Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 2919073, 12 pages
https://doi.org/10.1155/2021/2919073

Research Article

Expressway Traffic Flow Missing Data Repair Method Based on Coupled Matrix-Tensor Factorizations

Hui Jiang [1,2] and Hongxing Deng [1]

[1] School of Traffic and Transportation, Northeast Forestry University, Harbin 150040, China
[2] School of Civil Engineering and Transportation, Beihua University, Jilin 132013, China

Correspondence should be addressed to Hongxing Deng; 1969dhx@nefu.edu.cn

Received 5 August 2020; Revised 15 January 2021; Accepted 28 January 2021; Published 11 February 2021

Academic Editor: Frederic Kratz

Traffic flow data is the basis of traffic management, planning, control, and other forms of implementation. Once missing, it will directly affect the monitoring and prediction of expressway traffic status. Regarding this, this paper proposes a repair method for the traffic flow missing data of expressway, combined with the idea of coupled matrix-tensor factorizations (CMTF), to couple the auxiliary traffic flow data into the main traffic flow data and to construct the coupling matrix-tensor expression of traffic flow data, and the alternating direction multiplier algorithm is used to realize the repair of missing traffic flow data. Combined with the measured data of expressway traffic flow, the experimental results show that, under different missing data types and missing rates, the proposed method outperforms the methods lacking auxiliary traffic flow data and achieves a good repair effect, especially for high miss data rates.

1. Introduction

The traffic management department requires traffic flow data to be complete to realize accurate assessment and prediction of the traffic status. Thus, the repair of the missing data is crucial for effective management of the traffic flow.

In the past few decades, researchers have shown much interest in the problem of repairing the missing traffic flow data. The existing research on this area can be categorized into three types of data composition: vector, matrix, and tensor. Below, we elaborate on the existing studies to provide an overview of the state of the art.

First of all, researchers used the traffic flow data in the form of a one-dimensional vector to repair the missing parts based on spatial or temporal correlation. Gold et al. classified the reasons for the missing data based on the characteristics of expressway traffic flow data and repaired the missing data using nuclear regression, polynomial regression, and linear interpolation [1]. Based on the hotspot analysis of the spatial correlation of traffic data, Conklin et al. used the similarity of the adjacent lanes to repair the missing data [2]. Smith et al. utilized the temporal correlation of traffic data and repaired the missing parts by using the average value of the historical data at the corresponding time and the exponential smoothing value of data at adjacent periods from several days ago [3]. Based on the spatial and temporal correlation, Jiang and Gang predicted the value of data in previous periods and adjacent sections to repair the missing data [4]. Jin and Wang proposed a traffic flow parameter correction method based on correlation analysis from the perspective of statistical analysis of traffic flow data [5]. Gheyas and Smith mainly focused on the measured traffic flow data and utilized the methods of historical trend and exponential smoothing to repair the missing parts. By comparison, these two methods were proved to be more effective in dealing with continuous abnormal data [6]. In consideration of the spatial-temporal characteristics of the traffic flow data, Wang et al. proposed a repair method based on the 3D shape function and the spatial-temporal interpolation [7]. Chen et al. proposed a long short-term memory (LSTM) method, which was built on the normal time-series using the calculation error of the Gaussian Bayesian model. Experimental results on three real datasets showed the advantages of this method [8].

Then, the data is organized into a matrix to repair the missing data. Kurucz et al. proposed a low-rank matrix completion algorithm using the path norm minimization of the matrix to replace the minimization of the rank function [9]. Qu et al. performed (Bayesian) principal component analysis to reconstruct the traffic flow data into a matrix. The results showed the effectiveness of this method in data repair for low missing rates [10]. Tang et al. made use of fuzzy c-means clustering and genetic algorithm to repair the missing data based on a matrix. The repair effect of data on different time scales was quantified. The results revealed that the fusion model had a significant repair effect [11]. Lu et al. proposed a data repair model structured on an improved multiscale principal component analysis. Based on the spatial-temporal correlation characteristics of the traffic flow data, they calculated the correlation coefficient of the missing data to estimate its true value [12]. Similarly, Jiang proposed a data repair algorithm using the fusion matrix low-rank decomposition [13].

The data composition in the form of multidimensional tensor expands the vector and matrix to higher dimensions. In recent years, it has proliferated in machine learning (clustering and dimensionality reduction) and other fields [14]. This method can repair the missing data by organizing the traffic flow data into tensors. Acar et al. proposed an algorithm based on CP decomposition and gradient optimization and then proved that the algorithm had high flexibility [15]. Silva and Herrmann proposed a tensor repair method based on the hierarchical Tucker decomposition to repair the missing seismic data [16]. Tan et al. proposed a traffic data repair method based on tensor for the first time, which achieved good experimental results, especially in the case of high data missing rates. Subsequently, Tan et al. proposed several other tensor-based repair methods, which proved that the tensor-based method was still useful, especially for the cases with high missed data [17]. Liu et al. proposed a low-rank tensor repair algorithm to solve the problem of image repair and proved that their method was better than the tensor model using CP decomposition [18]. Based on the tensor theory, Xiang used the spatial-temporal characteristics of the traffic data for modelling, which significantly improved the accuracy of data repair [19]. Zhang focused on the microwave data of the urban expressway and repaired the data using a CP-ALS algorithm based on tensor decomposition. The results proved that their method could effectively improve data quality [20]. Zhang et al. proposed an iterative tensor decomposition (ITD) method to repair the lost data in the ANPR system. They used the multidimensional intrinsic correlation of the traffic flow data for the detection of missed parts and their repair. The results showed that this method could accurately detect and repair the missing data under various missing rates [21].

To summarize, the one-dimensional vector was highly dependent on the historical information, requiring stable traffic flow data within a period or space. Thus, it could not use the information collected after the data missed. Besides, if the data missed within a long period, the error in the repair results turned to be high. On the other hand, although the two-dimensional matrix made full use of the spatial-temporal correlation characteristics of the traffic flow data, it was limited to the 2D data. Furthermore, it could not simultaneously utilize the multimode correlation of the traffic flow data, which resulted in low repair accuracy. In recent years, the high-dimensional tensor model has attracted significant attention in data repair, and some progress has been achieved, especially for the repair of traffic flow missing data. However, most of the existing efforts have focused on the model construction of spatial and temporal correlation of the traffic flow data.

