

Research Article

Robust Missing Traffic Flow Imputation Considering Nonnegativity and Road Capacity

Huachun Tan,¹ Yuankai Wu,¹ Bin Cheng,¹ Wuhong Wang,¹ and Bin Ran²

¹ Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

² Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA

Correspondence should be addressed to Huachun Tan; tanhc00@gmail.com

Received 18 December 2013; Accepted 13 February 2014; Published 18 March 2014

Academic Editor: Huimin Niu

Copyright © 2014 Huachun Tan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are increasing concerns about missing traffic data in recent years. In this paper, a robust missing traffic flow data imputation approach based on matrix completion is proposed. In the proposed method, the similarity of traffic flow from day to day is exploited to impute missing data by the low-rank hypothesis of constructed traffic flow matrix. And the physical limitation of road capacity and nonnegativity is also considered through the optimization process, which avoids the possibility of producing negative and overcapacity values. Moreover, the proposed algorithm can impute missing data and recover outlier in a unify framework. The experiment results show that the proposed method is more accurate, stable, and reasonable.

1. Introduction

Traffic information collected by various kinds of sensors are a vital component of intelligent transportation system (ITS) which aims to influence travel behavior, reduce traffic congestion, improve mobility, and enhance air quality [1]. For example, the real-time traffic information can be provided to drivers before and during their travels for supporting their decision of route choice [2] and it is also an important guideline for modern traffic control system to adjust the signal timing [3]. Moreover, after proper preprocessing, the real-time traffic data can be used as the real-time traffic state estimation of transportation networks [4]. On the other hand, several data mining techniques have been applied to mine time related association rules from historical traffic databases and its results have been used for traffic prediction such as the works of Qiao et al. [5] and Zargari et al. [6].

However, the missing traffic data problems remain inevitable due to detector faults or transmission distortion in many places. About 10% of daily traffic flow is usually missing in Beijing, China [7]. Turner et al. [8] reported that almost a quarter of data from San Antonio, Texas, is missing; and more than 5% of data are lost within the PeMS traffic flow database [9]. The missing data adversely affect the

applications of intelligent transportation system; for example, the traffic control system requires sufficient traffic flow data (i.e., traffic volumes, occupancy rates, and flow speeds) to generate appropriate traffic management strategies [10, 11]. In traffic forecast area, if there exists missing data, the predicting performance will reduce sharply [12, 13]. Clearly, missing data problem is a large obstacle for any of the functions for which ITS data is to be used.

In the past decades, numerous imputation methods have been proposed to handle missing traffic data problem. These imputation methods can be roughly divided into two parts: interpolation based and inductive learning based methods.

Interpolation based methods always fill the missing data with a weighted average value calculated from part of known data. Yin et al. [14] use historical averages from the same detector at the same time period but in neighboring days to replace missing data. Zhong et al. [15] interpolate error traffic data by traffic data from similar daily flow variation patterns considering the error type and traffic condition. Such approaches only use part of the traffic information and always fail to accurately estimate missing value at high missing ratio.

On the other hand, inductive learning methods try to build imputation modeling from the a priori characteristics of traffic data. Most of these methods are based on some

TABLE 1: The possibilities of producing unreliable results.

Methods	MCI	PPCA	IALM	SVT
Frequency	0%	38.57%	7.14%	11.43%

assumptions of traffic flow data. Autoregressive Integrated Moving Average (ARIMA) is based on the assumption that the historical value and future value of traffic flow provide an indication of the missing value [16]. The probabilistic principle component analysis (PPCA) based methods assume that the basic characteristic of traffic flow variations can be captured by the probability distribution of PPCA [7, 17]. The recent tensor-based methods assume that traffic data are highly correlated in multimode (day, week, link) and construct traffic flow data into multiway array (tensor) to capture these correlations. By utilizing the essential characteristics of traffic flow information by the assumptions, these kinds of imputation methods often outperform traditional interpolation methods [7]. It can be concluded that most of the inductive missing traffic data imputation methods make the following assumptions.

Assumption 1. Traffic flow has a high similarity from day to day and week to week but also link to the neighboring link which can be utilized to impute missing traffic flow.

Assumption 2. The traffic flow data have not been spoiled by the outliers, which frequently occur in real-world traffic information system.

