Neural Networks 100 (2018) 39–48

Contents lists available at ScienceDirect

Neural Networks

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

Accelerated low-rank representation for subspace clustering and semi-supervised classification on large-scale data

Jicong Fan [a,*], Zhaoyang Tian [a], Mingbo Zhao [b], Tommy W.S. Chow [a]

[a] *Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region*
[b] *School of Information Science and Technology, Donghua University, Shanghai, PR China*

A R T I C L E   I N F O

Article history:
Received 9 August 2017
Received in revised form 19 January 2018
Accepted 26 January 2018
Available online 2 February 2018

Keywords:
Subspace clustering
Low-rank representation
Semi-supervised classification
Large-scale data
Nuclear norm
Matrix factorization

A B S T R A C T

The scalability of low-rank representation (LRR) to large-scale data is still a major research issue, because it is extremely time-consuming to solve singular value decomposition (SVD) in each optimization iteration especially for large matrices. Several methods were proposed to speed up LRR, but they are still computationally heavy, and the overall representation results were also found degenerated. In this paper, a novel method, called accelerated LRR (ALRR) is proposed for large-scale data. The proposed accelerated method integrates matrix factorization with nuclear-norm minimization to find a low-rank representation. In our proposed method, the large square matrix of representation coefficients is transformed into a significantly smaller square matrix, on which SVD can be efficiently implemented. The size of the transformed matrix is not related to the number of data points and the optimization of ALRR is linear with the number of data points. The proposed ALRR is convex, accurate, robust, and efficient for large-scale data. In this paper, ALRR is compared with state-of-the-art in subspace clustering and semi-supervised classification on real image datasets. The obtained results verify the effectiveness and superiority of the proposed ALRR method.

1. Introduction

Datasets from different areas such as computer vision, machine learning, and data mining, are generally organized into a matrix form. The corresponding matrices are often of low-rank (Fan & Chow, 2017b). The low-rank assumption and model provides considerable advantages for storing, processing, and analyzing. Representative low-rank techniques can be found in principle component analysis (PCA) (Cand, Li, Ma, & Wright, 2011; Jolliffe, 2002), dimensionality reduction (Fan, Chow, Zhao, & Ho, 2017), matrix completion (LRMC) (Cands & Recht, 2009; Fan & Cheng, 2018; Fan & Chow, 2017a, c; Wen, Yin, & Zhang, 2012), and low-rank representation (LRR) (Liu et al., 2013; Liu, Wang, Han, Fan, & Luo, 2017; Liu, Xu, Tang, Liu, & Yan, 2016; Peng, Lu, & Wang, 2015; Vidal & Favaro, 2014; Zhang, Yan, & Zhao, 2014). LRR is to represent a set of data points as the multiplication of a dictionary matrix and a low-rank coefficients matrix. In LRR, nuclear-norm minimization is used for rank-minimization, and singular value thresholding (Liu et al., 2013; Lu, Zhu, Xu, Yan, & Lin, 2015) is performed iteratively. Compared with PCA, which is a single-subspace method, LRR is able to segment multiple subspaces (Tang et al., 2016) because one data point can be effectively represented by the data points from a common subspace. LRR has been applied to subspace clustering (Kriegel, Kröger, & Zimek, 2009; Nie & Huang, 2016; Parsons, Haque, & Liu, 2004; Sim, Gopalkrishnan, Zimek, & Cong, 2013; Vidal, 2011; Zhu, Zhu, Hu, Zhang, & Zuo, 2017), motion segmentation (Panagiotakis, Pelekis, Kopanakis, Ramasso, & Theodoridis, 2012), and image denoising. Another counterpart of LRR is sparse representation (Elhamifar & Vidal, 2013; Fan & Chow, 2017d; Zhu, Zhu, Wang, Zuo, & Hu, 2017; Zhu, Zuo, Zhang, Hu, & Shiu, 2015), in which the coefficients matrix is sparse. Sparse representation was also applied to many problems such as subspace clustering (SSC) (Elhamifar & Vidal, 2013). Both LRR and SSC are robust to sparse noises and outliers (Elhamifar & Vidal, 2013; Gong, 2017; Liu et al., 2013). Many extensions of LRR and SSC have recently been proposed to improve performance (Liu & Yan, 2011; Patel & Vidal, 2014; Wang, Xu, & Leng, 2013; Xiao, Tan, Xu, & Dong, 2016). For example, latent LRR (Liu & Yan, 2011) was proposed to solve the problem when the observations are insufficient and/or grossly corrupted. Low-rank sparse subspace clustering (LRSSC) (Wang et al., 2013) was proposed to obtain a simultaneously low-rank and sparse coefficients matrix for clustering. Kernel SSC (Patel, Hien Van, & Vidal, 2015; Patel & Vidal, 2014) and robust kernel LRR (Xiao, Tan et al., 2016) were proposed for nonlinear subspace clustering.

