Hindawi
Journal of Sensors
Volume 2019, Article ID 7092713, 11 pages
https://doi.org/10.1155/2019/7092713

Research Article

Traffic Data Imputation Algorithm Based on Improved Low-Rank Matrix Decomposition

Xianglong Luo,[1,2] **Xue Meng**,[1] **Wenjuan Gan**,[1] **and Yonghong Chen**[1]

[1] *School of Information Engineering, Chang'an University, Xi'an 710064, China*
[2] *School of Highway, Chang'an University, Xi'an 710064, China*

Correspondence should be addressed to Xue Meng; 2017124037@chd.edu.cn

Received 16 January 2019; Revised 8 May 2019; Accepted 11 June 2019; Published 1 July 2019

Academic Editor: Eduard Llobet

Traffic data plays a very important role in Intelligent Transportation Systems (ITS). ITS requires complete traffic data in transportation control, management, guidance, and evaluation. However, the traffic data collected from many different types of sensors often includes missing data due to sensor damage or data transmission error, which affects the effectiveness and reliability of ITS. In order to ensure the quality and integrity of traffic flow data, it is very important to propose a satisfying data imputation method. However, most of the existing imputation methods cannot fully consider the impact of sensor data with data missing and the spatiotemporal correlation characteristics of traffic flow on imputation results. In this paper, a traffic data imputation method is proposed based on improved low-rank matrix decomposition (ILRMD), which fully considers the influence of missing data and effectively utilizes the spatiotemporal correlation characteristics among traffic data. The proposed method uses not only the traffic data around the sensor including missing data, but also the sensor data with data missing. The information of missing data is reflected into the coefficient matrix, and the spatiotemporal correlation characteristics are applied in order to obtain more accurate imputation results. The real traffic data collected from the Caltrans Performance Measurement System (PeMS) are used to evaluate the imputation performance of the proposed method. Experiment results show that the average imputation accuracy with proposed method can be improved 87.07% compared with the SVR, ARIMA, KNN, DBN-SVR, WNN, and traditional MC methods, and it is an effective method for data imputation.

1. Introduction

With the rapid development of the social economy, many kinds of the massive road infrastructure are implemented [1–4] but traffic congestion still exists in the highway. Therefore, it is necessary to collect the information of the highway for the convenience of people's travelling demand. With the development of information technology, the collection of highway information becomes possible, and the collection equipment used for highways includes Bluetooth sensor, remote traffic microwave sensor, video sensors, and loop detectors. However, traffic flow data are lost in different degrees due to sensor damage, malfunction, or transmission errors, etc. Missing data makes it difficult to extract valid information from traffic data. Meanwhile, the missing data is also an obstacle in the traffic and the travel time prediction field [5–8], and the integrity of traffic flow data is the premise of data analysis in ITS. Therefore, it is very important to put forward an effective traffic data imputation method. At present, various methods have emerged in the field of traffic flow data imputation. These imputation methods can be roughly divided into three categories: prediction methods, interpolation methods, and statistical learning methods [9].

Traffic flow prediction models [10–12] are critical for road traffic management in complex road networks. Prediction methods usually build predictive models with historical data and treat missing data as values to be predicted. There are many ways to build traffic flow prediction models, from a simple null value imputation to complex spatiotemporal imputation models [13]. The representative prediction methods include Autoregressive Integrated Moving Average Model (ARIMA) [14–16], Bayesian networks (BNs) [17–19], and support vector regression (SVR) [20, 21]. Elshenawy et al. [22] proposed an intelligent data imputation method

with ARIMA model and presented a mechanism based on Hyndman-Khandakar algorithm to determine ARIMA parameters. Sun et al. [23] partitioned a day into different time section and used SVR to forecast traffic flow data. Chen et al. [24] proposed an Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARIMA-GARCH) model for traffic flow prediction. However, these prediction methods failed to utilize the sensor information with missing data, which would affect data imputation accuracy.

Interpolation methods are divided into temporal-neighboring and pattern-neighboring [25]. Temporal-neighboring methods fill up the missing data by the known data from the same sensors at the same daily time but on some neighboring days [20, 26]. Pattern-neighboring methods use the similarity characteristics of the daily traffic flow data [27] and estimate missing data using historical data collected from the same sensors on different days [17, 20]. The typical pattern-neighboring methods include K-nearest neighbors (KNN) model [28, 29] and Local Least Squares (LLS) [30, 31] model, and the key difficulty of these methods is to determine the neighbors by an appropriate distance metric [32, 33]. Nguyen et al. [34] used the mean value of the historical data to estimate missing data. Smith et al. [35] used historical data or the data from surrounding periods and locations to impute the missing data. The interpolation model assumes that the daily traffic flow data are similar, but the actual traffic flow data fluctuates and changes with time. Therefore, it is impossible to obtain satisfactory imputation performances.

The method based on statistical learning has been developed in recent years. This method primarily assumed the probability distribution model of traffic data and used iterative methods to estimate the parameters of the probability distribution. Then the observed data was used to impute the missing data. The statistical learning methods include Probabilistic Principal Component Analysis (PPCA) [6, 9], Bayesian Principal Component Analysis (BPCA) [26], neural network method [36], and Markov Chain Monte Carlo (MCMC) [37]. The MCMC is a typical imputation method based on statistical learning. The basic idea of the MCMC method regards the missing data as the target parameter and estimate the parameter by the sample values of the parameter. Y Higashijima et al. [38] proposed a regression tree imputation method and used a preprocessing method to improve imputation accuracy. Wei et al. [39] proposed a data-driven imputation method and used k-means clustering to group the most correlated road segments; the trained model is able to estimate the missing data at multiple locations under a unified framework. Although the methods based on statistical learning have strong hypothesis about traffic data, their performance is superior to traditional imputation methods [40] because the assumed probability distribution captures the essentials of traffic flow.

The methods based on prediction and interpolation simply impute the data with the temporal or spatial correlation characteristic and only consider the information of historical data. The historical imputation methods fill the missing data with the known data point collected on the same sensors at the same daily time but from different days. These methods require higher stability of historical data, but traffic flow data is usually unstable and fluctuate to some extent in practical applications. The traditional imputation method sets all the missing data to zero and uses the data matrix with zero-padding to participate the operation for the data imputation, which cannot consider the impact of missing sensor data into the imputation result. Generally, the sensors including missing data have the highest correlation with final imputation results. However, the missing data is set to zero directly in the traditional imputation method, which ignores the effect of the missing data on the imputation results and reduces the accuracy of the imputation results. In order to address the above problems, a traffic data imputation method is proposed based on improved low rank matrix decomposition (ILRMD). Compared with the traditional imputation method, the ILRMD method fully considers the impact of missing data in the imputation results. In the process of data imputation, the ILRMD method does not directly discard the information of missing data, and the effect of missing data is reflected in the coefficient matrix. The reconstructed data matrix multiplied by the coefficient matrix, containing the missing data information, is the imputation result. The ILRMD method uses not only the traffic data around the sensor including missing data, but also sensor data with data missing. The information contained in the missing data is fully considered, and the spatiotemporal correlation characteristics of the traffic flow are adequately utilized. The tested results with traffic data collected from the Caltrans Performance Measurement System (PeMS) show that the proposed algorithm has superior imputation accuracy.

