Robust non-negative matrix factorization for subspace learning

Xiangguang Dai
School of Three Gorges Artificial Intelligence
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100
China

Yingyin Tao
School of Three Gorges Artificial Intelligence
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100
China

Wei Zhang
Chongqing Engineering Research Center of
Internet of Things and Intelligent Control Technology
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100
China

Yuming Feng*
Key Laboratory of Intelligent Information Processing and
Control of Chongqing Municipal Institutions of Higher Education
Chongqing Three Gorges University
Wanzhou, Chongqing, 404100, China
yumingfeng25928@163.com

Abstract. Conventional non-negative matrix factorization methods fail to clusetring due to many noisy data. Therefore, this paper proposes a novel non-negative matrix factorization method to learn a robust subspace from the noisy space. Considering matrix completion and non-negative matrix factorization, the proposed method can recover the contaminated data and learn a low-dimensional effective subspace. Experiments on the ORL face dataset with Salt & Pepper noise show that our proposed method is effective and robust.
Keywords: non-negative matrix factorization, matrix completion, robustness, clustering.

1. Introduction

Non-negative matrix factorization (NMF) [1] can discover a non-negative properties from a high-dimensional data space. Therefore, NMF has received considerable attention in data mining [2, 3], signal processing [4] and computer vision [5]. NMF aims to decompose the original data matrix into two low-dimensional

*. Corresponding author

matrices such that their product approximates the data matrix. However, NMF cannot learn an effective low-dimensional subspace due to the noisy data.

Many variant NMF models have shown the robustness and effectiveness for subspace learning. L_1-NMF [6] was proposed to handle the Laplace noise by minimizing the non-smooth L_1 norm function, however its related optimization has the defect of slow convergence. To solve this problem, Manhattan NMF [7] is proposed to optimize L_1-NMF by transforming the non-smooth function into a smooth one and solving it by Nesterov's optimization method [8]. Zhang et al. [9] proposed robust non-negative matrix factorization to recover the undamaged matrix by subtracting one error matrix from the damaged data matrix. Although previous variants have made contribution to handle noises, they fail to learn a robust subspace with considerable contaminated data.

Recently, with the popularity of compressed sensing (CS) theory [10, 11], matrix completion (MC) [12] has been paid considerable attention. When a few entries of the original data matrix are missing, corrupted or contaminated, MC can recover the damaged data from the other undamaged data. At present, MC has applied many areas such as image inpainting [13], data denosing [14] and recommended system [15].

In this paper, a novel method called non-negative matrix completion factorization (NMCF) is proposed to learn a robust and effective subspace from a contaminated face dataset. The advantages of NMCF are to recover the noisy data from MC and learn a robust subspace from NMF. The optimization algorithm of NMCF is designed according to the block-coordinate-descent-method [16] and Nesterov's optimization method [17]. Finally, the image inpainting and clustering on the contaminated ORL dataset demonstrate the effectiveness and robustness of NMCF.

2. Robust non-negative matrix factorization

2.1 Non-negative matrix factorization

Supposed that $\{v_i\}_{i=1}^{n} \in R^m$ is the feature description of i-th image and $V = [v_1, \cdots, v_n] \in R^{m\times n}$ is an image matrix. By NMF, V is decomposed into two non-negative matrices $W \in R^{m\times r}$ and $H \in R^{r\times n}$ such that their product approximates V. Commonly, there are many loss functions to minimize the difference, however it is not the main topic in this paper. We propose the Euclidean distance to measure the reconstruction error. Hence, the optimization function of NMF is as follows:

$$\begin{aligned} (1) \qquad & \min_{W,H} \quad F(W,H) = \frac{1}{2} \parallel V - WH \parallel_F^2 \\ & s.t. \quad W \geq 0, H \geq 0. \end{aligned}$$