It is important to note that LRR is rather computationally expensive making it unable to cope with large-scale data, for instance more than thousands of data points (Zhang, Sun, Liu, & Ma,

* Corresponding author.
E-mail addresses: jicongfan2-c@my.cityu.edu.hk (J. Fan), zytian3-c@my.cityu.edu.hk (Z. Tian), mzhao4@dhu.edu.cn (M. Zhao), eechow@cityu.edu.hk (T.W.S. Chow).

https://doi.org/10.1016/j.neunet.2018.01.014
0893-6080/

2014). The computational problem of LRR stemmed from the basic procedures of performing singular value decomposition (SVD) in each iteration, and SVD is well-known to be time-consuming when handling large matrix. To solve this problem, truncated SVD, rather than full SVD, can be used. But truncated SVD is still computationally expensive for large matrix. Recently, in Shen and Li (2016), instead of nuclear-norm minimization, matrix factorization is proposed to solve LRR. The low-rank coefficients matrix is replaced by the multiplication of two thin matrices that are to be optimized using Frobenius-norm minimization. The method is non-convex and hence called non-convex LRR (NLRR). As NLRR replaces SVD with matrix multiplication, it is considerably efficient compared with LRR. In Xiao, Li, Xu, and Tao (2015), a method called fast LRR (FaLRR) was proposed for large-scale data. FaLRR also replaces the coefficients matrix by the multiplication of two matrices, which include a fixed matrix obtained by truncated SVD of the data matrix, and another matrix that is to be optimized by nuclear-norm minimization. FaLRR converts the nuclear-norm minimization of a large and square matrix into the nuclear-norm minimization of a large but thin matrix. As a result, the computational cost is significantly reduced. Although NLRR and FaLRR have become more computationally feasible for handling large-scale data, their final results, i.e., the effectiveness of the representation coefficients, are regressed compared to those of LRR and SSC. This is possibly due to the following reasons. First, NLRR is non-convex and hence may suffer from local minima. Second, FaLRR is unable to exactly handle entry-wise or/and column-wise sparse corruptions, which will be discussed in later session of this paper. At last, the numbers of unknown variables to be optimized for the coefficients matrices in NLRR and FaLRR are relatively large leading to large complexities of the representation models. Despite all the above-mentioned, improving the computational efficiency of low-rank representation is still an important and possible task.

In this paper, we propose a novel method called accelerated LRR (ALRR) to improve the efficiency and accuracy of low-rank representation. The main idea of ALRR is to integrate matrix factorization with nuclear-norm minimization. Specifically, ALRR replaces the representation coefficients matrix by the multiplication of three matrices, of which the leftmost one is a thin matrix, the middle one is a small square matrix, and the rightmost one is a short matrix. Nuclear-norm minimization is implemented on the middle matrix whose size is not related to the number of data points. Therefore, the computation of ALRR is substantially faster than LRR. In addition, ALRR is robust to noises and outliers. In this study, ALRR is compared with SSC, LRR, NLRR, and FaLRR for subspace clustering and semi-supervised classification (Belkin, Niyogi, & Sindhwani, 2006; Gong et al., 2015; Gong, Tao, Liu, Liu, & Yang, 2017; Gong et al., 2016; Zhu, Ghahramani, & Lafferty, 2003) on the Extended Yale face dataset B (Kuang-Chih, Ho, & Kriegman, 2005) and MNIST handwritten digits dataset (Lecun, Bottou, Bengio, & Haffner, 1998). The experimental results show that ALRR is more efficient and accurate compared to other methods.

The contributions of this paper are as follows. First, we propose a new method ALRR for low-rank representation on large-scale data. ALRR is more efficient than SSC, LRR, NLRR, and FaLRR. ALRR often provides higher accuracy than other methods do in subspace clustering and semi-supervised classification. Second, we analyze the robustness of ALRR and use ALRR to reduce noises of images, which verify that ALRR is able to handle entry-wise and column-wise sparse corruptions. Finally, we thoroughly analyze and compare the floating-point operation counts of all related methods. It is verified that the computational complexity of ALRR is the lowest among the studied methods.

The remaining content of this paper are organized as follows. Section 2 gives the related work and the corresponding discussions. Section 3 elaborates the proposed method ALRR. Section 4 consists of the case studies of image denoising, subspace clustering, and classification. Section 5 is the conclusion of this study.

