10th International Conference on Information Science and Technology
Bath, United Kingdom, September 9-15, 2020

Weighted Non-negative Matrix Factorization for Image Recovery and Representation

Xiangguang Dai[1], Keke Zhang[1], Jiang Xiong[1], Xianxiu Zhang[1], Zhengwen Tu[1], and Nian Zhang[2,*]
[1]Chongqing Engineering Research Center of Internet of Things and Intelligent Control Technology ,
Chongqing Three Gorges University, Wanzhou, Chongqing, 404100, China
[2]Department of Electrical and Computer Engineering University of the District of Columbia
Washington, D.C., Washington 20008 USA
Emails: daixiangguang@163.com, xk_zhang0924@126.com,
jiangxiong1969cq@163.com, 253549484@qq.com,
zxx1234567@sina.com, tuzhengwen@163.com, nzhang@udc.edu

***Abstract*—Non-negative matrix factorization and its variants cannot learn an effective subspace from the dataset corrupted by outliers. In this article, we propose a robust non-negative matrix factorization approach, called weighted non-negative matrix factorization, which can both recover the corrupted data space and learn a more effective subspace from the corrupted data space. In the proposed method, introducing a weighted graph, which uses Boolean values to mark noise points, while a clean data space and subspace can be achieved by the unlabel data points. The proposed problem can be formulated as a non-convex optimization problem, which can be optimized by the multiplicative update methods. Taking the face data polluted by Salt and Pepper noise as an example, the effectiveness of the proposed method in image recovery and low-dimensional representations learning is verified.**

***Keywords*—robustness; image recovery; representation; non-negative matrix factorization**

I. Introduction

NMF (Non-negative Matrix Factorization) [1] is a well-known dimensionality-based strategy in which the original data matrix is converted into two low-dimensional non-negative matrices including the coefficient matrix (low-dimensional representation) and the basic matrix (partial representation of the original data). In view of its good method, NMF has been widely used in recommendation system [2], cluster analysis [3], semi-supervised learning [4], and other fields.

In recent years, many works have explored excellent nonnegative matrix factorization to deal with outliers in data sets [5]–[13]. Hamza and Brady [5] proposed that the Frobenius norm can substitute for the hypersurface cost function (HCNMF). Compared with NMF, HCNMF has better robust representation. Gao et al. [7] put forward the capped norm as the objective function and utilized a threshold to remove outliers. However, it's difficult to find a method to ascertain an accurate outlier threshold value. Guan et al. [10] came up with a three-sigma-rule to determine outliers and a Truncated Cauchy objective function to process outliers.

The aforementioned non-negative matrix factorization approaches have been used in signal handling [14], clustering [15] and image classification [16]. Nevertheless, they have several shortcomings: 1) Many approaches cannot remove non-Gaussian noise (i.e. Salt and Pepper noise). Therefore, the learned subspace cannot be applied to classification or clustering. 2) These approaches using distinct cost functions presumed that the decomposition error becomes smaller and smaller, it means that representation will be better and better. As far as we know, the optimization algorithms are more efficient complex and consume a lot of energy to realize matrix factorization.

Based on the recent studies, we put up with a significant matrix factorization framework, namely WNMF (weighted non-negative matrix factorization) to resolve the above-mentioned issues, which establishes the relationship between original data and outliers by constructing a weight map. Therefore, WNMF can recover data and learn a robust representation from the damaged data. Since the objective function of WNMF is non-convex, it is transformed into three convex optimization problems and solved in turn until convergence. The main contributions of this paper are as below:

- Recent NMF models are more applicable to deal with Gaussian noise. In this paper, in order to eliminate Salt and Pepper noise, we come up with a weighted non-negative matrix factorization framework.
- In comparison with recent NMF models, our proposed model can achieve data recovery, as well as learn a robust representation. In addition, our proposed algorithm can be implemented easily.

