Robust sparse coding for subspace learning

Xiangguang Dai
Yingyin Tao
Jiang Xiong
Yuming Feng*
School of Three Gorges Artificial Intelligence
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100
China
yumingfeng25928@163.com

Abstract. Conventional sparse coding fails to learn a robust subspace when the data is corrupted by noises or outliers. To remedy this problem, matrix completion is considered into sparse coding to recover the corrupted data by the normal data and the robust sparse representations can be learned from the recovered data and the normal data. Therefore, a robust sparse coding method, called RSC, is proposed to learn a low-dimensional subspace from the corrupted data. Experiments are carried out on the image dataset which is contaminated by noises or outliers. It is demonstrated that our proposed RSC is more effective and robust in subspace learning and image clustering than other dimensionality reduction methods.
Keywords: sparse coding, matrix completion, robustness, dimensionality reduction.

1. Introduction

Many applications including community detection, recommender system, semi-supervised learning and clustering [1, 2, 3] suffer from the processing of high-dimensional data. Actually, only a few data are meaningful and the remaining data are redundant or noisy. It is necessary to find a low-dimensional space to represent images. For this purpose, matrix factorization methods have been paid great attention in the last two decades[4, 5, 6, 7].

Conventional matrix factorization methods are to seek several matrices such that can best approximate the original matrix. Principal component analysis (PCA) [8] is a famous unsupervised dimensionality reduction method, which require the basis vectors to be orthogonal. The non-negative matrix factorization (NMF) [4] method decomposes the non-negative data matrix into two non-negative low-dimensional matrices, which can achieve parts-based representations. Sparse coding (SC) [9] requires the leaned representations to be more sparser. This means that each image can be represented by a fewer features. However, above-mentioned approaches have a common drawback that they fail

∗. Corresponding author

to learn a robust subspace when the original data are corrupted by noises or outliers.

In this paper, a novel robust sparse coding (RSC) method is proposed based on matrix completion for dimensionality reduction from the corrupted dataset. Due to the use of matrix completion, the corrupted data matrix can be firstly recovered. Then we decompose the repaired matrix into two low-dimensional matrices by utilizing sparse coding. Finally we repeat above two steps until a robust subspace is obtained. The contributions of this paper are summarized as follows:

- A robust sparse coding framework called RSC is proposed, and we develop an iterative optimization algorithm to solve RSC.
- We evaluate the performance of RSC by clustering on the ORL dataset [10] with Salt and Pepper noise or contiguous occlusion. Experiments demonstrate that the proposed RSC is more effective to learn a robust subspace.

2. Robust sparse coding

Suppose that each image $\vec{\xi} \in R^m$ can be represented by sets of basis vectors $\vec{b}_1, \cdots, \vec{b}_r \in R^m$. Hence, we obtain

$$\vec{\xi} = \sum_{i=1}^{r} \vec{b}_i s_i, \tag{1}$$

where $\vec{s} = [s_1 \cdots s_r] \in R^r$ is a coefficient vector. Given an image dataset $V = [\vec{\xi}_1, \cdots, \vec{\xi}_n] \in R^{m\times n}$, the basis matrix $W = [\vec{b}_1, \cdots, \vec{b}_r] \in R^{m\times r}$ and the coefficient matrix $H = [\vec{s}_1, \cdots, \vec{s}_n] \in R^{r\times n}$ are expected to represent V. Based on the previous analysis, sparse coding is mathematically summarized into the following optimization problem:

$$\min_{W,H} F(W,H) = \| V - WH \|_F^2 + \lambda \sum_{i=1}^{n} \| H_i \|_1, \tag{2}$$

where $\| \cdot \|_F$ and $\| \cdot \|_1$ denote the Frobenius norm and the L1 norm, separately. Each column of W is a basis vector that captures the effective features in the dataset, and each column of H can be viewed as the r-dimensional representation with respect to the new bases. Hence, sparse coding is also a dimensionality reduction method because the dimension can be reduced from m to r ($r \ll m$). Utilizing sparse coding, a satisfactory subspace can be achieved from the image dataset without noises or outliers. However, it fails when the dataset is contaminated by Salt and Pepper noise or contiguous occlusion. To address this problem, matrix completion is proposed to recover the corrupted data by the normal data.