It is obvious that (1) are not convex in optimizing W and H simultaneously. The common optimization structure for (1) is to optimize W and H iteratively until convergence. Based on this structure, Lee and Seung [18] provided a classical algorithm for (1). The Multipartite Update algorithm minimizing (1) can be summarized as

$$W_{ik} \leftarrow W_{ik}\frac{\left(VH^T\right)_{ik}}{\left(WHH^T\right)_{ik}}, \quad (2)$$

$$H_{kj} \leftarrow H_{kj}\frac{\left(W^TV\right)_{kj}}{\left(W^TWH^T\right)_{kj}}. \quad (3)$$

2.2 Matrix completion

Given a damaged image matrix $M \in R^{m\times n}$, each column of M is an image vector and most of pixels in each image may be damaged. MC aims to recover the damaged pixels from other undamaged pixels. MC can be viewed as an optimization problem in the following form:

$$\min_X \quad rank(X) \quad (4)$$

$$s.t. \quad P_\Omega(M) = P_\Omega(X), \quad (5)$$

where $X \in R^{m\times n}$ is the image inpainting matrix, Ω is the damaged area, and $P_\Omega(M_{ij})$ is defined by

$$P_\Omega(M_{ij}) = \begin{cases} M_{ij}, & (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

2.3 Non-negative matrix completion factorization

NMF is a dimensionality reduction method which is useful for clustering, signal processing and image recognition. However, it performs unsatisfactorily in dimensionality reduction when the original images are damaged. In this section, a novel matrix factorization method, called nonnegative matrix completion factorization (NMCF), which is used for dimensionality reduction. The proposed method can learn a good parts-based representation from the damaged images.

Consider a damaged grayscale image matrix $M \in R^{m\times n}$ consisting of n damaged images, among which the pixel values 0 or 255 are the damaged area. Hence, a weight matrix $S \in R^{m\times n}$ corresponding to the damaged or undamaged area can be described as

$$S_{ij} = \begin{cases} 0, & M(i,j) \text{ is the damaged pixel} \\ 1, & \text{otherwise.} \end{cases} \quad (7)$$

Supposed that each entry of an estimate recovered matrix $V \in R^{m\times n}$ can be recovered by M_{ij}. According to the definition of S in (7), the following term can be used to measure the approximation between M_{ij} and V_{ij}:

$$
\begin{aligned}
(8) \qquad R &= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(M_{ij}-V_{ij})^2 \times S_{ij} \\
&= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}((M_{ij}-V_{ij}) \times S_{ij})^2 \\
&= \frac{1}{2}\parallel (V-M)\otimes S \parallel_F^2 .
\end{aligned}
$$

By minimizing (8), we expect if any point M_{ij} is damaged, M_{ij} can be recovered to be V_{ij}. Given a damaged image matrix $M \in R^{m\times n}$, our proposed method aims to find its corresponding recovered matrix V, and to decompose V into two non-negative matrices W and H. Non-negative matrix completion factorization (NMCF) can be minimized by Euclidean distance as follows:

$$
(9) \qquad \begin{aligned}
\min_{V,W,H} \quad & F(V,W,H) = \frac{1}{2}(\parallel V-WH \parallel_F^2 + \parallel (V-M)\otimes S \parallel_F^2) \\
s.t. \quad & V \geq 0, W \geq 0, H \geq 0.
\end{aligned}
$$

3. Algorithm minimizing non-negative matrix completion factorization

In this section, we propose Nesterov's method to optimize (9) instead of the Multipartite Update method. Firstly, we propose the block-coordinate-descent method to divide (9) into three optimization problems. Secondly, the Nesterov's optimization theory is applied to optimize each sub-problem. Thirdly, we optimize each sub-problem alternatively until convergence.

