Neurocomputing 152 (2015) 261–273

Contents lists available at ScienceDirect

Neurocomputing

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

Folded-concave penalization approaches to tensor completion ☆

Wenfei Cao [a], Yao Wang [a,*], Can Yang [b], Xiangyu Chang [a], Zhi Han [c], Zongben Xu [a]

[a] School of Mathematics and Statistics and Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China
[b] Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong
[c] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

ARTICLE INFO

Article history:
Received 6 April 2014
Received in revised form
24 September 2014
Accepted 31 October 2014
Communicated by Su-Jing Wang
Available online 11 November 2014

Keywords:
Tensor completion
Nuclear norm
Folded-concave penalization
Local linear approximation
Sparse learning

ABSTRACT

The existing studies involving matrix or tensor completion problems are commonly under the nuclear norm penalization framework due to the computational efficiency of the resulting convex optimization problem. Folded-concave penalization methods have demonstrated surprising developments in sparse learning problems due to their nice practical and theoretical properties. To share the same light of folded-concave penalization methods, we propose a new tensor completion model via folded-concave penalty for estimating missing values in tensor data. Two typical folded-concave penalties, the minmax concave plus (MCP) penalty and the smoothly clipped absolute deviation (SCAD) penalty, are employed in the new model. To solve the resulting nonconvex optimization problem, we develop a local linear approximation augmented Lagrange multiplier (LLA-ALM) algorithm which combines a two-step LLA strategy to search a local optimum of the proposed model efficiently. Finally, we provide numerical experiments with phase transitions, synthetic data sets, real image and video data sets to exhibit the superiority of the proposed model over the nuclear norm penalization method in terms of the accuracy and robustness.

1. Introduction

It is well-known that *tensors* are the higher-order generalization of vectors and matrices. There are many applications of tensors in physics, imaging, and information sciences. Especially, the problem of estimating missing values in tensor visual data arises in a number of computer vision and graphics applications, such as image inpainting [1], video inpainting [2], scan completion [3], etc. All these applications can be formulated as the so-called *tensor completion problem*. The core of tensor completion problem lies in how to build up an relationship between the known and unknown elements.

Two kinds of relationships, the local relationship and the global relationship, lead to two typical types of approaches for solving the tensor completion problem. The local relationship approaches assume that the missing entries depend on their neighborhoods and locally estimate the unknown values on basis of some difference measure between the adjacent entries. In contrast, the global relationship approaches fill in the missing entries by using global information, which is the strategy we will adopt in this paper.

Following [4,5], we first fix our notations and define the terminology of tensors used in this paper. The upper case letters are for matrices, e.g., X, and the lower case letters are for the entries, e.g., $x_{i,j}$. $\sigma_i(X)$ denotes the i-th largest singular value and $\sigma(X)$ denotes the singular vector of matrix X. The Frobenius norm of the matrix X is defined by $\|X\|_F=(\Sigma_{i,j}|x_{i,j}|^2)^{1/2}$. And the nuclear norm is defined by $\|X\|_*=\Sigma_i\sigma_i(X)$. We denote the inner product of the matrix space as $\langle X,Y\rangle=\Sigma_{i,j}X_{i,j}Y_{i,j}$. An N-order tensor to be recovered is defined by $\mathcal{X}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, and its elements are denoted by $x_{i_1,\ldots,i_N}$, where $1\le i_k\le I_k$, $1\le k\le N$; and an observed N-order tensor is defined by $\mathcal{T}$. The "unfold" operation along the k-th mode on a tensor $\mathcal{X}$ is defined by $\mathrm{unfold}_k(\mathcal{X}):=\mathcal{X}_{(k)}\in\mathbb{R}^{I_k\times(I_1,\ldots,I_{k-1}I_{k+1}\cdots I_N)}$, and the opposite operation "fold" is defined by $\mathrm{fold}_k(\mathcal{X}_{(k)})=:\mathcal{X}$. We also denote $\|\mathcal{X}\|_F=(\sum_{i_1,\ldots,i_N}|x_{i_1,\ldots,i_N}|^2)^{1/2}$ as the Frobenius norm of a tensor $\mathcal{X}$. Denote r_i as the rank of $\mathcal{X}_{(i)}$. For more details of tensor, see an elegant review [5].