While the inductive based imputation methods achieve somewhat success by the assumptions for traffic flow data, there are still some shortcomings in these methods. Firstly, according to the traffic flow theory, the volume of traffic flow is a certain value from zero to the road-capacity (the maximum traffic flow obtainable on a given roadway using all available lanes). But most data-driven imputation methods ignore this limitation of traffic volume. Secondly, the traditional methods such as the PCA-based methods cannot work well with big outlier or errors without the preprocessing of corrupted traffic data [7]. In fact, we have proposed several tensor completion methods [17–20], which make full use of multimode correlations [21] of traffic flow data, to impute missing traffic data. However, in our former works, the nonnegativity (lower bound) and capacity (upper bound) of traffic flow are still ignored. As a result, it is possible that these methods will produce some unreliable results.

To tackle these shortcomings, this paper proposes a traffic flow data imputation methods based on matrix completion. In the proposed method, the traffic data are constructed into a day \times interval matrix. The similarity of day mode is captured by the assumption that the constructed matrix is low-rank. By adding a limitation in the objective function, the proposed method can restrict the reconstructed traffic flow data between zero and road capacity. Moreover, for the traffic flow data corrupted by outliers, the proposed method can simultaneously impute the missing data and recover the outliers by the sparse assumption. It should be noted that the

method we proposed here can be considered as an idealized version of robust PCA (RPCA) but different from the natural approaches to robust PCA [22]. The proposed approach needs not to preprocess the data and isolate the outlier before the imputation. This advantage allows the proposed matrix completion (MC) methods outperform traditional imputation methods especially for the traffic flow corrupted by outliers.

To give a detailed explanation of the proposed method, the rest of this paper is organized as follows. The methodology and algorithm of the proposed method are proposed in Section 2. Section 3 presents the imputing testing results including comparison with other methods. The conclusion and future works are conducted in Section 4.

2. Methodology

In this section, a brief description of the matrix completion is presented. Then, we described the proposed missing traffic flow data imputation method in detail.

2.1. Review of Matrix Completion Methods. Let M be an $m \times n$ matrix of rank r ($< m$ or n); the low matrix M has some available sampled entries $\{M_{ij} : (i, j) \in \Omega\}$ where Ω is a subset of sampled cardinality. Then [17] proves that most matrices M of rank r can be perfectly reconstructed by solving the optimization problem:

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t} \quad & X_{ij} = M_{ij} \quad (i, j) \in \Omega. \end{aligned} \quad (1)$$

In (1), the functional $\|X\|_*$ is the nuclear norm of the matrix M , which is the sum of its singular values.

In [23], the singular value thresholding (SVT) algorithm is used to solve an approximate optimization problem of (1):

$$\begin{aligned} \min \quad & \tau \|X\|_* + \frac{1}{2} \|X\|_F^2 \\ \text{s.t} \quad & P_\Omega(X) = P_\Omega(M), \end{aligned} \quad (2)$$

where P_Ω is the orthogonal projector onto the span of matrices vanishing outside of Ω so that the (i, j) th component of $P_\Omega(X)$ is equal to X_{ij} if $(i, j) \in \Omega$ and zero otherwise.

The Lagrange multiplier of (2) is

$$L(X, Y) = \tau \|X\|_* + \frac{1}{2} \|X\|_F^2 + \{Y, P_\Omega(M - X^k)\}, \quad (3)$$

where $\{A, B\} = \text{trace}\{AB^T\}$, with optimization variable $X \in R^{m \times n}$. Fix $\tau > 0$ and a sequence $\{\delta_k\}_{k \geq 1}$ of scalar step sizes. Then starting with $Y^0 = 0 \in R^{m \times n}$, the algorithm inductively defines

$$\begin{aligned} X^k &= D_\tau(Y^{k-1}), \\ Y^k &= Y^{k-1} + \delta_k P_\Omega(M - X^k), \end{aligned} \quad (4)$$

where

$$D_\tau(X) = US_\tau(\Sigma)V^T \quad \text{if } X = U\Sigma V^T. \quad (5)$$

TABLE 2: The recovery and imputation accuracy with different ratio of outliers.

