Neurocomputing 178 (2016) 3–10

Contents lists available at ScienceDirect

Neurocomputing

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

“Low-rank + dual” model based dimensionality reduction

Si-Qi Wang *, Xiang-Chu Feng, Wei-Wei Wang

School of Mathematics and Statistics, Xidian University, Xi'an 710126, China

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form
16 July 2015
Accepted 17 July 2015
Available online 29 November 2015

Keywords:
Dimensionality reduction
Background modeling
Singular value decomposition
Thresholding method
l_p-Minimization problem

ABSTRACT

We propose a novel “low-rank + dual” model for the matrix decomposition problems. Based on the unitarily invariant property of the Schatten p-norm, we prove that the solution of the proposed model can be obtained by an “$l_\infty + l_1$” minimization problem, thus a simple and fast algorithm can be provided to solve our new model. Furthermore, we find that applying “$l_\infty + l_1$” to any vector can achieve a shifty threshold on the values. Experiments on the simulation data, the real surveillance video database and the Yale B database prove the proposed method to outperform the state-of-the-art techniques.

1. Introduction

Dimensionality reduction is an extraordinary important tool for processing the high-dimensional data, such as video, image, audio and bio-medical signals. A well-known technique for dimension reduction is the low-rank approximation method, which is actually solving the rank minimization problems (RMP) [1–4]. And recently, RMP attract more and more attentions in many fields of science and engineering, such as computer vision [5–10], machine learning [7,8,11–13], and signal and image processing [2,5,8,9,14]. Due to the effectiveness of the rank minimization, many researchers have proposed various algorithms to solve the problem. An important classical model is reduced-rank regression, which limits the coefficient matrix to be low-rank [15–17]. A nuclear norm penalized least squares estimator has been proposed by Yuan et al. [18], and the estimator encourages the singular values sparsity and the simultaneous rank reduction [19,20]. Rohde and Tsybakov investigate the Schatten-q quasi-norm penalty and non-asymptotic bounds of prediction risk [21].

Reduced-rank is a very effective dimension reduction assumption, which connections with many popular tools including principal component analysis (PCA) and canonical correlation analysis (CCA), and it is extensively studied in matrix completion problems [5,17,22–24]. Actually, the rank function and nuclear norm penalized methods in robust PCA (RPCA) problems can be respectively viewed as l_0 and l_1 penalized methods in the singular value decomposition (SVD) domain respectively. Therefore, the general strategy is to solve the nuclear norm minimization problem (NNMP) instead of the RMP, but the solution to NNMP suffers from the high computation cost of multiple SVDs.

In this paper, a “low-rank + dual” model is formalized from the viewpoint of dimension complexity reduction, and an efficient iterative update algorithm is designed to solve the model. It is necessary to particularly mention that “dual” here refers to the duality between the norms in functional analysis [25], rather than the typical concept of dual problem in computational science. Dual norms to be used as regular items was originally proposed by Meyer in image decomposition problem [26]. A model with total variation (TV) norm and G-norm (the dual of the TV-norm) was introduced by Meyer, which can better capture the cartoon and texture part in the decomposition of image. Thus, we try to apply dual norms to the regularization terms in the background and foreground modeling problems. The novel model can limit the correlation between the components, and the unitarily invariant property of the Schatten p-norm ensures that the model can be replaced by a simplified “$l_\infty + l_1$” minimization problem.

In order to test the efficiency and effectiveness of the proposed model, we apply our method both in the simulation data and real database experiments. In the first simulation experiment, randomly generated square matrix, which is actually the sum of the low-rank component and the error interference component, will be used as the test data. We apply our proposed method and other five state-of-the-art methods on the simulation test data to recover the low-rank and the interference component. In the second simulation experiment, we test the performances of IALM, GoDec and our method in matrix completion tasks. We discard a few entries from a randomly generated low-rank matrix, and

* Corresponding author. Tel.: +86 29 88202860.
E-mail addresses: pearlwangxd@163.com (S.-Q. Wang), xcfeng@mail.xidian.edu.cn (X.-C. Feng), wwwang@mail.xidian.edu.cn (W.-W. Wang).

http://dx.doi.org/10.1016/j.neucom.2015.07.117
0925-2312/

recover the whole matrix through three different methods. Background modeling and specularity removing are two challenging task for computer vision applications, which are the fundamental of the follow-up work, such as motion tracking [27] and feature extraction [28]. Therefore, in the real database experiments, we apply our model to the background–foreground modeling and specularity removing problems, and compare the performance of the proposed method with two existing widely used methods. Experimental results demonstrate that the proposed algorithm is superior to other algorithms in most cases.

The rest of the paper is organized as follows. First, we briefly review the "low-rank + sparse" model and present our new "low-rank + dual" model in the next section. In Section 3, we simplify the problem into the vector level and propose the relevant algorithm in detail. Next in Section 4, we discuss the thresholding essence of our method, and compare the numerical results with other existing methods on some practical experiments. Finally, we give some concluding remarks in Section 5.

