2.3. Truncated singular value decomposition (SVD)

Tensor decomposition, as its high-dimensional structure, implies more complex interactions among decomposed core tensor and factor matrices, some latent features can also be well extracted by the decomposition meanwhile (Kolda and Bader, 2009). As illustrated by Fig. 2, Tucker decomposition decomposes a given tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ into a core tensor $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and factor matrices $U \in \mathbb{R}^{n_1 \times r_1}$, $V \in \mathbb{R}^{n_2 \times r_2}$ and $W \in \mathbb{R}^{n_3 \times r_3}$ in a sequence (Tucker, 1966), such that

$$\mathcal{X} \approx \mathcal{G} \times_1 U \times_2 V \times_3 W, \tag{3}$$

where $\times_q,q = 1,2,3$ is mode-q product between tensor and matrix. The Tucker decomposition can be regarded as a generalization of the CP decomposition (Kolda and Bader, 2009; Acar et al., 2011). By comparison with CP decomposition, Tucker decomposition with its flexible structure can achieve lower modeling errors (Acar et al., 2011). More discussions about tensor decomposition and its application can be referred to Kolda and Bader (2009).

The tensor $\mathcal{X}$ is actually decomposed into a core tensor multiplied by factor matrices. The factor matrices explain features along each dimension and the entries of core tensor $\mathcal{G}$ characterize the degree of interaction among different dimensions (Sun and Axhausen, 2016). Typically, the determination of core tensor size is critical for Tucker decomposition and it always influences the final results (Chen et al., 2015). However, the rank-(r_1,r_2,r_3) was always considered as a set of pre-determined parameters in the previous studies. Here, we devise a sensitivity-driven rank selection with consideration of the ratio threshold of singular values. The leading left singular vectors and the largest singular values of each unfolding are obtained by the truncated SVD. If $\widetilde{\mathcal{X}_{(1)}} = U\Sigma^{(1)}M^T$, $\widetilde{\mathcal{X}_{(2)}} = V\Sigma^{(2)}F^T$ and $\widetilde{\mathcal{X}_{(3)}} = W\Sigma^{(3)}Q^T$ are the SVDs of the tensor unfolding, then the leading left singular vectors and the largest singular values of $\widetilde{\mathcal{X}_{(1)}}$, $\widetilde{\mathcal{X}_{(2)}}$ and $\widetilde{\mathcal{X}_{(3)}}$ can be obtained through the truncated SVD (Golub and Van Loan, 2013). In detail, Algorithm 2 gives a case of $A = \widetilde{\mathcal{X}_{(1)}}$.

Algorithm 2. Truncated SVD (case: $A = \widetilde{\mathcal{X}_{(1)}}$)

1. **Input**: mode-1 unfolding $A = \widetilde{\mathcal{X}_{(1)}}$, ratio threshold p
2. $A = U\Sigma^{(1)}M^T$
3. **while** $p_{r_1} > p,p_{r_1-1} \leqslant p$ **do**
4. $U^{(SVD)} = U_{r_1} \in \mathbb{R}^{n_1 \times r_1}$ consists of the r_1 leading left singular vectors
5. **if** $r_1 < 2$ **do**
6. $r_1 = r_1 + 1$, $U^{(SVD)} = U_{r_1} \in \mathbb{R}^{n_1 \times r_1}$
7. **Return** $U^{(SVD)},r_1$

In the Algorithm 2, it is important to note that p_{r_1} is the ratio of largest r_1 singular values, that is

$$p_{r_1} = \sum_{i=1}^{r_1} \sigma_i^{(1)} / \sum_i \sigma_i^{(1)},\Sigma_{r_1}^{(1)} = \mathrm{diag}(\sigma_1^{(1)},\cdots,\sigma_{r_1}^{(1)}). \tag{4}$$

Correspondingly, parameter p is a ratio threshold to condition rank-r_1. With same parameter p, we can derive $V^{(SVD)} \in \mathbb{R}^{n_2 \times r_2}$ and $W^{(SVD)} \in \mathbb{R}^{n_3 \times r_3}$ for $\widetilde{\mathscr{X}_{(2)}}$ and $\widetilde{\mathscr{X}_{(3)}}$, respectively. The mode-1 unfolding of core tensor $\mathscr{G}^{(SVD)}$ is

$$\mathscr{G}_{(1)}^{(SVD)} = U^{(SVD)T} A (W^{(SVD)} \otimes V^{(SVD)}), \tag{5}$$

where $\otimes$ represents Kronecker product.

2.4. SVD-combined tensor decomposition (STD)

In the presence of missing entries, Tucker decomposition can be generally regarded as a non-constraint optimization problem to minimize the squared errors and an additive regularization (Wang et al., 2014), which is expressed as follows:

$$\min L = \frac{1}{2} ||\mathscr{S} * (\mathscr{X} - \mathscr{G} \times_1 U \times_2 V \times_3 W)||_F^2 + \frac{\lambda}{2} (||\mathscr{G}||_F^2 + ||U||_F^2 + ||V||_F^2 + ||W||_F^2), \tag{6}$$

where $*$ denotes element-wise product, and λ is the parameter of regularization term.

For the (i,j,k)th entry of the tensor $\mathscr{X}$, Tucker decomposition in Eq. (3) is given by

$$x_{ijk} \approx \sum_{m=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl} u_{im} v_{jn} w_{kl}. \tag{7}$$

Correspondingly, Eq. (6) is rewritten as

$$\begin{aligned} &\min L = L_1 + L_2, \\ &L_1 = \frac{1}{2} \sum_{(i,j,k) \in \Omega} \left(x_{ijk} - \sum_{m=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl} u_{im} v_{jn} w_{kl} \right)^2, \\ &L_2 = \frac{\lambda}{2} \left(\sum_{m=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl}^2 + \sum_{i=1}^{n_1} \sum_{m=1}^{r_1} u_{im}^2 + \sum_{j=1}^{n_2} \sum_{n=1}^{r_2} v_{jn}^2 + \sum_{k=1}^{n_3} \sum_{l=1}^{r_3} w_{kl}^2 \right). \end{aligned} \tag{8}$$

Let $e_{ijk} = x_{ijk} - \sum_{m=1}^{r_1} \sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl} u_{im} v_{jn} w_{kl}$, then we can compute the partial derivative of L with respect to the decision variable u_{im}:

$$\frac{\partial L}{\partial u_{im}} = - \sum_{j,k:(i,j,k) \in \Omega} \left\{ e_{ijk} \left(\sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl} v_{jn} w_{kl} \right) \right\} + \lambda u_{im}$$

where j,k: $(i,j,k) \in \Omega$ is the index set of observed entries of $\mathscr{X}(i,:,:)$. Analogously, with respect to the decision variables v_{jn}, w_{kl} and g_{mnl}, we have

$$\frac{\partial L}{\partial v_{jn}} = - \sum_{i,k:(i,j,k) \in \Omega} \left\{ e_{ijk} \left(\sum_{m=1}^{r_1} \sum_{l=1}^{r_3} g_{mnl} u_{im} w_{kl} \right) \right\} + \lambda v_{jn}$$

$$\frac{\partial L}{\partial w_{kl}} = - \sum_{i,j:(i,j,k) \in \Omega} \left\{ e_{ijk} \left(\sum_{m=1}^{r_1} \sum_{n=1}^{r_2} g_{mnl} u_{im} v_{jn} \right) \right\} + \lambda w_{kl}$$

$$\frac{\partial L}{\partial g_{mnl}} = - \sum_{(i,j,k) \in \Omega} e_{ijk} u_{im} v_{jn} w_{kl} + \lambda g_{mnl}$$

where i,k: $(i,j,k) \in \Omega$ is the index set of observed entries of $\mathscr{X}(:,j,:)$ and i,j: $(i,j,k) \in \Omega$ is the index set of observed entries of $\mathscr{X}(:,:,k)$.

As an example, using the gradient descent method, the update of the decision variable u_{im} can be formulated as

$$u_{im} \Leftarrow (1 - \alpha\lambda) u_{im} + \alpha \sum_{j,k:(i,j,k) \in \Omega} \left\{ e_{ijk} \left(\sum_{n=1}^{r_2} \sum_{l=1}^{r_3} g_{mnl} v_{jn} w_{kl} \right) \right\}.$$

where α is the learning rate of gradient decent method.

Take $\mathscr{E} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ to be a residual tensor for the only observed e_{ijk},$(i,j,k) \in \Omega$, and further define a binary tensor $\mathscr{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ where $s_{ijk} = 1$,$(i,j,k) \in \Omega$ and $s_{ijk} = 0$,$(i,j,k) \notin \Omega$ otherwise. As such, in terms of tensor unfolding and Kronecker product, the updates of decision variable u_{im} for all $i \in \{1,...,n_1\}$, $m \in \{1,...,r_1\}$ are thus reformulated as

$$U \Leftarrow (1 - \alpha\lambda) U + \alpha (\mathscr{S} * \mathscr{E})_{(1)} (W \otimes V) \mathscr{G}_{(1)}^T.$$

Likewise, we have

$$V \Leftarrow (1 - \alpha\lambda) V + \alpha (\mathscr{S} * \mathscr{E})_{(2)} (W \otimes U) \mathscr{G}_{(2)}^T,$$

$$W \Leftarrow (1 - \alpha\lambda) W + \alpha (\mathscr{S} * \mathscr{E})_{(3)} (V \otimes U) \mathscr{G}_{(3)}^T,$$

$$\mathcal{G} \Leftarrow (1-\alpha\lambda)\mathcal{G} + \alpha\cdot\mathcal{E} \times_1 U^T \times_2 V^T \times_3 W^T.$$

These updates allow many calculations to happen simultaneously with the use of intermediate variables (see Algorithm 3). Note that, an optimal initialization for tensor decomposition can provide a better performance over the final results (Acar et al., 2011), in the step 3 of Algorithm 3, the core tensor $\mathcal{G}^{(SVD)}$ and factor matrices $U^{(SVD)}$, $V^{(SVD)}$ and $W^{(SVD)}$ decomposed by the tensor $\widetilde{\mathcal{X}}$ are treated as an optimal initialization.