Outlier ratio	Missing data imputation		Outlier recovery	
	MAPE	MAE	MAPE	MAE
5%	13.81%	32.0799	14.89%	36.9587
10%	14.47%	33.8156	15.32%	39.1584
15%	15.95%	39.7424	15.76%	46.6486

Input: Observation samples D_{ij} , $i, j \in \Omega$, of matrix $D \in R^{m \times n}$

- (1) $Y_0, E_0 = 0, \mu_0, k = 0, C$ (road capacity)
- (2) while not converged do
- (3) $(U, S, V) = \text{svd}(D - E_k + \mu_k^{-1}Y_k)$,
- (4) $A_{k+1} = US_{\mu_k^{-1}}V^T$,
- (5) $E_{k+1} = P_{D-C \leq E \leq D}(P_{\Omega}(S_{\mu/\lambda}(D - A_{k+1} + \mu_k^{-1}Y_k)) + P_{\bar{\Omega}}(D - A_{k+1} + \mu_k^{-1}Y_k))$
- (6) $Y_{k+1} = Y_k + \mu_k(D - A_{k+1} - E_{k+1})$
- (7) updating μ_k
- (8) $k = k + 1$
- (9) end while
- (10) output A

ALGORITHM 1: Matrix completion based traffic data imputation method.

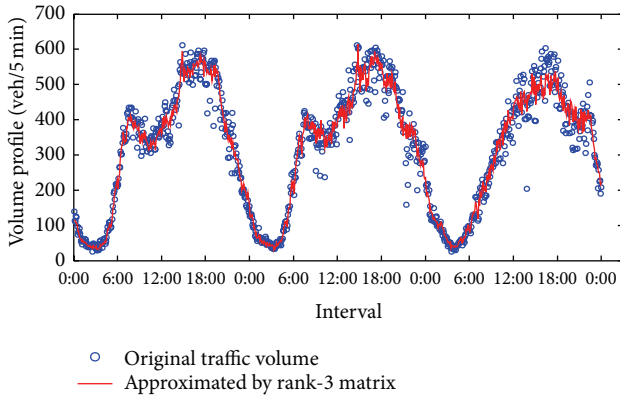


FIGURE 1: The original data (blue) and approximated low-rank data (red) from Detector 400141. The low-rank data basically keeps the characteristic of origin data.

$S_{\tau}(x)$ is the constriction factor of x , where

$$S_{\tau}(x) = \begin{cases} x - \tau & \text{if } x > \tau \\ x + \tau & \text{if } x < -\tau \\ 0 & \text{else.} \end{cases} \quad (6)$$

More details for SVT can be found in [24].

Literature [25] develops an augmented Lagrange multiplier method to solve MC. In their methods, the MC problem is formulated as

$$\begin{aligned} \min \quad & \|A\|_* \\ \text{s.t.} \quad & A + E = D \quad P_{\Omega}(E) = 0 \end{aligned} \quad (7)$$

as E will compensate for the unknown entries of D and the unknown entries of D are simply set as zeros. Then the partial augmented Lagrangian function is

$$L(A, E, Y, \mu) = \|A\|_* + \{Y, D - A - E\} + \frac{\mu}{2} \|D - A - E\|_F^2. \quad (8)$$

Then, A and E are updated according to the subproblems of (8), where A is updated by

$$\begin{aligned} \arg \min_A \|A\|_* + \frac{\mu}{2} \|D - A - E + \mu^{-1}Y\|_F^2 \\ = D_{\mu^{-1}}(D - E + \mu^{-1}Y). \end{aligned} \quad (9)$$

E is updated by

$$\arg \min_{P_{\Omega}(E)=0} \|D - A - E + \mu^{-1}Y\|_F^2 = P_{\Omega}(D - A + \mu^{-1}Y). \quad (10)$$

More detailed information of the augmented Lagrange multiplier (ALM) method can be found in literature [25].

2.2. The Proposed Algorithm. The goal of the proposed method is to impute the missed traffic data considering both the physical limitation of traffic flow data and the possible corruption by outliers. Firstly, the traffic flow data in a local place are formed into the matrix mode as follows:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}. \quad (11)$$

In this matrix a_{ij} represents the discretized volume on day j at time interval i within the given day. For the physical