II. Weighted Non-negative Matrix Factorization

Supposed that the corrupted matrix $M \in R^{m\times n}$, the decomposed matrices $W \in R^{m\times r}$ and $H \in R^{r\times n}$ and the error matrix $E \in R^{m\times n}$ are given. Traditional NMF and its variant problems can be reduced to the following problems

$$\min_{W\geq 0, H\geq 0, E} \quad loss(M, E, WH) + \lambda\Omega(E, W, H), \tag{1}$$

where the loss function is to measure the similarity among M, WH and E, λ is a non-negative parameter and the constraint term Ω is to constrain W, H and E. Thus, non-negative matrix factorization can summarized by

$$\min_{W\geq 0, H\geq 0} \parallel M - WH \parallel_F^2, \tag{2}$$

Most of NMF variants can remove Gaussian noise, however, they have some defects: 1) They cannot handle Salt and Pepper noise completely. 2) The related algorithms have a huge complexity such that they cannot learn effective representations from the data space. 3) few NMF variants can recover the corrupted data and achieve a clean subspace. To address above-mentioned problems, we mainly discuss the relation between Salt and Pepper noise and the dataset, and propose a more robust and effective non-negative matrix factorization for data recovery and representation.

A. Problem Formulation

Let $M \in R^{m\times n}$ be the damaged matrix and $V \in R^{m\times n}$ the recovered matrix. The similarity between M and V is measured by

$$\parallel (V - M) \otimes S \parallel_F^2, \tag{3}$$

where the weight matrix S labels the noisy points. Here, we propose S by

$$S_{ij} = \begin{cases} 0, & \text{if } S_{ij} \text{ is corrupted,} \\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

By optimizing (1), (2) and (3), we hope that V can be achieved from the noisy matrix E and the corrupted matrix M. Thus, the clean representation H can be learned from V. Combining (1), (2) and (3) results in weighted non-negative matrix factorization (WNMF).

Supposed that the damaged matrix and the error matrix denote by $M \in R^{m\times n}$ and $E \in R^{m\times n}$. WNMF is to decompose M into three non-negative matrices $W \in R^{m\times r}$, $H \in R^{r\times n}$ and $E \in R^{m\times n}$. In summary, WNMF can be formulated as

$$\begin{aligned} &\min_{W, H\geq 0, E\geq 0} F(W, H, E) \\ &= \parallel V - WH \parallel_F^2 + \lambda \parallel E \otimes S \parallel_F^2, \end{aligned} \tag{5}$$

where $V = M - E$ and λ is the hyper-parameter.

B. Optimization Algorithm

It is obvious that problem (5) is non-convex in simultaneously optimizing W, H and E. Hence, no algorithm can achieve a global solution. We propose a simple optimization scheme to search a local solution of problem (5).

Problem (5) can be transformed into three convex optimization problems. Supposed that the k-th solution of problem (5) is obtained. Thus, the $k+1$-th solution is generated by

$$E^{k+1} = \min_{E} \| M - W^k H^k - E \|_F^2 + \lambda \| E \otimes S \|_F^2 \tag{6}$$

and

$$W^{k+1} = \min_{W \geq 0} \| M - W H^k - E^{k+1} \|_F^2 \tag{7}$$

and

$$H^{k+1} = \min_{H \geq 0} \| M - W^{k+1} H - E^{k+1} \|_F^2 \cdot \tag{8}$$

Alternately solving (6), (7) and (8), the local optimal solution of problem (5) can be achieved. We can get the solution of problems (6), (7) and (8) by

$$E_{ij}^{k+1} = \frac{M_{ij}^k - (W^k H^k)_{ij}}{1 + \lambda S_{ij}}, \tag{9}$$

$$W_{ij}^{k+1} = W_{ij}^k \frac{((M^k - E^k) H^{k^T})_{ij}}{(W^k H^k H^{k^T})_{ij}}, \tag{10}$$

$$H_{ij}^{k+1} = H_{ij}^k \frac{(W^{k^T} (M^k - E^k))_{ij}}{(W^{k^T} W^k H^k)_{ij}}. \tag{11}$$

According to (9), (10) and (11), we can summarize the specific details in Algorithm 1.