In summary, it is necessary to introduce a data repair method, which makes full, reasonable, and effective use of multidimensional information between the traffic flow data. To do so, this work for the first time proposes a traffic flow missing data repair model based on the coupled matrix-tensor factorizations. Compared with the existing work, we make two contributions: (1) we aim to obtain better multidimensional information of traffic flow data, which is divided into main traffic flow data and auxiliary traffic flow data, and coupled the auxiliary traffic flow data into the main traffic flow data to construct the coupling matrix-tensor expression of traffic flow data. (2) In order to get better data repair effect, we used alternating direction multiplier (ADMM) algorithm to decompose large global problems into many smaller and easier local subproblems to solve the optimal solution. By comparison with the prior method, we validate the effectiveness and feasibility of the proposed methodology in data repair.

The rest of this paper is organized as follows. Firstly, the tensor theory is explained in Section 2. Then, the traffic flow data analysis is provided in Section 3. Section 4 presents the missing data repair method based on CMTF, where the results and discussion on the findings are elaborated in Section 5. Finally, Section 6 concludes the paper.

2. Tensor Theory

As an extension of vector and matrix to a high-dimensional space, tensor is a mathematical representation of a multidimensional data. The dimension of a tensor is also called the modulus or the order [22, 23]. Hence, if we consider a vector as a tensor of the first order and a matrix as a tensor of the second order, it refers to an array of three or more dimensions as a tensor of a high order, tensor for short. The third-order tensor $\chi \in \mathbb{R}^{I\times J\times K}$ with dimensions I, J, and K shown in Figure 1 is an example of this.

Tensors can be expressed in either a fibrous or a slice form [24, 25]. Any two dimensions in the third-order tensor are kept unchanged. In this case, the tensor can be expressed by a column fiber line $(x_{:\,jk})$, a fiber $(x_{i:\,k})$, and a pipe $(x_{ij:})$, where the colon represents the dimension all of elements. Alternatively, the tensor can be expressed in the form of a horizontal slice $(X_{i::})$, a side slice $(X_{:\,j:})$, and a front slice $(X_{:\,:\,k})$ when the other two dimensions are changed.

2.1. Tensor Matrixing. Tensor matrixing is a process, in which the elements of a tensor are rearranged into matrices. To be specific, tensor operations are simplified by expanding tensor χ into the matrix $X_{(n)}$ in the n modulus.

The CP decomposition is one of the earliest and most in-depth tensor decomposition methods. It can be interpreted as decomposing a tensor into the form of a sum of rank-1 tensors. For $\chi \in \mathbb{R}^{I\times J\times K}$, the CP decomposition can be expressed in the following matrix form:

$$\chi \approx [[A, B, C]] = \sum_{r=1}^{R} a_r \circ b_r \circ c_r = A \circ B \circ C, \tag{1}$$

where $A = [a_1, a_2, \ldots, a_r]$ is the vector a_r corresponding to rank-1 tensor. Similarly, B and C are the vectors b_r and c_r corresponding to rank-1 tensor. Thus, A, B, and C are the factor matrices decomposed by CP decomposition, where $A \in \mathbb{R}^{I\times R}$, $B \in \mathbb{R}^{J\times R}$, and $C \in \mathbb{R}^{K\times R}$. The slice of a tensor can be represented as follows:

$$\begin{aligned} X_{(1)} &\approx A(C \odot B)^T, \\ X_{(2)} &\approx B(C \odot A)^T, \\ X_{(3)} &\approx C(B \odot A)^T, \end{aligned} \tag{2}$$

where the symbol "$\odot$" can be specifically described as the output value of 1 when two input variables are the same, otherwise the output value is.

In practice, however, the column vectors in the factor matrix are usually normalized: $\|a_r\|_F = \|b_r\|_F = \|c_r\|_F = 1$. Thus, the CP decomposition can be expressed as

$$\chi \approx [[\lambda; A, B, C]] = \sum_{r=1}^{R} \lambda_r a_r \circ b_r \circ c_r, \tag{3}$$

where λ is the regularization parameter.

2.2. The Coupled Matrix-Tensor Factorizations. The joint analysis of data is quite useful in understanding the underlying structure of complex datasets. Here, the coupled matrix-tensor factorizations (CMTF) can capture the potential structural characteristics of heterogeneous data by coupling the matrix data into the high-order tensor model.

To explain this coupling model in detail, we assume that a third-order tensor $\chi \in \mathbb{R}^{I\times J\times K}$ and a matrix $U \in \mathbb{R}^{I\times M}$ are derived from different datasets, and U is coupled to a dimension of χ. The schematic diagram of the CMTF is shown in Figure 2.

The underlying structure of the dataset can be extracted by the CMTF, where the R-component of a tensor χ and a matrix U is defined as

$$f(A, B, C, V) = \frac{1}{2}\|\chi - [[A, B, C]]\|^2 + \frac{1}{2}\left\|U - AV^T\right\|^2, \tag{4}$$

where A, B, C are the factor matrices of the tensor χ extracted by CP decomposition, where $A \in \mathbb{R}^{I\times R}$, $B \in \mathbb{R}^{J\times R}$, and $C \in \mathbb{R}^{K\times R}$.

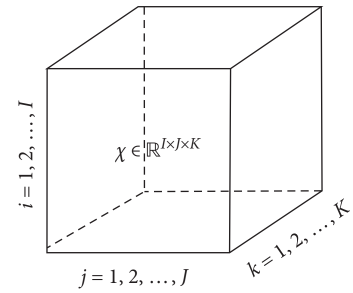

FIGURE 1: The third-order tensor $\chi \in \mathbb{R}^{I\times J\times K}$.

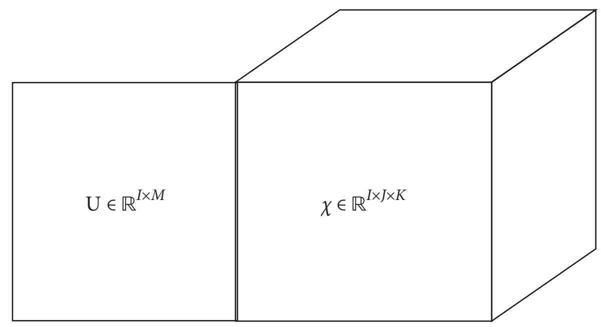

FIGURE 2: The CMTF model.

The factor matrices $A \in \mathbb{R}^{M\times R}$and $V \in \mathbb{R}^{M\times R}$ are extracted from matrix U, which can be represented as

$$U = AV^T, \tag{5}$$

where A is the common mode factor matrix of the tensor χ and the matrix U on a certain dimension (modulus).