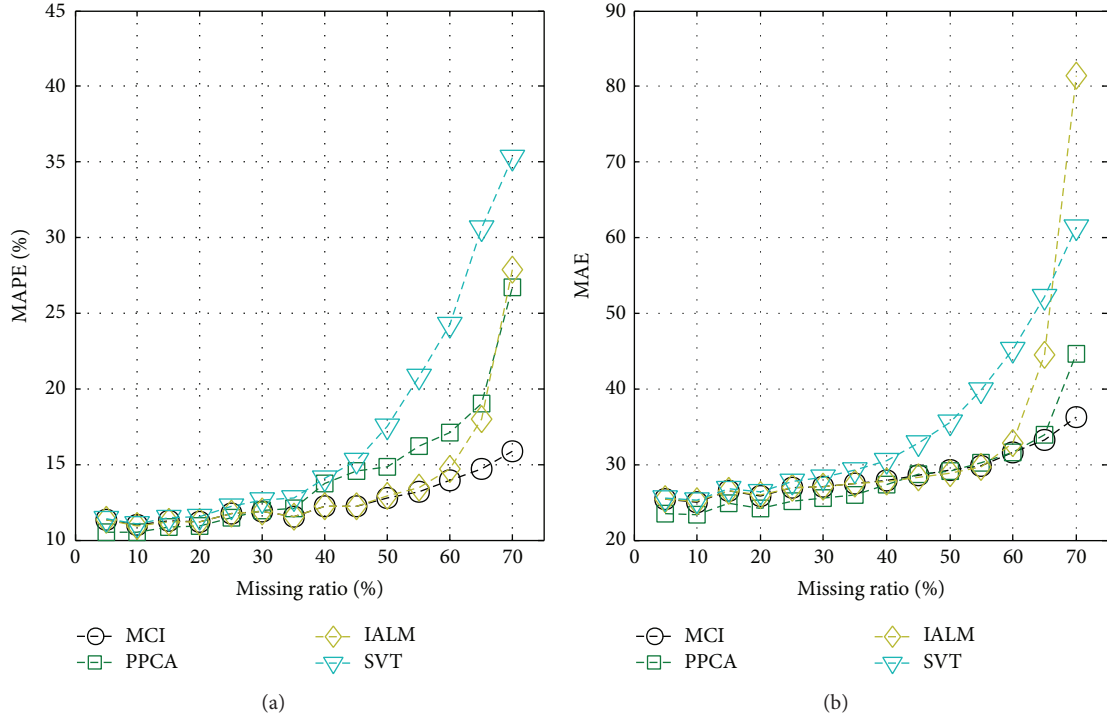


FIGURE 2: MAE and MAPE curves for three matrix completion methods and PPCA method.

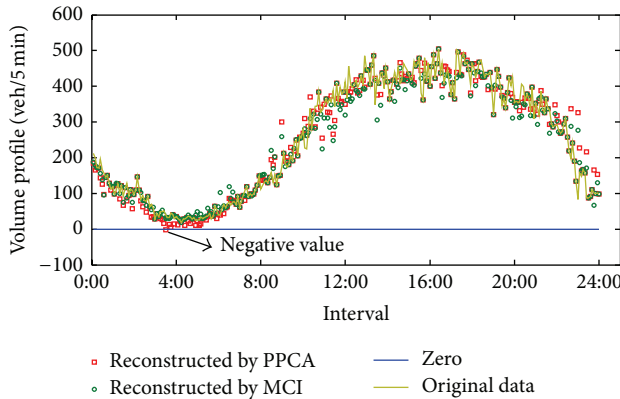


FIGURE 3: The negative value reconstructed by PPCA.

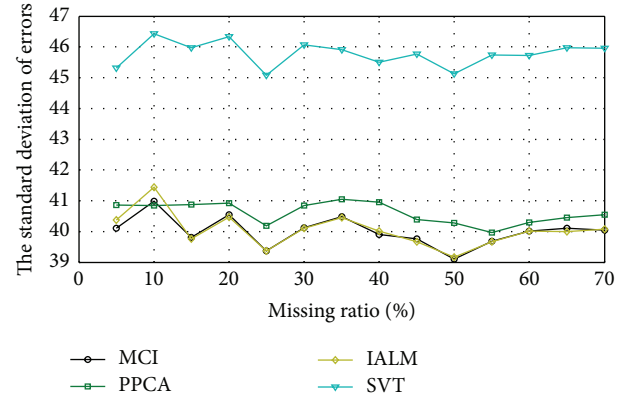


FIGURE 4: The standard deviation of errors (SDE) of the test methods.

limitation, a_{ij} changes in a particular range from zero to road capacity. The total number of days is n and each day is divided into m time intervals. Supposing the set of observed traffic volume data is Ω , and the traffic volume is corrupted by sparse outliers. Hence, the missing traffic data imputation problem is translated into a corrupted matrix completion problem [26–28]; the optimization problems can be described as

$$\begin{aligned} \min \quad & \text{rank}(\mathbf{A}) + \lambda \|P_{\Omega}(\mathbf{E})\|_0 \\ \text{s.t.} \quad & \mathbf{A} + \mathbf{E} = \mathbf{D}, \quad 0 \leq \mathbf{A} \leq \mathbf{C}, \end{aligned} \quad (12)$$

where $\|\cdot\|_0$ represents the number of the nonzero entries and P_{Ω} is the orthogonal projector onto the span of matrices vanishing outside of Ω .