Because (9) is not a convex problem, it is difficult to obtain the global solution. Fortunately, the block-coordinate-descent method can search a local solution. This approach alternatively optimizes one block of variables with the remaining blocks fixed. For NMCF, there are three block variables V, W and H. Hence, the block-coordinate-descent method alternatively minimizes the following three optimization problems:

$$
(10) \qquad \begin{aligned}
\min_{V} \quad & F(V) = \frac{1}{2}(\parallel V-WH \parallel_F^2 + \parallel (V-M)\otimes S \parallel_F^2) \\
s.t. \quad & V \geq 0
\end{aligned}
$$

and

$$
(11) \qquad \begin{aligned}
\min_{W} \quad & F(W) = \frac{1}{2}\parallel V-WH \parallel_F^2 \\
s.t. \quad & W \geq 0
\end{aligned}
$$

and

$$
(12) \qquad \begin{aligned}
\min_{H} \quad & F(H) = \frac{1}{2}\parallel V-WH \parallel_F^2 \\
s.t. \quad & H \geq 0.
\end{aligned}
$$

According to Lemma 1 and Lemma 2, (10), (11) and (12) can be easily solved by the optimal gradient method [17].

Lemma 1 ([17])**.** The objective functions $F(V)$, $F(W)$ and $F(H)$ are convex.

Proof. Given any V_1, $V_2 \in R^{m\times n}$ and $\lambda \in (0,1)$, we have

$$\begin{aligned} & F(V_1 + (1-\lambda)V_2) - (\lambda F(V_1) + (1-\lambda)F(V_2)) \\ = \quad & -\frac{\lambda(1-\lambda)}{2}(\| V_1 - V_2 \|_F^2 + \| (V_1 - V_2) \otimes S \|_F^2) \leq 0. \end{aligned}$$

Based on the definition of convex function, it is clear that $F(V)$ is convex. According to Lemma 1 in [17], we know that $F(W)$ and $F(H)$ are convex. This completes the proof.

Lemma 2 ([17])**.** The gradients of $F(V)$, $F(W)$ and $F(H)$ are differentiable and Lipschitz continuous. Their Lipschitz constants are $\| I + S \|_F$, $\| HH^T \|$ and $\| WW^T \|$, separately.

Proof. According to (10), we can get the gradient of $F(V)$ by

$$\nabla_V F(V) = V - WH + (V - M) \otimes S.$$

For any V_1 and V_2, we have

$$\begin{aligned} & \| \nabla_V F(V_1) - \nabla_V F(V_2) \|_F \\ = \quad & \| (V_1 - V_2) \otimes (I + S) \|_F \\ = \quad & \sum_{i=1}^{m}\sum_{j=1}^{n}\{(I+S)_{ij}(V_1 - V_2)_{ij}\} \\ \leq \quad & \sum_{i=1}^{m}\sum_{j=1}^{n}(I+S)_{ij}\sum_{i=1}^{m}\sum_{j=1}^{n}(V_1 - V_2)_{ij} \\ \leq \quad & \| I + S \|_F \| V_1 - V_2 \|_F, \end{aligned}$$

where $I \in R^{m\times n}$ is a matrix whose entries are all one. Hence, $F(V)$ is differentiable and Lipschitz continuous, and its Lipschitz constant is $\| I + S \|_F$. According to (11) and (12), we can get the gradients of $F(W)$ and $F(H)$ by

$$\nabla_W F(W) = WHH^T - VH^T,$$
$$\nabla_H F(H) = W^T WH - W^T V.$$

According to [17], $F(W)$ and $F(H)$ have been proved to be differentiable and Lipschitz continuous. This completes the proof.

Consider the following optimization problem:

$$\min_{X\geq 0} F(X), \tag{13}$$

where $X \in R^{m\times n}$ and $F : R^{m\times n} \to R$. Recent researches [8, 19] show that Nesterov's optimal gradient method optimizing problems can achieve the convergence rate $O(\frac{1}{k^2})$. Since $F(X)$ is convex and Lipschitz continuous, Nesterov's method can easily optimize (13). Generally, two sequences Y_k and X_k are alternatively updated in each iteration. For $k \geq 1$, we have