To ensure that the datasets are codecomposed and share the same potential space, the matrix U should have the same underlying low-rank structure as the tensor χ on at least one dimension.

3. Analysis of the Traffic Flow Data

3.1. Traffic Flow Missing Data: Causes and Types. The complete traffic flow data is key for the traffic management department to assess the traffic status. However, due to the influence/restriction of traffic conditions, meteorological environment, and many other factors, the traffic system has formed a complex operating environment. There are four main reasons of the missing data: detection equipment failure, power failure, communication network failure, and environmental factors.

The missing data types can be summarized as follows:

(1) *Missing Completely at Random (MCAR).* It refers to the type where the missing traffic flow data are completely random and independent of each other; that is, the probability of the missing data is not affected by other factors. The missing data location is represented as a random discrete distribution.

(2) *Missing at Random (MAR).* It refers to the type where the missing traffic flow data is not completely random. The missing data is not correlated with itself

but with other adjacent data points. The missing data location is a random continuous distribution.

(3) *Missing at Nonrandom (MANR).* It refers to the type where there is a certain correlation between the missing data and its characteristics, which also depends on other missing data. This loss is generally caused by the system itself.

One of the earlier studies in the field [26, 27] assumed that the MANR data was found and then deleted, where subsequent studies mainly used the MCAR and MAR. To comprehensively reflect the real traffic flow under the condition that the complexity of the data is missing, the follow-up work will be according to the three following case studies:

(1) The missing data type 100% are missing completely at random (MCAR)
(2) Lack of missing at random (MAR) data type 100%
(3) Lack of missing completely at random data type and random loss, 50% each, namely, lack of hybrid (MIX-M)

3.2. Analysis of the Traffic Flow Data. In analysis of traffic flow data variation, the traffic flow data is divided into main traffic flow data and auxiliary traffic flow data according to whether it can directly reflect the actual traffic status of the road.

3.2.1. Analysis of the Main Traffic Flow Data. This paper defines the main traffic flow data as the data that can directly reflect the state change of the system. Specifically, it is described as the index value that can represent the state of the traffic flow through the data, such as flow, speed, and occupancy, which plays a leading role in the repair of the missing data.

The phenomena of the main traffic flow data changes with time and space are called the spatial and temporal distribution of traffic flow, which has the characteristics of temporal and spatial correlation. The time correlation analysis of the traffic flow data describes the change characteristics in the time domain. The travellers determine the time regularity of the traffic flow. According to the traffic data of different times of a month, week, and day, the similarity of the different degrees is reflected. Considering the different time organization modes of traffic flow data, the time correlation analysis will be performed according to the correlation analysis of different weeks of a month, days of a week, and periods of a day. Some studies have revealed that the traffic flow data has strong daily and weekly correlations [3–5, 7, 16, 18, 28–31].

The spatial correlation analysis of the traffic flow data includes two main parts: (i) the data correlation analysis of a lane and its transverse adjacent lanes and (ii) the data correlation analysis of a lane and its longitudinal upstream and downstream sections of the same lane. Some studies have proved that the flow data of an urban expressway has a strong correlation with its horizontal and vertical spaces. In theory, the spatial and temporal correlation of the traffic flow data can be used to effectively repair the missing data.

3.2.2. Analysis of the Auxiliary Traffic Flow Data. The core of data repair is to establish a relationship between the known and missing data. The more information related to missing data is contained in the data repair model, the better the effect of data repair will be. The auxiliary traffic flow data refers to the fact that the traffic flow data itself does not have a direct response state of the quantitative indicators. However, in the process of data repair, it can provide certain information about road conditions and environmental characteristics. Thus, the acquisition and use of auxiliary data will greatly help to restore the missing data.

The importance of auxiliary traffic flow data in data repair is explained as follows: Firstly, different weather conditions may affect the operational behaviours of vehicles, for example, in rainy weather, due to the slippery road surface, which will directly affect the speed of vehicles, thus leading to similar running state of vehicles in similar weather conditions. Secondly, the analysis of the running speed of vehicles in different lanes of the same section shows that the vehicles using the middle lane will travel at different speeds compared to the ones in the edge lanes [32, 33]. The difference in the information added for the traffic flow data repair helps to improve the restoration precision of the traffic flow data.

4. The Missing Data Repair Method Based on CMTF

The joint analysis of multisource data improves the understanding of the underlying structure of complex datasets. In the existing research, the repair of traffic flow missing data is mostly based on spatiotemporal information, while the importance of auxiliary information in data repair is neglected. Furthermore, due to the heterogeneity of the latter, there are different quantitative standards in practical application. Therefore, a method is required to integrate the auxiliary information about traffic flow with the spatial-temporal information to repair the missing data.

The CMTF model can fit the matrix data into the tensor model through the coupling and capture the underlying structural characteristics of the heterogeneous data, realizing the joint analysis. Thus, combining with the CMTF model, this paper proposes a missing data repair method based on multisource information.

4.1. The CMTF Representation of the Traffic Flow Data. The schematic diagram of the CMTF model is shown in Figure 3. There are three datasets in the figure: the tensor with missing traffic flow data, the feature matrix of weather feature data, and the lane feature data matrix.

The third-order tensor $\chi \in \mathbb{R}^{I\times J\times K}$ is constructed from the perspective of the spatiotemporal correlation of the main traffic flow data, where I represents the longitudinal

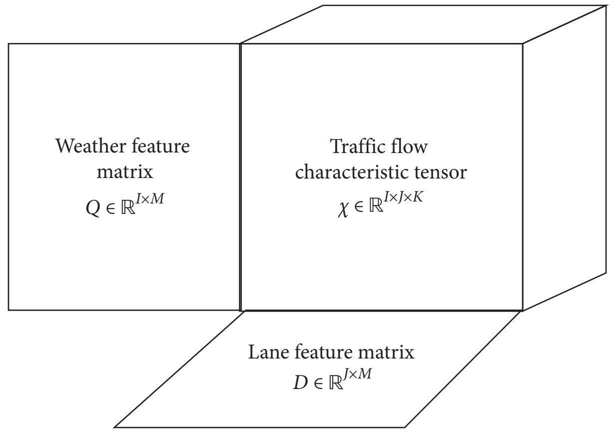

Figure 3: The CMTF model of the missing traffic flow data.

time-series data of the same day in I consecutive weeks, J represents the J lanes of the same detection section, and K represents the K transverse time-series data of a day.