The minimums of $\|\cdot\|_0$ (ℓ^0 -norm) and the rank of matrix are NP-hard problem [29]. To convert the objective function (12) into a convex optimization problem, the rank of \mathbf{A} is approximated by the nuclear norm (the sum of the singular values) of the matrix, and the $\|\cdot\|_0$ of $P_{\Omega}(\mathbf{E})$ is approximated by the ℓ^1 -norm of the matrix (the sum of the absolute value of its entries) [29] as follows:

$$\begin{aligned} \min \quad & \|\mathbf{A}\|_* + \lambda \|P_{\Omega}(\mathbf{E})\|_1 \\ \text{s.t.} \quad & \mathbf{A} + \mathbf{E} = \mathbf{D}, \quad 0 \leq \mathbf{A} \leq \mathbf{C}. \end{aligned} \quad (13)$$

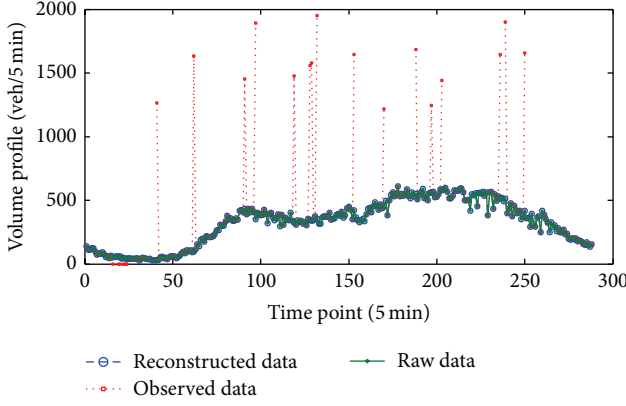


FIGURE 5: Comparisons with raw traffic volume data, data with 30% missing data and corrupted by outliers with 15% ratio, and data reconstructed by MCI.

Because the solution of E is easier than A , the function is converted into the following form:

$$\begin{aligned} \min \quad & \|A\|_* + \lambda \|P_\Omega(E)\|_1 \\ \text{s.t.} \quad & A + E = D, \quad D - C \leq E \leq D. \end{aligned} \quad (14)$$

Considering the faster computation speed and higher accuracy, the augmented Lagrangian method (ALM) [25] is employed to optimize the problem.

By introducing a Lagrange multiplier Y to remove the equality constraint, one has the Lagrangian function of (14):

$$\begin{aligned} L(A, E, Y, \mu) = & \|A\|_* + \lambda \|P_\Omega(E)\|_1 + \langle Y, D - A - E \rangle \\ & + \frac{\mu}{2} \|D - A - E\|_F^2. \end{aligned} \quad (15)$$

Lin et al. [25] proved that updating A and E once when solving this subproblem is sufficient for A and E to converge to the optimal solution of (15). A is updated by

$$\begin{aligned} \arg \min_A \|A\|_* + \frac{\mu}{2} \|D - A - E + \mu^{-1}Y\|_F^2 \\ = D_{\mu^{-1}}(D - E + \mu^{-1}Y), \end{aligned} \quad (16)$$

where

$$\begin{aligned} D_{\mu^{-1}}(D - E + \mu^{-1}Y) = US_{\mu^{-1}}(\Sigma)V^T \\ \text{if } D - E + \mu^{-1}Y = U\Sigma V^T. \end{aligned} \quad (17)$$

$S_{\mu^{-1}}(\cdot)$ is the constriction factor, where

$$S_\tau(x) = \begin{cases} x - \tau & \text{if } x > \tau \\ x + \tau & \text{if } x < -\tau \\ 0 & \text{else.} \end{cases} \quad (18)$$

E is updated by

$$\begin{aligned} \arg \min_{D-C \leq E \leq D} \lambda \|P_\Omega(E)\|_1 + \frac{\mu}{2} \|D - A - E + \mu^{-1}Y\|_F^2 \\ = P_{D-C \leq E \leq D} \left(P_\Omega \left(S_{\frac{\mu}{\lambda}} \left(D - A + \mu^{-1}Y \right) \right) \right. \\ \left. + P_\Omega(D - A + \mu^{-1}Y) \right). \end{aligned} \quad (19)$$

$P_{D-C \leq E \leq D}$ is the projector onto the span of matrices ranging from $D - C$ to D . This leads to a matrix completion based traffic data imputation method (MCI) described in Algorithm 1.

