Neurocomputing 223 (2017) 107–117

Contents lists available at ScienceDirect

Neurocomputing

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

Robust tensor decomposition based on Cauchy distribution and its applications

Yuankai Wu[a], Huachun Tan[a,*], Yong Li[b], Feng Li[c], Hongwen He[a]

[a] School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, PR China
[b] School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, PR China
[c] IBM Research China, Haidian District, Beijing, PR China

ARTICLE INFO

Communicated by Su-Jing Wang

Keywords:
Tensor decomposition
Cauchy distribution
Tensor completion and recovery

ABSTRACT

Tensor analysis has reached a celebrity status in the areas of machine learning, computer vision, and artificial intelligence. Completing and recovering tensor is an important problem for tensor analysis. It involves recovering a tensor from either a subset of its entries or the whole entries contaminated by noise. Classical tensor completion/recovery methods are based mainly on l_2-*norm* (Gaussian distribution noise) models, and are sensitive to noise of large magnitude. While l_1-*norm* (Laplacian distribution noise) based methods are robust to noise of large magnitude, they do not deal with dense noise effectively. In this paper, we present a novel Cauchy tensor decomposition method for simultaneously recovering and completing a low rank tensor with both missing data and complex noise. We utilize the Cauchy distribution to model noise and derive the objective function of Cauchy tensor decomposition under the maximum likelihood estimation (MLE) framework. Then we developed a robust tensor decomposition framework that used first-order optimization approaches to optimize the objective function. Extensive numerical experiments are conducted and show that our method is able to successfully recover and complete tensors with large/dense noise and missing data. We further demonstrate the usefulness of Cauchy tensor decomposition on three real-world applications, image inpainting, traffic data process and foreground/background separation. The experimental results show that the method is applicable to a wide range of problems.

1. Introduction

Tensors, which are the higher-order generalization of vectors and matrices, provide a useful representation of the existing real world data that has the natural multi-dimensional structure [1,2]. For instance, a video with multi-channels can be represented by a higher order tensor of dimensionality $time \times pixels \times pixels \times channel$; the traffic data can be grouped into a tensor of the form $interval \times location \times day \times week$. Tensor decomposition enables us to analyze the multi-mode structure by capturing the underlying low-dimensional structure of the tensor data. Tensor decomposition has been an active area during past decades (e g., [3]), and has been applied in various fields including face recognition [4–6], image compression [7], foreground segmentation [8], and missing traffic data imputation [9]. Among typical decomposition methods are CANDECOMP/PARAFAC (CP) decomposition and Tucker decomposition.

Many high-dimensional systems are suffering from missing data and/or irregular noise. Therefore a key problem in tensor analysis tasks is to find a suitable decomposition of the observed tensor when missing data and noise [10] exist, and hence tensor completion/recovery became a hot topic.

The problem of missing data comes up in many scientific areas, a great deal of effort has been made to develop tensor completion algorithms. Tensor completion methods can be roughly grouped into two classes: decomposition based methods and completion based methods. The decomposition based methods approximate a tensor with an estimated low rank or low-n-rank tensor when only part entries of a tensor are observed. In [11], the CP decomposition with missing data is formulated as a weighted least square problem and is applied to missing EEG (Electroencephalography) data imputation and network traffic data modeling. In light of [11], Tan et al. [9,12] developed Tucker decomposition based methods to impute missing traffic data. And a multi-tensor completion method based on Tucker decomposition has been proposed in [13]. In order to analyze tensors with missing data under statistical models, several Bayesian scheme based decomposition methods are also investigated and developed [14,15]. Decomposition methods successfully complete incomplete tensors contaminated by small and sparse noise, but may produce unreliable

* Corresponding author.
E-mail address: Tanhc@bit.edu.cn (H. Tan).

http://dx.doi.org/10.1016/j.neucom.2016.10.030
Received 5 January 2016; Received in revised form 2 September 2016; Accepted 24 October 2016
Available online 01 November 2016
0925-2312/

results in the presence of large and dense irregular noise. In another line of research, the completion scheme utilizes the tensor trace norm – the convex relaxation of a tensor's rank. Liu et al. [16] firstly defined the tensor trace norm as the weighted sum of traces norms of mode matrices. They converted tensor completion problem into a convex optimization problem and applied it to visual data inpainting. Followed in their works, a series of modified and similar methods have been proposed [17–20]. In a nutshell, completion based methods, which are an extended form of matrix completion [21], provide a reliable estimation of missing data with a convex optimization framework. However, completion-based methods cannot explicitly capture the underlying structure and factors of tensors, and the nuisance parameters of each mode are difficult to determine. This paper aims to develop a method that can not only recover multi-dimension data from missing data and complex irregular noise, but capture the underlying structure of tensor. So decomposition scheme is selected instead of completion scheme in this paper.

A number of works have been put forward to tensor recovery under irregular noise. The challenge of recovering a corrupted tensor lies in the understanding and modeling of irregular noise. From a statistical viewpoint, the convention methods including some tensor completion approaches that utilizing tensor l_2-*norm* can be roughly viewed as a maximum-likelihood estimation problem under the Gauss noise distribution. The conventional l_2-*norm*-based tensor decomposition methods are sensitive to gross outliers, because the Gauss distribution can sometimes amplify the negative effects of irregular noise with large magnitude. As an alternative, tensor recovery based on l_1-*norm* has been proposed [22–24]. The minimization problem utilizing the l_1-*norm* can be regarded as a maximum-likelihood estimation one under the Laplacian noise distribution, which is robust to gross outliers [25]. Li et al. [22] separated the observed tensor into a low dimensional structure with low-rank constraint and an additive (sparse) irregular pattern with a small l_0-*norm* value. Because the l_0-*norm* minimization problem is very difficult to solve [26], the objective function is then relaxed as a l_1-*norm* minimization problem. The Augmented Lagrange Multiplier optimization method for tensor recovery used l_1-*norm* is studied in [23]. Goldfarb and Qin [24] proposed a modified model and addressed the problem by nonconvex optimization. The experimental results show that the l_1-*norm* based methods can automatically exploit the low-rank structure of the tensor data from sparse noise.

