Transportation Research Part C 85 (2017) 348–362

Contents lists available at ScienceDirect

Transportation Research Part C

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

Traffic data imputation via tensor completion based on soft thresholding of Tucker core

J.H. de M. Goulart [a,c,*], A.Y. Kibangou [b], G. Favier [c]

[a] Univ. Grenoble Alpes, CNRS, GIPSA-lab, F-38000 Grenoble, France
[b] Univ. Grenoble Alpes, CNRS, Inria, GIPSA-lab, F-38000 Grenoble, France
[c] I3S Laboratory, CNRS, Univ. Côte D'Azur, 06900 Sophia-Antipolis, France

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form 11 September 2017
Accepted 12 September 2017

Keywords:
Intelligent transportation systems
Traffic data imputation
Tensor completion
Soft thresholding
Tucker model

ABSTRACT

Technological limitations and practical difficulties cause inevitable losses of traffic data in the typical processing chain of an intelligent transportation system. This has motivated the development of imputation algorithms for mitigating the consequences of such losses. As the involved datasets are usually multidimensional and bear strong spatio-temporal correlations, we propose for traffic data imputation a tensor completion algorithm which promotes parsimony of an estimated orthogonal Tucker model by iteratively softly thresholding its core. The motivation of this strategy is discussed on the basis of characteristics typically possessed by real-world datasets. An evaluation of the proposed method using speed data from the Grenoble south ring (France) shows that our algorithm outperforms other imputation methods, including tensor completion algorithms, and delivers good results even when the loss is severely systematic, being mostly concentrated in long time windows (of up to three hours) spread along the considered time horizon.

1. Introduction

Many tasks performed by an intelligent transportation system hinge upon a timely, reliable and continuous gathering of data by numerous sensors across several locations. This includes the monitoring and the prediction of traffic conditions, but also the provision of data to users, companies, relevant authorities and researchers for information, management, decision-making and scientific purposes.

Unfortunately, however, the data acquisition process is typically subject to failures which result in the loss of a significant amount of data. Common causes are the malfunctioning of sensing devices, of network links or processing units. This problem is quite pervasive and well documented. For instance, Zhong et al. (2004) have reported a study involving datasets provided by highway agencies in Canada and the United States in which missing data ratios range from 20% up to 90%. Similarly, Smith et al. (2003) mention the occurrence of up to 84% missing entries in datasets gathered by the Texas Transportation Institute.

Data imputation techniques have long played an important role in mitigating the consequences of such a (possibly quite severe) loss of information, Smith et al. (2003), Albright (1991). Many diverse strategies were proposed for this task, including simple *ad-hoc* approaches exploiting temporal or spatial correlation, standard statistical techniques such as regression

* Corresponding author at: I3S Laboratory, 2000 route des Lucioles, bât. Euclide B 06900 Sophia Antipolis, France.
E-mail addresses: jose-henrique.de-morais-goulart@gipsa-lab.fr (J.H. de M. Goulart), alain.kibangou@univ-grenoble-alpes.fr (A.Y. Kibangou), favier@i3s.unice.fr (G. Favier).

https://doi.org/10.1016/j.trc.2017.09.011
0968-090X/

and the expectation-maximization algorithm, neural networks and deep learning algorithms, Zhong et al. (2004), Li et al. (2013), Canudas de Wit et al. (2015), Tang et al. (2015), Duan et al. (2016).

Some of these methods rely on a matrix representation of the traffic data, as done, for instance, in Qu et al. (2009), Li et al. (2015), Asif et al. (2013). In particular, Asif et al. (2013) have employed a matrix completion algorithm, assuming the traffic data can be modeled by a low-rank matrix. In essence, this low dimensionality comes from the correlations contained in the dataset. However, approaches based on matrix models are limited in that only two modes or diversities can be jointly exploited (say, one spatial and one temporal mode). By contrast, methods based on higher-order tensors have no such limitation, allowing to capture the structure of the data more naturally by introducing several temporal and spatial modes. For instance, one can have modes associated with space, day, week, month, minute of the day, and so on.