The rest of this paper is organized as follows. Section 2 reviews the related work in traffic data imputation and gives a brief introduction. The traditional imputation approach is introduced in Section 3. Section 4 describes the ILRMD method proposed in this paper. Section 5 discusses the result analysis and method comparison. Section 6 makes the conclusion of this paper and gives some recommendations.

2. Related Work

With the rapid development of machine learning, pattern recognition, computer vision, and data mining, the processing of big data is becoming more and more important. The scale and growth rate of big data are continuously increasing, but large-scale high-dimensional data is often correlative and redundant. Therefore, it is necessary to perform reasonable compression processing on large-scale data. In order to reduce data redundancy, Candes [41] proposed the concept of low rank sparse matrix decomposition in 2009, which is also called Low-Rank Matrix Recovery (LRMR), Low-Rank Matrix Decomposition (LRMD), or Robust Principal Component Analysis (RPCA).

2.1. Low-Rank Matrix Decomposition. For a given data matrix $D \in R^{m\times n}$ distributed in a linear subspace with approximately low dimension, it can be decomposed into a low-rank matrix

A and a sparse matrix E [42].

$$\begin{aligned} \min_{A,E} \quad & \text{rank}\ (A) + \lambda \|E\|_0 \\ \text{s.t.} \quad & D = A + E \end{aligned} \tag{1}$$

where $\|E\|_0$ represents the L_0 norm of the matrix E and λ represents the compromise factor of matrices A and E.

Since the optimization problem of (1) is a NP-hard problem, it can be relaxed to the convex optimization problem [41–43], which is noted as follows:

$$\begin{aligned} \min_{A,E} \quad & \|A\|_* + \lambda \|E\|_{2,1} \\ \text{s.t.} \quad & D = A + E \end{aligned} \tag{2}$$

where $\|\mathrm{A}\|_*$ represents the nuclear norm of matrix A; $\|\mathrm{E}\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} \mathrm{e}_{ij}{}^2}$ is the L_{21} norm of the matrix E.

The low-rank characteristic of recovered matrix determines the matrix imputation performance. Therefore, choosing the suitable LRMD solution method is crucial. The main algorithms for solving LRMD problem include Iterative Threshold method [44, 45], the Dual Approach [46], Accelerated Proximal Gradient Algorithm [47], and Augmented Lagrange Multiplier method [48]. In this paper, Augmented Lagrange Multiplier method is used.

2.2. Matrix Imputation Based on Low-Rank Matrix Decomposition. Generally, we cannot recover all the data with partial sample data. But Candes [42] proved that the missing data can be recovered more accurately when data matrix is low or near low rank. From the Section 2.1, the low rank matrix A is acquired based on LRMD, which can be used to impute the missing data.

The model of matrix imputation can be noted as follows:

$$\begin{aligned} \min_{\mathrm{A}} \quad & \text{rank}\,(A) \\ s.t. \quad & P_\Omega\,(A) = P_\Omega\,(D) \end{aligned} \tag{3}$$

where Ω is the set of known element subscripts, and $\Omega \subseteq R^{\mathrm{m}\times\mathrm{n}}$, $P_\Omega : R^{\mathrm{m}\times n} \longrightarrow R^{m\times n}$ is a linear projection operator, which can be defined as follows:

$$P_\Omega\left(D_{ij}\right) = \begin{cases} D_{ij} & (i,j) \in \Omega \\ 0 & (i,j) \notin \Omega \end{cases} \tag{4}$$

The optimization problem of (3) is also a NP-hard problem, so it needs to be relaxed into a convex optimization problem:

$$\begin{aligned} \min_{A} \quad & \|\mathrm{A}\|_* \\ s.t. \quad & P_\Omega\,(A) = P_\Omega\,(D) \end{aligned} \tag{5}$$

2.3. Matrix Imputation Based on Low-Rank Matrix Representation. The low-rank matrix imputation method mentioned above directly minimizes the rank of imputed data. In order to improve imputation efficiency, a self-expression is applied to LRMD, which is called the low-rank matrix representation [49, 50]. The data matrix D is represented as a linear combination with a dictionary matrix B, that is, $D = BZ$. The matrix Z is the coefficient matrix, and it is expected to be low rank. Z can be obtained by solving the optimization problem in the following:

$$\begin{aligned} \min_{\mathrm{Z}} \quad & \text{rank}\,(Z) \\ s.t. \quad & D = BZ \end{aligned} \tag{6}$$

Equation (6) can be convexly relaxed to obtain the following:

$$\begin{aligned} \min_{\mathrm{z}} \quad & \|Z\|_* \\ \text{s.t.} \quad & D = BZ \end{aligned} \tag{7}$$

If the data matrix D is selected as the dictionary matrix, (7) can be noted as follows:

$$\begin{aligned} \min_{\mathrm{z}} \quad & \|Z\|_* \\ \text{s.t.} \quad & D = DZ \end{aligned} \tag{8}$$

In practical applications, the data matrix D may be disturbed by noise. In order to enhance the robustness, (8) can be revised as follows:

$$\begin{aligned} \min_{Z,E} \quad & \|Z\|_* + \lambda \|E\|_{2,1} \\ s.t. \quad & D = DZ + E \end{aligned} \tag{9}$$

A data matrix D is represented by a data dictionary B, and the coefficient matrix Z is sparser when D has higher similarity with B. But the stochastic noise is usually appended in data matrix D, which will influence the correlation within the data matrix. When the stochastic noise E is removed, the correlation of data matrix can be enhanced. D is selected as a dictionary, and its essence is to reveal the correlation within the data matrix. When the coefficient matrix Z is sparse, data columns in data matrix D are represented by each other's columns with few coefficients as possible. For the traffic flow data, it has high spatiotemporal correlation characteristics, but it is affected by the weather, holidays, and other factors, which makes the traffic flow data have stochastic volatility. Therefore, if the influence of this stochastic volatility on the traffic data is removed, the correlation between the traffic data will be enhanced. After removing the influence of stochastic noise, the correlation between the data itself is further explored, and the similarity between the data is expressed with as little information as possible. Then the internal correlation of traffic flow data is used to impute the data.

