COMSTA: 6397 Model 3Gsc pp. 1–15 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Bayesian robust principal component analysis with structured sparse component

Ningning Han[a], Yumeng Song[b,*], Zhanjie Song[a,c]

[a] School of Science, Tianjin University, Tianjin 300072, PR China

[b] School of Precision Instrument and Opto-electronics Engineering, Tianjin 300072, PR China

[c] Center for Applied Mathematics of Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form 22 October 2016
Accepted 6 December 2016
Available online xxxx

Keywords:
Robust principal component analysis
Low-rank component
Structured sparse component
Variational Bayesian inference
Structured sparsity

ABSTRACT

The robust principal component analysis (RPCA) refers to the decomposition of an observed matrix into the low-rank component and the sparse component. Conventional methods model the sparse component as pixel-wisely sparse (e.g., ℓ_1-norm for the sparsity). However, in many practical scenarios, elements in the sparse part are not truly independently sparse but distributed with contiguous structures. This is the reason why representative RPCA techniques fail to work well in realistic complex situations. To solve this problem, a Bayesian framework for RPCA with structured sparse component is proposed, where both amplitude and support correlation structure are considered simultaneously in recovering the sparse component. The model learning is based on the variational Bayesian inference, which can potentially be applied to estimate the posteriors of all latent variables. Experimental results demonstrate the proposed methodology is validated on synthetic and real data.

1. Introduction

High-dimensional data, such as image and video processing, documents, web search, and biological data often lie in a low-dimensional structure or manifold. As the fields of optimization and statistics develop, the problem of exploring and exploiting low-dimensional structure in high-dimensional data has been extensively studied.

Principal component analysis (PCA) (Serneels and Verdonck, 2008; Giordani and Kiers, 2006), as a classical and popular unsupervised dimensionality reduction approach, has wide applications in computer vision and pattern recognition. To a large extent, PCA efficiently develops the best ℓ_2-norm low-rank approximation of the observed data. However, ℓ_2-norm is sensitive to outliers which often appear in practical situations. Therefore, PCA may not get the optimal performance under a large corruption, even if the corruption affects only a small part of the data.

To address the brittleness of classical PCA with respect to outliers, robust principal component analysis (RPCA) (Candès et al., 2011) has been proposed to decompose the observed data matrix $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$ into a low-rank matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ (rank: $r \ll \min\{m,n\}$) and a sparse matrix $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ (with sparse outliers). This research seeks to recover the low-rank matrix $\boldsymbol{X}$ and the sparse matrix $\boldsymbol{E}$ from $\boldsymbol{Y}$ by solving the following optimization problem

$$\min_{\boldsymbol{X},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{X}) + \lambda\|\boldsymbol{E}\|_0, \quad s.t.\ \boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}, \tag{1}$$

* Corresponding author.
E-mail addresses: ning_ninghan@tju.edu.cn (N. Han), YumengSong@tju.edu.cn (Y. Song), zhanjiesong@tju.edu.cn (Z. Song).

http://dx.doi.org/10.1016/j.csda.2016.12.005
0167-9473/

Please cite this article in press as: Han, N., et al., Bayesian robust principal component analysis with structured sparse component. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.12.005

where $\|\boldsymbol{E}\|_0$ denotes the number of nonzero entries in $\boldsymbol{E}$ and λ is a positive weighting parameter. Unfortunately, (1) is a highly nonconvex optimization problem. However, one can obtain a tractable optimization problem by relaxing (1) that replaces the ℓ_0-norm with the ℓ_1-norm and the rank with the nuclear norm, yielding the following convex optimization problem

$$\min_{\boldsymbol{X},\boldsymbol{E}} \|\boldsymbol{X}\|_* + \lambda\|\boldsymbol{E}\|_1, \quad s.t.\ \boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}, \tag{2}$$

where $\|\cdot\|_*$ and $\|\cdot\|_1$ denote the matrix nuclear norm (the sum of its singular values) and the ℓ_1-norm (the sum of the absolute values of entries), respectively. Candès et al. (2011) showed that under broad conditions the underlying low-rank component $\boldsymbol{X}$ and the sparse component $\boldsymbol{E}$ can be exactly recovered with high probability by solving the above convex optimization problem. As a promising tool, RPCA has been successfully applied to many research fields, such as video surveillance (Zhou et al., 2013), face recognition (Wagner et al., 2012), subspace clustering (Liu et al., 2013) and image alignment (Peng et al., 2012).

