International Journal of Computational Intelligence Systems
Vol. 13(1), 2020, pp. 734–743
DOI: https://doi.org/10.2991/ijcis.d.200527.003; ISSN: 1875-6891; eISSN: 1875-6883
https://www.atlantis-press.com/journals/ijcis/

Special Issue

Weighted Nonnegative Matrix Factorization for Image Inpainting and Clustering

Xiangguang Dai[1,*], Nian Zhang[2], Keke Zhang[1], Jiang Xiong[1]

[1] Key Laboratory of Intelligent Information Processing and Control of Chongqing Municipal Institutions of Higher Education, Chongqing Three Gorges University, Chongqing 40044, China

[2] Department of Electrical and Computer Engineering, University of the District of Columbia, Washington, DC 20008, USA

ARTICLE INFO

Article History

Received 29 Feb 2020
Accepted 18 May 2020

Keywords

Recovery
Dimensionality reduction
Weighted nonnegative matrix factorization
Noise

ABSTRACT

Conventional nonnegative matrix factorization and its variants cannot separate the noise data space into a clean space and learn an effective low-dimensional subspace from Salt and Pepper noise or Contiguous Occlusion. This paper proposes a weighted nonnegative matrix factorization (WNMF) to improve the robustness of existing nonnegative matrix factorization. In WNMF, a weighted graph is constructed to label the uncorrupted data as 1 and the corrupted data as 0, and an effective matrix factorization model is proposed to recover the noise data and achieve clustering from the recovered data. Extensive experiments on the image datasets corrupted by Salt and Pepper noise or Contiguous Occlusion are presented to demonstrate the effectiveness and robustness of the proposed method in image inpainting and clustering.

1. INTRODUCTION

Non-negative matrix factorization (NMF) [1] is a popular dimensionality reduction method, which decomposes an original data matrix into two low-dimensional nonnegative matrices. Among the decomposed matrices, one is coefficient matrix to store a low-dimensional representation, and the other is a basis matrix which can be regarded as parts-based representations of the original data. Owing to the excellent presentation approach, NMF has been widely applied to clustering [2,3], recommender system [4], community detection [5], semi-supervised learning [6], and so on.

Recently, most of studies were seeking an effective NMF to handle outliers and noise in the dataset [7–18]. Hamza and Brady [7] was the first to replace the Frobenius norm with the hypersurface cost function (HCNMF). The main contribution of HCNMF is that it can achieve a more robust representation than NMF. Kong *et al.* [8] presented the $L_{2,1}$-norm ($L_{2,1}$NMF) as the cost function to remove outliers and noise. $L_{2,1}$NMF is less sensitive to outliers and noise than NMF and HCNMF, however, the related algorithm takes much time achieving factorization because of the nonsmooth loss function. Gao *et al.* [10] proposed a capped norm NMF to handle outliers by the outlier threshold, however, there is no approach to determine an exact outlier threshold value. Guan *et al.* [13] proposed the three-sigma-rule to detect outliers and a Truncated Cauchy loss (CauchyNMF) to handle outliers.

The above robust NMF variants have been utilized in signal processing [19], image processing [20], clustering [21] and image classification [22]. However, they have the following defects: (1) Most of methods can neither handle Salt and Pepper noise nor Contiguous Occlusion. In this case, the learned subspace is not suitable for clustering or classification. (2) Robust NMF methods using different loss functions supposed that the smaller factorization error and the better representation are achieved by a suitable loss function. To our knowledge, the proposed algorithms optimizing these loss functions are more complicated and take much time to complete matrix factorization.

Motivated by recent work, we propose an effective matrix decomposition framework, called weighted non-negative matrix factorization (WNMF) to overcome the abovementioned problems, which constructs the weighted graph to build the relation between the original data and outliers. Thus, WNMF can recover the corrupted data and achieve robust clustering. Because the objective function of WNMF is nonconvex, we propose an iterative algorithm to solve it and prove the convergence of the proposed optimization scheme. The main contributions of this paper can be summarized as follows:

- We propose a WNMF framework to handle outliers and noise, and we explain why the proposed model is effective and robust.
- Our proposed model can achieve data recovery and clustering from the original data corrupted by Salt and Pepper noise or Contiguous Occlusion.