Algorithm 3. SVD-combined tensor decomposition (STD)

1. **Input**: incomplete tensor $\mathcal{X}$, binary tensor $\mathcal{S}$
2. Set the learning rate α, the regularization parameter λ, and ε
3. $\mathcal{G} = \mathcal{G}^{(SVD)}, U = U^{(SVD)}, V = V^{(SVD)}, W = W^{(SVD)}$
4. $\mathcal{X}^{(0)} = \mathcal{G} \times_1 U \times_2 V \times_3 W$
5. $\mathcal{E} = \mathcal{S} * (\mathcal{X} - \mathcal{G} \times_1 U \times_2 V \times_3 W)$
6. $U^+ = (1-\alpha\lambda)U + \alpha(\mathcal{S} * \mathcal{E})_{(1)}(W \otimes V)\mathcal{G}_{(1)}^T$
7. $V^+ = (1-\alpha\lambda)V + \alpha(\mathcal{S} * \mathcal{E})_{(2)}(W \otimes U)\mathcal{G}_{(2)}^T$
8. $W^+ = (1-\alpha\lambda)W + \alpha(\mathcal{S} * \mathcal{E})_{(3)}(V \otimes U)\mathcal{G}_{(3)}^T$
9. $\mathcal{G}^+ = (1-\alpha\lambda)\mathcal{G} + \alpha\cdot\mathcal{E} \times_1 U^T \times_2 V^T \times_3 W^T$
10. Update $U \Leftarrow U^+$, $V \Leftarrow V^+$, $W \Leftarrow W^+$, $\mathcal{G} \Leftarrow \mathcal{G}^+$
11. $\mathcal{X}^{(1)} = \mathcal{G} \times_1 U \times_2 V \times_3 W$
12. Check the convergence condition, $\|\mathcal{X}^{(1)} - \mathcal{X}^{(0)}\|_F^2 < \varepsilon$
13. **while** not (convergence) **do**
14. $\mathcal{X}^{(0)} = \mathcal{X}^{(1)}$, execute step 5–12
15. **Return** $\mathcal{G}$, U, V, W, $\widehat{\mathcal{X}} = \mathcal{X}^{(1)}$

Instead, using statement "*Randomly initialize core tensor* $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$, *factor matrices* $U \in \mathbb{R}^{n_1 \times r_1}, V \in \mathbb{R}^{n_2 \times r_2}, W \in \mathbb{R}^{n_3 \times r_3}$" to replace the step 3 of Algorithm 3, we give a random initialization based tensor decomposition (RTD), where its other steps and core tensor size are same with the STD. In order to analyze the impacts of different initialization strategies for tensor decomposition, unlike the STD and RTD model, we consider another initialization strategy is that the factor matrices and core tensor for initialization are directly computed by sparse tensor $\mathcal{X}$. We thus call it HOSVD-initialized tensor decomposition (HTD, see Algorithm 4) for convenience.

Algorithm 4. HOSVD-initialized tensor decomposition (HTD)

1. **Input**: rank-(r_1, r_2, r_3) (returned by Algorithm 2)
   Incomplete tensor $\mathcal{X}$
   Binary tensor $\mathcal{S}$
2. Set the learning rate α, the regularization parameter λ, and ε
3. Compute factor matrices $U^{(SVD)}$, $V^{(SVD)}$ and $W^{(SVD)}$ via $\mathcal{X}$'s unfoldings
4. $\mathcal{G}^{(SVD)} = \mathcal{X} \times_1 U^{(SVD)T} \times_2 V^{(SVD)T} \times_3 W^{(SVD)T}$
5. Execute step 3–15 of Algorithm 3

A distinct difference between the STD and HTD is that the core tensor and factor matrices of Algorithm 4 are calculated by the incomplete tensor $\mathcal{X}$ rather than $\widetilde{\mathcal{X}}$. Therefore, the STD allows us to grapple with the multi-mode biases based traffic pattern. In addition, although the TDI (i.e., Tucker decomposition based imputation) algorithm employed by Tan et al. (2013) is very similar to the HTD, there are two differences as follows. Firstly, the HTD has considered the additive regularization term in the objective function (i.e., Eq. (6)) as the TDI does not have. Secondly, to make reasonable comparisons, the core tensor size is same with the STD.

3. Empirical studies

In this section, relying on a network traffic speed dataset, we experimentally investigate the performance of the STD, RTD and HTD models under the scenarios of both element-like missing and fiber-like missing.

3.1. Data, missing scenarios and performance measures

In the whole empirical studies, we select the traffic speed data released by the Communications Commission of Guangzhou Municipality, China as an experiment dataset (is available at: http://www.openits.cn/openData2/792.jhtml), and structure the speed data into an original tensor denoted by $\mathcal{X}$. The dataset contains 214 road segments (mainly consist of urban expressways and

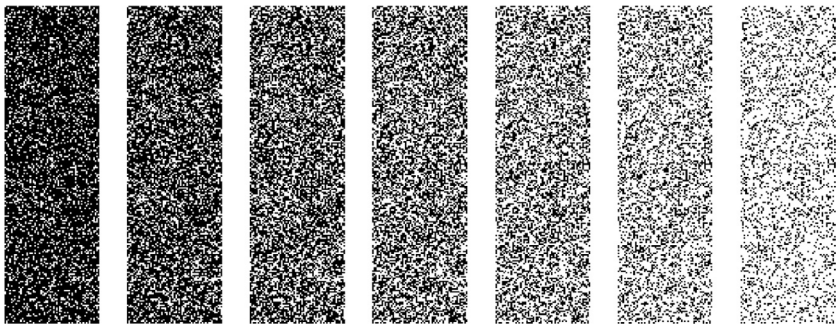


Fig. 3. The random binary matrix S, from left to right, the missing rate is increased from 20% to 80% (step is 10%) where the white squares indicate zeros, note that each row and column of the matrix S under different missing rates have at least one non-zero entry.

arterials) in urban areas from August 1, 2016 to September 30, 2016 (i.e., 61 days). During each day, the speed data of each road segment is aggregated with 10-min time window, in such case, there are 144 speed values everyday if observations are complete. The missing rate of original tensor $\mathscr{X} \in \mathbb{R}^{214\times61\times144}$ is 1.29% and its average is 39.01 km/h. The set $\Omega_1 = \{(i,j,k)|x_{ijk} \text{ is observed}\}$ is defined to represent the indices of observed entries.

In the previous studies, the common missing scenario is to remove some entries of the original tensor at random. Let $\Omega = \{(i,j,k)|s_{ijk}\cdot x_{ijk} \neq 0\}$,$\Omega \subset \Omega_1$ and $\mathscr{S}$ be a tensor with random entries 0 and 1, a new sparse tensor $\mathscr{X} \Leftarrow \mathscr{S}*\mathscr{X}$ can be obtained. However, this scenario (i.e., element-like missing) may not be consistent with the real-world missing, for example, the speed data of one specific road segment during one day may be completely lost. Thus, the experiment takes into account a special fiber-like random missing, that is, given a matrix $S \in \mathbb{R}^{n_1\times n_2}$ with random entries 0 and 1, we can obtain tensor $\mathscr{X} \Leftarrow \mathscr{S}*\mathscr{X}$ where $\mathscr{S}(i,j,k) = S(i,j)$ for all $i \in \{1,\cdots,n_1\}$, $j \in \{1,\cdots,n_2\}$ and $k \in \{1,\cdots,n_3\}$.

It should be emphasized that the missing rate is ranged from 20% to 80%. In the scenario of element-like random missing, first, tensor $\mathscr{A} \in \mathbb{R}^{214\times61\times144}$ with uniformly distributed random numbers between 0 and 1 is generated. Second, suppose a missing rate such as $\psi = 0.70$, $\mathscr{S} = round\{\mathscr{A}-0.20\}$ is filled with 0 and 1, where $round\{\mathscr{A}-0.20\}$ maps the entries to the nearest integer. Thereby, the missing rate of tensor $\mathscr{X}$ can be prescribed by $\mathscr{X} \Leftarrow \mathscr{S}*\mathscr{X}$. Alternatively, the matrix $A \in \mathbb{R}^{214\times61}$ with uniformly distributed random numbers between 0 and 1 is generated to create the fiber-like random missing. Set the missing rate as $\psi = 0.20,0.30,\ldots,0.80$ respectively, Fig. 3 shows the random binary matrix $S = round\{A + 0.5-\psi\} \in \mathbb{R}^{214\times61}$.