Algorithm 1 Weighted Non-negative Matrix Factorization (WNMF)

Require: $W \in R^{m \times r}$, $H \in R^{r \times n}$ and $E \in R^{m \times n}$, $M \in R^{m \times n}$, $iter$

Ensure: V, H

1: λ and S by (4)
2: **for** $k = 0$ to $iter$ **do**
3: $E_{ij}^{k+1} = \frac{M_{ij}^k - (W^k H^k)_{ij}}{1 + \lambda S_{ij}}$
4: $W_{ij}^{k+1} = W_{ij}^k \frac{((M^k - E^k) H^{k^T})_{ij}}{(W^k H^k H^{k^T})_{ij}}$
5: $H_{ij}^{k+1} = H_{ij}^k \frac{(W^{k^T} (M^k - E^k))_{ij}}{(W^{k^T} W^k H^k)_{ij}}$
6: **end for**
7: $V = WH$

III. Experiment results

The recovery and the clustering performances are explored on two face datasets (i.e. ORL and YALE) and compared it with CauchyNMF [10], RNMF [7], MahNMF [6] and NMF [1]. We propose two face datasets (i.e. ORL and YALE) and Salt and Pepper noise to validate the effectiveness of the above-mentioned NMF methods.

Salt and Pepper noise contaminates a few pixels of each image to be 0 or 255. We propose four level percentage of Salt and Pepper noise from p =10% to 25% with the step size 5% to corrupt each image. To validate the effectiveness of all NMF methods, we utilize Peak Signal-to-Noise Ratio (PSNR) and Accuracy (AC) and Normalized Mutual Information (NMI) [17] to test the recovery effect and the clustering performance.

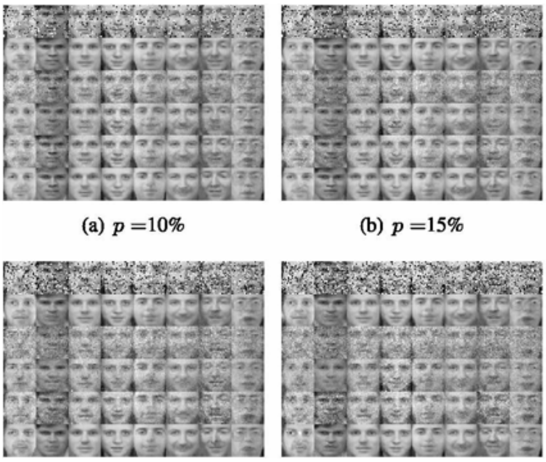
A. Visualization of Recovered Faces

Supposing $\lambda = 100$ and $iter = 200$, the recovered faces achieved by different NMF methods are presented in Figure 1 and 2. To validate whether all NMF methods achieve satisfactory face recovery, we list PSNRS achieved by differen NMF algorithms in Table I and II. From these comparisons, we can observe

- When p =10%, all NMF methods achieve face recovery from Salt and Pepper noise. As p varies, the PSNRs denote that NMF achieves the worse recovery effects than other NMF methods. WNMF and CauthyNMF can achieve face recovery. Above all, NMF cannot handle Salt and Pepper noise. WNMF and CauthyNMF can handle Salt and Pepper noise.
- WNMF obtains the highest PSNRs than other NMF methods. Although RNMF and CauthyNMF achieve satisfactory performances in the beginning, they perform worse as p varies. It is obvious that WNMF achieves the smallest objective value than other NMF methods.