Suppose that $M_i \in R^m$ and $V_i \in R^m$ are the corrupted image vector and the recovered image vector, separately. The conventional matrix completion to minimize the error between the corrupted dataset $M = [M_1, \cdots, M_n] \in R^{m\times n}$ and the recovered dataset $V = [V_1, \cdots, V_n] \in R^{m\times n}$ [11] can be summarized as

$$\begin{aligned} &\min_{V} \quad rank(V) \\ &s.t. \quad P_\Omega(M) = P_\Omega(V), \end{aligned} \tag{3}$$

where P is the projection operator and Ω is the corrupted area. If $(i, j) \in \Omega$, we obtain that

$$[P_\Omega(M)]_{ij} = \begin{cases} M_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where M_{ij} is the (i,j)-th element of M. It is obvious that problem (3) cannot be solved directly. Hence, many researchers proposed some related matrix completion methods [11, 12, 13, 14, 15]. According to previous researches, (3) can be rewritten into the following optimization problem:

$$\begin{aligned} &\min_{V} \quad \| V \|_F^2 \\ &s.t. \quad P_\Omega(M) = P_\Omega(V). \end{aligned} \tag{5}$$

Euclidean distance is proposed to measure $P_\Omega(M)$ and $P_\Omega(V)$. Problem (5) can be transformed into the following form:

$$\begin{aligned} &\min_{V} \quad R_2 = \beta \| V \|_F^2 + \alpha \| (V - M) \otimes S \|_F^2, \\ &S_{ij} = \begin{cases} 0, & \text{if } (i,j) \in \Omega, \\ 1, & \text{otherwise}, \end{cases} \end{aligned} \tag{6}$$

where $\alpha \geq 0$ and $\beta \geq 0$ are the regularization parameters, and S is the weight matrix that denotes the contaminated or uncontaminated positions.

Combining sparse coding and matrix completion, a novel robust sparse coding method (RSC) is proposed to handle Salt and Pepper noise and contiguous occlusion. Therefore, RSC can be summarized as follows:

$$\begin{aligned} \min_{V,W,H} \quad F(V,W,H) = &\frac{1}{2}(\| V - WH \|_F^2 + \beta \| V \|_F^2 \\ &+ \alpha \| (V - M) \otimes S \|_F^2) + \lambda \sum_{i=1}^{n} \| H_i \|_1 . \end{aligned} \tag{7}$$

3. Optimization schemes

In this section, we present necessary optimization theory to solve (7). Because of non-smooth and non-convex properties of (7), it is too complicated to obtain the global solution. Therefore, we transform (7) into several convex smooth problems and alternately solve them until convergence.

Problem (7) can be transformed into the following three convex problems:

$$\min_{V} \quad F(V) = \frac{1}{2}(\parallel V - WH \parallel_F^2 + \beta \parallel V \parallel_F^2 + \alpha \parallel (V - M) \otimes S \parallel_F^2) \tag{8}$$

and

$$\min_{W} \quad F(W) = \frac{1}{2} \parallel V - WH \parallel_F^2 \tag{9}$$

and

$$\min_{H} \quad F(H) = \frac{1}{2} \parallel V - WH \parallel_F^2 + \lambda \sum_{i=1}^{n} \parallel H_i \parallel_1 . \tag{10}$$

For problem (8), we obtain that its optimum solution as

$$V = (\alpha M \otimes S + WH) \oslash ((1 + \beta) \otimes I + \alpha S), \tag{11}$$

where $X \otimes Y$ and $X \oslash Y$ are the element-wise product and division, respectively. Similarly, the solution of (9) is

$$W = W - \frac{1}{L}\nabla F(H) = W - \frac{1}{L}(WHH^T - VH^T), \tag{12}$$

where the Lipschitz constant $L_W = \parallel H^T H \parallel$ and $\parallel \cdot \parallel$ denotes the L2 norm. The objective function of (10) is non-smooth, therefore, it should be transformed into smooth functions. Problem (10) can be re-written as

$$\begin{aligned} \min_{P,Q} \quad & F(P,Q) = \frac{1}{2}\|V - W(P - Q)\|_F^2 + \lambda \sum_{i,j} p_{ij} + \lambda \sum_{i,j} q_{ij} \\ \text{subject to} \quad & P \geq 0, Q \geq 0, \end{aligned} \tag{13}$$

where $P = [p_{ij}] = \max(H, 0)$, $Q = [q_{ij}] = \max(-H, 0)$ and $H = P - Q$. To minimize (13), we transform it into the following two problems:

$$\begin{aligned} \min_{P} \quad & F(P) = \frac{1}{2}\|V + WQ - WP\|_F^2 + \beta \sum_{i,j} p_{ij} \\ \text{subject to} \quad & P \geq 0 \end{aligned} \tag{14}$$

and

$$\begin{aligned} \min_{Q} \quad & F(Q) = \frac{1}{2}\|V - WP - (-W)(Q)\|_F^2 + \beta \sum_{i,j} q_{ij} \\ \text{subject to} \quad & Q \geq 0. \end{aligned} \tag{15}$$

The solution of (14) and (15) can be achieved by

$$P = [P - \frac{1}{L_H}\nabla F(P)]^+ = [P - \frac{1}{L_H}(W^TWP - (W^TV + W^TWQ) + \lambda)]^+ \tag{16}$$

and

$$Q = [Q - \frac{1}{L_H}\nabla F(Q)]^+ = [Q - \frac{1}{L_H}(W^TWQ - (-W^TV + W^TWP) + \lambda)]^+, \tag{17}$$

where the Lipschitz $L_H = \| W^TW \|$ and $[\cdot]^+$ makes the negative data zeros.