To test the accuracy of missing data recovery, we adopt three criteria to measure the performance, including RMSE (root mean square error), MAE (mean absolute error) and MRE (mean relative error). If y_i and z_i are the ith estimated and actual value, respectively, then three criteria can be written as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-z_i)^2},$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i-z_i|,$$

$$\text{MRE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i-z_i|}{z_i}.$$

3.2. Exploiting multi-mode biases based traffic pattern

In the first procedure of the framework, the method of missing data initialization is rather simple and intuitive (see Algorithm 1). Related multi-mode biases based approach takes into account the truth that the real-world traffic speed is highly dependent on the specific road segment and time period (i.e., day and time window). In Fig. 4, Algorithm 1 is indicated to be applicable and effective. For either element-like or fiber-like missing, the overall RMSE is close to 6.00 km/h and the MAE stays around 4.35 km/h with the 20%–80% missing.

Tables 1and 2 present the mean and standard deviation of $\boldsymbol{b}^{(q)}$,$q = 1,2,3$ under different missing rates. To avoid confusion, the 21.00% missing is viewed as 20% missing, the 30.87% missing as the 30% missing, etc. However, in the tables and figures, we keep the original missing rates. Recall that $\boldsymbol{b}^{(1)}$ is the bias vector for road segments, $\boldsymbol{b}^{(2)}$ is the bias vector for days and $\boldsymbol{b}^{(3)}$ is the bias vector for time windows. While the missing rates ranging from 20% to 80%, there are little changes of mean and standard deviation. Among different rows of Tables 1 and 2, we observe that, as $mean\{\boldsymbol{b}^{(q)}\}$,$q = 1,2,3$ are all extremely close to 0, the standard deviation of $\boldsymbol{b}^{(q)}$,$q = 1,2,3$ are remarkably different from each other. More specifically, the differences among standard deviation of biases can be described from following two aspects.

From a viewpoint of the information theory (Manning et al., 2008), the larger variance implies more information hidden in the data. As such, the road segment (i.e., first dimension) and time window (i.e., third dimension) are the most influential factors for the

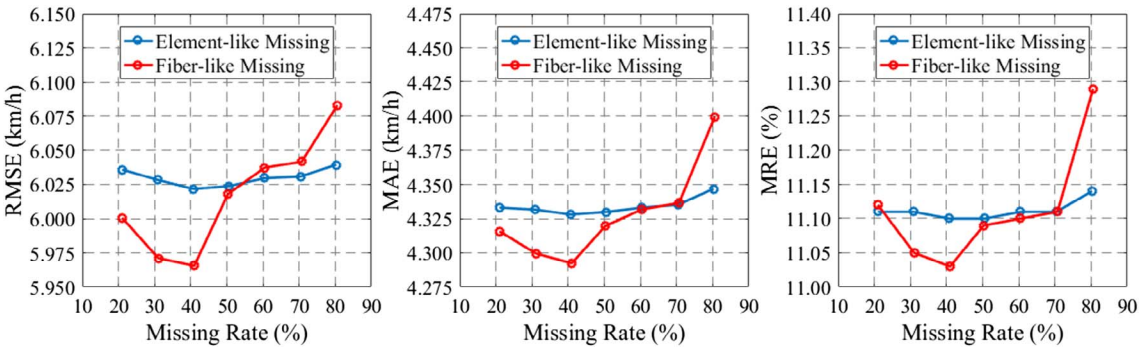


Fig. 4. Performance of missing data recovery using Algorithm 1 over missing rates ranging from 20% to 80%.

Table 1
The mean and standard deviation of three bias vectors with different missing rates under the element-like missing.

Statistics	21.00% missing	30.87% missing	40.76% missing	50.60% missing	60.47% missing	70.37% missing	80.24% missing
$mean\{\boldsymbol{b}^{(1)}\}$	0.0042	0.0080	0.0013	0.0096	0.0054	0.0087	0.0116
$mean\{\boldsymbol{b}^{(2)}\}$	−0.1130	−0.1118	−0.1091	−0.1107	−0.1068	−0.1003	−0.0924
$mean\{\boldsymbol{b}^{(3)}\}$	−0.0190	−0.0208	−0.0181	−0.0158	−0.0135	−0.0118	−0.0064
$std\{\boldsymbol{b}^{(1)}\}$	6.5948	6.5941	6.5920	6.5843	6.5781	6.5382	6.3542
$std\{\boldsymbol{b}^{(2)}\}$	1.5248	1.5286	1.5317	1.5391	1.5410	1.5403	1.5449
$std\{\boldsymbol{b}^{(3)}\}$	6.0336	6.0303	6.0315	6.0340	6.0323	6.0320	6.0025

Table 2
The mean and standard deviation of three bias vectors with different missing rates under the fiber-like missing.

Statistics	21.00% missing	31.03% missing	40.97% missing	50.30% missing	60.46% missing	70.71% missing	80.56% missing
$mean\{\boldsymbol{b}^{(1)}\}$	−0.0068	−0.0035	−0.0025	−0.0015	0.0059	0.0005	−0.0039
$mean\{\boldsymbol{b}^{(2)}\}$	−0.1342	−0.1358	−0.1387	−0.1397	−0.1102	−0.1220	−0.1550
$mean\{\boldsymbol{b}^{(3)}\}$	−0.0293	−0.0287	−0.0353	−0.0276	−0.0175	−0.0249	−0.0373
$std\{\boldsymbol{b}^{(1)}\}$	6.5899	6.5944	6.5771	6.5824	6.5611	6.4878	6.2132
$std\{\boldsymbol{b}^{(2)}\}$	1.5292	1.5355	1.5537	1.5587	1.5728	1.6003	1.6284
$std\{\boldsymbol{b}^{(3)}\}$	6.0340	6.0467	6.0720	6.0615	6.0592	6.0686	6.0521

traffic speed. With the concept of similarity (Wang et al., 2014), the large standard deviation of a bias vector means less similarities along corresponding dimension. Let vectors $\boldsymbol{b}^{(2)} \in \mathbb{R}^{61}$ and $\boldsymbol{b}^{(3)} \in \mathbb{R}^{144}$ by example, $std\{\boldsymbol{b}^{(2)}\} < std\{\boldsymbol{b}^{(3)}\}$ demonstrates that the speed differences along second dimension are smaller than differences along third dimension. Though, the differences among different days are not remarkable as differences among different road segments or among different time windows, the day dimension is essential for converting the time-series data into a tensor. Since the week information is always categorized by daily similarity, it is unnecessary to insert the week dimension and further structure original traffic speed data into a fourth-order tensor. Thus, the third-order tensor constructed in this study is rational.

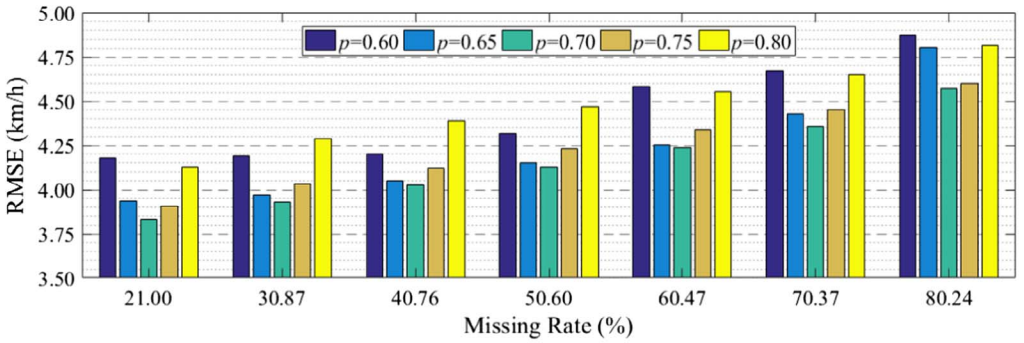
3.3. Sensitivity-driven rank selection

The ratio threshold p of singular values is critical for capturing the leading left singular vectors of tensor unfolding. Here, a smaller p may drop the main speed features hidden in the data, and a larger p will make no difference in data noise removal. Predefine a series of ratio thresholds ranging from 0.60 to 0.80 (step is 0.05), the RMSEs of the STD model under two missing scenarios are presented in Fig. 5. When the ratio threshold is $p = 0.70$ as shown in Fig. 5(a), for any missing rate ranging from 20% to 80%, the STD achieves the highest accuracy. Either a smaller or a larger ratio threshold tends to cause more errors for the final results. In Fig. 5(b), while missing rate ranges from 20% to 60%, the best ratio threshold is $p = 0.65$; however, for the 70% and 80% missing rates, the best ratio threshold is $p = 0.70$.

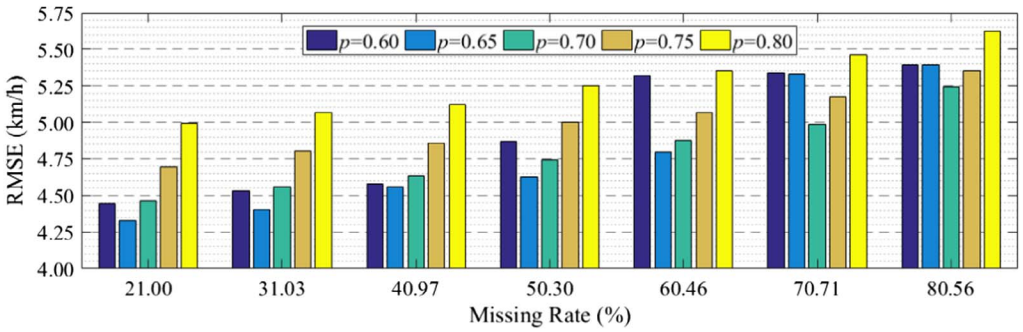
3.4. Overall performance among STD, RTD and HTD

3.4.1. Scenario of element-like random missing

The best ratio threshold $p = 0.70$ illustrated by Fig. 5(a) allows us to capture the main features hidden in the unfoldings $\widetilde{\mathcal{X}_{(1)}}$, $\widetilde{\mathcal{X}_{(2)}}$



(a) Scenario of element-like random missing



(b) Scenario of fiber-like random missing

Fig. 5. Performance of missing data recovery using the STD with different ratio thresholds.

and $\widetilde{\mathcal{X}_{(3)}}$. Recall that the tensor $\widetilde{\mathcal{X}}$ is filled with a combination of the average speed and additive biases. While the ratio threshold is larger than 0.70, hidden noises of $\widetilde{\mathcal{X}}$ may appear in the factor matrices $U^{(SVD)}$, $V^{(SVD)}$ and $W^{(SVD)}$. Adversely, with smaller ratio thresholds, part of the main features may be removed, and worse performance will be appeared unsurprisingly. As a result, the core tensor size with $p = 0.70$ is given by the last row of Table 3.