According to the auxiliary traffic flow data, also considering that the third-order tensor contains the same underlying low-rank structure with the construction of the third-order tensor, the two following feature matrices are constructed:

(1) *Weather Feature Matrix*. Different weather conditions, for example, rain, will affect the operational behaviours of vehicles. The wet roads and low visibility will directly affect the speed of vehicles, resulting in a similar running state for all vehicles. The weather data matrix $Q \in \mathbb{R}^{I\times M}$ is in the composition of feature dimension, where I represents the longitudinal time-series data of the same day in I consecutive weeks, and M represents the M transverse time-series data of a day. Here, the sunny, cloudy (foggy), and rainy/snowy days are assigned the values of 1, 2, and 3, respectively. Thus, the element x_{im} in the matrix is an integer varying from 1 to 3.

(2) *Lane Feature Matrix*. The running state of the vehicles traveling towards the same direction in the different lanes of the same section will have typical differences. For example, near the separator, the operating characteristics of the vehicle lane edge will be different, which will be considered in the traffic flow data repair. For the lane feature matrix $D \in \mathbb{R}^{J\times N}$, where J represents the J lanes of the same detection section and N represents the lane attribute values. Starting from the median, lane of attribute assignment is, respectively, 1, 2, 2, and 3. Thus, the elements in matrix y_{jn} are integers varying between 1 and 3.

The tensor containing missing data is constructed for the traffic flow spatiotemporal data, where the lane and weather feature matrices are coupled in lane and date dimensions, respectively. In this way, the data fusion problem is transformed into the CMTF decomposition problem. Thus, the missing data in the original tensor can be repaired by decomposing the tensor and matrix into multiple datasets and then reconstructing the tensor.

4.2. Objective Function Construction

4.2. Objective Function Construction. The factor matrices A_1, B_1, and C are obtained by the CP decomposition of tensor χ, where $A_1 \in \mathbb{R}^{I\times R}$, $B_1 \in \mathbb{R}^{J\times R}$, and $C \in \mathbb{R}^{K\times R}$. Decomposition of matrices Q and D, respectively, gives the factor matrices A_2, V, B_2, and W, where $A_2 \in \mathbb{R}^{I\times M}$, $V \in \mathbb{R}^{I\times M}$, $B_2 \in \mathbb{R}^{J\times N}$, and $W \in \mathbb{R}^{J\times N}$. Thus, the objective function of CMTF decomposition can be expressed as

$$f(A,B,C,V) = \frac{1}{2}\left\|\chi - [[A_1, B_1, C]]\right\|^2 + \frac{1}{2}\left\|Q - A_2 V^T\right\|^2 + \frac{1}{2}\left\|D - B_2 W^T\right\|^2. \tag{6}$$

The missing data in tensor χ can be repaired by coupling it with a known matrix.

Next, denote by F_x the set of the squared errors between the original tensor and the factor matrices A_1, B_1, and C, where the reconstruction tensor is obtained by CP decomposition. The smaller F_x is, the better the repair effect of missing data in the original tensor will be. R is the rank of the tensor, and λ is the regularization parameter. $\|A_1\|_F$, $\|B_1\|_F$, and $\|C\|_F$ are the norms corresponding to each factor matrix. Thus, the tensor χ is decomposed and expanded according to one module CP as

$$F_\chi = \frac{1}{2}\left\|X_{(1)} - A_1 (B_1 \odot C)^T\right\|_F^2 + \frac{\lambda_1}{2}\left(\left\|A_1\right\|_F^2 + \left\|B_1\right\|_F^2 + \|C\|_F^2\right), \tag{7}$$

where $X_{(1)}$ is the matrix representation of the tensor on one modulus. The sets F_Q and F_D represent the squared error between the original matrix and the matrix reconstructed from the factor matrix, which are obtained by matrix decomposition as

$$F_Q = \frac{1}{2}\left\|Q - A_2 V^T\right\|_F^2 + \frac{\lambda_2}{2}\left(\left\|A_2\right\|_F^2 + \|V\|_F^2\right),$$
$$F_D = \frac{1}{2}\left\|D - B_2 W^T\right\|_F^2 + \frac{\lambda_3}{2}\left(\left\|B_2\right\|_F^2 + \|W\|_F^2\right), \quad (8)$$
$$F_f = F_x + F_Q + F_D.$$

According to the analysis of the CMTF model, the matrix should have the same underlying low-rank structure as the tensor in at least one dimension. Therefore, the matrix and the tensor should have a common modulus factor matrix. The assumption is that A_1 and $A_2$are equal; and B_1 and B_2 are equal. Therefore, two global optimal variables $\overline{A}$ and $\overline{B}$ are introduced to satisfy the application premise of the CMTF model:

$$\begin{aligned} A_1 - \overline{A} &= 0,\\ A_2 - \overline{A} &= 0,\\ B_1 - \overline{B} &= 0,\\ B_2 - \overline{B} &= 0. \end{aligned} \quad (9)$$

Thus, the optimization problem of the objective function is described as

$$\min_{\Psi} \quad F_f = F_\chi + F_Q + F_D,$$
$$\text{s.t.} \quad \begin{cases} A_i - \overline{A} = 0 \\ B_i - \overline{B} = 0 \end{cases} \text{and } i = 1, 2, \quad (10)$$

where Ψ represents the factor matrix set: $\Psi = \{A_1, B_1, C, A_2, V, B_2, W\}$. Therefore, the problem of missing data repair is transformed into the problem of solving the optimal value of the objective function F_f.

4.3. Optimization Solution. For optimization, the alternating least square (ALS) and gradient descent method are commonly used in the literature [34–40]. Since both the matrix and the tensor decomposition are a nonconvex problem, the commonly used methods cannot achieve the global optimum. To resolve this issue, the alternating direction multiplier algorithm (ADMM) is suggested [41]. In this method, large global problems are decomposed into many smaller and easier local subproblems, obtaining large global optimal solutions.