* Corresponding author. Email: daixiangguang@163.com

2. RELATED WORKS

Suppose that the input data matrix $M = [m_{ij}] \in R^{m\times n}$, the decomposed matrices $W = [w_{il}] \in R^{m\times r}$, $H = h_{lj} \in R^{r\times n}$ and the noise error matrix $E = [e_{ij}] \in R^{m\times n}$ are given. Thus, existing robust NMF frameworks can be formulated into the following optimization problem:

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

where the first term is the criterion of the loss function to measure the approximation error, the second term is the constraint term on W, H or E, or both, and E and λ are the noise matrix and a tradeoff parameter. Standard NMF is to decompose a nonnegative matrix $M \in R^{m\times n}$ into two low-dimensional matrices $W^{m\times r}$ and $H^{r\times n}$. Generally, NMF utilizes the Frobenius norm as the cost function to measure the factorization error. Thus, standard NMF can be formulated by

$$\min_{W,H} \| M - WH \|_F^2 \qquad (2)$$
$$s.t. \quad W \geq 0, H \geq 0.$$

In [12], Zhang *et al.* proposed a robust model (RNMF)to handle outliers and noise as follows:

$$\min_{W,H,E} \| M - WH - E \|_F^2 + \lambda \| E \|_M \qquad (3)$$
$$s.t. \quad W \geq 0, H \geq 0.$$

Guan *et al.* [9] proposed Manhattan distance (MahNMF) to be the criterion of loss function. MahNMF can reduce the approximation error, which can be summarized into the following optimization problem:

$$\min_{W,H} \| M - WH \|_M \qquad (4)$$
$$s.t. \quad W \geq 0, H \geq 0.$$

Guan *et al.* [13] proposed the three-sigma-rule to detect outliers and a Truncated Cauchy loss (CauchyNMF) to remove outliers. CauchyNMF can be summarized as follows:

$$\min_{W\geq 0, H\geq 0} F(W,H) = \sum_{i=1}^{m}\sum_{j=1}^{n} g\left(\frac{(V-WH)_{ij}}{\gamma}\right), \qquad (5)$$

where $g(x) = \begin{cases} ln(1+x), & 0 \leq x \leq \sigma \\ ln(1+\sigma), & x > \sigma \end{cases}$; σ and γ denote the scale parameter and the truncation parameter. σ can be obtained by three-sigma-rule, and γ is given by the Nagy algorithm [13].

3. WEIGHTED NONNEGATIVE MATRIX FACTORIZATION

3.1. Model Formulation

Existing robust models have the following properties: (1) They can easily handle Gaussian noise, however, they fail to remove Salt and Pepper noise and Contiguous Occlusion. (2) The proposed algorithms of some robust models (e.g., RMahNMF and CauchyNMF) are too complicated to learn a robust low-dimensional subspace from the high-dimensional data. (3) Only RNMF can achieve data recovery and representation simultaneously. In the following, we investigate the relation between the noise distribution and the corrupted data, and propose a robust weighted NMF to achieve a clean data space and a robust low-dimensional representation from the corrupted data.

Suppose that $M_i \in R^m$ and $V_i \in R^m$ are the corrupted feature vector and the recovered feature vector, separately. The approximation error between $M = [M_1, \cdots, M_n] \in R^{m\times n}$ and $V = [V_1, \cdots, V_n] \in R^{m\times n}$ can be formulated as the following optimization problem:

$$\| (V - M) \otimes S \|_F^2, \qquad (6)$$

where S is a weighted matrix that denotes the contaminated or uncontaminated position that can be defined by

$$S_{ij} = \begin{cases} 0, & \text{if } (i,j) \in \Omega, \\ 1, & \text{otherwise}, \end{cases} \qquad (7)$$

where Ω is the corrupted area. Supposing $E = V - M$, we conclude that

$$\| E \otimes S \|_F^2 . \qquad (8)$$

By minimizing (8), (2) and (3), we expect that if the recovered data matrix is obtained from the corrupted image matrix M and the noise matrix E, and effective low-dimensional representation H will also be learned from the recovered data matrix. Combining (8), (2) and (3) results in our WNMF.