Place the same sized core tensor on the STD, RTD and HTD, their performance of missing data recovery is summarized in Table 3. Among three models, the STD yields the best performance, then HTD. Think of HTD in this study as the TDI (Tan et al., 2013), here, we also observe that the performance of the HTD is better than that of the RTD. In this vein, the initialization of STD captures more features and more accurate information than either HTD or RTD. Moreover, from Table 3, as the missing rate increases from 20% to 80%, the estimation errors become larger using the STD and HTD. This may be concluded as the fact that, using partially observed data to estimate missing values, the heavier missing implies less features can be captured or trained by tensor decomposition. In addition, back to Fig. 4, though the multi-mode biases based approach is rather simple, it performs slightly better than the RTD.

As exemplified by the 20%, 50% and 80% missing, in Fig. 6, we can explicitly see that the STD performs the best among three tensor decomposition models. In detail, some limitations of the RTD and HTD are shown as: (1) for the speed less than 20 km/h which is mainly caused by congestion, the RTD model is vulnerable; (2) for speed that is higher than 50km/h, the scatters of RTD and HTD are not close to the red line as the STD accomplished. No matter how low or high the speed is, the STD always performs better than other two models.

Table 3
Performance of missing data recovery using there models with scenario of element-like random missing.

Measures	Models	21.00% missing	30.87% missing	40.76% missing	50.60% missing	60.47% missing	70.37% missing	80.24% missing
RMSE (km/h)	STD	3.8308	3.9286	4.0265	4.1253	4.2355	4.3560	4.5708
	RTD	6.1167	6.1060	6.0985	6.1007	6.1049	6.1055	6.1092
	HTD	4.5966	4.6683	4.6875	4.7876	4.8897	5.4499	5.5282
MAE (km/h)	STD	2.5315	2.5823	2.6422	2.7063	2.7849	2.8764	3.0482
	RTD	4.3845	4.3821	4.3776	4.3794	4.3820	4.3828	4.3865
	HTD	3.1443	3.2024	3.2312	3.3053	3.3660	3.7723	3.8199
MRE (%)	STD	6.49	6.61	6.76	6.92	7.13	7.36	7.81
	RTD	11.24	11.23	11.22	11.22	11.23	11.23	11.24
	HTD	8.15	8.29	8.35	8.53	8.68	9.74	9.82
Rank-(r_1,r_2,r_3)		(70,16,34)	(71,16,37)	(71,15,39)	(70,14,40)	(67,13,39)	(62,10,36)	(52,5,29)

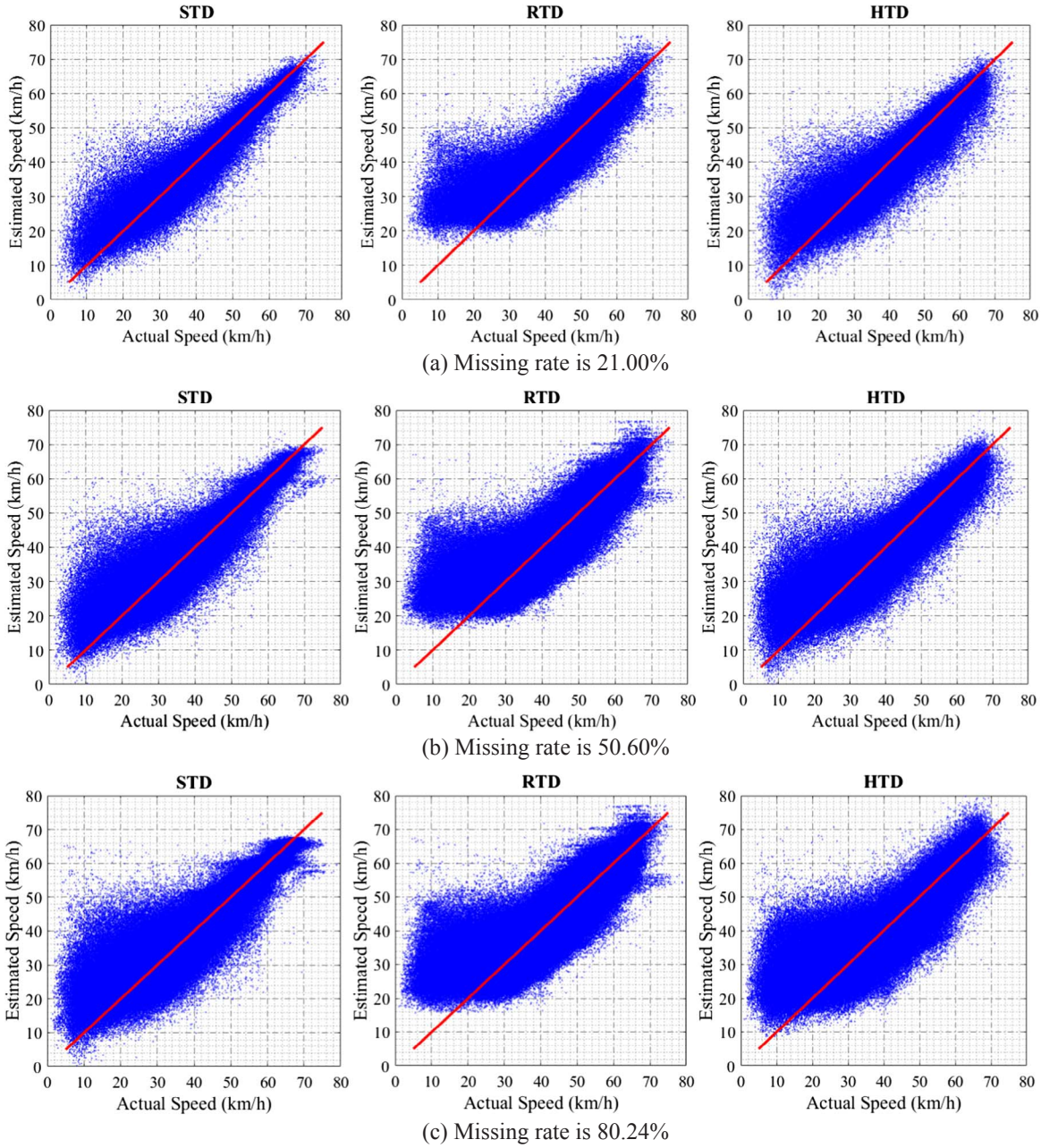


Fig. 6. Performance of the missing data recovery, herein, blue scatters represent coordinates $(x_{ijk},\widehat{x}_{ijk}),(i,j,k) \in \overline{\Omega} \cap \Omega_1$ and blue scatter at the red line implies that the estimated speed is equal to the actual speed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4.2. Scenario of fiber-like random missing

On a single road segment, the speed observations may be completely lost during several days. Nevertheless, the speed of other road segments is valuable for discovering traffic patterns and recovering the missing values. Thus, the fiber-like missing aims to simulate the scenario where the speed is randomly lost over road segments and day dimensions. On the one hand, from Table 4, we can summarize same conclusions such as the STD performs best for any missing rate. On the other hand, in terms of features' availability, for the same missing rate, the rank-(r_1,r_2,r_3) under the fiber-like missing is generally smaller than the element-like missing, and this implies that the fiber-like missing may keep less features. In terms of performance given by Tables 3 and 4, an accurate estimation to the fiber-like missing becomes more difficult comparing with the element-like missing. Previous studies (Acar et al., 2011) have also found that the scenarios of structured missing are generally more difficult than those of random missing.

Table 4
Performance of missing data recovery using there models with scenario of fiber-like random missing.

