Traffic Missing Data Completion With Spatial-temporal Correlations

Huachun Tan (corresponding author)
Department of Transportation Engineering,
Beijing Institute of Technology, Beijing 100084, CHINA
Tel: (86)-10-68914582
E-mail: tanhc@bit.edu.cn

Yuankai Wu
Department of Transportation Engineering,
Beijing Institute of Technology, Beijing 100084, CHINA
Tel: (86)-18301509972
5433809@bit.edu.cn

Jianshuai Feng
Institute of Automation,
Chinese Academy of Sciences, Beijing 100190,CHINA
Tel: 86-15312489827

Wuhong Wang
Department of Transportation Engineering,
Beijing Institute of Technology, Beijing 100084, CHINA
wangwh@bit.edu.cn

Bin Ran
Department of Civil and Environmental Engineering,
University of Wisconsin-Madison, Madison, USA
Tel: 1-608-332-0528
bran@wisc.edu

Submitted to the 93rd Transportation Research Board Annual
Meeting for possible presentation and publication

4,976 words + 9 figures + 1 tables

November 16, 2013

ABSTRACT

The missing data problem remains as a difficulty in transportation information system, which seriously restricted the application of intelligent transportation system, e,g. traffic control and traffic flow prediction. To solve this problem, numerous imputation methods had been proposed in the last decade. However, few existing studies had fully used the spatial correlation for traffic data imputation. In this paper, tensor based imputing method, which had been proven to be an effective imputation method, is applied to multi-detector missing data imputation for freeway corridor by constructing the traffic data into a 4-way spatial tensor. We make three main contributions in this paper: (a) Various tensor patterns are explored to model the traffic data, and take the multi-detectors into account. (b) Various tensor completion methods are explored and evaluated for missing traffic data imputation. Experiments show HaLRTC is more robust for missing traffic data than TDI. (c) The coefficient of the number of loop detectors used for missing traffic volume and speed data imputation is studied. Experiment results show the number of locations related to the spatial-temporal correlation of traffic data.

INTRODUCTION

With a steady increase of freeway traffic in the recent years worldwide, the traffic congestion of freeways becomes more serious. The freeway traffic congestion can no longer be dealt with simply by extending more highways for economical and environmental reasons (Kerner, 2009). As a consequence, the optimization of existing traffic network especially the freeway corridor control (Liu et al., 2011) has increasingly become a more desirable alternative for management of freeway traffic congestion. Intelligent transportation systems (ITS) play a significant role in optimizing the existing traffic network. Real-time traffic data is one of the key factor to ITS. It is evidently indicated that the conventional ITS will eventually evolve into a data-driven intelligent transportation system. And traffic data that are collected from multiple sources such as loop detector, GPS and video sensors will become more and more important in ITS. (Ran et al., 2012; Zhang et al., 2011)

Unfortunately, missing data problems are inevitable due to detector faults or transmission distortion (Lin & Chang, 2006; Faouzi et al., 2011), which seriously restricts the application and generalization of intelligent transportation systme. 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 (Carlson et al., 2010). In traffic forecast area, if there exists missing data, the predicting performance will reduce sharply (Xu et al., 2010; Van Lint et al., 2005). Without proper imputation methods, traffic counts with missing values are usually either discarded or simply estimated, which may seriously affect the performance of ITS. Consequently, it is very urgent to develop a method with better effect to estimating the missing data.

The frequently used imputation methods for missing traffic data are historical (neighboring) imputation methods(Ni et al., 2005), spline (including linear)/regression imputation methods (Chen & Shao, 2000), autoregressive integrated moving average (ARIMA) models (Zhong et al., 2004) and Probabilistic Principal Component Analysis (Qu et al., 2009). These methods focus on imputing missing data for a single loop detector, which often utilize the temporal correlations such as day mode periodicity, week mode periodicity and interval variation of traffic data to estimate missing data. Nevertheless, the traffic data are spatial-temporal correlated (Wu et al., 2012; Krawczyk et al., 2011) . Compared with temporal correlations, the spatial correlations of traffic data have not been fully utilized. The most state-of-art methods only use spatial information from neighbor detectors (Zhang, 2013; Zhang & Liu, 2009; Li et al., 2013). However, the traffic data are correlated not only in short-distance (Liu et al., 2009b), but also in a large area (Min & Wynter, 2011) especially in a freeway corridor (Van Lint & Hoogendoorn, 2010). As a result, only using neighbor detector information is not the best approach for imputation of missing traffic data.

Recently, a tensor (multi-way array) based method (Tan et al., 2013b; Huachun Tan & Zhang, 2013) has been applied to missing traffic data imputation and outlier traffic data recovery. The traffic data are modeled by multi-way matrix (tensor) pattern, and the missing traffic data are estimated by tensor completion method. Tensor completion allows for combining and utilizing the multi-mode temporal correlations (e.g., week-mode, day-mode, and interval-mode) to estimate the missing data, which has been proved to be a efficient tool to model traffic data for missing traffic data imputation. Despite the good results of tensor-based method, this work is still applied for single loop detector missing data imputation.