Given a corrupted data matrix $M \in R^{m\times n}$, WNMF aims to find three nonnegative matrices $E \in R^{m\times n}$, $W \in R^{m\times r}$ and $H \in R^{r\times n}$. Thus, WNMF can be described as the following optimization problem:

$$\min_{W,H,E} F(W,H,E) \qquad (9)$$
$$= \| M - WH - E \|_F^2 + \lambda \| E \otimes S \|_F^2$$
$$s.t. \quad W \geq 0, H \geq 0.$$

where the hyper-parameter λ can be utilized to balance the contribution from each term. Suppose that the entries of S are all zeros. This phenomenon denotes the proposed model (9) can be simplified to standard NMF in (2).

3.2. Robustness Analysis

In this subsection, we compare the robustness of RMahNMF with existing robust NMF models (e.g., NMF [1], MahNMF [9], RNMF [12] and CauchyNMF [13]) by utilizing a simple-weighted procedure. A robust NMF algorithm should produce a small weight to an entry of the training sample with large noise. We present some assumptions as follows:

- $F(WH)$ is a objective function and $f(t) = F(tWH)$.
- $f'(t)$ is the derivative of $f(t)$.
- $c(M_{ij}, WH) = (M - WH)_{ij}(-WH)_{ij}$ is the contribution of the j-th entry of the i-th sample to the optimization procedure.

- $e(M_{ij}, WH) = |M - WH|_{ij}$ represents the noise error to the (i,j)-th entry of M.

Thus, we should find the $f'(1)$ such that $f'(1) = 0$. We compare the robustness between WNMF and other four competing models in Table 1. According to the comparison results, we make the following statements: (1) Because NMF has constant weights, it is more sensitive to noise than other models. (2) RNMF is less sensitive to outliers and noise than MahNMF because RNMF utilizes the error noise matrix to adjust the weights. (3) The weight of CauchyNMF can drop to zero when the noise value is larger than a threshold. Thus, CauchyNMF is more robust than RNMF and MahNMF. (4) Combining NMF and RNMF, our proposed WNMF can not only drop to zero but also has different weighted strategies. Therefore, WNMF is more effective and robust than any other NMF models.

4. OPTIMIZATION ALGORITHM

Since problem (9) is nonconvex in optimizing W, H and E simultaneously, it cannot search the global optimal solution. Suppose that the solutions of W^k, H^k and E^k are obtained. We solve the following convex problems:

$$\begin{aligned} E^{k+1} &= \arg\min_E \parallel M - W^kH^k - E \parallel_F^2 \\ &+\lambda \parallel E \otimes S \parallel_F^2 \end{aligned} \tag{10}$$

and

$$\begin{aligned} &W^{k+1} = \arg\min_W \parallel M - WH^k - E^{k+1} \parallel_F^2 \\ &s.t. \quad W \geq 0 \end{aligned} \tag{11}$$

and

$$\begin{aligned} &H^{k+1} = \arg\min_H \parallel M - W^{k+1}H - E^{k+1} \parallel_F^2 \\ &s.t. \quad H \geq 0, \end{aligned} \tag{12}$$

until convergence. Thus, the local solution of (9) can be obtained. Most of NMF algorithms obey this optimization scheme. Based on this structure, we introduce the gradient method and KKT conditions to solve (9).