Measures	Models	21.00% missing	31.03% missing	40.97% missing	50.30% missing	60.46% missing	70.71% missing	80.56% missing
RMSE (km/h)	STD	4.3300	4.4029	4.5573	4.6291	4.7998	4.9902	5.2467
	RTD	6.0824	6.0479	6.0321	6.0846	6.1062	6.1064	6.1274
	HTD	4.5092	4.5048	4.8879	4.9457	5.0346	5.7542	5.8802
MAE (km/h)	STD	2.8351	2.8903	2.9873	3.0365	3.1734	3.2879	3.5265
	RTD	4.3608	4.3458	4.3328	4.3616	4.3773	4.3779	4.4092
	HTD	3.1073	3.1021	3.4568	3.4248	3.4839	4.1651	4.1594
MRE (%)	STD	7.29	7.42	7.66	7.78	8.12	8.40	9.04
	RTD	11.22	11.17	11.14	11.20	11.22	11.22	11.32
	HTD	8.09	8.05	9.05	8.89	9.02	11.05	10.84
Rank-(r_1,r_2,r_3)		(50,9,10)	(50,9,9)	(48,9,7)	(46,7,6)	(42,6,4)	(51,9,5)	(42,4,3)

3.5. Interpretation of factor matrices

In fact, the traffic patterns along day dimension can be roughly divided into weekday and weekend, and an evidence is that the traffic during weekday may be busier. As for time periods of a day, the traffic patterns can be further divided into morning peak hours, afternoon peak hours and off-peak hours. In this section, we intend to find some indications from the traffic speed data fitting real-world traffic. For the fact that factor matrices trained under different missing scenarios have similar leading factors, the fiber-like missing with smaller rank-(r_1,r_2,r_3) (i.e., less columns for factor matrices, see the last row of Tables 3 and 4) is chosen as an example.

3.5.1. Factor matrix $V \in \mathbb{R}^{n_2 \times r_2}$

Although the entries and total columns of factor matrix V trained under different missing rates are not identical, from the right three heat maps of Fig. 7, we can see that the second column of these factor matrices shares the analogous trend along 61 days. Explicitly, the traffic pattern during weekend (such as August 6th, August 7th, August 13th, August 14th, August 20th, and August 21st) has been discovered. In particular, the traffic pattern during three-day holiday (the traditional Middle-Autumn festival in China) from September 15 to September 17, 2016 has been automatically categorized as a regular pattern of weekend. Therefore, the second column of factor matrix V indicates the days without busy traffic and heavy congestion.

3.5.2. Factor matrix $W \in \mathbb{R}^{n_3 \times r_3}$

In Fig. 8, the first four columns of factor matrix are leading patterns. Specifically, for the first column, the striking similar values for all time windows imply that it is a common speed feature. Instead, the second and third column tend to emphasize the speed features during afternoon and morning, respectively. In the fourth column, it is evident that there are two special red areas with

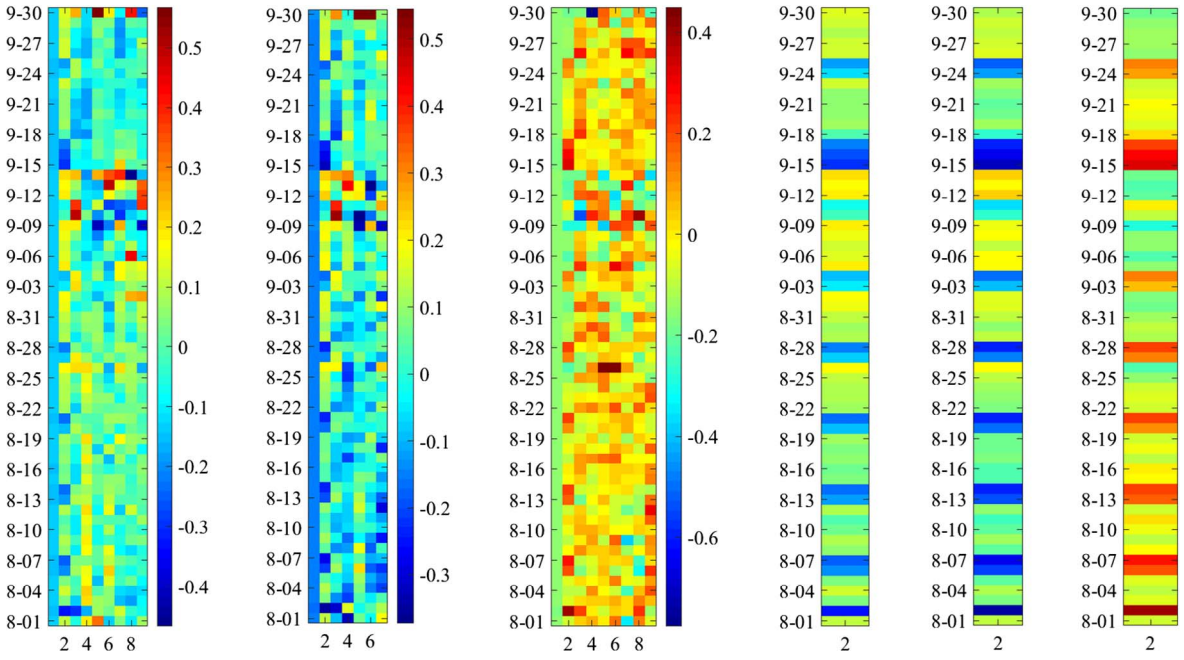


Fig. 7. The visualization of factor matrix $V \in \mathbb{R}^{n_2 \times r_2}$ trained by the STD, from left to right, the left three heat maps correspond to the 30%, 50% and 70% missing, and the right three heat maps are the second column of left three heat maps.

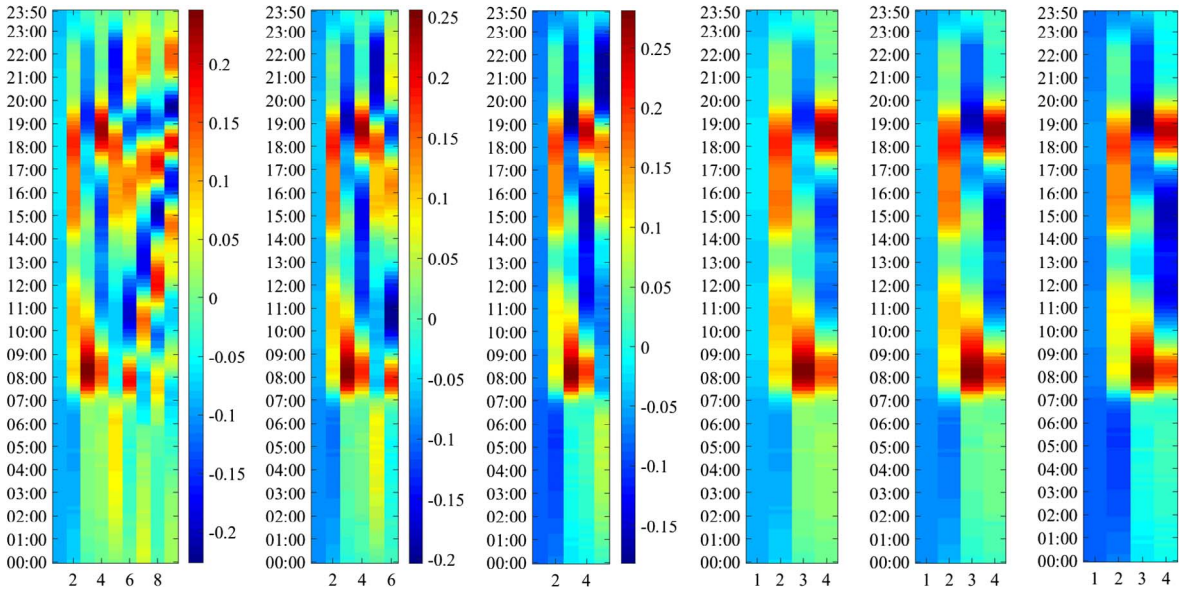


Fig. 8. The visualization of factor matrix $W \in \mathbb{R}^{n_3 \times r_3}$ trained by the STD, from left to right, the left three heat maps correspond to the 30%, 50% and 70% missing, and the right three heat maps are the first four columns of left three heat maps.

positive values during the periods of 7:00–9:00 and 17:30–19:30 approximately. Meanwhile, the periods often correspond to the morning peak hours and afternoon peak hours. Thus, the fourth column can be treated as a pattern to characterize the peak hours.

3.6. Time-series traffic speed analysis

Tensor decomposition methods using partially observed data to estimate the missing data are effective as analyzed above. However, despite of the multi-mode biases based traffic pattern, other patterns discovered by tensor decomposition are rather implicit to explain. In this vein, time-series traffic speed analysis (Li et al., 2015) is an explicit way to seek out what patterns the related methods have achieved.

3.6.1. Scenario of element-like random missing

Instead of presenting the whole recovery performance by Table 3 and Fig. 6, we select a number of speed data of a single road segment from August 1, 2016 to August 7, 2016 as an example. The time-series traffic speed curves and their estimates are plotted as shown in Fig. 9. Intuitively, we can find that the RTD is vulnerable for the abnormal speed during August 2, 2016 and the speed which is lower than 25km/h. The estimation of time-series traffic speed through the STD is more accurate than the other two models. Moreover, even though the original time-series observations have fluctuations, the STD can still grasp the time-evolving traffic patterns from partially observed data accurately.

Using the STD model, while the missing rate is 50.60%, Fig. 10 gives further insight to describe what the patterns are hidden in the factor matrix $W \in \mathbb{R}^{n_3 \times r_3}$. During the night (period of 20:00–06:00), there are little color changes among rows, and slight differences are indicated among different time windows. However, during the period of 7:00–9:00 (morning peak hours) or 17:30–19:30 (evening peak hours), we observe that, among different time windows, there exist remarkable differences. In such case, these indications are consistent with the real-world traffic and can be interpreted as follows. (1) During the off-peak hours, the traffic flow of all road segments may be fluent, and thereby, the changes of time-series traffic speed are very little. (2) During the peak hours, the congestion among different road segments may vary significantly, for instance, some road segments meet congestion earlier than others, and congestion durations of some road segments may be more than one hour while others are only few minutes. Thus, the tensor decomposition model can describe the inherent patterns hidden in the time-series traffic speed data.