2. Related work and discussion

SSC. Denoting a dataset of m features and n samples drawn from multiple subspaces by $X \in \mathbb{R}^{m\times n}$, sparse subspace clustering (SSC) (Elhamifar & Vidal, 2013) first considers the following sparse regression problem:

$$\begin{aligned}&\min_{C,E} \|C\|_1 + \lambda\|E\|_1,\\ &s.t.\ X = XC + E,\ diag(C) = 0\end{aligned} \tag{1}$$

where $C \in \mathbb{R}^{n\times n}$ is the representation coefficients matrix, $E \in \mathbb{R}^{m\times n}$ is the representation errors matrix, and $\|\cdot\|_1$ is the ℓ_1 norm of matrix to obtain entry-wise sparsity. The affinity matrix is given as $A = |C| + |C|^T$ and used for spectral clustering (Li, Xia, Shan, & Liu, 2015; Vidal, 2011) to divide the data into different groups.

LRR. Low-rank representation (LRR) (Liu et al., 2013) solves the following optimization problem:

$$\begin{aligned}&\min_{C,E} \|C\|_* + \lambda\|E\|_{21},\\ &s.t.\ X = XC + E.\end{aligned} \tag{2}$$

In (2), $\|\cdot\|_*$ is the nuclear-norm of matrix, i.e. $\|C\|_* = \sum_{i=1}^{min(m,n)} \sigma_i(C)$, where $\sigma_i(C)$ is the ith singular value of C. $\|\cdot\|_{21}$ is the ℓ_{21} norm of matrix to obtain column-wise sparsity. An affinity matrix is formed by $[A]_{ij} = ([U\Sigma^{\frac{1}{2}}]_{ij})^q$, where U and Σ are given by the skinny SVD of $C = U\Sigma V^T$ and q can be chosen from $\{2, 4, 6, 8, 10\}$ to enhance the sparsity of A. The affinity matrix A is used for spectral clustering (Li et al., 2015; Lu, Yan, & Lin, 2016; Vidal, 2011) to segment different subspaces of X. In recent years, a few extensions of LRR and SSC have been proposed (Fan & Chow, 2017b, d; Li & Vidal, 2016; Liu & Yan, 2011; Wang et al., 2013; Xiao, Tan et al., 2016). For instance, in Wang et al. (2013), it was proposed to construct a simultaneously low-rank and sparse affinity matrix for subspace clustering. In Xiao, Tan et al. (2016), a robust kernel LRR (Xiao, Tan et al., 2016) was proposed for nonlinear subspace clustering with outliers. In Fan and Chow (2017d) and Li and Vidal (2016), the problem of subspace clustering on incomplete data was studied. The computational complexity of these extensions is higher than that of LRR.

NLRR. Non-convex LRR (NLRR) (Shen & Li, 2016) replaces nuclear-norm minimization with matrix factorization (Shiga & Mamitsuka, 2015) and solves the following problem:

$$\begin{aligned}&\min_{U,V,E} \tfrac{\beta}{2}\|X - XUV^T - E\|_F^2\\ &+ \tfrac{1}{2}\|U\|_F^2 + \tfrac{1}{2}\|V\|_F^2 + \lambda\|E\|_1,\end{aligned} \tag{3}$$

where $U \in \mathbb{R}^{n\times d}$ and $V \in \mathbb{R}^{n\times d}$ are two thin matrices. The representation coefficients matrix is given by $C = UV^T$ and d is the pre-defined rank of C. Because of the non-convexity, NLRR may suffer from the problem of local minima.

FaLRR. Fast LRR (FaLRR) (Xiao, Li et al., 2015) first solves the following problem

$$\min_{W} \|W\|_* + \lambda\|S_d(V_d^T D - W)\|_{21}, \tag{4}$$

and then sets $C = V_d W$ and $E = XD - XV_d W$, where $W \in \mathbb{R}^{d\times n}$. In (4), D is a pre-defined matrix, S_d is a diagonal matrix formed by the largest d singular values of $X = USV^T$, and V_d is formed by the corresponding right singular vectors. If D is an identity matrix and $d = min(m, n)$, (4) will be the same as LRR.

However, we find that FaLRR is unable to exactly handle entry-wise or column-wise sparse corruptions. In Xiao, Li et al. (2015), it was claimed that $\|S_d(V_d^T D - W)\|_{21} = \|E\|_{21}$. We see that

$$\begin{aligned}&\|S_d(V_d^T D - W)\|_{21} = \|U_d S_d(V_d^T D - W)\|_{21}\\ &= \|U_d S_d V_d^T D - U_d S_d W\|_{21} = \|U_d S_d V_d^T D - XV_d W\|_{21}.\end{aligned} \tag{5}$$