In this paper, we focus on the missing traffic data completion for multi-loop detectors on freeway corridor. Motivated by the power of tensor pattern in modeling multi-correlations of traffic data and the reliable performance of tensor completion in missing traffic data imputation, this paper explores the ability of tensor based method for multi-loop detector's missing data imputation. The traffic data are constructed into various 4-way spatial tensor, which covers the spatial information of the freeway corridor. Two tensor completion methods, including HaLRTC (Liu et al., 2009a) and TDI (Tan et al., 2013b), are explored to mine the underlying spatial-temporal information and impute the missing traffic data. Experimental results on missing traffic volume and speed data show that the 4-way tensor considering the spatial information is better than 3-way tensor without spatial correlation. Tensor completion method based on trace norm -

HaLRTC outperforms the method base on tensor decomposition - TDI. The best number of loop detectors for missing traffic data completion is also studied. Experiment results show the spatial-temporal correlation of traffic data related to the number of loop detectors.

This paper is organized as follows: The necessary knowledge about tensor and tensor completion are given in section 2. The tensor model for freeway corridor is conducted in section 3. In section 4, the experiment results are given. The conclusion and future works are discussed in section 5.

TENSOR BASIC AND TENSOR COMPLETION

Notation and Tensor

Tensor which is also called the multidimensional array is the higher-order generalization of vector and matrix. In this paper, the nomenclatures and the notations in (Acar et al., 2011; Tan et al., 2013a) on tensor are partially adopted. Scalars are denoted by lowercase letters $(a, b, c \ldots)$, vectors by bold lowercase letters $(\mathbf{a}, \mathbf{b}, \mathbf{c} \ldots)$,and matrices by uppercase letters $(\mathbf{A}, \mathbf{B}, \mathbf{C} \ldots)$. Tensors are written as calligraphic letters $(\mathcal{A}, \mathcal{B}, \mathcal{C} \ldots)$.

N-mode tensors are denoted as $\mathcal{A} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_N}$. Its elements are denoted as $a_{i_1 \ldots i_k \ldots i_n}$, where $1 \leq i_k \leq I_k, 1 \leq k \leq N$. The mode-n unfolding (also called matricization or flattening) of a tensor $\mathcal{A} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is defined as $unfold(\mathcal{A}, n) = \mathbf{A}_n$. The tensor element $(i_1, i_2, , i_N)$ is mapped to the matrix element (i_n, j), where

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k, \quad with \;\; J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m. \tag{1}$$

Therefore, $\mathbf{A}_n \in R^{I_n \times J}$ where $J = \prod_{\substack{k=1 \\ k \neq n}}^{N} I_k$. The $n - rank$ of a N-dimensional tensor $\mathcal{A} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_N}$, denoted by r_n, is the rank of the mode-n unfolding matrix $\mathbf{A}_{(n)}$.

$$r_n = rank_n(A) = rank(A_{(n)}). \tag{2}$$

The inner product of two same-size tensors $\mathcal{A}, \mathcal{B} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is defined as the sum of the products of their entries, i.e.

$$(\mathcal{A}, \mathcal{B}) = \sum_{i_1} \ldots \sum_{i_k} \ldots \sum_{i_N} a_{i_1 \ldots i_k \ldots i_N} b_{i_1 \ldots i_k \ldots i_N} \tag{3}$$

Given two tensors $\mathcal{A}$ and $\mathcal{B}$ of same size $I_1 \times I_2 \times \ldots \times I_N$, their Hadamard (element wise) product is denoted by $\mathcal{A} * \mathcal{B}$, is defined as

$$(\mathcal{A} * \mathcal{B})_{i_1 \ldots i_k \ldots i_N} = a_{i_1 \ldots i_k \ldots i_N} b_{i_1 \ldots i_k \ldots i_N} \tag{4}$$

The corresponding Frobenius norm is $\|\mathcal{A}\|_F = \sqrt{(\mathcal{A}, \mathcal{A})}$. For any $1 \leq n \leq N$. The n-mode (matrix) product of a tensor $\mathcal{A} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_N}$ with a matrix $\mathbf{M} \in \mathcal{R}^{J \times I_N}$ is denoted by $\mathcal{A} \times_n \mathbf{M}$ and is of size $I_1 \times \ldots \times I_n - 1 \times J \times I_n + 1 \times \ldots \times I_N$. In terms of flattened matrix, the n-mode product can be expressed as

$$\mathcal{Y} = \mathcal{A} \times_n \mathbf{M} \iff \mathbf{Y}_(n) = \mathbf{M}\mathbf{A}_{(n)} \tag{5}$$

The definition of the trace norm for the general tensor case is

$$\|\mathcal{A}\|_* = \sum_{i=1}^{N} \alpha_i \|\mathbf{X}_{(i)}\|_* \tag{6}$$

where α_i are constants satisfying $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i = 1$. In essence, the trace norm of a tensor is a convex combination of the trace norms of all matrices unfolded along each mode.

Low-n-rank tensor completion for traffic data

As (Tan et al., 2013b) analyzed, As shown in Fig.1, N days' traffic data of this detector can be formulated as a three-way data tensor $\mathcal{A} \in \mathcal{R}^{N \times 24 \times 12}$ according to the multi mode temporal correlations of traffic data. Traffic data are highly correlated in multiple modes. As a result, the missing data in the traffic tensor can be estimated by tensor completion.