3. Experiments

The evaluation of an imputation method's performance is a multiobjective problem. In this section, four key performance indicators of the proposed method are discussed, which are the accuracy, stability, robustness, and computation complexity.

3.1. The Test Data. To evaluate the proposed method, traffic flow datasets from PeMS [9] open database are used. The dataset is collected from Detector 400141. The detector is located at north bound freeway I880. The freeway has four lanes under surveillance. The sampling period is between July 11, 2013 and July 30, 2013. The data are almost all observed with a 99.9% observed ratio.

The assumption is an important premise for the missing traffic data imputation methods based on inductive learning. For the proposed MCI, the traffic volume matrix is assumed to be low-rank. The correctness of this assumption is validated by the low-rank approximation of the original data. The low-rank approximation data is computed by singular value decomposition according to Eckart-Young theorem [30]. If we consider singular value decomposition (SVD) of the constructed traffic volume matrix M_G , we get

$$M_G = U_G \nabla_G V_G^T, \quad (20)$$

where columns of U_G and V_G^T are left-singular vectors and right-singular vectors of M_G , respectively. The diagonal entries of ∇_G are equal to the singular values of M_G .

The full-rank matrix M_G can be approximated as a low-rank matrix \widehat{M}_G by the SVD of M_G , namely,

$$\widehat{M}_G = U_G \widehat{\nabla}_G V_G^T, \quad (21)$$

where $\widehat{\nabla}_G$ is the same matrix as ∇_G except that it contains only the r largest singular values (the other singular values are replaced by zero). The low-rank approximation of the selected traffic data is given in Figure 1.

As we can see from Figure 1, the approximated low-rank data basically keeps the characteristic of origin data. From the results, it can be concluded that the low-rank hypothesis of traffic volume matrix is reasonable.

3.2. *Quantitative Measures.* The set of measures including MAE, MAPE, and SDE allows one to directly evaluate the performance of multiple imputation techniques

3.2.1. *Accuracy.* In this paper, the mean absolute percentage error (MAPE) is used to evaluate the performance of missing traffic data imputation. However, the MAPE will be lower if the traffic volumes are higher [31]. In observance of this phenomenon, this paper also applies the mean absolute error (MAE) as a complementary measure for MAPE.

The mean absolute error (MAE) is defined to be

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |x_t - M_t|. \quad (22)$$

The mean absolute percentage error (MAPE) is defined to be

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - M_t}{x_t} \right| \times 100\%, \quad (23)$$

where n is the total number of missing data, x_t is the observed value, and M_t is the reconstructed value.

3.2.2. *Stability.* The standard deviation of errors (SDE) of the test methods is evaluated. The smaller SDE means that the errors are tightly clustered around the mean value [29]:

$$\text{SDE} = \sqrt{\text{var}(M_t - x_t)}. \quad (24)$$

3.2.3. *Robustness.* The robustness is evaluated by the accuracy on the dataset added outlier under different missing ratio.

3.3. *The Results without Outlier.* In this part, we evaluate the performance of MCI algorithm and compare it with other state-of-the-art algorithms including PCA-based PPCA [7], SVT [22], and IALM [23] on random missing case without outlier. For MCI, the tolerance on the $\| \cdot \|_f$ of $D - A - E$ divided by $\| \cdot \|_f$ of D in the gradient is set to 0.01, and the maximum number of iterations is set to $10^3 \lambda$ which is set to 150. For SVT and IALM, the tolerance on the $\| \cdot \|_f$ is set to 0.01, and the maximum number of iterations is also set to 10^3 . For PPCA, similar to [7], the tolerance is set to 0.01, and the latent space is set to 15.

In order to better verify the change in imputation performance, the total missing ratio (the number of missing data points divided by the total number of data points) is set from 5% to 70%. The MAE and MAPE curves are shown in Figure 2.

In Figure 2, all the methods achieved equal results under missing ratio lower than 30%. However, the performances of other methods except for MCI degrade sharply when the missing ratio is higher than 50% except MCI. The reason may be that the MCI can utilize the physical limitation of traffic flow in the imputation process while the other methods ignored the physical limitation of traffic flow data.