Thus, when F_f is finding the optimal solution, it can be interpreted by solving the local F_χ, F_Q, and F_D optimal solution problems. The objective function of this problem can be given as

$$L_\rho\left(\Psi, \Theta_A^1, \Theta_A^2, \Theta_B^1, \Theta_B^2, \overline{A}, \overline{B}\right) = F_f + \sum_{i=1}^{2}\left(\operatorname{tr}\left(\left[\Theta_A^i\right](A_i - \overline{A})\right) + \frac{\rho}{2}\left\|A_i - \overline{A}\right\|_F^2\right) + \sum_{i=1}^{2}\left(\operatorname{tr}\left(\left[\Theta_B^i\right](B_i - \overline{B})\right) + \frac{\rho}{2}\left\|B_i - \overline{B}\right\|_F^2\right), \quad (11)$$

where $\Theta_{A_1}^1, \Theta_{A_2}^1, \Theta_{B_1}^1, \Theta_{B_2}^1$ are the Lagrangian multipliers [31], ρ is the penalty coefficient (>0), and tr is the trace of the matrix, that is, the sum of the diagonal elements in the matrix. The iterative calculation formula of (9) is

$$\begin{aligned} \Psi^{k+1} &\leftarrow \arg\min_{\Psi} L_\rho\left(\Psi, \left(\Theta_A^1\right)_k, \left(\Theta_A^2\right)_k, \left(\Theta_B^1\right)_k, \left(\Theta_B^2\right)_k, \overline{A}_k, \overline{B}_k\right)\\ \overline{A}^{k+1}, \overline{B}^{k+1} &\leftarrow \arg\min_{\overline{AB}} L_\rho\left(\Psi^{k+1}, \left(\Theta_A^1\right)_k, \left(\Theta_A^2\right)_k, \left(\Theta_B^1\right)_k, \left(\Theta_B^2\right)_k, \overline{A}, \overline{B}\right)\\ \Theta_A^i &= \Theta_A^i + \rho\left(A_i - \overline{A}\right) \quad i = 1, 2,\\ \Theta_B^i &= \Theta_B^i + \rho\left(B_i - \overline{B}\right) \quad i = 1, 2. \end{aligned} \quad (12)$$

Concerning L_ρ, the partial derivative of each factor matrix can be obtained as

$$\frac{\partial L_\rho}{\partial A_1} = -\left[X(1) - A_1\left(C \odot B_1\right)^T\right]\left(C \odot B_1\right) + \left(\lambda_1 + \rho\right)A_1 + \Theta_A^1 - \rho\overline{A}. \quad (13)$$

By setting $(\partial L_\rho / \partial A_1) = 0$,

$$A_1 = \left[X_{(1)}\left(C \odot B_1\right) + \rho\overline{A} - \Theta_A^1\right]\left[C^T C * B_1^T B_1 + \left(\lambda_1 + \rho\right)I_R\right]^{-1}. \quad (14)$$

Similarly,

$$B_1 = \left[X_{(2)}\left(C \odot A_1\right) + \rho\overline{B} - \Theta_B^1\right]\left[C^T C * A_1^T A_1 + \left(\lambda_1 + \rho\right)I_R\right]^{-1}, \quad (15)$$

$$C = \left[X_{(3)}\left(B_1 \odot A_1\right)\right]\left[B_1^T B_1 * A_1^T A + \lambda_1 I_R\right]^{-1}, \quad (16)$$

$$A_2 = \left[QV + \rho\overline{A} - \Theta_A^2\right]\left[V^T V + \left(\lambda_2 + \rho\right)I_R\right]^{-1}, \quad (17)$$

$$B_2 = \left[DW + \rho\overline{B} - \Theta_B^2\right]\left[W^T W + \left(\lambda_3 + \rho\right)I_R\right]^{-1}, \quad (18)$$

$$V = \left(Q^T A_2\right)\left[A_2^T A_2 + \lambda_2 I_R\right]^{-1}, \quad (19)$$

$$W = \left(D^T B_2\right)\left[B_2^T B_2 + \lambda_3 I_R\right]^{-1}. \quad (20)$$

Similarly, it can be inferred further that

$$\begin{aligned} \overline{A} &= \frac{\left(\Theta_A^1 + \Theta_A^2\right)}{2\rho} + \frac{\left(A_1 + A_2\right)}{2},\\ \overline{B} &= \frac{\left(\Theta_B^1 + \Theta_B^2\right)}{2\rho} + \frac{\left(B_1 + B_2\right)}{2}. \end{aligned} \quad (21)$$

Table 1: The main steps of CMTF.

Input: the tensor $\chi \in \mathbb{R}^{I\times J\times K}$ with missing traffic flow data, matrices $Q \in \mathbb{R}^{I\times M}$ and $D \in \mathbb{R}^{J\times N}$, tensor rank R, initial iteration number *iter*, maximum iteration number *MaxIter*
Output: completed tensor $\chi \in \mathbb{R}^{I\times J\times K}$
(1) Initialize $\rho, \lambda_1, \lambda_2, \lambda_3 R, A_1, B_1, CA_2, V, B_2, W\,(\Theta_A^1, \Theta_A^2, \Theta_B^1, \Theta_B^2) = 0$, iter = 0
(2) Calculate $\overline{A}$ using (22), and $\overline{B}$ using (23)
(3) If $iter < MaxIter$, repeat Step 4 ~ Step 8
(4) Using (14)–(16), calculate and update A_1, B_1, C, respectively
(5) Using (17)–(20), calculate and update A_2, V, B_2, W, respectively
(6) Using (22), calculate and update $\overline{A}$. Similarly, calculate and update $\overline{B}$ using (23)
(7) Using (12), calculate and update $\Theta_{A_1}^1, \Theta_{A_2}^1, \Theta_{B_1}^1, \Theta_{B_2}^1$
(8) Set $iter + 1 = iter$. If the algorithm reaches the convergence standard, the loop is broken. Otherwise, return to Step 3
(9) Return to $\chi = A_1 \circ B_1 \circ C$

If set $(\Theta_A^i)_0 = 0, i = 1, 2,$ then $\sum_{i=1}^{2} (\Theta_A^i)_k = 0, k = 1, 2, \ldots, I_{\max}$, and it can be inferred further that

$$\overline{A} = \frac{(A_1 + A_2)}{2}, \tag{22}$$

$$\overline{B} = \frac{(B_1 + B_2)}{2}. \tag{23}$$

Based on the above-given analysis, the main steps of the CMTF-based traffic flow missing data repair method are shown in Table 1.

5. Results and Discussion

5.1. Data Source and Analysis

5.1.1. Data Sources. The main section of an urban expressway is selected as the area to be analysed. The section includes five detection sections (in the east and west) with four lanes in each section and a total of 40 coil detectors. The data that will be collected by the detectors is shown in Table 2, where "TY" indicates the code of the detection section, "D" refers to the east, and "10 (2)" means Lane 2 of the detection section #10. The data is sampled from August 27, 2018, to September 29, 2018, with a sampling interval of 5 min. There are 161280 groups of data.

Table 2: Data types.