B. Clustering Performances

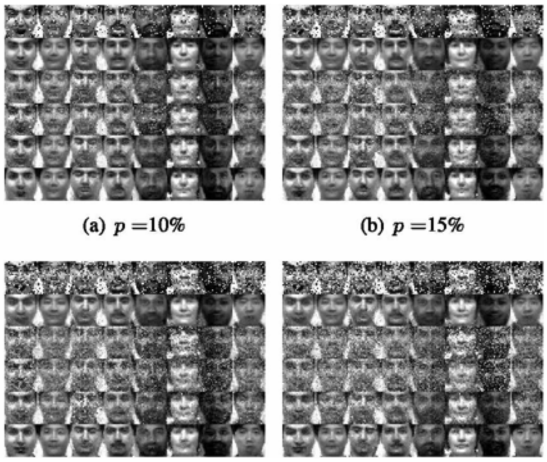
Figure 3 presents the clustering results on ORL corrupted by Salt and Pepper noise. According to the clustering results, we conclude that



(a) p =10% (b) p =15%

(c) p =20% (d) p =25%

Fig. 1: Recovered faces from ORL with Salt and Pepper noise. For each image, the images in the first row is sample images with different corrupted percentages, and the images in the last five rows are the recovered faces by WNMF, NMF, MahNMF, RNMF and CauchyNMF.



(a) p =10% (b) p =15%

(c) p =20% (d) p =25%

Fig. 2: Recovered faces from YALE with Salt and Pepper noise. For each image, the images in the first row is sample images with different corrupted percentages, and the images in the last five rows are the recovered faces by WNMF, NMF, MahNMF, RNMF and CauchyNMF.

TABLE I: PSNRs on ORL with Salt and Pepper.

$p(\%)$	10	15	20	25
WNMF	25.28	25.17	25.05	24.89
NMF	20.83	19.44	18.38	17.55
MahNMF	22.82	21.68	20.79	19.98
RNMF	23.47	22.32	21.16	20.04
CauchyNMF	23.76	23.8	23.93	24.09

TABLE II: PSNRs on YALE with Salt and Pepper.

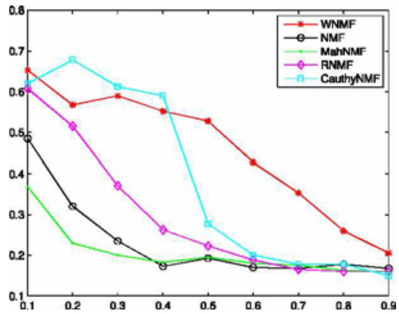
$p(\%)$	10	15	20	25
WNMF	20.28	20.38	20.17	19.92
NMF	17.82	16.41	15.31	14.50
MahNMF	17.73	16.34	15.31	14.38
RNMF	20.26	19.19	18.17	16.96
CauchyNMF	18.67	18.65	18.62	18.30

- WNMF and CauthyNMF can achieve the better clustering ACs and NMIs than other NMF methods, which indicates they can learn an effective subspace when ORL is corrupted outliers.
- When p is smaller, CauthyNMF achieves satisfactory clustering results. However, CauthyNMF achieves poor performances as p varies.
- WNMF has the better clustering performances as p varies. In the other word, WNMF can not only remove outliers in the subspace, but also learn better presentations than other NMF methods.

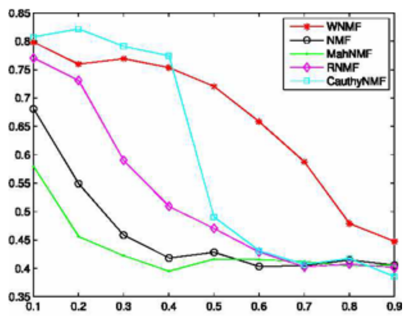
IV. Conclusion and Future Work

In this paper, we proposed a weighted NMF method (WNMF) to learn a more robust subspace from the data space contaminated by Salt and Pepper noise. The main advantages of WNMF are as follows: (1) WNMF can remove Salt and Pepper noise more effectively than other NMF methods. (2) WNMF can not only achieve a cleaner data space, but also learn a more robust subspace when the data space is corrupted by outliers. Two questions are proposed to be discussed as follows:

- Some other noises (i.e. Gaussian noise, Poisson noise, etc) can be considered into NMF.



(a) Clustering ACs



(b) Clustering NMIs

Fig. 3: Clustering ACs and NMIs when p varies from 10% to 90%.

Therefore, the relation between the noise distribution and NMF should be discussed.