By (11), (12), (16) and (17), the local solution of (7) can be easily searched. The related algorithm is summarized in Algorithm 1.

Algorithm 1 RSC

Input: $V \in R^{m\times n}$, $M \in R^{m\times n}$, $W \in R^{m\times r}$, $H \in R^{r\times n}$, α, β, λ, $iter$, $number$

Oupt: $V \in R^{m\times n}$, $W \in R^{m\times r}$, $H \in R^{r\times n}$

Calculate S by (6)
for $it = 0$ to $number$ **do**
  $V = (\alpha M \otimes S + WH) \oslash ((1+\beta)\otimes I + \alpha S)$
  $W = W - \frac{1}{L}\nabla F(H) = W - \frac{1}{L_W}(WHH^T - VH^T)$
  for $k = 0$ to $iter$ **do**
    $P = [P - \frac{1}{L_H}\nabla F(P)]^+ = [P - \frac{1}{L_H}(W^TWP - (W^TV + W^TWQ) + \lambda)]^+$
    $Q = [Q - \frac{1}{L_H}\nabla F(Q)]^+ = [Q - \frac{1}{L_H}(W^TWQ - (-W^TV + W^TWP) + \lambda)]^+$
  end for
end for
$V = WH$
return V,W,H

4. Experiments

We explore the clustering performance of RSC on the ORL dataset, by comparing with three dimensionality reduction methods: NMF [4], PCA [8] and SC [9]. In this experiment, two corruptions, i.e., Salt and Pepper noise and contiguous occlusion are proposed to test the effectiveness and robustness of above-mentioned methods.

The ORL dataset contains 400 images of different 40 individuals. Each person has 10 different images with different facial expressions, facial details (without-glasses or with-glasses) and lighting. Each face image is a grayscale image with the 32×32 pixel array.

We complete the cluttering experiments by the following steps:

- 90% face images are corrupted by Salt and Pepper noise or contiguous occlusion. The 360 normal images and 40 corrupted images can be represented by a matrix $M \in R^{1024\times 400}$.

- Given the factorization rank r, M can be decomposed into two low-dimensional matrices $W \in R^{1024\times 100}$ and $H \in R^{100\times 400}$ by different dimensionality methods.
- We utilize the K-means algorithm to cluster the subspace H, and present the clustering performances measured by two metrics including Accuracy (AC) and Normalized Mutual Information (NMI) [16].
- All methods are implemented in matlab R2012b and their related algorithms are all terminated in 200 iterations.
- Because the objective function of RSC, NMF and SC is non-convex, the related algorithms cannot search the global solution easily. Each algorithm is implemented 50 times and we report the average clustering results.

4.1 Salt and Pepper noise

For each image, Salt and Pepper noise is utilized to corrupt some pixels, and the corrupted pixel value is 0 or 255. In this subsection, we verify whether RSC can obtain a robust subspace from the ORL dataset contaminated by Salt and Pepper noise. To demonstrate the effectiveness and robustness of RSC, we propose different corrupted percentages for the ORL dataset, i.e., p =10%, 15%, 20%, 25% and 30%. Let $r = 100$, $\alpha = 100$, $\lambda = 100$, $\beta = 0$, $iter = 20$ and $number = 200$.

Table 1 shows the clustering results in different corrupted cases. From the comparisons, we observe that: (1) Our proposed RSC achieves the best clustering results than other methods. (2) SC performs better than NMF and PCA as the corrupted percentage increases. The clustering results demonstrate that RSC can learn a robust and effective subspace on the ORL dataset contaminated by Salt and Pepper noise.

Table 1: Clustering AC and NMI with different corrupted percentages from 10% to 30%(the best PSNR in bold).

p(%)	10	15	20	25	30
RSC	**63.62 (79.93)**	**60.03 (77.71)**	**57.18 (77.16)**	**61.57 (77.48)**	**61.70 (78.40)**
NMF	48.65 (67.47)	33.70 (56.10)	25.70 (47.14)	21.85 (43.79)	21.95 (43.05)
PCA	37.75 (58.28)	26.50 (45.75)	22.75 (42.60)	22.25 (42.27)	18.00 (40.66)
SC	55.07 (72.76)	43.95 (64.90)	37.60 (56.75)	30.10 (52.03)	22.92 (45.36)

4.2 Contiguous occlusion

In this subsection, contiguous occlusion is applied to simulate extreme outliers. We randomly corrupt a $b \times b$ pixel block on each image, and the corrupted pixel blocks are filled with the pixel value 255. To discuss the influence of different block sizes, we propose the block size b from 10 to 14 with the step size 1.