In general, there are two methods to estimate missing data including methods based on tensor decomposition (TDI) (Tan et al., 2013b) and trace norm based tensor completion method (LRTC) (Liu et al., 2009a, 2013).

In this paper, the high accuray low-rank tensor completion algorithm is applied to multi-loop missing traffic data completion. The method is to use the trace norm $\| \|_*$ to approximate the n-rank of tensors. The advantage of the trace norm is that it is the tightest convex envelop for the rank of matrices (Srebro & Shraibman, 2005). By introducing additional ralaxation factor matrices $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$, the minimum of tensor $\mathcal{A}$ trace norm can be solved independently. And it has been proved that the trace-norm based methods outperform tensor decomposition based method for missing data in literature. The optimization problem is:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{M}_i} : \sum_i \alpha_i \|\mathbf{M}_{(i)}\|_* + \frac{\beta_i}{2} \|\mathcal{X}_{(i)} - \mathbf{M}_i\|_F{}^2 \\ s.t : \mathcal{X} * \mathcal{W} = \mathcal{A} * \mathcal{W} \end{aligned} \tag{7}$$

$\mathcal{W}$ is a nonnegative weight tensor with the same size as $\mathcal{A}$ to indicate where missing data happen. Formally, it can be defined as

$$\mathcal{W}_{i_1 i_2 i_3} = \begin{cases} 1 & if\ a_{i_1 i_2 i_3}\ is\ known \\ 0 & if\ a_{i_1 i_2 i_3}\ is\ missing \end{cases} \tag{8}$$

Then, the block coordinate descent(BCD) is employed to optimize the problem. The basic idea of block coordinate descent is to optimize a group of variables while fixing the other groups (Tseng, 2001).The accuracy of low-n-rank tensor completion can be promoted by using ADMM framework. The low-n-rank tensor completion algorithm employing ADMM is called HaLRTC. More detailed discussion can be found in (Liu et al., 2009a).

TENSOR MODEL FOR FREEWAY CORRIDOR TRAFFIC DATA

We apply tensor completion to freeway corridor traffic data missing problem by taking the spatial-temporal correlation into account in this section. The traffic tensor of freeway corridor is conducted by 12 locations in a freeway corridor from PeMS (2013) database. As shown in Fig.2, these loop detectors are located at south bound freeway SR58. The sampling period is between May 13. 2013 and July 21. 2013. The data are nearly all observed with a very low missing ratio (about 2%), which have been imputed by built-in imputation methodology of PeMS (Crossroads, 2008). Due to the low missing ratio, the data set is regarded as an approximate complete data set.

Intuitively, both speed and volume are highly spatial-correlated in a freeway corridor as shown in Fig.3 and 4. To use this strong correlation to handle multi-loop detectors missing data in a freeway corridor by tensor completion, a location dimension is added to single loop traffic 3-way tensor to construct a new 4-way tensor $\mathcal{A} \in \mathcal{R}^{24(hour) \times 12(points) \times 70(days) \times 12(locations)}$. The 4-way tensor is shown in Fig.4.

Some previous works (Signoretto et al., 2011) show that the multi-mode correlations of data have a great effect on the performance of the tensor completion. Obviously, quantitative analysis of the traffic tensor multi-mode correlations not only helps to choose the number of loop detectors constructing tensor, but also helps to determine the parameters of tensor completion methods. Formally, the correlations of traffic data are measured by similarity coefficient:

$$s_m = \frac{\sum_{n_m \geq i \geq j \geq 1} R_m(i,j)}{n_m(n_m-1)/2} \tag{9}$$

where n_m refers to the number of data points; $R_m(i,j)$ refers to the entry of the correlation coefficient matrix of the m-mode unfolded matrix of the tensor.

The hour mode, interval mode, day mode and link mode correlations of speed and volume traffic data tensor ($\mathcal{A} \in \mathcal{R}^{12\times 24\times 70\times 12}$) for the freeway corridor are given in Table.1

Table.1 shows that both volume and speed are highly correlated in day mode and interval mode with a very high coefficient over 0.9. The speed data are stronger correlated in hour mode than volume data while the correlation of speed is very low with such a tensor size in location mode.

Furthermore, the relation between length of loop dimension (number of loop detectors) and multi-mode correlations are tested. Along the direction from west to east (upstream to downstream) in Fig.3, different number from 2 to 12 of loop detectors are used to construct 11 tensor models. The multi-mode correlations of these tensor models are given in Fig.5.

The results reflect different tendencies of each mode correlation with location of loop dimension length. In the general trend, volume tensor is stronger spatial-correlated while weaker temporal-correlated especially in interval mode with a longer loop dimension length. On the contrast, speed tensor is stronger temporal-correlated while weaker spatial-correlated with the increase of loop dimension length. Traffic volume are strongest temporal-correlated when only using 2 locations of loop detector and strongest spatial-correlated when using 7 loop detectors. Traffic speed are strongest temporal-correlated when using 9 locations of loop detectors and strongest spatial-correlated when using 4 locations of loop detectors. The results provide a reference for constructing the spatial-traffic-data tensor while doing missing traffic data imputation and selecting the parameters of tensor completion algorithm.