In many real data corruption applications, there exist both mixing noise patterns [27] and missing data. In particular, some noise patterns are dense and of large magnitude. In general, l_2-*norm* (Gauss distribution) based approaches are only suitable for small noise, while l_1-*norm* (Laplacian distribution) based approaches are only suitable for sparse noise. The drawbacks of l_1 norm and l_2 norm based tensor recovery methods have caught many researches' attentions, and numerous methods to handle the complex noise have been proposed [28–30]. However, these approaches cannot deal effectively with missing data completion problems.

In addition to Gauss distribution and Laplacian distribution, Cauchy distribution based filter can handle impulsively sparse noise more accurately than Gaussian and Laplace models for signal filter [31,32]. It can also easily control both the sparse large noise and the dense small noise [33]. Because of the strong ability to deal with noise with mixing pattern, the Cauchy distribution has been applied to one-way Compress Sensing [34] and two-way principal component analysis (PCA) [35]. In [34], the authors present Cauchy-derived reconstruction algorithms addressing the reconstruction for signals in heavy-tailed impulsive environments. The experimental results show that the Cauchy representation of noise offers a robust framework for CS. Xie and Xing [35] propose a Cauchy principle component analysis. They utilize Cauchy distribution to model dense and large noise, and derive Cauchy PCA under the maximum likelihood estimation (MLE) framework with low rank constraint. Except one-way and two-way data, the Cauchy distribution is applied to only a few multi-way tensor cases.

In this article, we propose novel robust CP and Tucker decomposition algorithms that can impute the missing data and recover the tensor from mixing irregular pattern via Cauchy representation of noise. Firstly, we use Cauchy distribution to model noise and formulate the optimization problem under a maximum likelihood estimation framework with tensor decomposition. Then, the problem is optimized by nonlinear conjugate gradient and limited memory BFGS methods. The contributions of this paper are in two aspects, that is, Cauchy distribution, which can efficiently handle mixed noise pattern, was introduced to tensorial data analysis, and a novel framework for tensor completion and recovery that can simultaneously address missing data and noise was proposed.

The rest of this paper is organized as follows. Section 2 presents preliminaries. In Section 3, we discuss the detailed process of proposed method. Section 4 reports experimental results of our method on simulated data and several applications. Finally, Section 5 provides some concluding remarks.

2. Preliminaries

2.1. Tensor basics

Multiway arrays, also referred to as tensors, are higher-order generalizations of vectors and matrices. A brief overview of tensor decomposition and its application can be found in [3]. Higher-order arrays are represented as $\mathcal{X} \in R^{I_1 \times I_2 \times \ldots I_N}$, where the order of $\mathcal{X}$ is N. Each dimension of a multiway array is called a mode. The mode-n unfolding (also called matricization or flattening) of a tensor $\mathcal{X} \in R^{I_1 \times I_2 \times I_N}$ is defined as $\mathrm{unfolding}(\mathcal{X}, n) = X_{(n)}$, where the tensor element $(i_1, i_2, \ldots, i_n)$ is mapped to the matrix element (i_n, j), where

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

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

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

$$(\mathcal{A}, \mathcal{B}) = \sum_{i_1} \sum_{i_2} \cdots\cdots \sum_{i_n} a_{i_1 \ldots i_k \ldots i_n} b_{i_1 \ldots i_k \ldots i_n}. \tag{2}$$

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

$$\mathcal{Y} = \mathcal{A} \times_n M \Longleftrightarrow Y_{(n)} = MA_{(n)}. \tag{3}$$

Let $\mathcal{A}^{\mathcal{W}}$ be the $I_1 \times I_2 \times \ldots I_N$ observed tensor that stores all the observed values, such that

$$\mathcal{A}^{\mathcal{W}} = \begin{cases} a_{i_1 \ldots i_k \ldots i_n} & \text{if } (i_1 \ldots i_k \ldots i_n) \in \mathcal{W} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The CP decomposition decomposes a tensor into a sum of component rank-one tensors. The CP decomposition of N-dimensional tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \ldots I_N}$ can be concisely expressed as

$$\mathcal{X} = \sum_{r=1}^{R} a_{(1)}^r \circ a_{(2)}^r \cdots \circ a_{(N)}^r = [A_{(1)}, A_{(2)}, \ldots, A_{(N)}], \tag{5}$$

where R is the rank of tensor. $\bigcirc$ denotes the vector outer product. $A_{(i)} = [a_{(i)}^1, a_{(i)}^2, \ldots, a_{(i)}^R]$ denotes the factor matrix of the i-th mode. Fig. 1 gives a CP decomposition of 3-way tensor.

The Tucker decomposition naturally generalizes the orthonormal subspaces corresponding to the left/right singular matrix computed by the matrix SVD [36]. The n-dimensional tensor $\mathcal{X} \in R^{I_1 \times I_2 \times \ldots I_N}$ can be