2.4. The Solution of the Coefficient Matrix. In order to obtain the solution of (9), a variable J is introduced and let $J = Z$ to separate the variable Z. The coefficient matrix Z can be

calculated with the Augmented Lagrange Multiplier method, and the optimization model becomes the following:

$$\begin{aligned} \min_{z} \quad & \|J\|_* + \lambda \|D - DZ\|_{2,1} \\ s.t. \quad & J = Z \end{aligned} \tag{10}$$

Construct an Augmented Lagrange function as (11), where Y is a Lagrange Multiplier, $\|\cdot\|_{\mathbf{F}}^2$ is Fibonacci norm, which represents the sum of the absolute squares of elements, and μ is a weight to tune the error term $\|Z - J\|_F^2$.

$$\begin{aligned} L(Z, J) = \|J\|_* + \frac{1}{2}\|D - DZ\|_F^2 + \langle Y, Z - J\rangle \\ + \frac{\mu}{2}\|Z - J\|_F^2 \end{aligned} \tag{11}$$

The Exact Augmented Lagrange Multiplier (EALM) method is used to solve the matrices J and Z according to the following:

$$\begin{aligned} J^{k+1} &= \arg\min_{J} L\left(Z^k, J\right) \\ &= \arg\min_{J} \|J\|_* + \langle Y, Z - J\rangle + \frac{\mu}{2}\|Z - J\|_F^2 \\ &= \arg\min_{J} \|J\|_* + \frac{\mu}{2}\left\|Z - J + \frac{Y}{\mu}\right\|_F^2 \end{aligned} \tag{12}$$

$$\begin{aligned} Z^{k+1} &= \arg\min_{Z} L\left(Z, J^{k+1}\right) \\ &= \arg\min_{Z} \frac{1}{2}\|D - DZ\|_F^2 + \langle Y, Z - J\rangle \\ &\quad + \frac{\mu}{2}\|Z - J\|_F^2 \\ &= \arg\min_{Z} \frac{1}{2}\|D - DZ\|_F^2 + \frac{\mu}{2}\left\|Z - J + \frac{Y}{\mu}\right\|_F^2 \end{aligned} \tag{13}$$

The updating of the coefficient matrix Z is as follows. Firstly, a projection matrix W is used to express the unmissing position of the matrix D, and $D = W.*(D)$. For convenience, set $G = J - Y/\mu$ and (13) can be expressed as follows:

$$\min_{Z} \frac{1}{2}\|W.*(D) - W.*(DZ)\|_F^2 + \frac{\mu}{2}\|Z - G\|_F^2 \tag{14}$$

In order to get a derivative about Z in (14), the cross product should be changed to inner product. The matrices of (14) are spread in column as follows:

$$\begin{aligned} \min_{Z} \frac{1}{2}\sum_{i=1}^{n}\left\|\mathbf{w}_i \cdot * \left(\mathbf{d}_i\right) - \mathbf{w}_i \cdot * \left(D\mathbf{z}_i\right)\right\|_2^2 \\ + \frac{\mu}{2}\sum_{i=1}^{n}\left\|\mathbf{z}_i - \mathbf{g}_i\right\|_2^2 \end{aligned} \tag{15}$$

where $\mathbf{w}_i$, $\mathbf{d}_i$, and $\mathbf{z}_i$ are, respectively, the ith column of matrices W, D, and Z.

Change vector $\mathbf{w}_i$ to a diagonal matrix, i.e., $\widehat{W}_i = \mathrm{diag}(\mathbf{w}_i)$ and $\mathbf{w}_i.*\mathbf{d}_i = \widehat{W}_i * \mathbf{d}_i$. Therefore, (15) can be expressed as follows:

$$\min_{Z} \frac{1}{2}\sum_{i=1}^{n}\left\|\widehat{W}_i * \left(\mathbf{d}_i\right) - \widehat{W}_i * \left(D\mathbf{z}_i\right)\right\|_F^2 + \frac{\mu}{2}\sum_{i=1}^{n}\left\|\mathbf{z}_i - \mathbf{g}_i\right\|_2^2 \tag{16}$$

For simplifying (16), $\widehat{W}_i * \mathbf{d}_i$ is denoted as K_i, and $\widehat{W}_i * D$ is denoted as H_i. Then (16) can be simplified as follows:

$$\min_{Z} \frac{1}{2}\sum_{i=1}^{n}\left(\left\|K_i - H_i\mathbf{z}_i\right\|_F^2 + \mu\left\|\mathbf{z}_i - \mathbf{g}_i\right\|_2^2\right) \tag{17}$$

For (17), $\mathbf{z}_i$ can be updated by the following:

$$\mathbf{z}_i = \left(H_i^T H_i - \mu I\right)^{-1}\left(H_i^T K_i - \mu\mathbf{g}_i\right) \tag{18}$$

Then repeat the above process until the objective function convergence. The coefficient matrix Z can be obtained when the termination condition is met, and it is expressed as follows:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nn} \end{bmatrix} \in \mathfrak{R}^{n\times n} \tag{19}$$

3. Traditional Imputation Method with LRMD

The traditional method imputed the missing data by zero-padding operation. For an original matrix $D_o = [d_1, d_2, \cdots d_n] \in R^{m\times n}$, suppose that d_p is missing, where d_p represent p^{th} column in D_0. The missing column of the matrix D_0 is imputed by 0, which can be represented as a matrix $\widehat{D}_1$:

$$\widehat{D}_1 = \left[\widehat{d}_1, \widehat{d}_2, \cdots, \widehat{d}_{p-1}, 0, \widehat{d}_{p+1}, \cdots \widehat{d}_n\right] \in R^{m\times n} \tag{20}$$

where $\widehat{d}_{ij} = \begin{cases} 0 & j=p \\ d_{ij} & j\in\{1,2,\cdots n\}\&j\neq p \end{cases}$ $(i = 1, 2, \cdots m)$ is the specific elements in the matrix $\widehat{D}_1$.

Multiplying $\widehat{D}_1$ by the pth column of coefficient matrix Z, $\widehat{d}_p = [\widehat{d}_{1p}, \widehat{d}_{2p}, \cdots, \widehat{d}_{mp}]^T$ can be recovered by the following:

$$\begin{aligned} \widehat{d}_{ip} &= \sum_{k=1}^{n}\widehat{d}_{ik}\cdot z_{kp} \\ &= d_{i1}z_{1p} + d_{i2}z_{2p}\cdots + d_{i(p-1)}z_{(p-1)p} \\ &\quad + d_{i(p+1)}z_{(p+1)p} + \cdots + d_{in}z_{np} \end{aligned} \tag{21}$$

The zero-padding operation is used for the traditional matrix imputation method to filling the missing column. Then the reconstructed matrix is multiplied by the corresponding column of the coefficient matrix Z; the imputed data of the missing column is obtained. This method only uses the data around the missing column to impute the

missing data; that is to say, the missing column does not contribute to the imputation result. Generally, the sensors including missing data have the highest correlation with final imputation results. However, the missing data is set to zero directly in the traditional imputation method, which ignores the effect of the missing data on the imputation results and reduces the accuracy of the imputation results.

4. Traffic Data Imputation with ILRMD

The missing data generally can be divided into three different types: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing at Determinate (MAD). This paper mainly deals with the problem of determinate missing. In road networks, traffic data was collected by various types of sensors, which usually demonstrated high temporal-spatial correlation characteristics; that is, traffic data have low-rank characteristic.

In a road network, suppose that there are m sensors and each sensor has n data samples, which can be denoted as a data matrix $D_{m\times n}$. This paper assumed that the data in the pth sensor is missing in $D_{m\times n}$. The traditional imputation method based on LRMD failed to consider the impact of missing data columns on imputation results. In order to address this shortcoming and combine the temporal-spatial correlation characteristics of traffic flow, this paper proposes a data imputation method based on ILRMD.