3.6.2. Scenario of fiber-like random missing

In the context of using partially observed data, Fig.11(a) shows that the STD model can recover the missing data (i.e., gray area) accurately. As the missing rate changes from 30% to 70% (see Fig. 11(a) and (b)), the incremental residuals reveal that the less partial observations are the less informative they are. While the missing rate is 70%, the extreme case where the observations of the road segment within four days are completely lost can also be solved. Among (b), (c) and (d) of Fig. 11, the RTD and HTD are indicated to be inferior to the STD.

In the matter of missing values at random, as long as each slice of sparse tensor $\mathscr{X}$ has at least one non-zero entry, we still have a chance to train the factor matrices U,V,W and core tensor $\mathscr{G}$ (Acar et al., 2011). Consequently, even for fiber-like missing, the tensor models can still work. However, in the matrix completion problems, when losing an entire row or column, it is impossible to recover

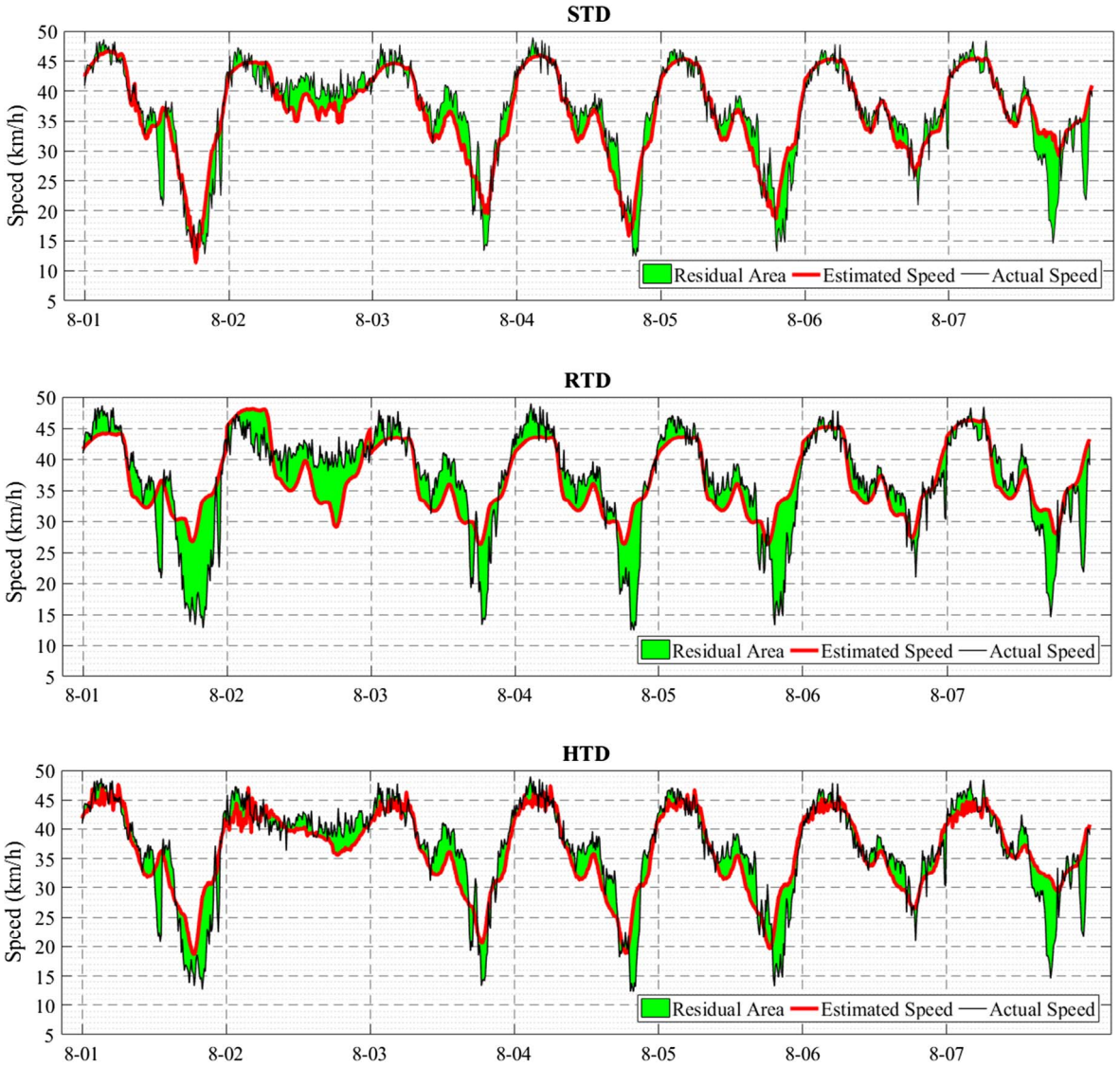


Fig. 9. The estimated speed and corresponding actual values of a road segment under the 50.60% missing, note that the green area (residual area) is only used to express the estimation performance, and which does not represent the cumulative residual.

the missing entries of the matrix. In brief, the structures of tensor can exhibit some advantages for dealing with high-dimensional problem, and the proposed STD can recover missing entries of incomplete tensor for both the element-like missing and the fiber-like missing.

Further, traffic speed has dimensions of road segment, day and time window, thus, this makes the recovery task a high-dimensional problem. Compared with the matrix based models, such tensor models allow us to describe a richer algebraic structure and thereby encode more information. In such case, matrix based models such as matrix decomposition (Koren et al., 2009; Asif et al., 2016), PCA-based (Qu et al., 2009; Li et al., 2013, 2014; Asif et al., 2016; Goulart et al., 2017) and other matrix completion methods may be weak to deal with high-dimensional element-like missing or fiber-like missing.

Thus, the main advantages of STD can be concluded as following three categories. (1) As its core, the high-dimensional tensor decomposition allows us to grapple with latent factors and multi-relations. (2) The multi-mode biases based traffic pattern is taken into account by missing data initialization, and it matches the truth that the real-world traffic speed is highly related to the road segment, day and time window. (3) The main features captured by the truncated SVD serve as an optimal initialization for tensor decomposition, where this initialization benefits the final results.

4. Concluding remarks

Missing traffic data recovery is a crucial task of data quality enhancement in the data-driven intelligent transportation systems.

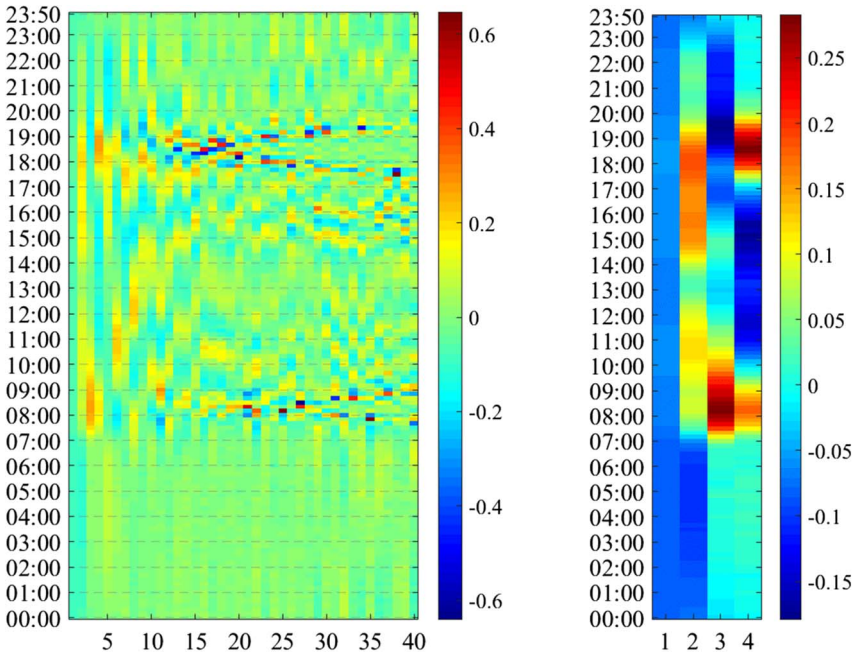


Fig. 10. The visualization of factor matrix $W \in \mathbb{R}^{n3\times r3}$ trained by the STD, the right heat map is the first four columns of left heat map.

Although, Asif et al. (2016) and Goulart et al. (2017) have worked on the incomplete traffic speed data by applying various matrix and tensor completion algorithms, and tensor completion methods have been previously used to solve the problem of missing traffic data (Tan et al., 2013; Asif et al., 2016; Ran et al., 2016; Goulart et al., 2017), the performance of different initialization strategies for Tucker decomposition under scenarios of element-like and fiber-like missing are not well analyzed. As most studies tend to bridge the gap between practical problems and tensor models, few studies described interpretability of tensor models and what latent traffic patterns have been discovered by tensor models. Take into account multi-mode biases based traffic pattern and further combine it with tensor decomposition, there may still need some comprehensive analysis about missing traffic data recovery.