NUMERICAL EXPERIMENTS AND ANALYSIS

In this paper, the proposed spatial-tensor is tested on two kinds of missing patterns as follows:
1) Missing Completely at Random (MCR), in which the missing points are completely independent of each other.
2) Missing at Random (MR), in which the missing points are related to the neighboring points. Thus, they usually appear as a small group of sequential points lost at one time, but the groups are randomly scattered (Qu et al., 2009)

We assess the performance of proposed spatial-tensor in terms of its ability to reconstruct the missing data. The spatial-tensor is tested on HaLRTC algorithm (Liu et al., 2009a, 2013). The performance are compared with 3-way temporal-tensor imputing missing data by TDI (Tan et al., 2013b) and HaLRTC.

For the 4-way tensor. The selection and parameter setting of HaLRTC are: The weighted coefficient α_i is set to [0.19,0.27,0.27,0.27] for volume data, [0.32,0.32,0.32,0.04] for speed data, The maximum iterations are set to 500, the tolerance of the relative difference of outputs of two neighbor iterations are set to 10^{-5}. Under 3-way case, the n-rank of TDI is set to [3,3,3] for volume, [3,3,2] for speed. The weighted coefficient α_i of HaLRTC is set to [1/3,1/3,1/3] for volume and [0.39,0.39,0.22] for speed.

Evaluation criteria of missing imputing performance

The imputing performance is evaluated by the root mean squared error (RMSE) between the estimated missing points test and the original data points t_{real}. RMSE is a commonly used error criteria, which reflects the average performance for the missing data imputing.

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(t_{real} - t_{east})_m^2} \tag{10}$$

where $(t_{real} - t_{east})_m$ are the m-th error between the known real value and the estimated value, M is the number of missing data, which can be used to calculate the missing ratio, as follows,

$$r = M/N \times 100\% \tag{11}$$

where r means the ratio of missing data; N means the total data number of test data.

MR is generated with respect to the observed patterns as follows: The location of the first missing data point in each missing group is generated to uniformly be distributed. The length of the missing data series is modeled as a normal distributed number between 0 and 20.

All the methods were performed using Matlab on a Windows Workstation with a Dual-Core Intel(R) Core(TM) 2.50 GHZ CPU and 4GB RAM.

Experiment results

To verify the advantage of 4-way tensor, the traffic data is formed into a 4-way tensor with size of $12 \times 24 \times 70 \times 12$ and the missing data is estimated by HaLRTC in a unify framework. The results of 4-way tensor are compared with the results of estimating missing traffic data in 12 different 3-way tensor with size of $12 \times 24 \times 70$ by TDI and HaLRTC. The missing ratio is ranging from 0.1 to 0.6. The experiment results are shown in Fig.6 and 7.

In Fig 6 and 7, the 4-way tensor outperforms 3-way tensor at almost all the missing rate. The reason is that the 4-way tensor can utilize the information of spatial modes simultaneously, while 3-way tensor only mines temporal correlation and independently estimating missing data without consideration about spatial correlation. It indicates that it is necessary to use spatial information when dealing with missing traffic data and it is far from enough that only using temporal information to impute missing traffic data.

3-way HalRTC outperforms 3-way TDI except when estimating missing volume under MR case. It indicates that low-n-rank tensor completion may be more suitable for traffic data than tensor decomposition based method.

However, it is not easy to determine the best parameter for tensor completion since different traffic data set encode different mode structure characteristics. In this paper, the parameter are firstly set by empirical hypothesis, then best parameter can be easily found by fine-tuning according to the experiment results.

The above results indicate that estimate missing data of each location in a unify spatial-tensor is more reliable than estimating them independently without consideration of spatial correlation when missing data happens in the whole detectors of freeway corridor. Sometimes missing data only occur in a single location of loop detectors. We also studied the appropriate number of location to use under this situation. We set the upstream detector in fig.2 contains missing data with ratio ranging from 10% to 60%, then using loop detectors with number ranging from 1 to 12 to construct the tensor. The results are shown in Fig 8 and 9.

Experiment shows the appropriate number of locations to construct the traffic tensor when missing data only occurs in a loop detector. Both under MCR and MR, the imputation performance can be promoted by using more than 4 loop detectors at the downstream of the loop detector contained missing data to construct traffic tensors. While for speed data, the tensor with length of 3 in loop dimension that only using correlations from 2 downstream locations achieve the highest accuracy. The reason may lie in that speed get strongest spatial correlation with 3 loop detectors. While volume are higher spatial correlated with more than 5 locations as analyzed. The results also show that impute missing volume data, only using adjacent loop detectors is far from enough.

However. It is not easy to determine the best parameter group for HaLRTC since different traffic data set encode different mode structure characteristics. Fortunately, the best parameter group can be easily found by fine-tuning.

