Abstracts of the Ninth Workshop on Simulation (2018) 1-2

Block tensor train decomposition for missing data estimation

Namgil Lee [1] Jong-Min Kim [2]

We propose a new method for imputation of missing values in large scale matrix data based on a low-rank tensor approximation technique called the block tensor train (TT) decomposition. Given sparsely observed data points, the proposed method iteratively computes the soft-thresholded singular value decomposition (SVD) of the underlying data matrix with missing values. The SVD of matrices is performed based on a low-rank block TT decomposition, by which storage and time complexities can be reduced dramatically for large scale data matrices with a low-rank tensor structure. We implemented an iterative soft-thresholding algorithm for missing data estimation [6], and the SVD with block TT decomposition was computed based on alternating least squares iteration. Experimental results on simulated data demonstrate that the proposed method can estimate a large amount of missing values accurately compared to a matrix-based standard method.

1 Introduction

A tensor refers to a multi-dimensional array, which can be considered as a generalization of vectors and matrices. Tensor decomposition, like matrix SVD, has been developed for a wide scope of applications in signal processing, machine learning, chemometrics, and neuroscience [3]. Traditional tensor decompositions include Candecomp/Parafac (CP) decomposition and Tucker decomposition; see, e.g., [3]. Modern tensor decompositions have been developed more recently to cope with the problem called as the curse-of-dimensionality, which means an exponential rate of increase in the storage and computational costs as the dimensionality of tensors increases [2]. The tensor train (TT) decomposition is one of the modern tensor decompositions which generalize the matrix SVD to higher-order (i.e., multi-dimensional) tensors [7]. Modern tensor decompositions such as the TT decomposition applies not only to higher order tensors, but also to large scale vectors and matrices, by transforming the vectors and matrices into higher-order tensors via a tensorization process [1]. Once the large scale vectors and matrices have been decomposed by TT decomposition, algebraic operations such as the matrix-by-vector multiplication can be performed much efficiently with logarithmically scaled computational costs [7].

[1]Department of Information Statistics, Kangwon National University, Chuncheon, Gangwon 24341, South Korea, E-mail: namgil.lee@kangwon.ac.kr

[2]Statistics Discipline, Division of Sciences and Mathematics, University of Minnesota-Morris, Morris, MN 56267-2134, USA, E-mail: jongmink@morris.umn.edu

In this work, we consider the singular value decomposition (SVD) of a large matrix $\widehat{\mathbf{Y}} \in \mathbb{R}^{I \times J_1 J_2 \cdots J_N}$, where the goal is to compute the R_X largest singular values and the corresponding left/right singular vectors as

$$\widehat{\mathbf{Y}} \approx \mathbf{U}\mathbf{S}\mathbf{V}^\top, \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{I \times R_X}$, $\mathbf{S} = \operatorname{diag}(s_1, \ldots, s_{R_X})$, and $\mathbf{V} \in \mathbb{R}^{J_1 \cdots J_N \times R_X}$. We consider that a large "tall-and-skinny" matrix $\mathbf{V} \in \mathbb{R}^{J_1 \cdots J_N \times R_X}$ is reshaped and permuted into a tensor $\mathcal{V}$ of size $J_1 \times \cdots \times J_n \times R_X \times J_{n+1} \times \cdots \times J_N$. The block-$n$ tensor train (TT) decomposition of $\mathbf{V}$ is defined by a product of a series of low-order tensors as

$$\mathbf{V} \approx \mathcal{V} = \mathcal{V}_1 \bullet \mathcal{V}_2 \bullet \cdots \bullet \mathcal{V}_N, \tag{2}$$

where $\mathcal{V}_m \in \mathbb{R}^{R_{m-1} \times J_m \times R_m}$ $(m \neq n)$ are third-order tensors, $\mathcal{V}_n \in \mathbb{R}^{R_{n-1} \times J_n \times R_X \times R_n}$ is a fourth-order tensor. The tensors $\mathcal{V}_1, \ldots, \mathcal{V}_N$ are called the TT-cores and $R_1, \ldots, R_{N-1}$ are called the TT-ranks. We assume that $R_0 = R_N = 1$. Note that when the large matrix $\mathbf{V}$ is decomposed by the block TT decomposition, the storage cost reduces from $\mathcal{O}(J^N R)$ to $\mathcal{O}(NJR^2)$, where $J = \max(\{J_m\})$ and $R = \max(\{R_m\}, R_X)$. See, e.g., [4, 5, 7], for further properties of TT decomposition and algebraic operations.

References

[1] O. Debals and L. De Lathauwer. Stochastic and deterministic tensorization for blind signal separation. In E. Vincent, A. Yeredor, Z. Koldovsky, and P. Tichavský, editors, *Proceedings of the 12th International Conference on Latent Variable Analysis and Signal Separation*, pages 3–13. Springer International Publishing, 2015.

[2] B. N. Khoromskij. Tensors-structured numerical methods in scientific computing: Survey on recent advances. *Chemometr. Intell. Lab. Syst.*, 110:1–19, 2012.

[3] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51:455–500, 2009.

[4] N. Lee and A. Cichocki. Estimating a few extreme singular values and vectors for large-scale matrices in tensor train format. *SIAM Journal on Matrix Analysis and Applications*, 36(3):994–1014, 2015.

[5] N. Lee and A. Cichocki. Fundamental tensor operations for large-scale data analysis using tensor network formats. *Multidimensional Systems and Signal Processing*, March 2017. Published online.

[6] R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research*, 11:2287–2322, 2010.

[7] I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.