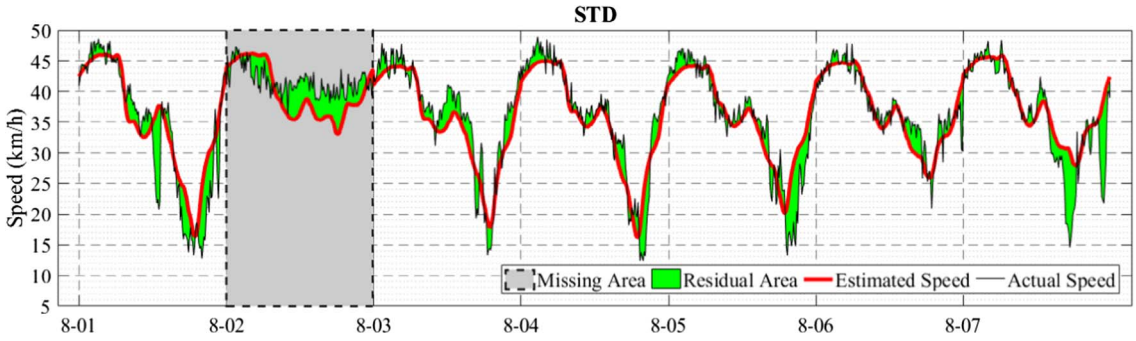
To this end, in this study, we treat missing traffic data recovery as the high-dimensional problem of tensor completion, and propose a three-procedure framework which allows for discovering traffic patterns to solve this problem. Incorporated with multi-mode biases based traffic pattern, the missing data initialization attempts to simply reconstruct missing entries of the incomplete tensor. Then, by means of the truncated SVD, main features along each dimension are captured as hidden noises are removed. Thereby, the proposed STD model uses these main latent features to estimate missing entries eventually. In the empirical study, we illustrate that the missing data initialization can estimate missing entries of incomplete tensor by multi-biases based approach efficiently. Furthermore, the truncated SVD can not only find the best core tensor size (i.e., rank-(r_1,r_2,r_3)), but also capture main features along each dimension in advance. Using the STD, the results demonstrate that the fore two procedures have a great beneficial impact on the performance of missing data recovery.

The impacts of three different initialization strategies under two missing scenarios are also investigated. By comparisons with the RTD and HTD, while the missing rate ranges from 20% to 80% under different missing scenarios, the results indicate the advantages of the STD model. Further, initialization strategy for tensor decomposition makes significant influence on the final recovery quality and an optimal initialization is beneficial to recovery performance. In the aspect of different missing scenarios, the experimental results demonstrate that the fiber-like missing is more difficult to recover than the element-like missing. However, for the observations of road segments within some days are completely lost, the STD model can still provide a promising solution.

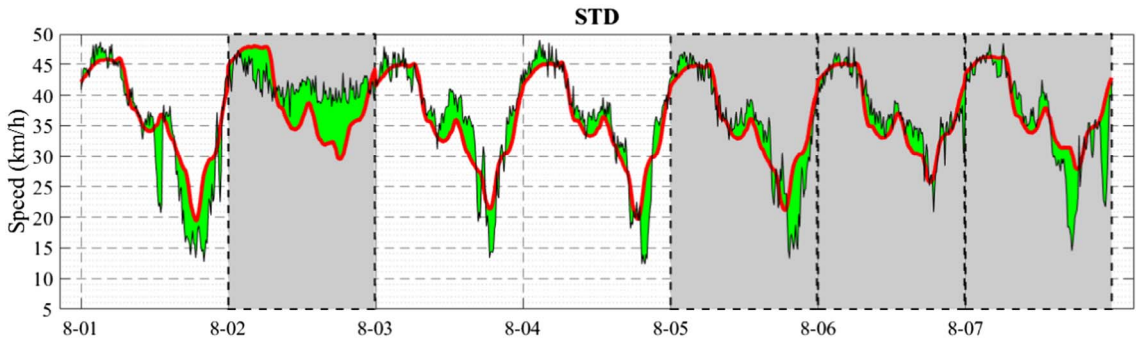
In the future, we plan to utilize multi-dimensional nature of traffic data to further discover spatial-temporal traffic pattern, improve data quality and scheme multi-sources data fusion, particularly, different traffic parameters such as volume and speed would be considered to be fused towards their multi-relations.

Acknowledgement

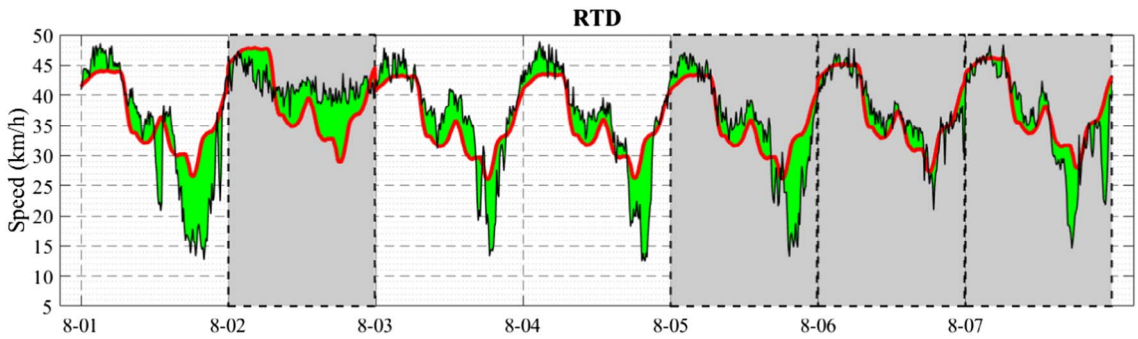
This work was supported by the Science and Technology Planning Project of Guangdong province, China under Grant 2014B010118002.



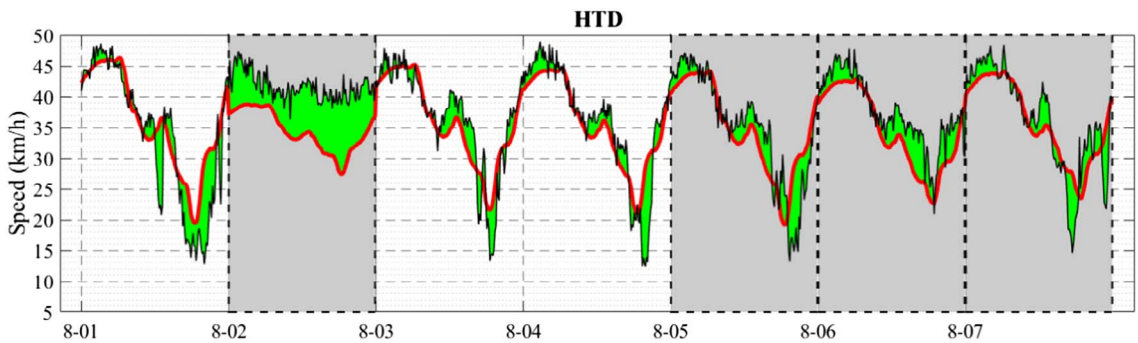
(a) Missing rate is 31.03%



(b) Missing rate is 70.37%



(c) Missing rate is 70.37%



(d) Missing rate is 70.37%

Fig. 11. The estimated speed and corresponding actual values of a road segment, note that the missing area (gray area) represents the fiber-like missing.

Appendix A

A.1. Convergence condition, iterations and performance measures

In the tensor decomposition algorithm, convergence condition of tensor decomposition is given by $||\mathscr{X}^{(1)}-\mathscr{X}^{(0)}||_F^2 < \varepsilon$, moreover, we define loss $= \sqrt{\frac{1}{|\Omega|}\sum_{(i,j,k)\in\Omega}(x_{ijk}-\hat{x}_{ijk})^2}$ which is equal to RMSE for $(i,j,k) \in \Omega$. Select smaller ε, in Fig. A1, *loss* and RMSE have little decrease whereas the iteration increase significantly. Therefore, to enable efficiency, we only choose $\varepsilon = 0.01$ be convergence condition in our empirical studies.

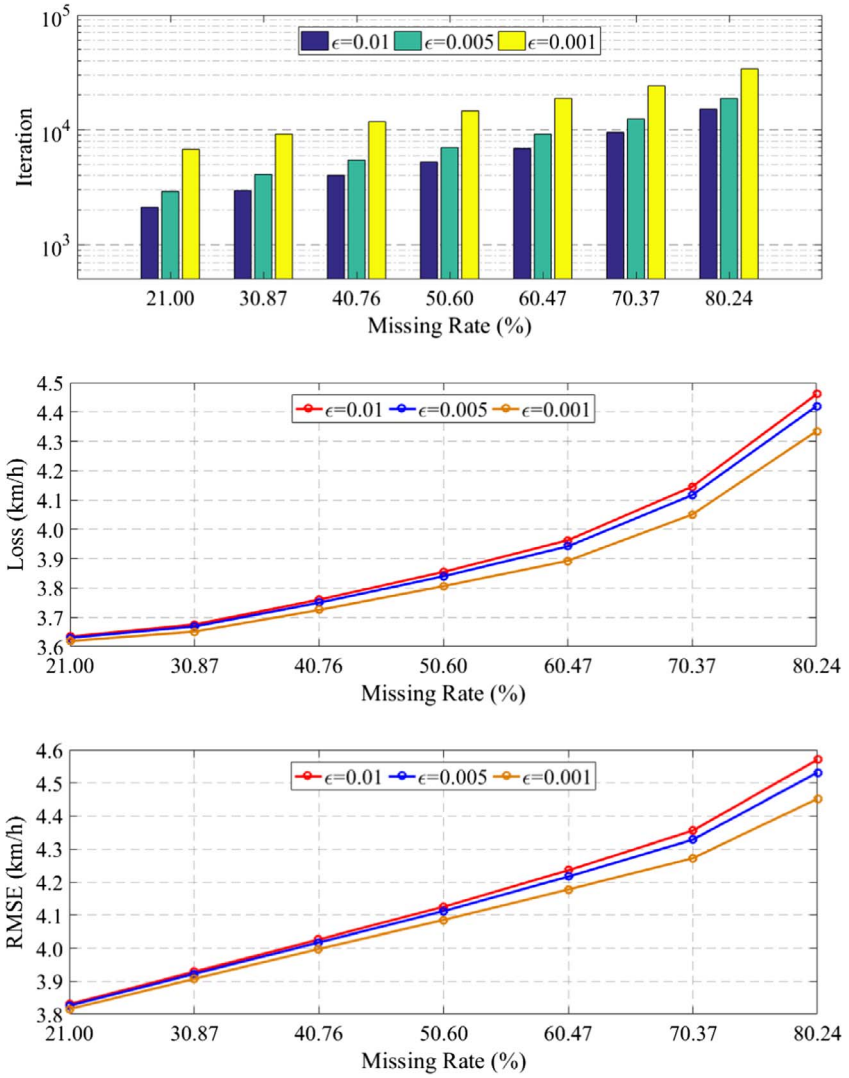


Fig. A1. STD: Convergence condition versus measures including iteration, loss and RMSE under the scenario of element-like missing.

