Transportation Research Part C 86 (2018) 59–77

Contents lists available at ScienceDirect

Transportation Research Part C

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

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

Xinyu Chen, Zhaocheng He*, Jiawei Wang

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

A R T I C L E I N F O

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

A B S T R A C T

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

1. Introduction

1.1. Motivation

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

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

https://doi.org/10.1016/j.trc.2017.10.023
Received 29 June 2017; Received in revised form 24 October 2017; Accepted 27 October 2017

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

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

1.2. Related works

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

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

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

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

1.3. Challenges and contributions

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