2. Related works

2.1. "Low-rank + sparse" model

Suppose $Y \in \mathbb{R}^{m\times n}$ is a large matrix whose columns are arranged by the data set of the observational images, the size of each image is $m_1 \times n_1 = m$ and the number of images is n. Then the low-rank matrix approximation problem is to efficiently and accurately obtain the low-rank matrix X from the measurement Y which has been corrupted by the errors E, $X, E \in \mathbb{R}^{m\times n}$. The common "low-rank + sparse" model formulates as follows:

$$\begin{cases} \min_{X,E} & \lambda\, Rank(X) + \|E\|_0 \\ \text{s.t.} & Y = X + E, \end{cases} \tag{1}$$

where $Rank(\cdot)$ represents the rank function of the desired matrix; $\|\cdot\|_0$ is the regularization term for promoting sparsity; and λ is a positive weighting parameter providing a trade-off between the sparse and low-rank components. Model (1) suggests to seek the lowest-rank matrix X subject to the sparse constraint E, we hope to exactly obtain the pair (X,E) that could have generated the data matrix Y. Nevertheless, due to the discrete of the rank function and the highly non-convex of the l_0-norm, (1) is NP-hard and no efficient solution is known to its sub-problems.

A tractable approach for solving (1) is to relax the model into the following convex optimization problem:

$$\begin{cases} \min_{X,E} & \lambda \|X\|_* + \|E\|_1 \\ \text{s.t.} & Y = X + E, \end{cases} \tag{2}$$

where $\|\cdot\|_*$ is the nuclear norm defined by the sum of all singular values, which is the convex hull of the matrix rank; and $\|\cdot\|_1$ is the l_1-norm defined by the component-wise sum of absolute values of all entries, which is the relaxation of the l_0-norm. Moreover, recent advances have proposed a variety of different methods to solve (2): the iterative thresholding approach [14], the accelerated proximal gradient approach [29], the exact and inexact augmented Lagrange multiplier approaches [3], etc. However, almost all the available algorithms of the "low-rank + sparse" model are plagued by the complex calculation accompanied with SVD.

2.2. "Low-rank + dual" model

In this paper, we propose a novel "low-rank + dual" model:

$$\begin{cases} \min_{X,E} & \lambda \|X\|_{\sigma,1} + \|E\|_{\sigma,\infty} \\ \text{s.t.} & Y = X + E. \end{cases} \tag{3}$$

Here $\|\cdot\|_{\sigma,p}$ is the Schatten p-norm, which is actually defined as the l_p-norm of the singular value vector.[1] To illustrate the advantages of our model, we need to recall the definition of Schatten p-norm first.

Definition 1. If σ_i denote the singular values of the matrix $A \in \mathbb{R}^{m\times n}$ $(m \ge n)$, then the Schatten p-norm can be defined as $\|A\|_{\sigma,p} = (\sum_{i=1}^n \sigma_i^p)^{1/p}$. All Schatten norms are sub-multiplicative and unitarily invariant, which means that $\|A\| = \|UAV^T\|$ holds for any matrix A and unitary matrices U, V.

From the definition, we can know that the most familiar case $p=1$ of the Schatten norm is the sum of all singular values, which yields the nuclear norm and associates with low-rank. Moreover, $\|\cdot\|_{\sigma,1}$ and $\|\cdot\|_{\sigma,\infty}$ in (3) are dual to each other, then due to the property of dual norms, we should have $|\langle X,E\rangle| \le \|X\|_{\sigma,1}\|E\|_{\sigma,\infty}$ holds for the matrix X and E. Thus, minimizing the sum of $\|X\|_{\sigma,1}$ and $\|E\|_{\sigma,\infty}$ should achieve a smaller inner product $|\langle X,E\rangle|$, which leads to the less correlated decomposition X, E of the data Y.

The varichange of the regularization term is inspired by duality theory applied to the image decomposition problems [26,30]. Through such improvement, it has no longer need to limit the non-low-rank part of the matrix to be "sparse", dual regularization terms in the model can directly lead to the decomposition of the data matrix with two less correlated components. More importantly, the particular norm constraint enables the novel "low-rank + dual" model to be solved in the vector field. A simple and fast algorithm of this model will be demonstrated both in the theoretical and experimental fields. Numerical experiment results show that our "low-rank + dual" model performs well in the matrix decomposition problems, and the time consumption of the corresponding algorithm is much less than other "low-rank + sparse" methods. The theoretical basis and the relevant algorithm will be provided in the following section.

3. Problem simplification and theoretical derivation

Our approach minimizes the model (3) as follows, we substitute the equation constrain into the minimization function to get the following form:

$$\min_X \|Y - X\|_{\sigma,\infty} + \lambda \|X\|_{\sigma,1}, \tag{4}$$

$Y \in \mathbb{R}^{m\times n}, m \ge n$. We confirm (4) as our original problem in this paper, and all the matrices involved in the minimization problem now are not square matrix. In the following sections, we will show the simplification of the original problem in details.

3.1. Simplified to vector level

Let $Y = L\Sigma R^T$ be the singular value decomposition of the data matrix Y, where L, R are unitary matrix of size $m\times m$, $n\times n$ and satisfy $L^TL = LL^T = I, R^TR = RR^T = I$. Then

$$\|Y - X\|_{\sigma,\infty} + \lambda\|X\|_{\sigma,1} = \left\|L(\Sigma - L^TXR)R^T\right\|_{\sigma,\infty} + \lambda\left\|LL^TXRR^T\right\|_{\sigma,1}. \tag{5}$$

Due to the unitarily invariant of the Schatten norms, we have

$$\left\|L(\Sigma - L^TXR)R^T\right\|_{\sigma,\infty} + \lambda\left\|LL^TXRR^T\right\|_{\sigma,1} = \left\|\Sigma - L^TXR\right\|_{\sigma,\infty} + \lambda\left\|L^TXR\right\|_{\sigma,1}. \tag{6}$$

[1] Notice that this is not an equivalence conception for the singular value vector and the singular vectors – the former, a unique vector, means the vector be composed of the elements of the matrix singular values, while the latter refers to the columns of the unitary matrices in the singular value decomposition.