4.1. The Proposed ILRMD Model. In (9), it is assumed that d_{ij}, e_{ij} are the elements of the jth $(1 < j < n)$ observed sensor at the ith $(1 < i < m)$ time, respectively, existing in the observed matrix $D \in R^{m\times n}$ and the noise matrix $E \in R^{m\times n}$. z_{ij} is the element of the coefficient matrix Z, and the coefficient matrix $Z \in R^{n\times n}$. According to the multiplication rule, the following is obtained:

$$\begin{aligned} d_{ij} &= \left(\sum_{k=1}^{n} d_{ik} z_{kj}\right) + e_{ij} \\ &= \left(d_{i1}z_{1j} + d_{i2}z_{2j} + \cdots + d_{in}z_{nj}\right) + e_{ij} \end{aligned} \tag{22}$$

Then, (22) can be transformed into the following:

$$\begin{aligned} d_{ij} &= \frac{d_{i1}z_{1j} + \cdots + d_{i(j-1)}z_{(j-1)j} + \cdots + d_{in}z_{nj}}{1 - z_{jj}} \\ &\quad + \frac{e_{ij}}{1 - z_{jj}} \end{aligned} \tag{23}$$

The coefficient matrix of the jth observed sensor can be expressed as follows:

$$Z_{0j} = \left[\frac{z_{1j}}{1 - z_{jj}} \cdots \frac{z_{(j-1)j}}{1 - z_{jj}} \frac{z_{(j+1)j}}{1 - z_{jj}} \cdots \frac{z_{nj}}{1 - z_{jj}}\right]^T \tag{24}$$

The final coefficient matrix Z_0 of all observed sensors is described as follows:

$$\begin{aligned} Z &= \left[Z_{01}, Z_{02}, \ldots, Z_{0n}\right] \\ &= \begin{bmatrix} \frac{z_{21}}{1-z_{11}} & \frac{z_{12}}{1-z_{22}} & \cdots & \frac{z_{1j}}{1-z_{jj}} & \cdots & \frac{z_{1n}}{1-z_{nn}} \\ \frac{z_{31}}{1-z_{11}} & \frac{z_{32}}{1-z_{22}} & \cdots & \frac{z_{2j}}{1-z_{jj}} & \cdots & \frac{z_{2n}}{1-z_{nn}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{z_{(j-1)1}}{1-z_{11}} & \frac{z_{(j-1)2}}{1-z_{22}} & \cdots & \frac{z_{(j-1)j}}{1-z_{jj}} & \cdots & \frac{z_{(j-1)n}}{1-z_{nn}} \\ \frac{z_{j1}}{1-z_{11}} & \frac{z_{j2}}{1-z_{22}} & \cdots & \frac{z_{(j+1)j}}{1-z_{jj}} & \cdots & \frac{z_{jn}}{1-z_{nn}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{z_{n1}}{1-z_{11}} & \frac{z_{n2}}{1-z_{22}} & \cdots & \frac{z_{nj}}{1-z_{jj}} & \cdots & \frac{z_{(n-1)n}}{1-z_{nn}} \end{bmatrix}_{(n-1)\times n} \end{aligned} \tag{25}$$

Assuming that D_2 represents the matrix D that removes the pth column. According to the matrix multiplication rule, the matrix $D_2 = [d_1, d_2, \cdots, d_{p-1}, d_{p+1}, \cdots d_n] \in R^{m\times(n-1)}$ is multiplied by the pth column of the coefficient matrix Z_0. The value $\tilde{d}_p = [\tilde{d}_{1p}, \tilde{d}_{2p}, \cdots, \tilde{d}_{mp}]^T$ is obtained and can be noted as follows:

$$\begin{aligned} \tilde{d}_{ip} &= \sum_{k=1}^{p-1} \tilde{d}_{ik} \cdot \frac{z_{kp}}{1 - z_{pp}} + \sum_{k=p+1}^{n} \tilde{d}_{ik} \cdot \frac{z_{kp}}{1 - z_{pp}} \\ &= \frac{d_{i1}z_{1p} + \cdots + d_{i(p-1)}z_{(p-1)p} + d_{i(p+1)}z_{(p+1)p} + \cdots + d_{in}z_{np}}{1 - z_{pp}} \end{aligned} \tag{26}$$

The ILRMD method proposed in this paper assumes that a certain column of data in the matrix is lost and then multiplies the matrix by the coefficient matrix Z_0 to recover the missing data. The influence of all observed sensors is considered including the sensor with missing data. In (24), if the value Z_{jj} is zero, the data of the surrounding sensors is used for imputation. If the value Z_{jj} is not zero, both the data of the surrounding sensors and the sensor including missing data are used.

The differences between the ILRMD method and the traditional imputation method are discussed as follows. The traditional imputation method performs the zero-padding operation on the missing column and then is directly multiplied by the corresponding column of the coefficient matrix Z. The traditional imputation method utilizes the data collected from the surrounding sensors to recover the matrix and ignores the effect of the sensors including missing data. The ILRMD method assumes that the pth column of the data is completely missing and the matrix D_2 represents the matrix D after removing the pth column data. Then after the conversion, the weight that is most relevant to each sensor itself is expressed in another form, in order to reduce the effect of the most relevant weight to the imputation result. From (22)-(24), a coefficient matrix Z_0 is obtained. The coefficient matrix Z_0 considers not only the surrounding sensors, but also the influence of the sensor including missing

data. Ultimately the matrix D_2 is multiplied by the coefficient matrix Z_0 for obtaining imputation result.

The main steps of the proposed imputation method are as follows.

Step 1. The traffic flow data is preprocessed by smoothing and filtering, and the complete traffic flow data of one day is randomly selected to construct the training matrix D.

Step 2. The preprocessed matrix D is decomposed into the low-rank matrix A and the sparse matrix E according to (1).

Step 3. According to (9), matrix A is decomposed into D and Z, and, from (10) to (20), the coefficient matrix Z is solved.

Step 4. Construct test matrix D and set matrix D_2 as the dictionary matrix. D_2 represents the matrix D that removes the pth column.

Step 5. The coefficient matrix Z_0 is obtained according to (25) and the missing data which need to be imputed is obtained by (26).

4.2. Performance Evaluation Criteria. The evaluation criteria to measure the error of the imputed data included root mean square error (RMSE), mean absolute error (MAE), mean squared percentage error (MSPE), and mean absolute percentage error (MAPE). The RMSE and MAPE are selected in this paper. The formulas are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}$$
$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{27}$$

where N is the total number of the missing data, y_i is the actual value of the ith missing data point, and $\hat{y}_i$ is the corresponding estimated value.