Detector number	Attribute	Detection time	Flow	Speed	Occupancy
TYD10(2)	Valid	2018-8-27 0:05	32	0	4
TYX10(1)	Invalid	2018-8-27 0:05	0	0	13
TYD12(1)	Lost	2018-8-27 0:10	0	0	0

5.1.2. Judgment Criteria for Data Repair. The existence and reparability of the missing flow data have to be judged before starting a repair process. To determine whether anything is missing, the traffic flow data can be used. For example, if the flow data is displayed as "0," it might be due to no vehicle passing through the detector or might be due to some missed data.

Under normal traffic conditions, the missing traffic flow data can be repaired according to the changing trend in the traffic flow. However, under abnormal conditions, such as traffic accidents conditions, once the data is "0," it is impossible to judge the existence and (if exists) the reparability of the data.

Following the continuous progress in detection technology, the detectors have become more capable of realizing the preliminary discrimination of data anomalies. For example, the location detector can classify the data as valid, invalid, and/or lost. For repair judgment, this paper only considers the invalid data with the value of "0," which is denoted as the missed data that will be repaired.

5.2. Experimental Scheme Design.

(1) *Data Representation.* Let us construct the tensor $\chi \in \mathbb{R}^{5\times 4\times 288}$ for the target data, where "5" represents "Day-mode," which means the longitudinal time-series data of five consecutive Mondays, "4" represents "Lane-mode," which means the transverse spatial sequence data of four lanes in the same detection section, and "288" represents "Time-mode," which means the 288 transverse time-series data obtained in one day. The weather feature matrix is $Q \in \mathbb{R}^{5\times 288}$, where "5" refers to five longitudinal time-series data of Monday, and "288" represents the 288 transverse time-series data obtained in one day. The lane feature matrix is $D \in \mathbb{R}^{4\times 3}$, where "4" refers to the data of four lanes of the same detection section, and "3" represents the lane attribute values.

(2) The evaluation index of the missing data repair effect. The root mean square error (RMSE) and mean absolute percentage error (MAPE) are used as evaluation indexes [14]. It should be noted that the smaller the difference between the repaired and actual data is, the smaller the RMSE and MAPE will be. These errors are calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}},$$
$$\text{MAPE} = \frac{\sum_{i=1}^{n} \left|(\hat{y}_i - y_i/y_i)\right|}{n} \times 100\%, \tag{24}$$

where y_i is the actual data, $\hat{y}_i$ is the repaired data, and n is the number of repaired pieces of data.

(3) *Missing Data Type Design*. To illustrate the data repair performance of the CMTF, three types of missing data were manually set up: MCAR, MAR, and MIX-M. Then, the data missing rates of each type are set to 10%, 20%, . . ., 90%. In order to avoid the influence of the randomly set data missing position on the repair result, ten settings were made for different data missing types, and the average value of the 10 repair effects (RMSE and MAPE) corresponding to each method was selected as the final repair result.

5.3. Analysis of the Experimental Results. The experiment environment is designed using Matlab-R2014a software, where the data repair based on the CMTF algorithm is realized by the CMTF_Toolbox_v1_1toolbox [15]. The relevant parameters of the model are obtained by cross-validation.

The TYD12(2) detector with relatively complete flow data on August 27, 2018, was selected for tests, which were loaded with three missing data types (MCAR, MAR, and MIX-M) and nine missing rates (10% to 90%).

Three methods, namely, matrix repair method [13] (Method 1), tensor repair algorithm (Method 2) [20], and the optimization algorithm of CMTF based on alternating least square (Method 3), are compared with the algorithm proposed in this paper (Method 4). The specific analysis of the data repair effect is as follows.

5.3.1. Repair Effect Evaluation of the MACR. For the MACR, the repair results are shown in Tables 3 and 4.

The results show that the RMSE and MAPE of all methods are relatively small and close to each other for low data missing rates. However, with the gradual increase of the missing rate, the RMSE and MAPE for all methods also increase gradually. For high missing rates, the RMSE of Method 3 and Method 4 remains around 15, which is significantly lower than that of the other two methods. This indicates that the multidimensional flow data has a significant effect on the repair of the missing data. The RMSE and MAPE of Method 4 are generally lower than the corresponding values of Method 3, indicating that the local optimization solution based on ADMM is better than the one based on ALS.

5.3.2. Repair Effect Evaluation of the MAR. For the MAR, the repair results are shown in Tables 5 and 6.

The results show that the RMSE and MAPE of all methods are relatively small for low missing rates similar to MACR. However, with the gradual increase of the miss rate, the RMSE and MAPE values also increase gradually, where Method 4 has the slowest growth rate. For high missing rates, the RMSE of Method 3 and Method 4 is still lower than 16.5%, which is significantly lower than the other two methods. This indicates that the multisource flow data has a more significant effect on the repair.

Table 3: The RMSE of the methods under the MACR.

MACR	Missing rates (%)	RMSE			
		Method 1	Method 2	Method 3	Method 4
1	10	3.47	2.32	1.566	1.336
2	20	5.39	2.98	1.991	1.791
3	30	8.04	6.15	5.89	5.174
4	40	10.71	7.99	7.87	7.337
5	50	12.11	11.46	10.767	9.767
6	60	13.35	12.75	11.252	10.902
7	70	15.033	14.233	13.354	12.654
8	80	17.61	17.62	13.96	13.236
9	90	20.95	18.47	15.19	14.422

The repair effect of Method 4 is lower than that of Method 3, indicating that the local optimization solution based on ADMM is better than the one based on ALS. Compared to the MACR, the RMSE and MAPE of all methods under the MAR are relatively larger. This is because the missing data is continuous, which will affect the repair effect to some extent.

5.3.3. Repair Effect Evaluation of the MIX-M. For the MIX-M, the repair results are shown in Tables 7 and 8.

The overall trend is similar to that of MACR and MAR, but the RMSE and MAPE of all methods are in the middle. This is because the missing data in MIX-M includes both the discrete missing data of MACR and the continuous missing data of MAR. Thus, the existence of continuous missing data will affect the repair process.

The MIX-M can reflect the lack of traffic flow data. From Figure 4, for the missing rate of 50%, the MAPE of all methods is lower than 21%. When the missing rate exceeds 60%, the MAPE difference between the methods becomes larger. Here, Method 4 still has the best repair effect for high data missing rates.

Under the three types of data missing, all methods can repair the missing data at certain levels, where the repair effect of the MACR is the best, and the MAR is the worst.

In comparison, the repair effect of Method 2 is better than that of Method 1. This indicates that the tensor algorithm based on multidimensional data can improve the repair effect.

The repair effect of Method 3 and Method 4 is better than the other two, indicating that the addition of auxiliary data is useful in data repair.