$$X_k = \arg\min_{X\geq 0}\{\phi_k(Y_k, X)$$

$$= F(X) + < \nabla F(X), X - Y_k > + \frac{L}{2} \| X - Y_k \|_F^2\}, \tag{14}$$

$$Y_{k+1} = X_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(X_k - X_{k-1}), \tag{15}$$

where $\phi_k(Y_k, X)$ is the proximal function, L is the Lipschitz constant of $F(X)$, $< \cdot, \cdot >$ denotes the sum of the element-wise multiplication of two matrices and $\alpha_{k+1} = \frac{1+\sqrt{4\alpha_k^2+1}}{2}$. Using the Lagrange multiplier method and K.K.T. conditions, (14) can be solved by

$$X_k = \max(Y_k - \frac{1}{L}\nabla F(X), 0). \tag{16}$$

By updating X_k, α_{k+1} and Y_{k+1} iteratively until convergence, the optimal solution of (13) can be obtained. Based on above analysis, we summarize Nesterov's method to optimize (9) in Algorithm 1.

4. Experiments

In this section, two experiments are implemented on the ORL face image database [20] to demonstrate the effectiveness and robustness of our proposed method. Firstly, Salt & Pepper noise is considered to corrupt $\%p$ pixels in all face images ($p = 5, 20, 35, 50$). Secondly, we apply NMCF and NMF to recover the corruptions. Thirdly, we evaluate the effectiveness of NMCF and NMF by clustering the low-dimensional subspace of the damaged images.

The first experiment is to evaluate the robustness of NMCF and NMF in recovering the damaged images. We verify whether NMCF and NMF have the capacity to deal with Salt & Pepper noise. Let $m = 1024, n = 400, r = 50, n = 20$ and $number = 100$. Figure 1 mainly shows contaminated face images and recovered images by NMCF and NMF. In each sub-figure, the first two groups of images are the example images and contaminated images, and the last two groups of images are the recovered images by NMCF and NMF, separately. By contrast, NMCF can recover images when different percentages of pixels are contaminated. However, NMF fails to recover contaminated images when face images have more contaminated pixels.

The second experiment is to evaluate effectiveness of parts-based representation. After dimensional reduction by NMCF and NMF, we can achieve low-dimensional features for clustering. We compare the clustering effect by Kmeans