A.2. Run times and iterations of STD, RTD and HTD

As exemplified by the element-like missing scenario (see Fig. B1), the convergence condition of the STD, RTD and HTD is given by $\varepsilon = 0.01$. The learning rate of the STD is $\alpha = 2 \times 10^{-10}$, the RTD is prescribed by $\alpha = 1 \times 10^{-8}$, and the HTD is $\alpha = 1 \times 10^{-10}$. All algorithms were tested using MATLAB R2016b on a Windows Workstation with a Quad-Core Intel (R) Core (TM) 3.30 GHz CPU and 8.00 GB RAM.

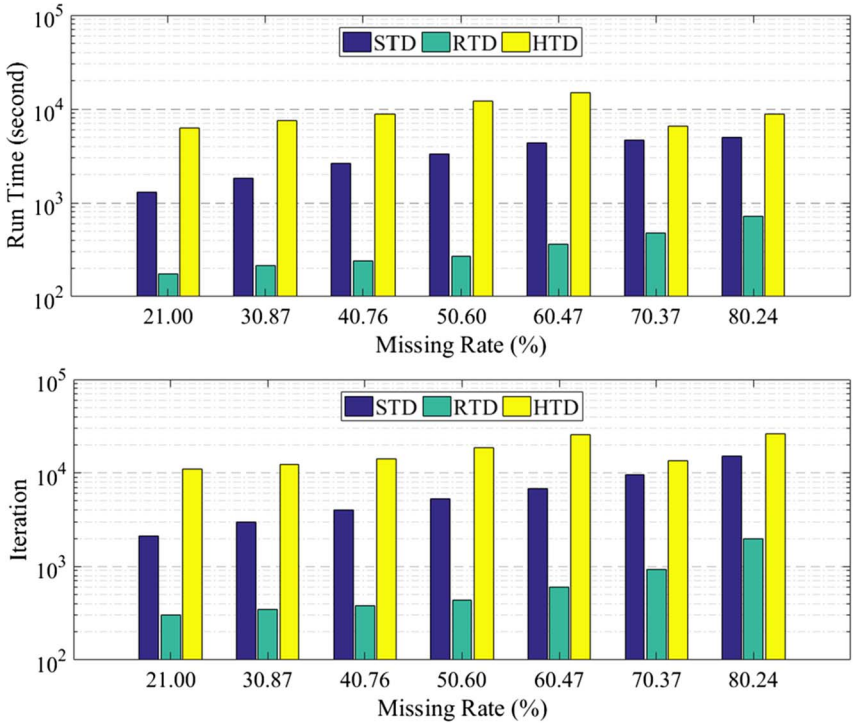


Fig. B1. Run times and iterations of the STD, RTD and HTD with missing rate ranging from 20% to 80%.

References

Acar, E., Dunlavy, D.M., Kolda, T.G., Mørup, M., 2011. Scalable tensor factorizations for incomplete data. Chenoment. Intell. Lab. Syst. 106 (1), 41–56.

Anandkumar, A., Ge, R., Hsu, D., Kakade, S.M., Telgarsky, M., 2014. Tensor decompositions for learning latent variable models. J. Mach. Learn. Res. 15 (1), 2773–2832.

Asif, M.T., Kannan, S., Dauwels, J., Jaillet, P., 2013. Data compression techniques for urban traffic data. In: Computational Intelligence in Vehicles and Transportation Systems (CIVTS). 2013 IEEE Symposium. IEEE, pp. 44–49.

Asif, M.T., Srinivasan, K., Mitrovic, N., Dauwels, J., Jaillet, P., 2015. Near-lossless compression for large traffic networks. IEEE Trans. Intell. Transp. Syst. 16 (4), 1817–1826.

Asif, M.T., Mitrovic, N., Dauwels, J., Jaillet, P., 2016. Matrix and tensor based methods for missing data estimation in large traffic networks. IEEE Trans. Intell. Transp. Syst. 17 (7), 1816–1825.

Chen, B., Li, Z., Zhang, S., 2015. On optimal low rank Tucker approximation for tensors: the case for an adjustable core size. J. Glob. Optim. 62, 811–832.

Carroll, J.D., Chang, J.J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35, 283–319.

Duan, Y., Lv, Y., Liu, Y.-L., Wang, F.-Y., 2016. An efficient realization of deep learning for traffic data imputation. Transp. Res. Part C: Emerg. Technol. 72, 168–181.

Golub, G.H., Van Loan, C.F., 2013. Matrix Computations. Johns Hopkins Univ. Press, Baltimore, MD.

Goulart, J.H.de.M., Kibangou, A.Y., Favier, G., 2017. Traffic data imputation via tensor completion based on soft thresholding of Tucker core. Transp. Res. Part C: Emerg. Technol. 85, 348–362.

Han, Y., Moutarde, F., 2016. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. Int. J. ITS Res. 14, 36–49.

Hitchcock, F.L., 1927a. The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. 6, 164–189.

Hichcock, F.L., 1927b. Multiple invariants and generalized rank of a p-way matrix or tensor. J. Math. Phys. 7, 39–79.

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

Koren, Y., Bell, R.M., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. IEEE Compt. 42 (8), 30–37.

Lathauwer, L.D., Moor, B.D., Vandewalle, J., 2000. A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. 20 (4), 1253–1278.

Li, L., Li, Y., Li, Z., 2013. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. Transp. Res. Part C: Emerg. Technol. 34, 108–120.

Li, L., Su, X., Zhang, Y., Lin, Y., Li, Z., 2015. Trend modeling for traffic time series analysis: An integrated study. IEEE Trans. Intell. Transp. Syst. 16 (6), 3430–3439.

Li, X., Li, M., Gong, Y.-J., Zhang, X.-L., Yin, J., 2016. T-DesP: Destination prediction based on big trajectory data. IEEE Trans. Intell. Transp. Syst. 17 (8), 2344–2354.

Li, Y., Li, Z., Li, L., 2014. Missing traffic data: comparison of imputation methods. IET Intel. Transp. Syst. 8 (1), 51–57.

Liu, J., Musialski, P., Wonka, P., Ye, J., 2013. Tensor completion for estimating missing values in visual data. IEEE Trans. Pattern Anal. Mach. Intell. 35 (1), 208–220.
Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK.
Paterek, A., 2007. Improving regularized singular value decomposition for collaborative filtering. In: Proceeding of KDD Cup and Workshop, pp. 5–8.
Qu, L., Li, L., Zhang, Y., Hu, J., 2009. PPCA-based missing data imputation for traffic flow volume: A systematical approach. IEEE Trans. Intell. Transp. Syst. 10 (3), 512–522.
Ran, B., Tan, H., Wu, Y., Jin, P.J., 2016. Tensor based missing traffic data completion with spatial-temporal correlation. Phys. A 446, 54–63.
Schifanella, C., Candan, K.S., Sapino, M.L., 2014. Multiresolution tensor decompositions with mode hierarchies. ACM Trans. Knowl. Discovery from Data 8 (2), 10.
Soriguera, F., 2016. Highway Travel Time Estimation With Data Fusion. Springer, Berlin.
Sun, L., Axhausen, K.W., 2016. Understanding urban mobility patterns with a probabilistic tensor factorization framework. Transp. Res. Part B: Methodol. 91, 511–524.
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: Emerg. Technol. 28, 15–27.
Tan, H., Feng, J., Chen, Z., Yang, F., Wang, W., 2014. Low multilinear rank approximation of tensors and application in missing traffic data. Adv. Mech. Eng. 2014.
Tan, H., Wu, Y., Shen, B., Jin, P.J., Ran, B., 2016. Short-term traffic prediction based on dynamic tensor completion. IEEE Trans. Intell. Transp. Syst. 17 (8), 2123–2133.
Tucker, L., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31 (3), 279–311.
Wang, Y., Zheng, Y., Xue, Y., 2014. Travel time estimation of a path using sparse trajectories. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'14. ACM, pp. 374–383.
Zhao, Q., Zhang, L., Cichocki, A., 2015. Bayesian CP factorization of incomplete tensors with automatic rank determination. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1751–1763.