Traffic volume must be nonnegative and less than the value of road capacity. The PPCA imputation strategy is a

kind of statistical method which imputes data through the a priori statistical characteristics of data. As shown in Figure 3, the PPCA method may give a negative value of volume during the low flow rate interval. The phenomenon also can be found in the other two matrix completion strategies without the constraints of nonnegativity and road capacity in their objective function. Our proposed MCI can tackle this shortcoming by adding the limits to the algorithm. The negative and overcapacity value is not observed in the experiments of MCI. The possibilities of four methods that produce unreliable results in our experiments are given in Table 1 (the frequency of experiment results with unreliable results: negative or overcapacity).

In the above experiments, the accuracy of MCI has been tested. Then, we will test the stability of imputation methods by SDE under different missing ratio. As shown in Figure 4, the SDEs of MCI and IALM using augmented Lagrangian function are lower than PPCA and SVT. It suggests that the MCI not only can compute missing data more accurate but also more stable by employing the augmented Lagrangian function.

3.4. *Missing Data Imputation with Outlier.* The above experiments assume that the data have not been spoiled by outliers. However, the traffic flow series are often corrupted by the outliers which are caused by numerous reasons [32]. Unfortunately, these outliers are usually not easy to be isolated by the traditional missing traffic data imputation approaches. Thus, the recovery of outlier and imputation of missing data are often completed in different frameworks separately [7, 29].

For the problem, the MCI algorithm makes it possible to impute missing data and recover outlier in a unify framework by adding the sparse matrix E .

There are various kinds of outlier in traffic data. Here, we only consider two common scenarios of outliers:

- (a) volume out of range (VOR): percentage of the detector records with volumes larger than 1000 v/5 min;
- (b) volume repeating zero (VRZ): percentage of the detector records with repeating zero volumes for 30 min.

It is hard to enumerate all the situations with different mixing ratios of the two outliers' scenarios. In the experiments, the methods are tested on a typical situation by assuming that the mixed VOR and VRZ data have a ratio of 1:1. Ratios of outlier data are set from 5% to 15% and the outlier data are produced randomly. The missing data ratio is set to 30%. All the results are averaged by 10 instances. The MAE and MAPE for missing data and outlier recovery are both given in Table 2. Figure 5 presents the part of traffic volume data and reconstructed volume data. The results show that MCI could impute the missing data and recover the traffic volume outlier data with a reliable performance.

3.5. *Computation Complexity of MCI Approach.* As the same as IALM [25], it is not necessary to compute the full SVD in MCI. By using Lansvd [21], a fast SVD method that only computes singular values larger than a particular threshold

and their corresponding singular vectors, the complexity of the singular value decomposition is not a problem for MCI. And the computation speed of MCI is faster than traditional matrix completion based methods such as SVT by utilizing the augmented Lagrangian function [25].

It is not easy to choose the parameter λ which is the weight parameter between the rank of matrix and the number of sparse outlier. For the traffic data without outliers, setting λ larger than 100 can obtain a good performance. But for data corrupted by outliers, a proper lower value of λ will achieve better results. In this paper, we suggest $\lambda = 150$ for real application for data without corruption of outlier and $\lambda = 0.05$ for corrupted data.

4. Conclusion and Future Works

In this paper, a matrix completion method which fully utilizes the physical limitation of traffic volume and the day mode similarity has been proposed dealing with missing traffic flow problem. The experiment shows that the proposed method is more reasonable, accurate, and stable than the state-of-art methods for traffic flow data. Moreover, the proposed MCI can impute missing data and recover the outlier in a unify framework with a reliable performance.

Future research should look into missing traffic data imputation method that incorporates spatial and temporal correlations among adjacent detectors to improve imputation accuracy. In addition, future studies may evaluate the performance of MCI on other parameters such as speed and occupancy. It still needs more researches on the appropriate choice of parameter for the MCI.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The research was supported by NSFC (Grant nos. 61271376, 51308115, and 91120010), National Basic Research Program of China (973 Program no. 2012CB725405), and Beijing Natural Science Foundation (4122067).