CONCLUSION

Freeway traffic control plays a key role on the alleviation of freeway traffic congestion. The traffic control system requires sufficient traffic flow data. The missing data in traffic information system poses a serious challenge for the alleviation of freeway traffic congestion. In previous work, most missing data imputation methods focus on a single loop detector traffic data imputation or only using a limited number of loop detectors to impute missing data.

In this work we have shown an alternative approach to work around the multi-loop missing data problem by tensor completion. We construct a freeway corridor traffic data into a 4-way spatial-tensor. The results shown estimating missing data in a unify framework by tensor completion is more reliable than imputing every single location's missing data independently without consideration about spatial correlation. This provides further evidence that not only missing data in the freeway corridor but also the large-scale area such may can be imputed by a multi-way tensor completion.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (61271376, 61171118 and 91120015), and Beijing Natural Science Foundation (4122067). The authors would like to thank experienced anonymous reviewers for their constructive and valuable suggestions for improving the overall quality of this paper.

REFERENCES

Acar, E., Dunlavy, D. M., Kolda, T. G., & Mørup, M. (2011). Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems*, *106*, 41–56.

Carlson, R. C., Papamichail, I., Papageorgiou, M., & Messmer, A. (2010). Optimal mainstream traffic flow control of large-scale motorway networks. *Transportation Research Part C: Emerging Technologies*, *18*, 193–212.

Chen, J., & Shao, J. (2000). Nearest neighbor imputation for survey data. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-*, *16*, 113–132.

Crossroads, U. (2008). Inc. pems data extraction methodology and execution technical memorandum for the southern california association of governments. 2006.

Faouzi, N.-E. E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges–a survey. *Information Fusion*, *12*, 4–10.

Huachun Tan, G. F. W. W., Jianshuai Feng, & Zhang, Y.-J. (2013). Traffic volume data outlier recovery via tensor model. *Mathematical Problems in Engineering*, *2013*.

Kerner, B. S. (2009). *Introduction to modern traffic flow theory and control: the long road to three-phase traffic theory*. Springer.

Krawczyk, M. J., Ruiz, C. B., & Kułakowski, K. (2011). Situations in trafficąłhow quickly they change. *Central European Journal of Physics*, *9*, 1452–1457.

Li, L., Li, Y., & Li, Z. (2013). Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. *Transportation Research Part C: Emerging Technologies*, *34*, 108–120.

Lin, P.-W., & Chang, G.-L. (2006). Modeling measurement errors and missing initial values in freeway dynamic origin–destination estimation systems. *Transportation Research Part C: Emerging Technologies*, *14*, 384–402.

Liu, J., Musialski, P., Wonka, P., & Ye, J. (2009a). Tensor completion for estimating missing values in visual data. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 2114–2121). IEEE.

Liu, J., Musialski, P., Wonka, P., & Ye, J. (2013). Tensor completion for estimating missing values in visual data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*, 208–220.

Liu, L., Khalilia, M., Tan, H., & Zhuang, P. (2009b). Traffic pattern forecasting using time series analysis between spatially adjacent sensor clusters. In *Machine Learning and Cybernetics, 2009 International Conference on* (pp. 3155–3160). IEEE volume 6.

Liu, Y., Chang, G.-L., & Yu, J. (2011). An integrated control model for freeway corridor under nonrecurrent congestion. *Vehicular Technology, IEEE Transactions on*, *60*, 1404–1418.

Min, W., & Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, *19*, 606–616.

Ni, D., Leonard, J. D., Guin, A., & Feng, C. (2005). Multiple imputation scheme for overcoming the missing values and variability issues in its data. *Journal of transportation engineering*, *131*, 931–938.

PeMS (2013). California performance measurement system. <`http://pems.eecs.berkeley.edu/`>.

Qu, L., Hu, J., Li, L., & Zhang, Y. (2009). Ppca-based missing data imputation for traffic flow volume: a systematical approach. *Intelligent Transportation Systems, IEEE Transactions on*, *10*, 512–522.

Ran, B., Jin, P. J., Boyce, D., Qiu, T. Z., & Cheng, Y. (2012). Perspectives on future transportation research: Impact of intelligent transportation system technologies on next-generation transportation modeling. *Journal of Intelligent Transportation Systems*, *16*, 226–242.

Signoretto, M., Van de Plas, R., De Moor, B., & Suykens, J. A. (2011). Tensor versus matrix completion: a comparison with application to spectral data. *Signal Processing Letters, IEEE*, *18*, 403–406.

Srebro, N., & Shraibman, A. (2005). Rank, trace-norm and max-norm. In *Learning Theory* (pp. 545–560). Springer.

Tan, H., Cheng, B., Feng, J., Feng, G., Wang, W., & Zhang, Y.-J. (2013a). Low-n-rank tensor recovery based on multi-linear augmented lagrange multiplier method. *Neurocomputing*, .

Tan, H., Feng, G., Feng, J., Wang, W., Zhang, Y.-J., & Li, F. (2013b). A tensor-based method for missing traffic data completion. *Transportation Research Part C: Emerging Technologies*, *28*, 15–27.

Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, *109*, 475–494.

Van Lint, J., Hoogendoorn, S., & van Zuylen, H. J. (2005). Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*, *13*, 347–369.