5. Experiment Results

5.1. Data Description. The data used to evaluate the performance of the proposed model was collected in mainline detectors provided by the PeMS database, which includes more than 39,000 individual sensors that span the highway system in all major metropolitan areas of California. In this paper, 46 mainline sensors numbered from 1108512 to 1221232 are selected to perform data imputation test from April 1st, 2018, to April 30th, 2018. The traffic flow data is aggregated at 5-minute intervals and generate 288 data points for the daily flow. The data of 1 day, 7 days, and 14 days are, respectively, selected to construct the training matrix; however, the experimental results show that the improvement of the imputation accuracy is not obvious when the training samples become larger and larger. Therefore, the traffic flow data on April 23th, 2018, is used as training data, and the data on April 30th, 2018, is used as test data. The data in sensor numbered 1108512 is assumed to be missing, which needs to be imputed. According to the analysis of the spatial-temporal correlation characteristics of traffic flow, the traffic flow data on the same day in different consecutive weeks have high regularity and relevancy. Therefore, this paper selects traffic flow data from the same day on consecutive weeks (two Mondays) to perform the experiment. The traffic flow data of 46 observed sensors on April 23th, 2018, are selected as training matrix, and the data in sensor numbered 1108512 on April 30th, 2018, is assumed to be missing, which needs to be imputed.

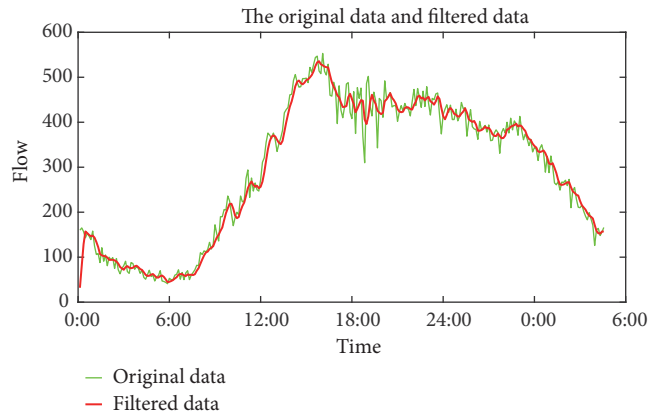


Figure 1: The original data and filtered data.

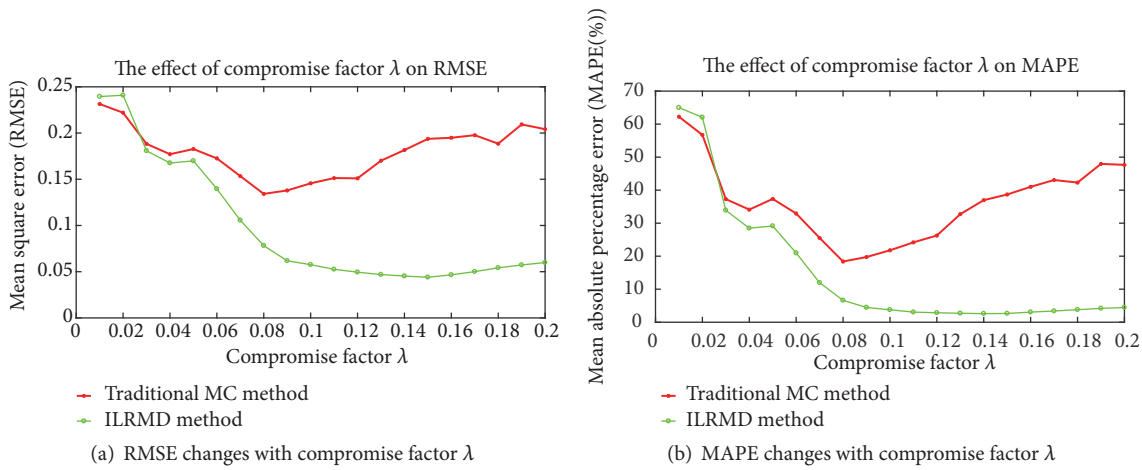
Due to the influence of people's willing for a trip, weather, and other factors, the traffic flow data presents certain stochastic fluctuation and abrupt. In order to reduce the impact of stochastic fluctuation of traffic flow data on imputation results, a five-point smoothing filtering method was used to preprocess the data. The original and filtered data, in the sensor numbered 1108512 on April 8, 2018, are shown in Figure 1.

From Figure 1, it can be seen that the filtered data intuitively reflects the regularity of the traffic data, and the abrupt points are effectively filtered out in the original traffic flow data.

In this paper, the training data and the test data are all preprocessed with a smoothing filtering method at first, which can remove the abnormal points in the sensor data. Then we randomly assume that a sensor data is missing and then impute the missing sensor data with the proposed model.

5.2. Results and Performances Analysis

5.2.1. Influence of Parameter λ. The compromise factor λ is an important parameter of low rank matrix decomposition, and the different λ values have an important impact on the performance of data imputation. In order to verify the effectiveness of ILRMD method, the influence of parameter λ is analyzed. The RMSE and MAPE of imputation results changes with the compromise factor λ are, respectively, shown in Figures 2(a) and 2(b).



(a) RMSE changes with compromise factor λ

(b) MAPE changes with compromise factor λ

Figure 2: The effect of compromise factor λ on error function.

From Figure 2, we can see that, for the traditional MC method, both RMSE and MAPE gradually decrease with the increase of the compromise factor λ. After RMSE and MAPE reach the minimum value ($\lambda = 0.08$), which increase again. For the ILRMD method, RMSE and MAPE all decrease with the change of λ. When $\lambda = 0.15$, they reach the minimum and then increase slowly. In any case, the traditional MC method is far less effective than ILRMD method. Therefore, in order to compare the imputation results of the two methods in the best state, λ is set as 0.08 for traditional MC method and 0.15 for ILRMD method in this paper.

5.2.2. The Selection of the Training Data. Due to traffic flow has high spatial-temporal correlation characteristics, it is necessary to analyze the effect of different training data to imputation results. However, the selection of training data has little influence on the performance of the proposed ILRMD method. In order to show that the performance of the proposed method is not sensitive to the time, the traffic flow data of four days (April 21th, 2018, April 22th, 2018, April 23th, 2018, and April 24th, 2018) are randomly selected as training data to impute the data of April 30th, 2018. The experimental results are shown in Figures 3(a), 3(b), 3(c), and 3(d).

It can be seen from Figure 3 that the proposed ILRMD method always has good performance and is not sensitive to the selection of training data. And the imputation performance of different training data is shown in Table 1.

It can be seen from Table 1 that the proposed method always has good performance although the different training data is used. The results indicate that the selection of time has little influence on the proposed ILRMD method. Therefore, we only select the traffic flow data of one day (April 23th, 2018) to verify the proposed model in the paper.

