Neurocomputing 167 (2015) 218–229

Contents lists available at ScienceDirect

Neurocomputing

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

Robust orthogonal matrix factorization for efficient subspace learning

Eunwoo Kim, Songhwai Oh *

Department of Electrical and Computer Engineering, ASRI, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-744, South Korea

ARTICLE INFO

Article history:
Received 19 January 2015
Received in revised form
15 March 2015
Accepted 28 April 2015
Communicated by X. Gao
Available online 8 May 2015

Keywords:
Low-rank matrix factorization
l_1-norm
Subspace learning
Augmented Lagrangian method
Rank estimation

ABSTRACT

Low-rank matrix factorization plays an important role in the areas of pattern recognition, computer vision, and machine learning. Recently, a new family of methods, such as l_1-norm minimization and robust PCA, has been proposed for low-rank subspace analysis problems and has shown to be robust against outliers and missing data. But these methods suffer from heavy computation loads and can fail to find a solution when highly corrupted data are presented. In this paper, a robust orthogonal matrix approximation method using fixed-rank factorization is proposed. The proposed method finds a robust solution efficiently using orthogonality and smoothness constraints. The proposed method is also extended to handle the rank uncertainty issue by a rank estimation strategy for practical real-world problems. The proposed method is applied to a number of low-rank matrix approximation problems and experimental results show that the proposed method is highly accurate, fast, and efficient compared to the existing methods.

1. Introduction

Low-rank matrix approximation has attracted much attention in the areas of data reconstruction [1], image denoising [2–4], collaborative filtering [5–7], background modeling [8,9], structure from motion or motion estimation [1,10–12,3], and photometric stereo [13,12], to name a few. It is usually assumed that the rank of a matrix is fixed or known beforehand.

Although real-world data is usually high dimensional, it can be well-represented with fewer parameters in many cases. Hence, reducing the data dimension to a small number of dominating principal components is desirable to reduce the computation time and also to remove unwanted noisy components. A popular method for addressing this issue is principal component analysis (PCA) [14]. PCA transforms data to a low-dimensional subspace which maximizes the variance of the data based on the l_2-norm. To handle the missing data, Srebro [5] solved a weighted low-rank matrix approximation problem using the expectation–maximization (EM) algorithm. In addition, there is a common technique which adds regularization terms to prevent data overfitting and solves the problem using semidefinite programming (SDP) [13]. These conventional l_2-norm based approximation methods have been utilized in many problems but it is known that they are sensitive to outliers and missing data because the l_2-norm amplifies the negative effects of corrupted data.

As an alternative, low-rank matrix approximation methods using the l_1-norm have been proposed for robustness against outliers [1–3,12,15–19] and extended to low-rank tensor approximation problems [20–22]. These techniques assume a Laplacian noise model instead of a Gaussian noise model. In addition, there have been several probabilistic extensions of low-rank matrix factorization for robust approximation [23,24].

Ke and Kanade [1] presented low-rank matrix approximation methods by alternatively minimizing an l_1-norm based cost function using convex programming. Eriksson and Hengel [3] proposed a weighted low-rank matrix approximation method using the l_1-norm in the presence of missing data. However, these methods require a heavy computational load since a large number of iterations is required to find a good solution using convex programming for optimizing a nonconvex and nonsmooth l_1 cost function [12]. Kwak [15] proposed PCA-l_1 to find successive principal components in the feature space based on the l_1-norm. However, it results in degradation of robustness against corruptions since it uses a modified cost function which is different from the original l_1 cost function given in [1,3].

Robust principal component analysis (RPCA) has been proposed to solve a low-rank and sparse matrix decomposition problem and successfully applied to a number of problems [16,8,6,17,25]. RPCA uses recent advances in rank minimization using the nuclear norm with an l_1-norm regularized term for a non-fixed rank matrix approximation problem. They have shown that the RPCA methods are suitable for problems, such as background modeling,

* Corresponding author. Tel.: +82 2 880 1511.
E-mail addresses: eunwoo.kim@cpslab.snu.ac.kr (E. Kim), songhwai.oh@cpslab.snu.ac.kr (S. Oh).

http://dx.doi.org/10.1016/j.neucom.2015.04.074
0925-2312/

corruption removal, and collaborative filtering [8,6,17]. However, they have high computational complexity, especially for large-scale problems, because it performs singular value decomposition (SVD) at each iteration. To overcome this complexity issue, Liu et al. [18] proposed a fast algorithm based on the tri-factorization approach with an orthogonality constraint for low-rank matrix recovery and completion problems. Recently, Shu et al. [26] proposed efficient low-rank recovery methods based on a new rank measure and showed their efficiency compared to conventional RPCA methods.

Recently, many efficient matrix factorization algorithms using the l_1-norm under the augmented Lagrangian framework have been proposed [2,12,27,19]. Shen et al. [2] proposed a low-rank matrix approximation method using the l_1-norm based on the augmented Lagrangian alternating direction method (ALADM). Zheng et al. [12] proposed a practical weighted low-rank approximation method with an orthogonality constraint using ALM (Regl_1-ALM). Liu and Yan [28] presented an active subspace method for a scalable nuclear-norm regularized optimization problem. Cabral et al. [27] proposed an approach, which unifies bilinear factorization and nuclear-norm minimization by utilizing an alternative definition of the nuclear-norm. Kim et al. [19] proposed two l_1-norm based alternating rectified gradient methods to obtain efficient l_1-norm based solutions with a convergence guarantee. These l_1-norm based methods have been successfully applied to fixed-rank factorization problems in the presence of missing data and outliers, outperforming RPCA methods.