Van Lint, J., & Hoogendoorn, S. P. (2010). A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil and Infrastructure Engineering*, *25*, 596–612.

Wu, Y.-J., Chen, F., Lu, C., Smith, B., & Chen, Y. (2012). Traffic flow prediction for urban network using spatio-temporal random effects model. In *91st Annual Meeting of the Transportation Research Board (TRB)*.

Xu, J.-r., Li, X.-y., & Shi, H.-j. (2010). Short-term traffic flow forecasting model under missing data. *Journal of Computer Applications*, *30*, 1117–1120.

Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., & Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *Intelligent Transportation Systems, IEEE Transactions on*, *12*, 1624–1639.

Zhang, Y. (2013). Freeway short-term traffic flow forecasting by considering traffic volatility dynamics and missing data situations, .

Zhang, Y., & Liu, Y. (2009). Missing traffic flow data prediction using least squares support vector machines in urban arterial streets. In *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on* (pp. 76–83). IEEE.

Zhong, M., Lingras, P., & Sharma, S. (2004). Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. *Transportation Research Part C: Emerging Technologies*, *12*, 139–166.

LIST OF TABLES AND FIGURES

List of Tables

List of Figures

TABLE 1 the multi-mode correlations of traffic data for the freeway corridor

Data type	Hour mode	Interval mode	Day mode	Loop mode
Volume	0.6844	0.9501	0.9155	0.8977
Speed	0.8454	0.9607	0.9191	0.3686