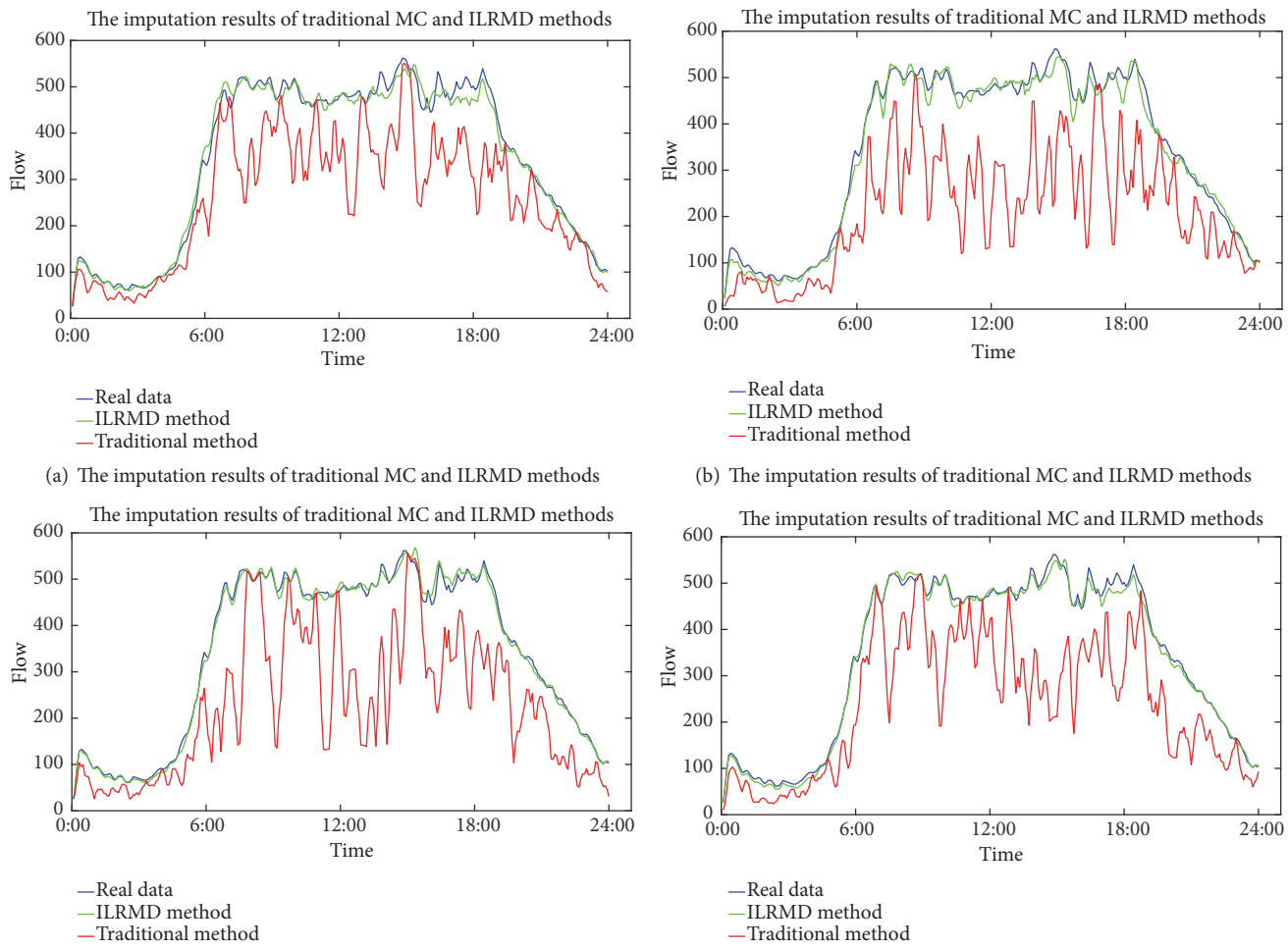
5.2.3. Comparison of Imputation Results. For the purpose of verifying the performances of ILRMD method, the proposed method is compared with the traditional method. The imputation results of the ILRMD method under the best condition ($\lambda = 0.15$) and the traditional method under the best condition ($\lambda = 0.08$) are shown in Figures 4(a) and 4(b).

Table 1: The performance comparison of different training data.

Training data	MAPE	RMSE
April 21th, 2018	0.0294	0.0454
April 22th, 2018	0.0364	0.0588
April 23th, 2018	0.0260	0.0453
April 24th, 2018	0.0207	0.0409

From Figure 4, it can be seen that the imputation results of traffic flow data through the ILRMD are more accurate than the traditional MC method. Although the imputation result is obtained in the optimal compromise factor λ with the traditional MC method, there is a big deviation between the imputation result and the real data, and the ILRMD method still recovers the missing traffic data more accurately. When compromise factor λ is set as the optimal value for the ILRMD method, the imputation result is almost identical with the real value, but there are more deviations in traditional methods. It is observed that the imputation results of the proposed ILRMD method have similar traffic patterns with the real traffic flow, especially in morning and evening peak hours.

5.2.4. The Comparison of ILRMD and Other Imputation Methods. In order to evaluate the advantages of our proposed approach, the ARIMA, SVR, DBN-SVR, WNN, KNN, and Traditional MC imputation methods are selected under the premise of testing with the same experimental data. In the ARIMA model, the orders of autoregressive p, moving average q, and difference d are, respectively, set as 5, 5, and 1. In the SVR model, the nuclear function is configured as "*rbf*", the number of iterations is 10,000, and the penalty factor is taken as 0.01. In the WNN model, the number of iterations is 1000, the number of the hidden layer nodes is 3. In the DBN-SVR model, the number of network layers in the DBN model is set as 3 and the number of iterations is 200. The ILRMD model proposed in this paper is compared with



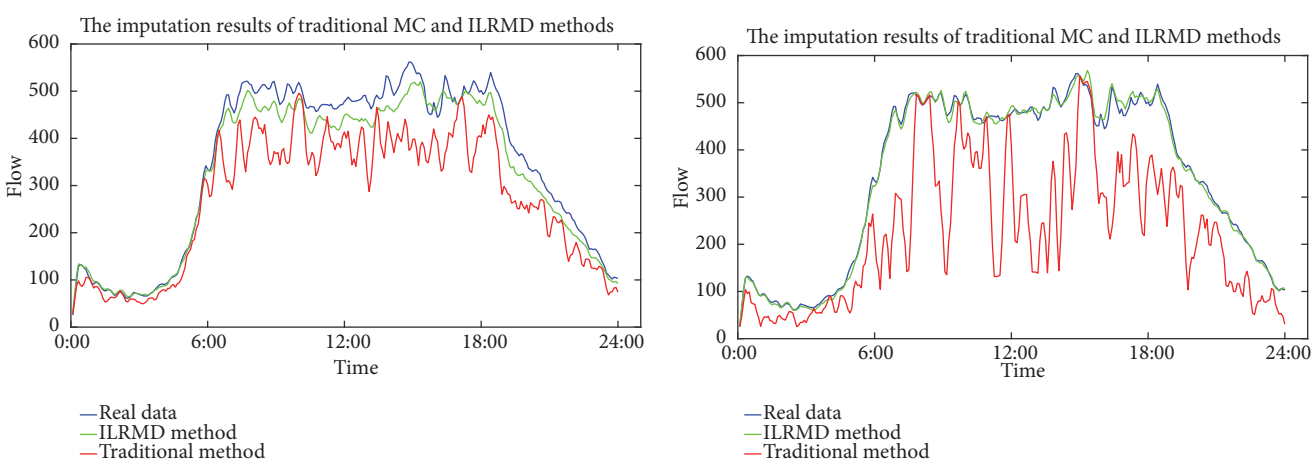
(a) The imputation results of traditional MC and ILRMD methods

(b) The imputation results of traditional MC and ILRMD methods

(c) The imputation results of traditional MC and ILRMD methods

(d) The imputation results of traditional MC and ILRMD methods

Figure 3: The imputation results of traditional MC and ILRMD methods.