We first discuss transform the objective function of (9) as follows:

$$\begin{aligned} &F(M, W, H) \\ =\ &tr((M - WH - E)(M - WH - E)^T) \\ &+\lambda tr((E \otimes S)(E \otimes S)^T) \\ =\ &tr(M^TM) - 2tr(H^TW^TM) - 2tr(E^TM) \\ &+2tr(H^TW^TE) + tr(E^TE) + tr(H^TW^TWH) \\ &+\lambda tr((E \otimes S)(E \otimes S)^T), \end{aligned} \tag{13}$$

where tr is the trace of a matrix. Suppose that $\Psi = [\psi_{il}]$ and $\Phi = [\phi_{lj}]$ are the lagrange multiplier for the constraint of W and H, respectively. Thus, the Lagrange function is

$$\begin{aligned} &L(E, W, H) \\ =\ &tr(M^TM) - 2tr(H^TW^TM) - 2tr(E^TM) \\ &+2tr(H^TW^TE) + tr(E^TE) + tr(H^TW^TWH) \\ &+\lambda tr((E \otimes S)(E \otimes S)^T) + tr(\Psi W^T) + tr(\Phi H^T) \end{aligned} \tag{14}$$

The partial derivatives of $L(V, W, H)$ with respect to V, W and H are

$$\frac{\partial L}{\partial E} = 2M - 2WH - 2\lambda E \otimes S \tag{15}$$

$$\frac{\partial L}{\partial W} = -2MH^T + 2EH^T + 2WHH^T + \Psi \tag{16}$$

$$\frac{\partial L}{\partial H} = -2W^TM + 2W^TE + 2W^TWH + \Phi. \tag{17}$$

The gradient method and KKT conditions are utilized to solve (15), (16) and (17). Based on the gradient method, the solution of (15) can be obtained

$$e_{ij} \leftarrow \frac{m_{ij} - (WH)_{ij}}{1 + \lambda s_{ij}}. \tag{18}$$

Table 1 Robustness comparison results between WNMF and other NMF models.

NMF Methods	Objective Function $F(WH)$	Derivative $f'(1)$
NMF	$\parallel M - WH \parallel_F^2$	$\sum_{ij} 2c(M_{i,j}, WH)$
MahNMF	$\parallel M - WH \parallel_M$	$\sum_{ij} \frac{1}{\lvert M-WH\rvert_{ij}} c(M_{i,j}, WH)$
RNMF	$\parallel M - E - WH \parallel_F^2 + \parallel E \parallel_M$	$\sum_{ij} 2\left(1 - \frac{E_{ij}}{(M-WH)_{ij}}\right) c(M_{i,j}, WH)$
CauchyNMF	$\sum_{ij} g\left(\left(\frac{M-WH}{\gamma}\right)^2_{ij}\right) g(x) = \begin{cases} ln(1+x), 0 \leq x \leq \sigma \\ ln(1+\sigma), x > \sigma \end{cases}$	$\sum_{ij} \begin{cases} \frac{2}{\gamma^2 + (M-WH)^2_{ij}} c(M_{i,j}, WH), \lvert M - WH\rvert_{ij} \leq \gamma\sqrt{\sigma} \\ 0.c(M_{i,j}, WH), \text{otherwise} \end{cases}$
WNMF	$\parallel M - WH - E \parallel_F^2 + \lambda \parallel E \otimes S \parallel_F^2$	$\sum_{ij} \begin{cases} 0.c(M_{ij}, WH), S_{ij} = 0, E_{ij} = (M - WH)_{ij}, \\ 0.c(M_{ij}, WH), S_{ij} = 1, E_{ij} = (M - WH)_{ij}, \\ 2c(M_{ij}, WH), S_{ij} = 1, E_{ij} = 0. \end{cases}$

Note: WNMF, weighted nonnegative matrix factorization; NMF, nonnegative matrix factorization.

Based on the KKT conditions $\psi_{il} w_{il} = 0$ and $\phi h_{lj} h_{lj} = 0$, we can obtain the following equations:

$$(-(MH^T)_{il} + (EH^T)_{il} + (WHH^T)_{il}) w_{il} = 0 \quad (19)$$

$$(-(W^T M)_{lj} + (W^T E)_{lj} + (W^T WH)_{lj}) h_{lj} = 0. \quad (20)$$

Equations (19) and (20) can lead to the solution of (16) and (17) as follows:

$$w_{il} \leftarrow w_{il} \frac{(MH^T)_{il} - (EH^T)_{il}}{(WHH^T)_{il}}, \quad (21)$$

$$h_{lj} \leftarrow h_{lj} \frac{(W^T M)_{lj} - (W^T E)_{lj}}{(W^T WH)_{lj}}. \quad (22)$$

According to above analysis, we summarize the update rules (18), (21) and (22) in Algorithm 1.