- WNMF is a unsupervised problem and cannot be utilized to solve semi-supervised problems. Hence, manifold learning or label information should be considered into WNMF.

ACKNOWLEDGMENT

Our work is sponsored by National Science Foundation (NSF) grant #1505509 and DoD grant #W911NF1810475, Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant Nos. KJQN201901218 and KJ1710248) and Natural Science Foundation of Chongqing (Grant No. cstc2019jcyj-bshX0101).

REFERENCES

[1] D. D. Lee and H. S. Seung, Learning the Parts of Objects by Non-negative Matrix Factorization, *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] H. Li, K. Li, J. An, W. Zhang and K. Li, An Efficient Manifold Regularized Sparse Non-negative Matrix Factorization Model for Large-scale Recommender Systems on GPUs, *Information Sciences*, DOI: 10.1016/j.ins.2018.07.060,2018.

[3] W. Wu, Y. Jia, S. Kwong and J. Hou, Pairwise Constraint Propagation-Induced Symmetric Nonnegative Matrix Factorization, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 12, pp. 6348–6361, 2018.

[4] X. Peng, D. Chen and D. Xu, Hyperplane-Based Nonnegative Matrix Factorization with Label Information, *Information Sciences*, vol. 493, pp. 1–9, 2019.

[5] A. B. Hamza and D. J. Brady, Reconstruction of Reflectance Spectra Using Robust Nonnegative Matrix Factorization, *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 37637–3642, 2006.

[6] N. Guan, D. Tao, Z. Luo and J. Shawetaylor, MahNMF: Manhattan Non-negative Matrix Factorization, *Journal of Machine Learning Research*, arXiv:1207.3438v1, 2012.

[7] H. Gao, F. Nie, W. Cai and H. Huang, Robust Capped Norm Nonnegative Matrix Factorization, *ACM International on Conference on Information and Knowledge Management*, pp. 871–880, 2015.

[8] L. Du, X. Li and Y. Shen, Robust Nonnegative Matrix Factorization via Half-Quadratic Minimization, *IEEE International Conference on Data Mining*, pp. 201–210, 2012.

[9] L. Zhang, Z. Chen, M. Zheng and X. He, Robust Nonnegative Matrix Factorization, *Frontiers of Electrical and Electronic Engineering in China*, vol. 6, no. 2, pp. 192–200, 2015.

[10] N. Guan, T. Liu, Y. Zhang, D. Tao and L. Davis, Truncated Cauchy Non-negative Matrix Factorization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 246–259, 2018.

[11] Q. Gu, and J. Zhou, Local Learning Regularized Nonnegative Matrix Factorization, *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009. Morgan Kaufmann Publishers Inc*, 2009.

[12] C. Peng, Z. Kang, C. Chen and Q. Cheng, Nonnegative Matrix Factorization with Local Similarity Learning, *arXiv preprint arXiv:1907.04150*, 2019, doi:10.1109/tcyb.2018.2887094.

[13] Z. Kang, H. Pan, S. Hoi and Z. Xu, Robust graph learning from noisy data, *IEEE transactions on cybernetics*, 2019.

[14] J. Zhou, R. Liang , L. Zhao, L. Tao and C. Zou, Unsupervised Learning of Phonemes of Whispered Speech in a Noisy Environment Based on Convolutive Non-negative Matrix Factorization, *Information Sciences*, vol. 257, pp. 115–126, 2014.

[15] Y. Lu, Z. Lai, Y. Xu, J. You, X. Li and C. Yuan, Projective Robust Nonnegative Factorization, *Information Sciences*, vol. 364, pp. 16-32, 2016.

[16] Z. Li, J. Tang and X. He, Robust Structured Nonnegative Matrix Factorization for Image Representation, *IEEE*

Transactions on Neural Networks and Learning Systems, vol. 29, no. 5, pp. 1947–1960, 2018.
[17] D. Cai, X. He and J. Han, Document Clustering Using Locality Preserving Indexing, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 12, pp. 1624–1637, 2005.