Tensor completion via the global relationship approach assumes that the tensor $\mathcal{X}$ is *sparse* in the sense that each unfolding matrix $\mathcal{X}_{(k)}$ is low rank. Mathematically, tensor completion can be formulated as the following optimization problem:

$$\min_{\mathcal{X}} \quad \sum_{i=1}^{N}\alpha_i Rank(\mathcal{X}_{(i)})$$

☆This work was supported in part by National 973 Project of China under Grant number 2013CB329404 and Natural Science Foundation of China under Grant numbers 61273020 and 61303168.

* Corresponding author.
E-mail addresses: caowenf2006@163.com (W. Cao), yao.s.wang@gmail.com (Y. Wang), eeyangc@gmail.com (C. Yang), xianyuchang@gmail.com (X. Chang), han.zhi.hunt@gmail.com (Z. Han), zbxu@mail.xjtu.edu.cn (Z. Xu).

http://dx.doi.org/10.1016/j.neucom.2014.10.069
0925-2312/

$$\text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega, \tag{1.1}$$

where $\mathcal{X}$ and $\mathcal{T}$ are N-order tensors, $\mathcal{X}_{(i)}$ is a mode-i unfolding matrix, α_i's are constants satisfying $\alpha_i \geq 0$ and $\sum_{i=1}^{N}\alpha_i = 1$, and the elements of $\mathcal{X}$ and $\mathcal{T}$ in the set Ω are observed while the remainings are missing. The missing elements of $\mathcal{X}$ will be completed by making the rank of each mode-i unfolding matrix as small as possible. It is not hard to see that the tensor completion problem is a natural generalization of the well-studied *matrix completion problem* [6,7]. Because such optimization problem is a difficult nonconvex problem due to the combination nature of the rank function, Liu et al. [4] applied the matrix nuclear norm $\|\cdot\|_*$ to approximate the rank of unfolding matrices. The advantage of such replacement is that the nuclear norm is the tightest convex envelop of the rank of matrices. This leads to the following convex optimization problem for tensor completion:

$$\min_{\mathcal{X}} \quad \sum_{i=1}^{N}\alpha_i\|\mathcal{X}_{(i)}\|_* \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{1.2}$$

Liu et al. [4] also defined the tensor nuclear norm by $\|\mathcal{X}\|_* := \sum_{i=1}^{N}\alpha_i\|\mathcal{X}_{(i)}\|_*$. It is obvious that the tensor nuclear norm is a convex combination of the nuclear norms of all matrices unfolded along each mode. With this new notion, (1.2) then can be rewritten as

$$\min_{\mathcal{X}} \quad \|\mathcal{X}\|_* \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{1.3}$$

Following Liu's seminal work [4], an increasingly growing interest concentrates on the tensor completion study. Overall, the new published literature can be coarsely classified into two aspects, i.e., the algorithm design and the model innovation. From the algorithm design point of view, many new algorithms have been developed for the tensor completion model (1.3) and its variants. To name a few, Gandy et al. [8] developed two typical algorithms based on the Douglas–Rachford splitting technique and its dual variant, the alternating direction method of multipliers (ADMM). Tomioka et al. [9] discussed the different variants for the tensor nuclear norm based on the matrix nuclear norm and designed the corresponding algorithms based on ADMM. Yang et al. [10] employed the variable splitting technique and the augmented Lagrangian method to design a new algorithm and gave a rigid convergence analysis. In [11], they further extended fixed-point continuation method from matrix completion to tensor completion. Xu et al. [12] extended LMaFit for matrix completion to tensor completion. Kressner et al. [13] proposed a new algorithm that performs Riemannian optimization techniques on the manifold of tensors of fixed multi-linear rank. Signoretto et al. [14] studied a learning framework with tensors and developed a hard completion algorithm for the tensor completion problem. Krishnamurthy and Singh [15] developed an efficient algorithm for tensor completion using the adaptive sampling technique. Although many excellent algorithms currently have been designed for the tensor completion model (1.3) and its variants, this model and its variants claim many drawbacks. Romera-Paredes and Pontil [16] proposed a new tighter convex relaxation for tensor completion. Yuan and Zhang [17] in the recent work proposed a new formulation of tensor nuclear norm based on the duality of spectral norm and made an elegant theoretical analysis. Zhang et al. [18] proposed a new definition of tensor nuclear norm based on t-SVD decomposition. Based on Kronecker tensor decomposition (KTD), Phan et al. [19] proposed a novel recovery model for tensor completion. Rauhut et al. [20] adopted an efficient algorithm based on truncated hard thresholding in hierarchical tensor representations for tensor completion and provided partial convergence results. Considering the fact that matrix nuclear norm is prone to over-penalize those large singular values, and thus usually leads to a biased estimation, we will utilize the folded-concave penalization techniques to partially overcome this shortcoming. Our work will complement the current research for tensor completion from the viewpoint of model.