(a) The imputation results of two imputation methods ($\lambda = 0.08$)

(b) The imputation results of two imputation methods ($\lambda = 0.15$)

Figure 4: The imputation results of traditional MC and ILRMD methods.

Table 2: The performance comparison of data imputation models.

Missing sensor ID	1108512		1119921	
Model	MAPE	RMSE	MAPE	RMSE
KNN	0.1338	0.5022	0.1095	0.4260
SVR	0.1024	0.5396	0.1401	0.2424
WNN	0.1442	0.6660	0.3994	0.1229
DBN-SVR	0.1129	0.6402	0.0994	0.7280
ARIMA	0.2078	0.6065	0.2047	0.5968
Traditional MC	0.3717	0.1830	0.1350	0.1099
ILRMD	0.0260	0.0453	0.0495	0.0672

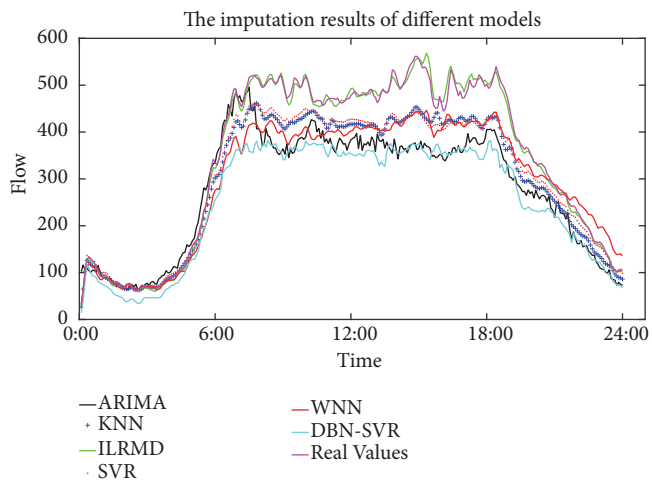


Figure 5: The imputation results of different models.

these imputation methods; the imputation results of different models and real traffic flow are shown within one day in Figure 5.

It can be seen from Figure 5, the imputation traffic flow has similar traffic patterns with the real traffic flow. The DBN-SVR model has the worst imputation performance; the ARIMA, SVR, KNN, and WNN are better than the DBN-SVR, while they show weakness compared with the ILRMD method. The imputation value of the proposed ILRMD model is almost coincided with the measured data. It is observed that the proposed ILRMD model has better imputation performance.

The error analysis test is conducted using two error evaluation criteria, which is expressed in Table 2. In order to more precisely verify the performance of the proposed model, another sensor numbered 1119921 is randomly selected to perform the test. In Table 2, the sensors numbered 1108512 and 1119921 are, respectively, assumed to be imputed to verify the performance of the proposed model. It can be seen from Table 2, when the sensors numbered 1108512 and 1119921 are assumed to be imputed, the proposed ILRMD models all have the best performance compared with other approaches. These experiments can verify that the ILRMD model proposed in this paper is an effective method for data imputation.

From Table 2 of the first condition (1108512 sensor), it can be seen that the imputation accuracy of the ILRMD model, respectively, improves 93.01%, 74.61%, 95.96%, 80.57%, 96.30%, and 81.97% compared with the traditional MC, SVR, ARIMA, KNN, DBN-SVR, and WNN methods. The average imputation accuracy is 87.07% higher than other imputation methods. Results demonstrate that the proposed ILRMD model has the best performance compared with other approaches, and it is an effective method for data imputation.

6. Conclusions and Recommendations

In the paper, a data imputation method is proposed to impute the missing traffic flow data. Different from the most known traffic flow data imputation methods, the ILRMD model makes an effective use of the information of missing sensors and takes full advantage of the high spatiotemporal correlation characteristics of traffic flow data. The experiment result shows that the proposed imputation method is superior to other methods. However, this paper focuses on dealing with the missing traffic data at a single sensor; we only considered one observed sensor with missing data. In practical terms, the missing traffic data is always distributed on multisensors.

In our future research, the missing data analysis on multisensors is being studied. The concept of missing rate can be introduced, and the more effective data imputation

method can be proposed for different degrees of missing data in order to improve the imputation accuracy.

Data Availability

The data used in this paper are collected from the Caltrans Performance Measurement System (PeMS) in 46 sensors numbered from 1108512 to 1221232 on 04/01/2018~04/27/2018. If any researcher requests for these data, he can log into the website: http://pems.dot.ca.gov/.

Conflicts of Interest

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

Acknowledgments

This research was partly supported by the National Key R&D Program of China (2018YFC0808706) and the National Natural Science Foundation of China (Grant no. 5157081053). The authors are also grateful to the PeMS for providing the data.

References

[1] Q. Wang and H. Sun, "Traffic structure optimization in historic districts based on green transportation and sustainable development concept," *Advances in Civil Engineering*, vol. 2019, Article ID 9196263, 18 pages, 2019.

[2] L. Duan, Y. Zhang, and J. Lai, "Influence of ground temperature on shotcrete-to-rock adhesion in tunnels," *Advances in Materials Science and Engineering*, vol. 2019, Article ID 8709087, 16 pages, 2019.

[3] X. Wang, J. Lai, R. S. Garnes, and Y. Luo, "Support system for tunnelling in squeezing ground of qingling-daba mountainous area: a case study from soft rock tunnels," *Advances in Civil Engineering*, vol. 2019, Article ID 8682535, 17 pages, 2019.

[4] L. Duan, W. Lin, J. Lai, P. Zhang, and Y. Luo, "Vibration characteristic of high-voltage tower influenced by adjacent tunnel blasting construction," *Shock and Vibration*, vol. 2019, Article ID 8520564, 16 pages, 2019.

[5] N. Zhao, Z. Li, and Y. Li, "Improving the traffic data imputation accuracy using temporal and spatial information," in *Proceedings of the 7th International Conference on Intelligent Computation Technology and Automation, ICICTA 2014*, pp. 312–317, China, October 2014.

[6] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "The retrieval of intra-day trend and its influence on traffic prediction," *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 103–118, 2012.

[7] J. W. C. van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 5-6, pp. 347–369, 2005.

[8] H. Sun, Q. P. Wang, P. Zhang et al., "Spatial-temporal characteristics of tunnel traffic accidents in China from 2001 to present," *Advances in Civil Engineering*, Article ID 4536414.

[9] L. Qu, L. Li, Y. 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.

[10] X. Luo, L. Niu, and S. Zhang, "An algorithm for traffic flow prediction based on improved SARIMA and GA," *KSCE Journal of Civil Engineering*, pp. 1–9, 2018.

[11] X. Luo, D. Li, and S. Zhang, "Traffic flow prediction during the holidays based on DFT and SVR," *Journal of Sensors*, Article ID 6461450.