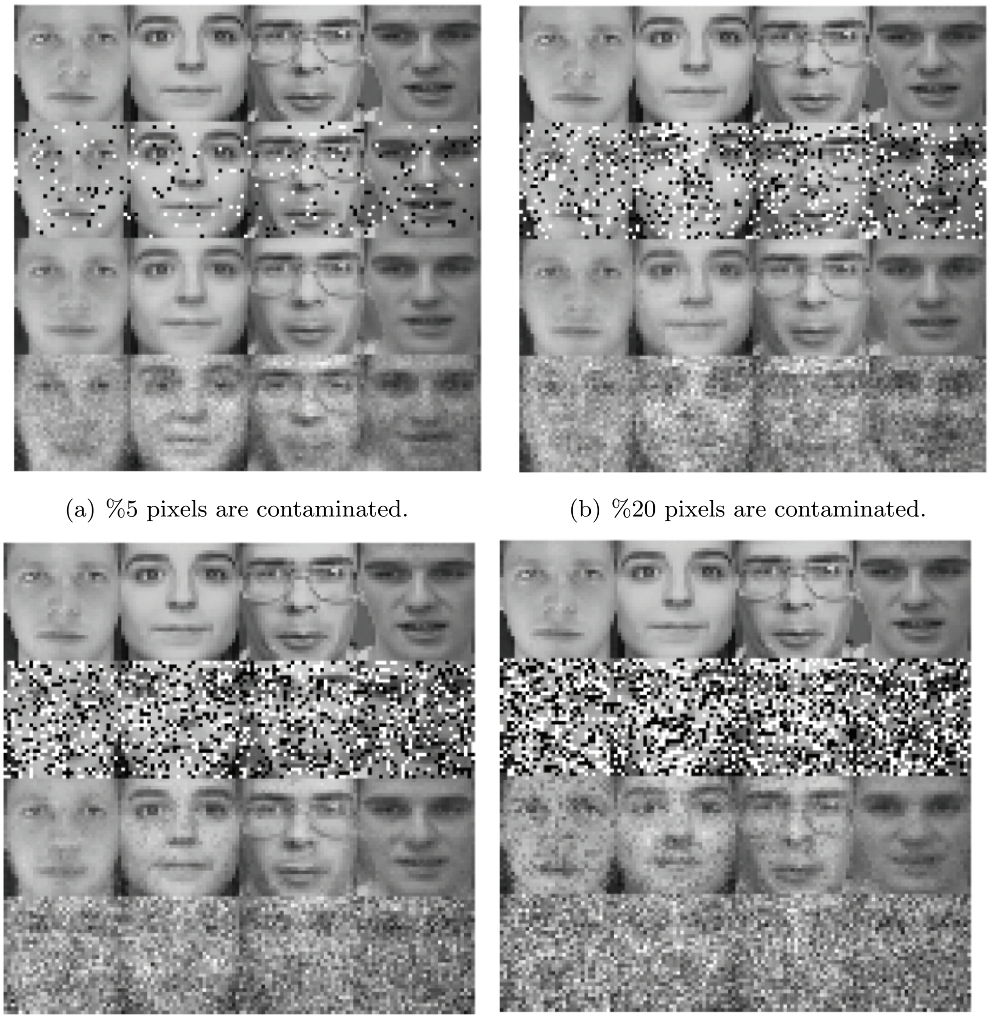
Algorithm 1 NMCF

Input: V,W,H,n,$number$
Output: V^*, W^*, H^*
for $i = 0$ to number **do**
1. $Y_0 = V, \alpha_0 = 1, L = \| I + S \|_F$
for $k = 0$ to n **do**
2. $V_k = max(Y_k - \frac{1}{L}\nabla_V F(V), 0)$
3. $\alpha_{k+1} = \frac{1+\sqrt{4\alpha_k^2+1}}{2}$
4. $Y_{k+1} = V_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(V_k - V_{k-1})$
end for
5. $V = V_k$
6. $Y_0 = W, \alpha_0 = 1, L = \| HH^T \|_2$
for $k = 0$ to n **do**
7. $W_k = max(Y_k - \frac{1}{L}\nabla_W F(W), 0)$
8. $\alpha_{k+1} = \frac{1+\sqrt{4\alpha_k^2+1}}{2}$
9. $Y_{k+1} = W_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(W_k - W_{k-1})$
end for
10. $W = W_k$
11. $Y_0 = H, \alpha_0 = 1, L = \| WW^T \|_2$
for $k = 0$ to n **do**
12. $H_k = max(Y_k - \frac{1}{L}\nabla_H F(H), 0)$
13. $\alpha_{k+1} = \frac{1+\sqrt{4\alpha_k^2+1}}{2}$
14. $Y_{k+1} = H_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(H_k - H_{k-1})$
end for
15. $H = H_k$
end for
16. $V^* \leftarrow V$, $W^* \leftarrow W$, $H^* \leftarrow H$

,NMCF and NMF in terms of two metrics including Normalized Mutual Information (NMI) and Accuracy (AC) [21]. The final clustering results on the contaminated dataset are presented in Table 1. It is clear that our proposed NMCF performs more robustly and effectively when the percentage of the contaminated pixels is increasing.

5. Conclusion

This paper proposes a non-negative matrix completion factorization method to learn a robust and effective subspace from the contaminated data. The loss function based on NMF and MC is optimized by Nesterov's methods, hence the convergence speed of the proposed optimization problem is very fast. According to experiments on the ORL face dataset with Salt & Pepper noise, there are two



(a) %5 pixels are contaminated. (b) %20 pixels are contaminated.

(c) %35 pixels are contaminated. (d) %50 pixels are contaminated.