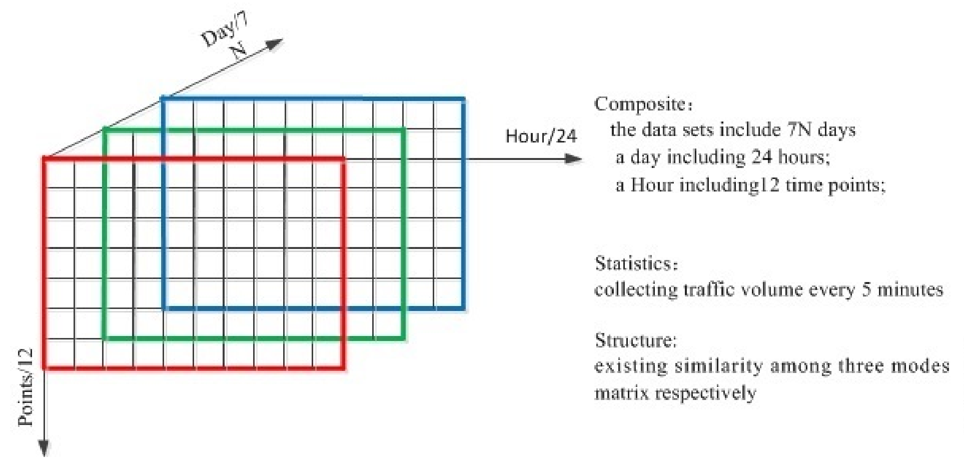

FIGURE 1 Three way traffic data for one loop detector traffic data

FIGURE 2 The Detectors From E-SR58

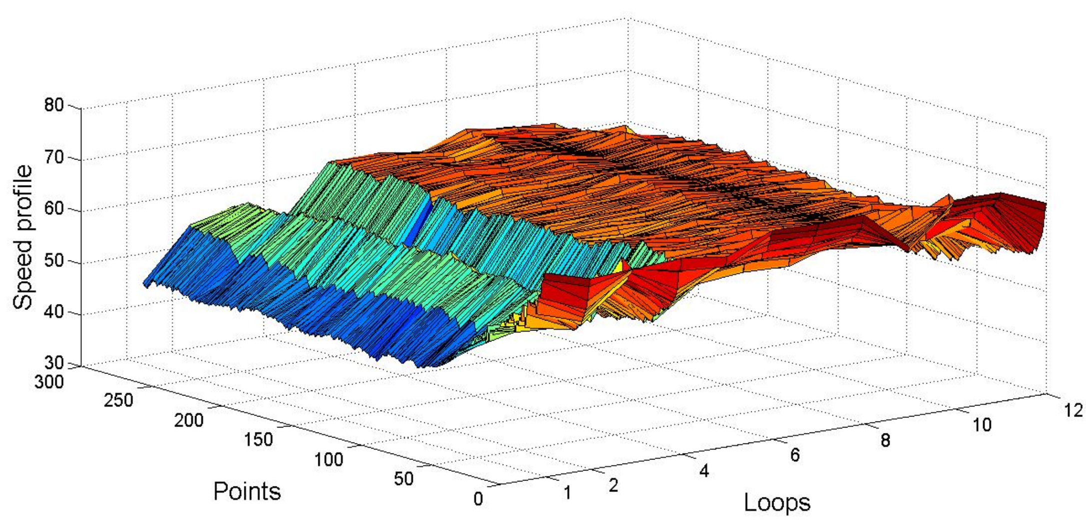

FIGURE 3 The speed data in a freeway corridor

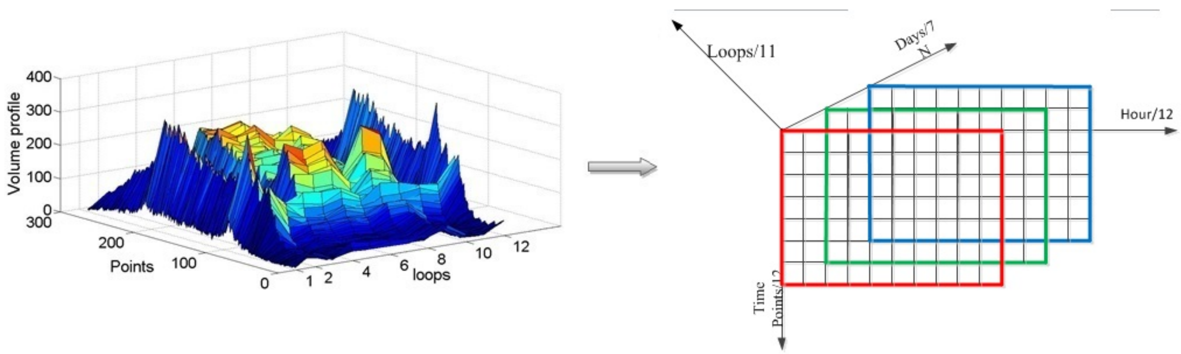

FIGURE 4 The Freeway Corridor Traffic Tensor

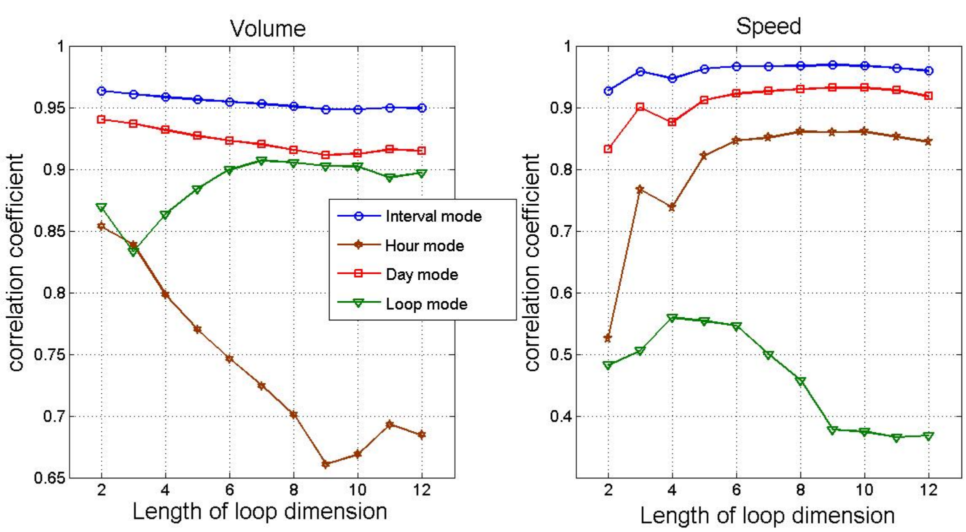

FIGURE 5 The relation between Loop dimension Length and multi-mode correlations

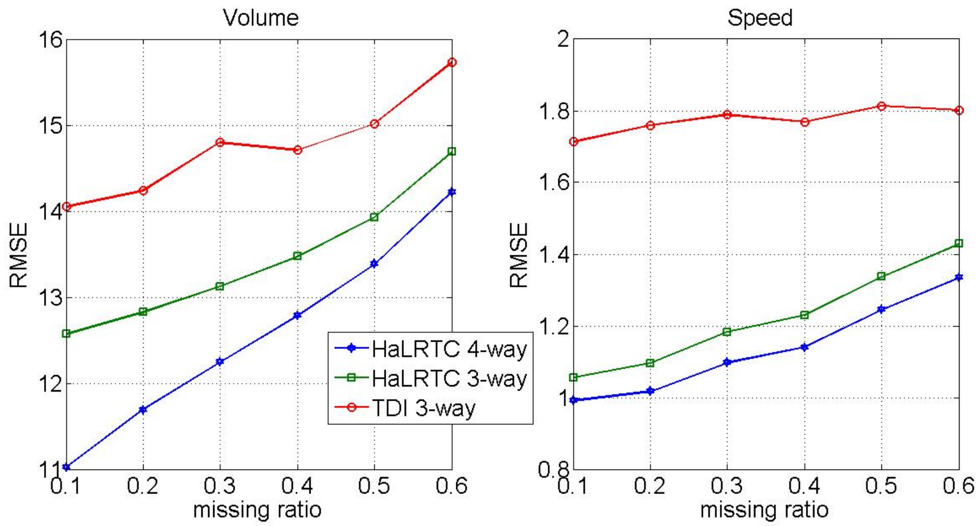