The repair effect of Method 4 is better than that of Method 3; that is, the optimal solution of ADMM is better than the traditional solution that ALS offers.

Our proposed method has a best repair effect, which has the lowest RMSE and MAPE among all methods. Even for

Table 4: The MAPE of the methods under the MACR.

MACR	Missing rate (%)	MAPE			
		Method 1 (%)	Method 2 (%)	Method 3 (%)	Method 4 (%)
1	10	5.10	2.90	2.50	1.80
2	20	7.28	3.98	3.70	2.98
3	30	9.63	8.97	8.20	6.25
4	40	14.18	13.78	12.80	10.80
5	50	19.45	17.44	14.11	12.31
6	60	21.61	19.12	15.47	13.33
7	70	25.70	19.98	15.95	14.75
8	80	27.80	21.26	16.21	15.10
9	90	32.43	25.67	17.45	15.97

Table 5: The RMSE of the methods under the MAR.

MAR	Missing rates (%)	RMSE			
		Method 1	Method 2	Method 3	Method 4
1	10	4.71	3.2	2.73	2.331
2	20	7.09	4.12	3.78	3.054
3	30	10.42	9.43	7.97	6.28
4	40	11.4	11.72	10.21	9.631
5	50	15.27	14.98	14.05	11.536
6	60	17.07	16.27	14.28	13.417
7	70	20.33	18.56	15.59	14.89
8	80	23.61	19.37	16.01	15.733
9	90	30.86	21.3	16.46	16.193

Table 6: The MAPE of the methods under the MAR.

MAR	Missing rates (%)	MAPE			
		Method 1 (%)	Method 2 (%)	Method 3 (%)	Method 4 (%)
1	10	6.88	3.77	3.26	2.26
2	20	8.77	4.81	4.83	4.18
3	30	10.97	9.09	9.55	7.99
4	40	15.88	14.26	13.77	11.89
5	50	21.99	19.87	17.27	14.56
6	60	24.85	20.33	18.63	15.71
7	70	28.37	22.31	19.37	16.45
8	80	30.55	24.69	19.93	17.97
9	90	34.69	30.71	20.11	19.54

Table 7: The RMSE of the methods under the MIX-M.

MIX-M	Missing rates (%)	RMSE			
		Method 1	Method 2	Method 3	Method 4
1	10	4.05	2.63	2.05	1.412
2	20	6.05	3.41	2.92	2.141
3	30	9.55	8.73	7.88	5.976
4	40	11.38	10.5	9.51	8.3
5	50	14.6	14.3	13	10.324
6	60	16.04	15.78	13.92	12.875
7	70	19.4	16.46	14.36	13.684
8	80	22.5	18.75	15.51	14.2
9	90	27.42	20.31	15.98	14.996

Table 8: The MAPE of the methods under the MIX-M.

MIX-M	Missing rates (%)	MAPE			
		Method 1 (%)	Method 2 (%)	Method 3 (%)	Method 4 (%)
1	10	6.90	3.42	2.91	1.94
2	20	8.23	4.76	3.89	3.28
3	30	9.97	8.88	8.65	7.47
4	40	13.88	14.11	13.14	11.76
5	50	21.00	18.56	16.72	12.99
6	60	23.55	19.76	17.43	14.81
7	70	26.57	20.12	18.97	15.95
8	80	29.65	22.25	19.00	16.77
9	90	35.07	27.81	19.99	17.04

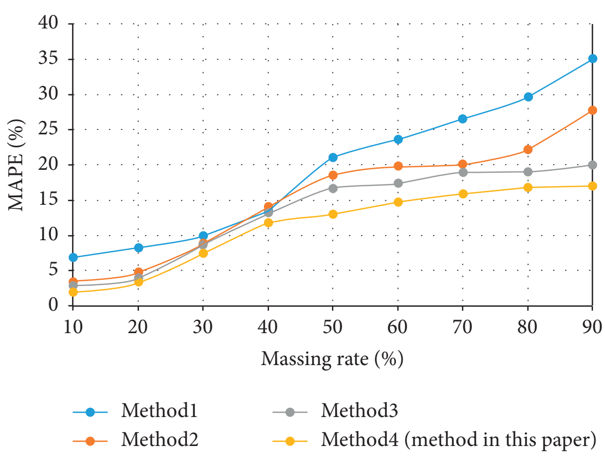

Figure 4: The MAPE versus missing rate for all methods under the MIX-M.

the MAR with a miss rate of more than 80%, the RMSE and MAPE of our method are still around 16.5 and 20%, respectively. This shows that the accuracy of traffic flow missing data repair can be effectively improved by adding auxiliary data based on sufficient and reasonable use of the spatial and temporal correlation between the traffic flow data.

6. Conclusions

To improve the repair effect of traffic flow missing data, this paper proposes a CMTF-based repair method. Different from the existing methods, the proposed method can make full and reasonable use of the spatial and temporal data of traffic flow to further increase the auxiliary traffic flow data and use the ADMM algorithm for optimization solution. The results showed that the proposed algorithm outperformed the other three comparison methods; that is, it has a better repair effect under three data deletion modes and nine data deletion rates. It can also achieve the best repair accuracy under high data deletion rates. The acquisition of complete traffic flow data realized by the proposed method will be highly beneficial in traffic management and control.

Since this paper only considered the coupling of a 3D tensor and a 2D matrix, future research will focus on the coupling of multiple matrices based on the multidimensional tensor. While doing that, we will also consider the total time cost of the running model.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51638004).

References

[1] D. L. Gold, S. M. Turner, and B. J. Gajewski, *Imputing Missing Values in its Data Archives for Intervals under 5 Minutes*, pp. 4804–4612, Transportation Research Board 80th Annual Meeting, Washington, DC, USA, 2001.

[2] J. H. Conklin and B. L. Smith, "Use of local lane distribution patterns to estimate missing data values from traffic monitoring systems," *Transportation Research Board*, vol. 1811, no. 1, pp. 50–56, 2002.

[3] B. L. Smith, W. T. Scherer, and J. H. Conklin, "Exploring imputation techniques for missing data in transportation management systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1836, no. 1, pp. 132–142, 2003.

[4] G. Y. Jiang and L. H. Gang, "Dynamic traffic data fault diagnosis and repair method," *Journal of Transportation Engineering*, vol. 4, no. 1, pp. 121–125, 2004.

[5] Y. W. Jin and J. C. Wang, "Research on traffic flow parameter correction method based on correlation analysis," *Control Engineering of China*, vol. 15, no. 0, pp. 142–144+147, 2008.