The convergence condition of Algorithm 1 can be summarized as follows:

$$\frac{|F(W^{k+1}, H^{k+1}, E^{k+1}) - F(W^k, H^k, E^k)|}{F(W^{k+1}, H^{k+1}, E^{k+1})} \leq \epsilon, \quad (23)$$

where the precision ϵ can be set as $1e^{-3}, 1e^{-4}, 1e^{-5}$ or $1e^{-6}$.

5. CONVERGENCE PROOFS

Definition 1. [23] Suppose that $G(x, x')$ is defined to be an auxiliary function for the objective function $F(x)$. Thus, the auxiliary function should satisfy the following conditions:

$$G(x, x') \geq F(x), G(x, x) = F(x) \quad (24)$$

Lemma 1. [23] Let $G(x, x')$ be an auxiliary function of $F(x)$. $F(x)$ is nonincreasing under the update

$$x^{t+1} = \arg\min G(x, x^t), \quad (25)$$

where x^t is the t-th solution of $F(x)$.

Algorithm 1: Weighted Nonnegative Matrix Factorization

Require: $M \in R^{m \times n}$
Ensure: $E \in R^{m \times n}$, $W \in R^{m \times r}$, $H \in R^{r \times n}$
1: Initializing $k = 0, \lambda \geq 0$, $W_{ij}^0 \in (0, 1]$, $H_{ij}^0 \in (0, 1]$, $E_{ij}^0 \in [-1, 1]$ and S by (7)
2: **while** true **do**
3: $E_{ij}^{k+1} = \frac{M_{ij}^k - (W^k H^k)_{ij}}{1 + \lambda S_{ij}}$
4: $W_{il}^{k+1} = W_{il}^k \frac{((M^k - E^k) H^{k\,T})_{il}}{(W^k H^k H^{k\,T})_{il}}$
5: $H_{lj}^{k+1} = H_{lj}^k \frac{(W^{k\,T} (M^k - E^k))_{lj}}{(W^{k\,T} W^k H^k)_{lj}}$
6: Check convergence
7: $k = k + 1$
8: **end while**
9: $\hat{M} = WH$

Lemma 2. [23] The following function

$$\begin{aligned} G(h, h_{ab}^t) = & F_{ab}(h_{ab}^t) + F'_{ab}(h_{ab}^t)(h - h_{ab}^t) \\ & + \frac{(W^T WH)_{ab}}{h_{ab}^t} (h - h_{ab}^t)^2 \end{aligned} \quad (26)$$

is an auxiliary function of (12).

Proof. It is obvious that $G(h, h) = F_{ab}(h)$. We should prove that $G(h, h_{ab}^t) \geq F_{ab}(h)$. Thus, we utilize the Taylor series expansion of $F_{ab}(h)$ as follows:

$$\begin{aligned} F_{ab}(h) = & F_{ab}(h_{ab}^t) + F'_{ab}(h_{ab}^t)(h - h_{ab}^t) \\ & + (W^T W)_{aa} (h - h_{ab}^t)^2. \end{aligned} \quad (27)$$

According to Definition 1, $G(h, h_{ab}^t) \geq F_{ab}(h)$ is equivalent to the following inequation:

$$\frac{(W^T WH)_{ab}}{h_{ab}^t} \geq (W^T W)_{aa}. \quad (28)$$

We can obtain

$$(W^T WH)_{ab} = \sum_{l=1}^{r} (W^T W)_{al} h_{lb}^t \geq (W^T W)_{aa} h_{ab}^t. \quad (29)$$

Therefore, (28) holds and $G(h, h_{ab}^t) \geq F_{ab}(h)$.