In fact, it is an identifiability issue to perfectly disentangle the low-rank and the sparse components. To make the problem meaningful, we first need to impose that the low-rank component is not sparse. As done in Candès et al. (2011), the singular value decomposition of the real low-rank component $\boldsymbol{X}_0$ is written as

$$\boldsymbol{X}_0 = \boldsymbol{U}\Sigma\boldsymbol{V}^*, \tag{3}$$

where $\boldsymbol{U}$, $\boldsymbol{V}$ are the left-singular and the right-singular matrices. Suppose that the rank of $\boldsymbol{X}_0$ is r, the incoherence condition with parameter μ is defined as

$$\max_i \|\boldsymbol{U}^* e_i\|^2 \le \frac{\mu r}{m}, \qquad \max_i \|\boldsymbol{V}^* e_i\|^2 \le \frac{ur}{n}, \qquad \|\boldsymbol{U}\boldsymbol{V}^*\|_\infty \le \sqrt{\frac{\mu r}{mn}}, \tag{4}$$

where $\|\boldsymbol{M}\|_\infty = \max_{i,j} |\boldsymbol{M}_{ij}|$ and e_i is the unit vector that has a one in the ith element and zeros everywhere else. As discussed in Candès et al. (2011), the incoherence condition asserts that for small values of μ, the singular vectors of $\boldsymbol{X}_0$ spread out. In other words, the low-rank component $\boldsymbol{X}_0$ is not sparse. To avoid another meaningless situation, we assume that the sparsity pattern of the real sparse component is selected uniformly at random (the sparse component is not low-rank). Under these assumptions that the rank of the low-rank component is not too large and the sparse component is reasonably sparse, the model (2) can perfectly recover the low-rank and the sparse components. The result is given in the following theorem.

Theorem 1 (*Candès et al., 2011*). *Suppose $\boldsymbol{X}_0$ obeys (4). Fix any $m \times n$ matrix Ψ of signs. Suppose that the support set Ω of $\boldsymbol{E}_0$ is uniformly distributed among all sets of cardinality w, and that $sgn([\boldsymbol{E}_0]_{ij}) = \Psi_{ij}$ for all $(i,j) \in \Omega$. Then, there is numerical constant c such that with probability at least $1 - cn_1^{-10}$ (over the choice of support of $\boldsymbol{E}_0$, $n_1 = max(m,n)$), the model (2) with a range of correct values of λ can exactly recover the real low-rank and the real sparse components, provided that $rank(\boldsymbol{X}_0) \le \rho_r n_2 \mu^{-1}(\log n_1)^{-2}$ and $w \le \rho_s mn$. In this equation, ρ_r and ρ_s are positive numerical constants and $n_2 = min(m,n)$.*

Algorithms developed for the RPCA problem are often intuitive extensions of low-rank matrix recovery, therefore share a similar trajectory. Among the different methods proposed are heuristic deterministic approaches based on nuclear norm relaxation, such as singular value thresholding (SVT) (Cai et al., 2010), singular value projection (SVP) (Jain et al., 2010), an accelerated proximal gradient algorithm (APG) (Toh and Yun, 2010), the augmented lagrange multiplier method (ALM) (Lin et al., 0000) etc. Although these algorithms may work well theoretically, they have limited reconstruction effects since the nuclear norm may not be a good surrogate to the rank function. To get a more accurate and robust approximation to the rank function, Hu et al. proposed a novel method called truncated nuclear norm regularization (TNNR) (Hu et al., 2013) which only minimized the smallest p singular values to recover the low-rank component. Noted that all the existing nonconvex penalty functions were concave and their gradients were decreasing functions, an iteratively reweighted nuclear norm (IRNN) was suggested in Lu et al. (2014). Inspired by the paradigm of ℓ_p-norm in compressive sensing (Ince et al., 2013), some try to expand this concept to the traditional nuclear norm (Lu, 2014; Lu et al., 2015; Nie et al., 2012; Peng et al., 2014), which can approximate the rank function better.

In addition to the above deterministic methods, some statistical algorithms express the RPCA as the solution of a Bayesian inference problem and apply statistical tools to solve it. The statistical procedures (Manteiga and Vieu, 2007), recently gaining popularity, offer several advantages over deterministic methods. First, prior knowledge about the rank of matrix is not necessary, and the way to estimate the unknown rank is similar to the automatic relevance determination strategy in machine learning. On the other hand, algorithmic parameters are treated as stochastic quantities so that it is insensitive to the initialization of parameters. In Ding et al. (2011), the authors modeled the singular values of low-rank component $\boldsymbol{X}$ and the entries of sparse component $\boldsymbol{E}$ with beta-Bernoulli priors, and the resulting algorithm used a Markov Chain Monte Carlo (MCMC) sampling scheme with high computational complexity for inference. Babacan et al. (2012) adopted sparse Bayesian learning principles to recover the sparse component and the low-rank component, which started from a matrix factorization formulation and enforced the low-rank constraint in the process through the sparsity constraint. The complex noise, as considered in Zhao et al. (2014), results in representing data noise as a mixture of Gaussian, which could fit a wide range of noises and provide high recovery performance.

However, the present methods obviously impose the sparsity constraint on individual coefficient of $\boldsymbol{E}$, and ignore the spatial connection of nonzero coefficients. In many practical scenarios, the distributions of coefficients in the sparse

Please cite this article in press as: Han, N., et al., Bayesian robust principal component analysis with structured sparse component. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.12.005