In this paper, we present a new robust orthogonal matrix approximation method using fixed-rank factorization based on the l_1-norm for low-rank subspace learning problems in the presence of various corruptions. We introduce an efficient Frobenius-norm regularizer to prevent the overfitting problem which can arise from an alternative minimization algorithm and an orthogonality constraint to reduce the solution space for faster convergence. The proposed regularized optimization problem is constructed under the augmented Lagrangian framework and solved using an alternating direction approach which optimizes the cost function with respect to one variable at a time while fixing the other variables. We also present a rank estimation strategy for the propose method without increasing the computational complexity to overcome the disadvantage of fixed-rank factorization and the parameterization issue when the exact rank of a problem is unknown. We demonstrate that the performance of the proposed method in terms of the reconstruction error and computational speed in the presence of corruptions using well-known benchmark datasets from non-rigid motion estimation, background modeling, and collaborative filtering.

This paper is organized as follows. In Section 2, we briefly review the l_1-norm based low-rank matrix approximation and introduce approaches based on different regularization methods and describe the difference between fixed-rank matrix factorization and RPCA methods. The proposed algorithm and its extension to rank estimation based method are described in Section 3. In Section 4, we present various experimental results to evaluate the proposed method with respect to other well-known low-rank matrix factorization methods and RPCA methods.

2. Preliminaries

2.1. Robust low-rank matrix factorization

We briefly describe a fixed-rank matrix factorization problem based on the l_1-norm and discuss its related work. The problem arises in a number of problems in computer vision, pattern recognition, and machine learning to handle missing data and outliers and obtain robust and exact solutions, such as rigid and non-rigid motion estimation [29,11], collaborative filtering (CF) [5–7], and background modeling [8,9,27], to name a few. A minimization problem based on the l_1-norm can be regarded as a maximum likelihood estimation problem under the Laplacian noise distribution [1,19].

We first consider an approximation problem for vector $\mathbf{y} = [y_1\ y_2\ \dots\ y_m]^T$ by a multiplication of vector $\boldsymbol{x} \in \mathbb{R}^m$ and scalar α, i.e.,

$$\mathbf{y} = \alpha \boldsymbol{x} + \boldsymbol{\delta}, \tag{1}$$

where $\boldsymbol{\delta}$ is a noise vector whose elements have the independently and identically distributed Laplacian distribution [19]. The probability model for (1) can be written as

$$p(\mathbf{y}|\boldsymbol{x}) \sim \exp\left(-\frac{\|\mathbf{y}-\alpha\boldsymbol{x}\|_1}{s}\right), \tag{2}$$

where $\|\cdot\|_1$ denotes the l_1-norm, and $s > 0$ is a scaling constant [1]. Maximizing the log likelihood of the observed data is equivalent to minimizing the following cost function for given $\boldsymbol{x}$:

$$J(\alpha) = \|\mathbf{y}-\alpha\boldsymbol{x}\|_1. \tag{3}$$

The problem (1) can be generalized to the problem of matrix approximation. Let us consider the l_1 approximation of matrix Y such that

$$\min_{P,X} J(P,X) = \|Y - PX\|_1, \tag{4}$$

where $Y \in \mathbb{R}^{m\times n}$, $P \in \mathbb{R}^{m\times r}$, and $X \in \mathbb{R}^{r\times n}$ are the observation, projection, and coefficient matrices, respectively. Here, r is a predefined parameter less than $\min(m,n)$ and PX is a low-rank approximation of Y. In addition, since it is difficult to obtain observations for all entries of the observation matrix in practice, this problem can be considered as the following weighted low-rank matrix approximation problem to consider unknown entries:

$$\min_{P,X} \|W \odot (Y - PX)\|_1, \tag{5}$$

where W is a weight or mask matrix, whose element w_{ij} is 1 if y_{ij} is known and 0 if y_{ij} is unknown, and $\odot$ is the component-wise multiplication or the Hadamard product.

Despite the robustness against outliers, the discussed l_1-norm based methods require a heavy computational load for finding a solution using linear or quadratic programming [1], which requires a large number of iterations to obtain a reasonable solution, making them applicable only for small-scale problems. To overcome the computational complexity issue, methods based on an augmented Lagrangian method (ALM) have been proposed [2,27] and it solves the problem using an alternating minimization technique, which minimizes the cost function with respect to one target variable while other variables are held fixed. In addition, a nuclear-norm regularized l_1-norm minimization method (Regl_1-ALM) has been proposed to improve convergence by introducing an implicit rank constraint into the cost function via the bilinear form of PX [12,28]. Unlike the ALM based methods, Kim et al. [19] proposed two l_1-norm based alternating rectified gradient methods to achieve robustness. They rectified the gradient using QR factorization for quickness and found a step size using the modified weighted median approach [19]. However, it is difficult for a matrix factorization method to find the global optimal solution because the considered problem is non-convex. Furthermore, when the rank of the data matrix is unknown, the problem becomes more challenging.

2.2. Robust principal component analysis (RPCA)

Low-rank matrix approximation finds a low-rank matrix representation of an observation or data matrix, such that the difference between the estimated low-rank matrix and the observation matrix is