Theorem 1. *w_{il} and h_{lj} under the update rules (21) and (22) are nonnegative.*

Proof. Suppose that the k-th iteration W^k and H^k are nonnegative. In the update rule (21), if $M - E > 0$ holds, then w_{il} under the update rule (21) is nonnegative. Substituting (18) into (21), we can obtain that

$$\begin{aligned} m_{ij} - e_{ij} & = m_{ij} - \frac{m_{ij} - (WH)_{ij}}{1 + \lambda s_{ij}} \\ & = \frac{m_{ij} \lambda S_{ij} + (WH)_{ij}}{1 + \lambda S_{ij}} \geq 0. \end{aligned} \quad (30)$$

Similarly, h_{lj} under the update rules (22) is nonnegative.

Theorem 2. *The objective function in (9) is nonincreasing with the abovementioned update rules (18), (21) and (22). $F(W, H, E)$ is invariant under these updates if and only if E, W and H are at a stationary point.*

Proof. The update rule in (18) is obtained by the gradient method. Obviously, $F(E)$ is nonincreasing under the update rule. In the following, we need prove that $F(W)$ and $F(H)$ are nonincreasing under the update rules (21) and (22). According to Lemmas 1 and 2, we conclude that

$$\begin{aligned} h_{ab}^{t+1} & = h_{ab}^t - h_{ab}^t \frac{F'_{ab}(h_{ab}^t)}{2(W^T WH)_{ab}} \\ & = h_{ab}^t \frac{(W^T M)_{ab} - (W^T E)_{ab}}{(W^T WH)_{ab}}. \end{aligned} \quad (31)$$

Therefore, $F(H)$ is nonincreasing for the update rule (22). Fortunately, the objective functions of $F(W)$ and $F(H)$ are symmetric. By reversing H in Lemmas 1 and 2, $F(W)$ can accordingly proved to be nonincreasing under the update rule (21). According to above proofs, $F(W, H, E)$ is nonincreasing under the prosed update rules.

6. EXPERIMENTAL RESULTS

We explore the recovery and the clustering performance of WNMF on the ORL and YALE face dataset and compare it with four NMF models (i.e., NMF [1], RNMF [10], MahNMF [9] and CauchyNMF [13]). In the experiments, Salt and Pepper noise and Contiguous Occlusion are proposed to evaluate the effectiveness and robustness of the abovementioned NMF models.

Salt and Pepper noise randomly generates a portion of white and black pixels. To test the recovery effect of WNMF, we propose four level percentages of corrupted pixels (i.e., $p = 10\%$, 15%, 20% and 25%). To demonstrate the clustering performance of WNMF, we vary the corrupted percentage from 10% to 90%. Contiguous Occlusion generates a $b \times b$ size pixel block in each image and the block is filled with the pixel value 255. We propose four block sizes b (i.e., b = 10, 12, 14 and 16) to test the recovery effect and vary the block size from 1 to 20 to demonstrate the clustering performance.

The ORL dataset includes 400 face images of different 40 individuals. There are 10 images of each person with different facial expressions, facial details (without-glasses or with-glasses) and lighting. Each image is a 32×32 pixel grayscale array and it can be normalized to a vector. The YALE dataset contains 165 face images of various 15 persons. There are 11 32×32 pixel images of each person with different facial expression or configuration (i.e., center-light, without-glasses or with-glasses, happy, left-light or right-light, sad, sleepy, surprised and wink).

To demonstrate the recovery effects and the clustering performances of all NMF models, we propose two metrics as follows:

- Peak Signal-to-Noise Ratio (PSNR) is proposed to evaluate the recovery effect, which can be defined by

$$PSNR = 20log_{10}\frac{255}{\sqrt{Error}}, \qquad (32)$$

  where $Error = \frac{1}{m \times n} \parallel M - \hat{M} \parallel_F^2$ and $\hat{M} = WH$.
- Accuracy (AC) and Normalized Mutual Information (NMI) [24] are proposed to test the clustering effect. Due to the nonconvexity of all NMF models, 30 random initial W and H are proposed and the average ACs and NMIs are reported.