Recently, the use of low-rank tensor completion (TC) methods has been proposed for addressing the imputation of traffic data, Tan et al. (2013), Asif et al. (2013), Tan et al. (2014), Ran et al. (2015, 2016). The idea is to capitalize on the multidimensional structure of traffic datasets in order to fill in the missing entries by fitting the data to a low-rank tensor model. Similarly to the approach based on matrix completion, this is motivated by the fact that low-rank tensors are quite often able to accurately model this kind of data, because of the strong spatial and temporal correlations that are typically present: different sensors tend to have similar readings in the same time of the day; the readings made by a sensor tend to be highly similar for different days; the trends observed along different weeks tend to resemble each other. The interest in this kind of approach is in part spurred also by the success of matrix completion techniques in many different applications, Candés and Plan (2010).

In this spirit, Tan et al. (2013) propose a first-order algorithm for optimizing the components of a Tucker model having low multilinear rank, henceforth abbreviated as mrank (see Section 2.1 for definitions of these concepts). The same model is adopted by Tan et al. (2014), who apply a Riemannian optimization algorithm to search for a minimum along the product of Grassmann manifolds associated with the model factors. It also underlies the approach used by Ran et al. (2015), Ran et al. (2016), who employ an algorithm based on the minimization of the sum of nuclear norms (SNN) of the model's matrix unfoldings (see Section 2.1) named HaLRTC, Liu et al. (2013). A dynamic tensor completion method was also recently proposed by Tan et al. (2016) for short-term flow prediction using low-rank matrix factorizations of matrix unfoldings of the data tensor.

Instead of exhibiting an ideally low mrank, traffic data tensors (or, more generally, real-world tensors, for that matter) typically possess matrix unfoldings having decaying singular values, which can be seen as a kind of *compressibility* (in the sense used in compressive sensing, Foucart and Rauhut (2013)). The decay rate, however, is often not fast enough to allow accurate modeling with a Tucker model of considerably low mrank. In this situation, algorithms imposing a hard constraint of exactly low mrank often yield unsatisfying results. On the other hand, the estimate obtained by minimizing the SNN of matrix unfoldings is equally unsatisfying in many cases, which is explained by the suboptimality of this approach in terms of sampling requirements, Mu et al. (2014).

As an alternative, this paper proposes an approach for TC based on employing an iterative single imputation scheme which promotes parsimony of an orthogonal Tucker model by softly thresholding its core. When completing real-world data, this strategy indeed outperforms algorithms imposing a hard constraint of exactly low mrank quite often, as well as that based on SNN minimization. We explain how our method can be interpreted in light of the compressibility property above mentioned, arguing that the latter is related to the compressibility of the core of the higher-order singular value decomposition (HOSVD), De Lathauwer et al. (2000). Performance improvement techniques involving an exponentially decaying threshold and a feedback mechanism are also proposed for our algorithm, which brings benefits both in terms of convergence speed and of accuracy.

Our approach is validated by means of simulations performed with data acquired by the Grenoble Traffic Lab (GTL) system, Canudas de Wit et al. (2015). Specifically, we simulate the occurrence of systematic data loss caused by permanent sensor malfunctioning, with missing data windows of up to three hours. The obtained results show that our approach is able to well reconstruct the data, being quite robust to these severely long data losses. It also outperforms four other TC algorithms and a matrix completion one.

The rest of the paper is organized as follows. In Section 2, some tensor basics are recalled, and the tensor completion problem is briefly introduced. Then, in Section 3, we derive our proposed tensor completion method, discuss the relevance of its rationale in the context of traffic data estimation and present the employed performance improvement techniques. The performance of the proposed solution is evaluated in Section 4 through experiments conducted with real traffic data of GTL, which involve comparison with several other imputation techniques. The paper is concluded in Section 5, where some perspectives are drawn for future work.

2. Problem statement and review of existing approaches

2.1. Tensor basics

Before formalizing our problem of interest, let us recall some definitions and tensor decompositions.

Scalars, column vectors, matrices, and tensors of order higher than two are denoted by lower-case, bold lower-case, bold upper-case, and upper-case calligraphic letters, *e.g.*, a, $\mathbf{a}$, $\mathbf{A}$, $\mathcal{A}$, respectively. The operator $\mathrm{vec}(\cdot)$ vectorizes its matrix (or tensor) argument, and the inverse of the vectorization operator is denoted $\mathrm{unvec}(\cdot)$. The symbol $\odot$ represents the Hadamard (element-wise) product.