FIGURE 6 The performance of tensor completion under MCR case

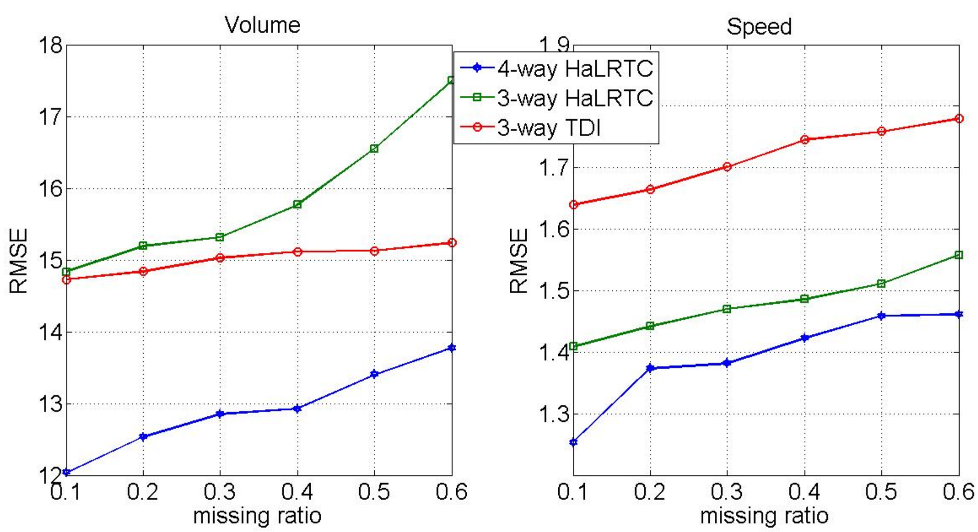

FIGURE 7 The performance of tensor completion under MR case

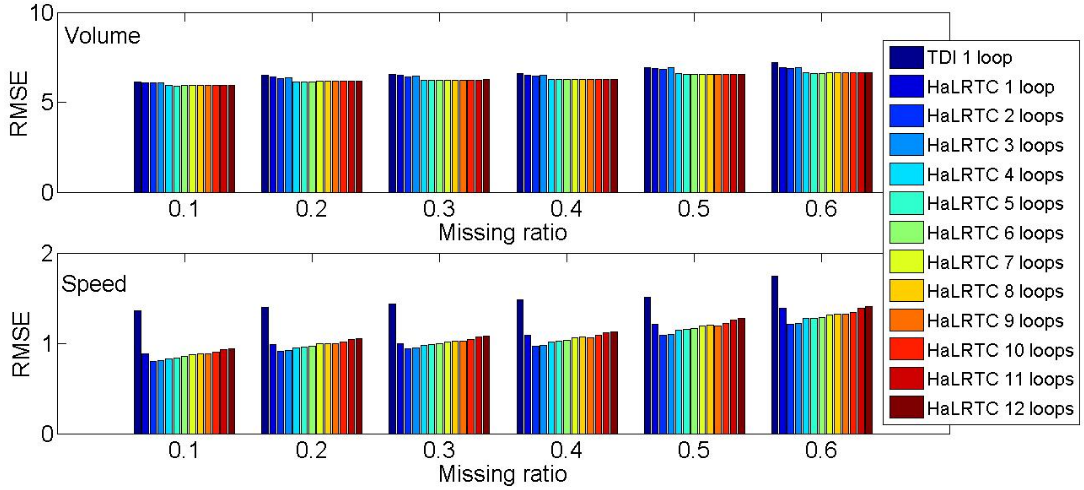

FIGURE 8 The performance of tensor completion under MCR when only one detector contain missing data

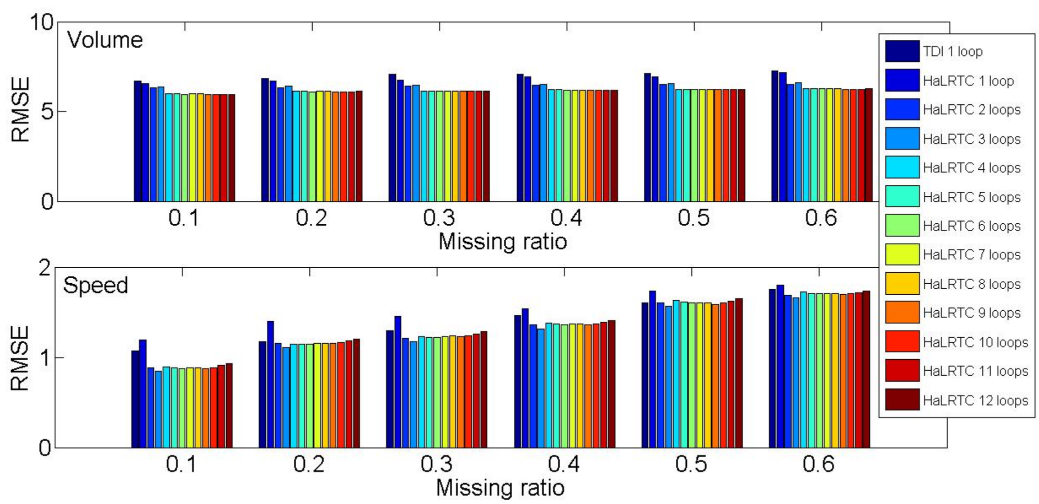

FIGURE 9 The performance of tensor completion under MR when only one detector contain missing data