References

- [1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [2] R. C. Jou and K. H. Chen, "A study of freeway drivers'demand for real-time traffic information along main freeways and alternative routes," *Transportation Research C*, vol. 31, pp. 62–72, 2013.
- [3] E. Azimirad, N. Pariz, and M. B. N. Sistani, "A novel fuzzy model and control of single intersection at urban traffic network," *IEEE Systems Journal*, vol. 4, no. 1, pp. 107–111, 2010.
- [4] C. P. I. J. van Hinsbergen, T. Schreiter, F. S. Zuurbier, J. W. C. van Lint, and H. J. van Zuylen, "Localized extended kalman filter for scalable real-time traffic state estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 385–394, 2012.
- [5] W. Qiao, A. Haghani, and M. Hamed, "A nonparametric model for short-term travel time prediction using bluetooth data," *Journal of Intelligent Transportation Systems*, vol. 17, no. 2, pp. 165–175, 2013.
- [6] S. A. Zargari, S. Z. Siabil, A. H. Alavi, and A. H. Gandomi, "A computational intelligence-based approach for short-term traffic flow prediction," *Expert Systems*, vol. 29, no. 2, pp. 124–142, 2012.
- [7] 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.
- [8] S. Turner, L. Albert, B. Gajewski, and W. Eisele, "Archived intelligent transportation system data quality: preliminary analyses of San Antonio TransGuide data," *Transportation Research Record*, vol. 1719, pp. 77–84, 2000.
- [9] PeMS, "California performance measurement system," <http://pems.dot.ca.gov/>.
- [10] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal mainstream traffic flow control of large-scale motorway networks," *Transportation Research C*, vol. 18, no. 2, pp. 193–212, 2010.
- [11] L. H. Nguyen and W. T. Scherer, "Imputation techniques to account for missing data in support of intelligent transportation systems applications," Tech. Rep. UVACTS-13-0-78, University of Virginia, Center for Transportation Studies, 2003.
- [12] J. R. Xu, X. Y. Li, and H. J. Shi, "Short-term traffic flow forecasting model under missing data," *Journal of Computer Applications*, vol. 30, pp. 1117–1120, 2010.
- [13] 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 C*, vol. 13, no. 5-6, pp. 347–369, 2005.
- [14] W. 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*, vol. 16, no. 3, pp. 159–176, 2012.
- [15] M. Zhong, P. Lingras, and S. Sharma, "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques," *Transportation Research C*, vol. 12, no. 2, pp. 139–166, 2004.
- [16] L. Li, Y. Li, and Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence," *Transportation Research C*, vol. 34, pp. 108–120, 2013.
- [17] H. Tan, G. Feng, J. Feng, W. Wang, Y. J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research C*, vol. 28, pp. 15–27, 2013.
- [18] H. Tan, J. Feng, Z. Chen, F. Yang, and W. Wang, "Low multilinear rank approximation of tensors and application in missing traffic data," *Advances in Mechanical Engineering*, vol. 2014, Article ID 157597, 12 pages, 2014.
- [19] H. Tan, B. Cheng, J. Feng, G. Feng, W. Wang, and Y. J. Zhang, "Low-n-rank tensor recovery based on multi-linear augmented Lagrange multiplier method," *Neurocomputing*, vol. 119, pp. 144–152, 2013.
- [20] H. Tan, C. Bin, W. Wuhong, Z. Yu-Jin, and R. Bin, "Tensor recovery via multi-linear augmented lagrange multiplier

- method,” in *Proceedings of the 6th International Conference on Image and Graphics (ICIG '11)*, pp. 141–146, August 2011.
- [21] X. Yuan and Y. Junfeng, “Sparse and low-rank matrix decomposition via alternating direction methods,” preprint 2009.
- [22] F. de la Torre and M. J. Black, “A framework for robust subspace learning,” *International Journal of Computer Vision*, vol. 54, no. 1–3, pp. 117–142, 2003.
- [23] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [24] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [25] Z. Lin, M. Chen, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” 2010, <http://arxiv.org/abs/1009.5055>.
- [26] E. J. Candès and Y. Plan, “Matrix completion with noise,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.
- [27] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *Journal of the ACM*, vol. 58, no. 3, article 11, 2011.
- [28] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [29] P. Indyk and M. Ružić, “Near-optimal sparse recovery in the L1 norm,” in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS '08)*, pp. 199–207, October 2008.
- [30] G. W. Stewart, “On the early history of the singular value decomposition,” *SIAM Review*, vol. 35, no. 4, pp. 551–566, 1993.
- [31] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, “Detecting errors and imputing missing data for single-loop surveillance systems,” *Transportation Research Record*, no. 1855, pp. 160–167, 2003.
- [32] S. Chen, W. Wang, and H. van Zuylen, “A comparison of outlier detection algorithms for ITS data,” *Expert Systems with Applications*, vol. 37, no. 2, pp. 1169–1178, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