More precisely, note that the rank function of a matrix can be related to the l_0-norm of a vector and the nuclear norm of a matrix can be related to the l_1-norm. It is commonly known that, the l_0-norm is the essential measure of sparsity, and the l_1-norm can only be seen as its best convex relaxation form. Several studies [21–23] have shown that the l_1-norm (or LASSO) penalty over-penalizes large entries of vectors, and usually cannot avoid modeling bias. Moveover, the relationship between the l_1-norm and the nuclear norm implies also that the nuclear norm over-penalizes large singular values, and thus the modeling bias phenomenon also exists in low rank structure estimation with nuclear norm penalty [24]. Even though a line of works [4,8] illustrate the power of the nuclear norm penalization for the low-rank tensor completion problem, there are some specific applications that the nuclear norm penalization method cannot provide desirable results. As shown in Fig. 1, when the missing entries is relatively large, the tensor nuclear norm penalization method cannot obtain good results. A natural question is that can we find an alternative penalization approach to overcome this drawback? Fig. 1 (f) provides an affirmative answer. It is shown that the SCAD based approach could obtain a near-optimal recovery result.

Recently, nonconvex penalties have been drawn more and more attention in sparse learning problems, because people believe that one of possible solutions of nonconvex penalization problem could overcome the drawbacks of the unique solution of convex penalization problem. As a common practice, the l_1-norm can be replaced by the l_q-norm with $0 < q < 1$ if a more sparse solution is expected to be obtained [25,26]. However, no theoretical guarantee with l_q-norm is made for reducing the modeling bias of LASSO. Fortunately, the folded-concave penalization such as SCAD and MCP have been proposed and show nice nearly unbiased property through numerous numerical and theoretical studies [23,27–29]. Furthermore, [30,31] have used the MCP penalization approach in various low-rank structure learning applications, where comprehensive numerical results demonstrated the outperformance over the nuclear norm penalization approach.

In this paper, encouraged by the nice properties of folded-concave penalty, we propose a general nonconvex formulation for the tensor completion problem on basis of the folded-concave penalization. This new formulation naturally extends the folded-concave penalty from vector cases to tensor cases. We denote by $P_\lambda(\cdot)$ a general folded-concave penalty function. More details of this function will be presented in Section 2. As such, we suggest the use of the following model for tensor completion:

$$\min_{\mathcal{X}} \quad \sum_{i=1}^{N}\alpha_i\|\mathcal{X}_{(i)}\|_{P_\lambda} \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega, \tag{1.4}$$

where $\|\mathcal{X}_{(i)}\|_{P_\lambda} = \sum_{j=1}^{r_i}P_\lambda(\sigma_j(\mathcal{X}_{(i)}))$, here $\sigma_j(\mathcal{X}_{(i)})$ is the j-th nonzero singular value of mode-i unfolding matrix $\mathcal{X}_{(i)}$. Similarly, we define the tensor folded-concave norm by $\|\mathcal{X}\|_{P_\lambda} := \sum_{i=1}^{N}\alpha_i\|\mathcal{X}_{(i)}\|_{P_\lambda}$. It is clear that (1.4) is a nonconvex optimization problem, and thus there exist multiple local minimizers in general. However, based on the "bet-on-folded-concave-penalization" principle [29], the local linear approximation (LLA) algorithm can find a good estimator of (1.4) as long as there exists a reasonable initial estimator. Therefore, by combining the LLA algorithm and the augmented Lagrange multiplier (ALM) method, we can derive an efficient algorithm for computing a specific local solution of the optimization problem (1.4). The contributions of this work are summarized as

- To alleviate the modeling bias of the nuclear norm, we propose a general folded-concave penalization approach for the tensor