[12] X. L. Luo, D. Y. Li, Y. Yang, and S. R. Zhang, "Spatiotemporal Traffic Flow Prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, Article ID 4145353, 10 pages, 2019.

[13] I. Laña, I. Olabarrieta, M. Vélez, and J. Del Ser, "On the imputation of missing data for road traffic forecasting: new insights and novel techniques," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 18–33, 2018.

[14] N. L. Nihan, "Aid to determining freeway metering rates and detecting loop errors," *Journal of Transportation Engineering*, vol. 123, no. 6, pp. 454–458, 1997.

[15] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1678, pp. 179–188, 1999.

[16] M.-C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.

[17] L. Li, Y. B. Li, and Z. H. Li, "Missing traffic data: comparison of imputation methods," *IET Intelligent Transport Systems*, vol. 8, no. 1, pp. 51–57, 2014.

[18] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[19] A. Pascale and M. Nicoli, "Adaptive Bayesian network for traffic flow prediction," in *Proceedings of the 2011 IEEE Statistical Signal Processing Workshop, SSP 2011*, pp. 177–180, France, June 2011.

[20] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[21] D. H. Zeng, J. M. Xu, J. W. Gu, L. Liu, and G. Xu, "Short term traffic flow prediction based on online learning SVR," in *Proceedings of the 2008 Workshop on Power Electronics and Intelligent Transportation System, PEITS 2008*, pp. 616–620, China, August 2008.

[22] M. Elshenawy, M. El-darieby, and B. Abdulhai, "Automatic imputation of missing highway traffic volume data," in *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 373–378, Athens, March 2018.

[23] Z. Q. Sun, J. S. Pan, and Q. Duan, "Study on a New Traffic Flow Forecasting Method," in *Proceedings of the International Conference on Natural Computation*, pp. 349–353, 2008.

[24] C. Chen, J. Hu, Q. Meng, and Y. Zhang, "Short-time traffic flow prediction with ARIMA-GARCH model," in *Proceedings of the 2011 IEEE Intelligent Vehicles Symposium, IV'11*, pp. 607–612, Germany, June 2011.

[25] W. H. Yin, P. Murray-Tuite, and H. Rakha, "Imputing erroneous data of single-station loop detectors for nonincident conditions: Comparison between temporal and spatial methods," *Journal*

of Intelligent Transportation Systems: Technology, Planning, and Operations, vol. 16, no. 3, pp. 159–176, 2012.

[26] L. Qu, Y. Zhang, J. Hu, L. Jia, and L. Li, "A BPCA based missing value imputing method for traffic flow volume data," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '08)*, pp. 985–990, June 2008.

[27] D. H. Ni, J. D. Leonard, A. Guin et al., "Multiple imputation scheme for overcoming the missing values and variability issues in its data," *Journal of Transportation Engineering*, vol. 131, no. 12, pp. 931–938, 2005.

[28] Z. B. Liu, S. Sharma, and S. Datla, "Imputation of missing traffic data during holiday periods," *Transportation Planning & Technology*, vol. 31, no. 5, pp. 525–544, 2008.

[29] B. Sun, W. Cheng, P. Goswami, and G. Bai, "Short-term traffic forecasting using self-adjusting k-nearest neighbours," *IET Intelligent Transport Systems*, vol. 12, no. 1, pp. 41–48, 2018.

[30] G. Chang, Q. Y. Wu, and L. Luo, "Missing data imputation for traffic flow based on weighted local least squres," in *Proceedings of the International Conference on Automatic Control and Artificial Intelligence*, pp. 1351–1354, 2012.

[31] G. Chang, Y. Zhang, and D. Yao, "Missing data imputation for traffic flow based on improved local least squares," *Tsinghua Science and Technology*, vol. 17, no. 3, pp. 304–309, 2012.

[32] Y. J. Duan, "A deep learning based approach for traffic data imputation," in *Proceedings of the IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 896–901, 2014.

[33] Q. Shang, Z. Yang, S. Gao, and D. Tan, "An imputation method for missing traffic data based on FCM optimized by PSO-SVR," *Journal of Advanced Transportation*, vol. 2018, Article ID 2935248, 21 pages, 2018.

[34] L. N. Nguyen and W. T. Scherer, *Imputation Techniques to Account for Missing Data in Support of Intelligent Transportation Systems Applications*, University of Virginia, Charlottesville, Va, USA, 2003.

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

[36] C. Goves, R. North, R. Johnston, and G. Fletcher, "Short term traffic prediction on the UK motorway network using neural networks," *Transportation Research Procedia*, vol. 13, pp. 184–195, 2016.

[37] D. H. Ni and J. D. Leonard, "Markov chain monte carlo multiple imputation for incomplete its data using bayesian networks," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1935, no. 1, pp. 57–67, 2005.

[38] Y. Higashijima, A. Yamamoto, T. Nakamura, M. Nakamura, and M. Matsuo, "Missing data imputation using regression tree model for sparse data collected via wide area ubiquitous network," in *Proceedings of the 2010 10th Annual International Symposium on Applications and the Internet, SAINT 2010*, pp. 189–192, Republic of Korea, July 2010.

[39] W. C. Ku, G. R. Jagadeesh, A. Prakash, and T. Srikanthan, "A clustering-based approach for data-driven imputation of missing traffic data," in *Proceedings of the 2016 IEEE Forum on Integrated and Sustainable Transportation Systems, FISTS 2016*, pp. 1–6, China, July 2016.

[40] L. Li, Y. Li, and Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence," *Transportation Research Part C: Emerging Technologies*, vol. 34, no. 9, pp. 108–120, 2013.

[41] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[42] F. Zhang and J. Yang, "A linear subspace learning approach via low rank decomposition," in *Proceedings of the 2011 2nd International Conference on Innovations in Bio-inspired Computing and Applications, IBICA 2011*, pp. 81–84, China, December 2011.

[43] J. Wright, A. Ganesh, S. Rao et al., "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, pp. 2080–2088, 2009.

[44] J. F. Cai, E. J. Candès, and Z. W. Shen, "A singular value thresholding algorithm for matrix imputation," *Siam Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[45] K. H. Cho and N. Reyhani, "An iterative algorithm for singular value decomposition on noisy incomplete matrices," in *Proceedings of the 2012 Annual International Joint Conference on Neural Networks, IJCNN 2012, Part of the 2012 IEEE World Congress on Computational Intelligence, WCCI 2012*, pp. 1–6, Australia, June 2012.

[46] A. Ganesh, Z. C. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast algorithms for recovering a corrupted low-rank matrix," in *Proceedings of the 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2009*, pp. 213–216, Netherlands, December 2009.

[47] T. Kimchuan and Y. Sangwoon, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 3, pp. 615–640, 2010.

[48] Z. C. Lin, M. M. Chen, and Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low rank matrices, pp. 9, https://arxiv.org/abs/1009.5055, 2010.

[49] R. Du, Y. Zhang, B. Y. Wang et al., "Low-rank representation based traffic data imputation method," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 5127–5134, 2016.

[50] G. C. Liu, Z. C. Lin, S. C. Lin et al., "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.