Table 2 present the clustering results measured by AC and NMI. According to Table 2, RSC achieves the best clustering performances in different block sizes, while other algorithms fail to clustering. It is obvious that RSC is more effective to learn a robust subspace in large number of outliers on the ORL dataset.

Table 2: Clustering AC and NMI with different block sizes from 10 to 14 (the best PSNR in bold).

b	10	11	12	13	14
RSC	**60.80** (**77.80**)	**56.68** (**77.34**)	**63.05** (**78.45**)	**61.95** (**78.80**)	**64.60** (**79.23**)
NMF	17.62 (40.38)	16.97 (39.30)	16.92 (38.97)	17.47 (40.43)	17.40 (39.90)
PCA	18.75 (39.09)	16.75 (37.25)	16.50 (38.84)	15.50 (38.29)	16.00 (38.09)
SC	17.92 (39.62)	16.50 (39.58)	16.95 (39.5)	17.17 (40.14)	17.20 (40.03)

5. Conclusion and future work

This paper proposed a novel robust sparse coding method (RSC) for dimensionality reduction. Based on matrix completion, RSC is more effective to handle Salt and Pepper noise and contiguous occlusion than NMF, PCA and SC. Experiments demonstrate that our proposed RSC can learn more robust low-dimensional representations and achieves better clustering results from the corrupted ORL dataset.

Two topics can be discussed in the future work:

- RSC, NMF, PCA and SC are unsupervised methods, hence semi-supervised or supervised methods should be considered to handle noises or outliers.
- Based on matrix completion, most of dimensionality reduction models can be improved to achieve a robust subspace.

Acknowledgment

This work is supported by Foundation of Chongqing Municipal Key Laboratory of Institutions of Higher Education ([2017]3), Foundation of Chongqing Development and Reform Commission (2017[1007]), Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant Nos. KJQN201901218 and KJ1710248), Natural Science Foundation of Chongqing (Grant No. cstc2019jcyj-bshX0101).

References

[1] G. Casalino and D. B. Nicoletta and C. Mencar, *Subtractive clustering for seeding non-negative matrix factorizations*, Information Sciences, 257 (2014), 369-387.

[2] H. Che and J. Wang, *A nonnegative matrix factorization algorithm based on a discrete-time projection neural network*, Neural Networks, 103 (2018), 63-71.

[3] J. Fan and J. Wang, *A collective neurodynamic optimization approach to nonnegative matrix factorization*, IEEE Transactions on Neural Networks and Learning Systems, 28 (2016), 2344-2356.

[4] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1998), 788.

[5] P. N. Belhumeur and J. P. Hespanha and D. J. Kriegman, *Eigenfaces vs. fisherfaces: recognition using class specific linear projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 7 (1997), 711-720.

[6] S. T. Roweis and L. K. Saul, *Nonlinear dimensionality reduction by locally linear embedding,* Science, 290 (2000), 2323-2326.

[7] X. He and S. Yan and Y. Hu and P. Niyogi and H. Zhang, *Face recognition using Laplacianfaces*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 3 (2005), 328-340.

[8] M. A. Turk and A. P. Pentland, *Face recognition using eigenfaces,* International Conference on Computer Research and Development, 2011.

[9] H. Lee and A. Battle and R. Raina and A. Y. Ng, *Efficient sparse coding algorithms*, Advances in Neural Information Processing Systems, 19 (2007), 801-808.

[10] F. S. Samaria and A. C. Harter, *Parametrisation of a stochastic model for human face identification*, IEEE Workshop on Applications of Computer Vision, 1994.

[11] E. J. Candes and B. Recht, *Exact matrix completion via convex*, Optimization, 9 (2009), 717 pages.

[12] J. Cai and E. J. Candes and Z. Shen, *A singular value thresholding algorithm for matrix completion*, Siam Journal on Optimization, 20 (2008), 1956-1982.

[13] S. Ma and D. Goldfarb and L. Chen, *Fixed point and Bregman iterative methods for matrix rank minimization,* Mathematical Programming, 128 (2009), 321-353.

[14] Y. Liu and L. C. Jiao and F. Shang, *A fast tri-factorization method for low-rank matrix recovery and completion,* Pattern Recognition, 46 (2013), 163-173.

[15] F. Nie and H. Wang and H. Huang and C. Ding, *oint Schatten p-norm and L_p-norm robust matrix completion for missing value recovery,* Knowledge and Information Systems, 42 (2015), 525-544.

[16] D. Cai and X. He and J. Han, *Document clustering using locality preserving indexing*, IEEE Transactions on Knowledge and Data Engineering, 17 (2005), 1624-1637.

Accepted: 29.06.2019