[6] I. A. Gheyas and L. S. Smith, "A neural network-based framework for the reconstruction of incomplete data sets," *Neurocomputing*, vol. 73, no. 6, pp. 3539–3560, 2010.

[7] W. Wang, Z. Y. Cheng, M. Y. Liu et al., "Traffic flow fault data repair method based on temporal and spatial correlation," *Journal of Zhejiang University (Engineering Science)*, vol. 51, no. 9, pp. 1727–1737, 2017.

[8] Z. Chen, L. Zhang, G. Tian, and E. A. Nasr, "Economic maintenance planning of complex systems based on discrete artificial bee colony algorithm," *IEEE Access*, vol. 8, no. 8, pp. 108062–108071, 2020.
[9] M. Kurucz, A. A. Benczur, and K. Csalogany, "Methods for large scale SVD with missing values," in *Proceedings of KDD Cup and Workshop*, pp. 31–38, 2007.
[10] L. Q. Qu, L. Y. Li, J. Zhang, and J. Hu, "PPCA-Based missing data imputation for traffic flow volume: a systematical approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 512–522, 2009.
[11] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 29–40, 2015.
[12] B. C. Lu, G. L. Guo, W. Q. Xiao et al., "Fault diagnosis and repair of dynamic traffic data based on multi-scale principal component analysis," *Natural Science*, vol. 1, pp. 134–174, 2016.
[13] Y. Q. Jiang, "An algorithm of imputing missing network data based on matrix low-rank decomposition," Dissertation, Chang'an University, Xi'an, China, 2018.
[14] J. Wang, W. Deng, and Y. Guo, "New Bayesian combination method for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 1, pp. 79–94, 2014.
[15] E. Acar, T. G. Kolda, and D. M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," *KDD Workshop on Mining and Learning with Graphs*, Sandia National Laboratories, Livermore, CA, USA, 2011.
[16] C. D. Silva and F. J. Herrmann, "Hierarchical tucker tensor optimization-application to tensor completion," in *Proceedings of the 10th International Conference on Sampling Theory and Application*, pp. 384–387, Bremen, Germany, July 2013.
[17] H. Tan, J. Feng, Z. Chen et al., "Low multilinear rank approximation of tensors and application in missing traffic data," *Advances in Mechanical Engineering*, vol. 6, pp. 1–12, 2015.
[18] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.
[19] Y. Xiang, "Traffic data recovery based on tensor filling and tensor matrix correlation analysis," M. S. thesis, Yunnan University, Kunming, China, 2019.
[20] C. H. Zhang, "Research on data quality evaluation and repair method of urban expressway traffic flow," M. S. thesis, China Beijing Jiaotong University, Beijing, China, 2019.
[21] 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, no. 3, pp. 337–355, 2019.
[22] B. Yang, "Research and application of machine learning methods based on tensor data," M. S. thesis, China Agricultural University, Beijing, China, 2014.
[23] X. Luo, H. Wu, H. Yuan, and M. Zhou, "Temporal pattern-aware QoS prediction via biased non-negative latent factorization of tensors," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1798–1809, 2020.
[24] C. Leng, H. Zhang, G. Cai, I. Cheng, and A. Basu, "Graph regularized Lp smooth non-negative matrix factorization for data representation," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, pp. 584–595, 2019.
[25] Y. Zhang, P. Zhou, G. M. Cui et al., "Multi-model based PSO method for burden distribution matrix optimization with expected burden distribution output behaviors," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1506–1512, 2019.
[26] X. Chen, Z. Wei, Z. Li, J. Liang, Y. Cai, and B. Zhang, "Ensemble correlation-based low-rank matrix completion with applications to traffic data imputation," *Knowledge-Based Systems*, vol. 132, no. 11, pp. 249–262, 2017.
[27] L. Li, J. Zhang, Y. Wang, and B. Ran, "Missing value imputation for traffic-related time series data based on a multi-view learning method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2933–2943, 2019.
[28] Y. R. Huang, Z. Xiao, D. Wang, H. B. Jiang, and D. Wu, "Exploring individual travel patterns across private car trajectory data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 168394–168404, 2020.
[29] M. Tang, Z. Li, G. Tian et al., "A Data-Driven-Based wavelet support vector approach for passenger flow forecasting of the metropolitan hub," *IEEE Access*, vol. 7, no. 2, pp. 7176–7183, 2019.
[30] L. Chao, B. W. Du, Y. G. Wang et al., "Estimation of missing values in heterogeneous traffic data: application of multimodal deep learning model," *Knowledge-Based Systems*, vol. 194, no. 22, 2020.
[31] H. Jiang and H. Deng, "Traffic incident detection method based on factor Analysis and weighted random forest," *IEEE Access*, vol. 8, pp. 168394–168404, 2020.
[32] J. J. Tang, X. S. Zhang, W. Q. Yin et al., "Missing data imputation for traffic flow based on combination of fuzzy neural network and rough set theory," *Journal of Intelligent Transportation Systems*, 2020.
[33] X. Y. Chen, J. M. Yang, L. J. Sun et al., "A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 117, 2020.
[34] M. Safaeian, A. M. Fathollahi-Fard, G. Tian, Z. Li, and H. Ke, "A multi-objective supplier selection and order allocation through incremental discount in a fuzzy environment," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 1, pp. 1435–1455, 2019.
[35] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 601–614, 2019.
[36] Z. Shi and Z. Wang, "Optimal control for a class of complex singular system based on adaptive dynamic programming," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 1, pp. 188–197, 2019.
[37] W. J. Wang, G. D. Tian, M. N. Chen et al., "Dual-objective program and improved artificial bee colony for the optimization of energy-conscious milling parameters subject to multiple constraints," *Journal of Cleaner Production*, vol. 245, no. 1, pp. 124–135, 2020.
[38] X. Luo, M. C. Zhou, Y. Xia et al., "Generating highly accurate predictions for missing QoS-data via aggregating non-negative latent factor models," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 579–592, 2016.
[39] Z. M. Lu, L. Q. Wang, Z. Y. Han et al., "Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 838–849, 2019.

[40] X. Luo, M. Zhou, Z. Wang, Y. Xia, and Q. Zhu, “An effective scheme for QoS estimation via alternating direction method-based matrix factorization,” *IEEE Transactions on Services Computing*, vol. 12, no. 4, pp. 503–518, 2019.

[41] B. Stephen, P. Neal, C. Chu et al., “Distributed optimization and statistical learning via the alternating irection method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.