Figure 1. Contaminated and recovered samples.

Table 1. Clustering performance on contaminated images by Kmeans, NMCF and NMF

	NMI				Accuracy			
	p=5	p=20	p=35	p=50	p=5	p=20	p=35	p=50
Kmeans	76.28	70.95	56.96	46.02	57.00	51.50	35.00	22.00
NMCF	77.42	78.10	75.44	68.49	60.50	63.50	58.00	48.25
NMF	76.90	51.46	40.19	39.60	62.00	63.50	21.00	18.75

significant advantages in our proposed method. Firstly, the contaminated pixels can be repaired from other uncontaminated pixels. Secondly, a low-dimensional effective subspace from the contaminated face dataset is learned and the clustering effect is better than NMF and Kmeans.

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), Foundation of Chongqing Three Gorges University.

References

[1] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), 788-791.

[2] V. P. Pauca, F. Shahnaz, M. W. Berry and R. J. Plemmons, *Text mining using non-negative matrix factorizations,* Siam International Conference on Data Mining, 2004.

[3] W. Xu, X. Liu and Y. H. Gong, *Document clustering based on non-negative matrix factorization,* Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003, 267-273.

[4] A. Cichocki, R. Zdunek and S. Amari, *New algorithms for non-negative matrix factorization in applications to blind source separation,* 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006.

[5] P. O. Hoyer, *Non-negative sparse coding,* Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing, 2002, 557-565.

[6] E. Y. Lam, *Non-negative matrix factorization for images with laplacian noise,* APCCAS 2008-2008 IEEE Asia Pacific Conference on Circuits and Systems, 2008, 798-801.

[7] N. Guan, D. Tao, Z. Luo and J. Shawetaylor, *MahNMF: Manhattan non-negative matrix factorization,* Journal of Machine Learning Research, 2012.

[8] Y. Nesterov, *Smooth minimization of non-smooth functions,* Mathematical Programming, 103 (2005), 127-152.

[9] L. Zhang, Z. Chen, M. Zheng and X. He, *Robust non-negative matrix factorization,* Frontiers of Electrical and Electronic Engineering in China, 6 (2015), 192-200.

[10] D. L. Donoho, *Compressed sensing,* IEEE Transactions on Information Theory, 52 (2006), 1289-1306.

[11] E. J. Candes, J. Romberg and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,* IEEE Press, 2006.

[12] E. J. Candes and B. Recht, *Exact matrix completion via convex optimization,* Foundations of Computational Mathematics, 9 (2009), 717-772.

[13] W. Li, L. Zhao, Z. Lin, D. Xu and D. Lu, *Non-local image inpainting using low-rank matrix completion,* Computer Graphics Forum, 2015, 111-122.

[14] E. J. Candes and Y. Plan, *Matrix completion with noise,* Proceedings of the IEEE, 98 (2010), 925-936.

[15] A. Gogna and A. Majumdar, *Matrix completion incorporating auxiliary information for recommender system design,* Expert Systems with Applications, 42 (2015), 5789-5799.

[16] P. Tseng, *Convergence of a block coordinate descent method for nondifferentiable minimization,* Journal of Optimization Theory and Applications, 109 (2001), 475-494.

[17] N. Guan, D. Tao, Z. Luo and B. Yuan, *NeNMF: an optimal gradient method for nonnegative matrix factorization,* IEEE Transactions on Signal Processing, 60 (2012), 2882-2898.

[18] D. D. Lee and H. S. Seung, *Algorithms for non-negative matrix factorization,* Advances in Neural Information Processing Systems, 2001, 556-562.

[19] Y. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $O(1/k^2)$, Soviet Mathematics Doklady, 27 (1983).

[20] F. S. Samaria and A. C. Harter, *Parameterisation of a stochastic model for human face identification,* IEEE Workshop on Application and Computer Vision, Sarasota, FL, 1994, 138-142.

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

Accepted: 2.04.2019