6.1. Parameter Selection

Our WNMF model has one essential parameter λ. In this subsection, we investigate how to choose a suitable λ when the input data is corrupted by Salt and Pepper noise or Contiguous Occlusion. Let $p = 0.25$, $b = 12$, $r = 50$ and $\epsilon = 1e^{-3}$. The PSNRs from the ORL and YALE datasets are presented in Figure 1. We can observe (1) two datasets are proposed to evaluate PSNRs. Obviously, λ is irrelevant to the datasets. (2) The PSNRs are mainly affected by λ, and the smaller λ leads to the worse PSNRS. Therefore, in the following experiments, we suppose that $\lambda = 100$.

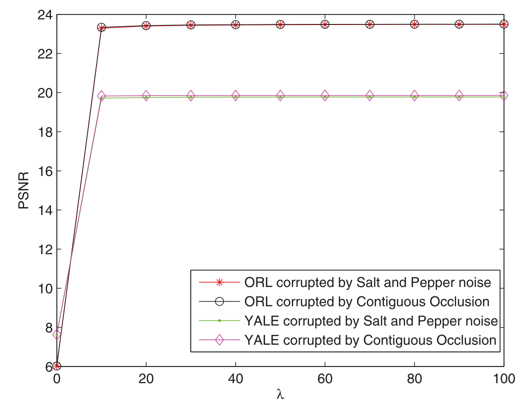


Figure 1 | The peak signal-to-noise ratios (PSNRs) of weighted nonnegative matrix factorization (WNMF) vs. parameter varies from 10 to 100.

6.2. Salt and Pepper Noise

6.2.1. Visualization of recovered faces

Recovered face images of the ORL and YALE datasets corrupted by Salt and Pepper noise are shown in Figure 2. The PSNRs between the face images contaminated by Salt and Pepper noise and the recovered face images are presented in Table 2. From the comparisons, we observe that

- Traditional NMF achieves the smallest PSNRs and the worse recovery performances than other NMF models. Therefore, NMF is more sensitive to Salt and Pepper noise. For the small corrupted percentage (i.e., $p = 10\%$), all NMF models have satisfactory face recovery performances. However, only WNMF and CauthyNMF achieve face recovery as the corrupted percentage varies. These phenomenons denote that WNMF and CauthyNMF can remove Salt and Pepper noise.
- According to comparisons of PSNRs, WNMF remains the highest PSNRs. CauthyNMF and RNMF perform satisfactorily in the beginning, however, they slow down as the corrupted percentage varies. In summary, WNMF can achieve the smallest factorization error than other NMF models.

6.2.2. Clustering

Figure 3 shows the clustering performances on the ORL and YALE datasets contaminated by Salt and Pepper noise. From the comparisons, the interesting observations are

- WNMF and CauthyNMF have the better clustering results, which indicates they can learn a more robust subspace for clustering.
- CauthyNMF performs satisfactorily in the beginning, however, it achieves poor performances as the corrupted percentage increases.

(a) p =10%

(b) p =15%

(c) p =20%

(d) p =25%

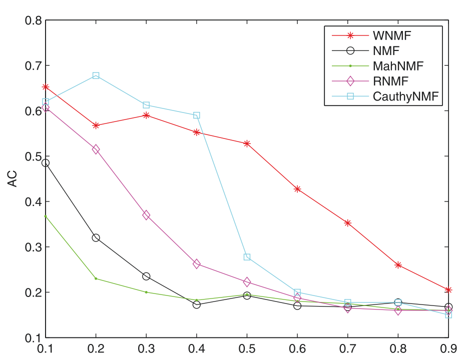
(e) p =10%

(f) p =15%

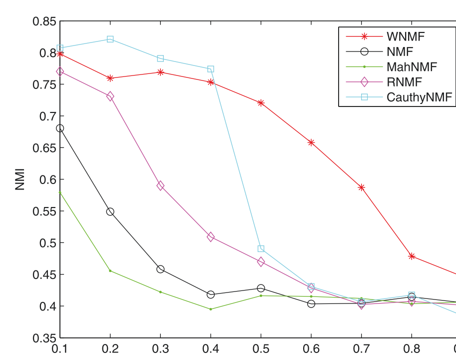
(g) p =20%

(h) p =25%

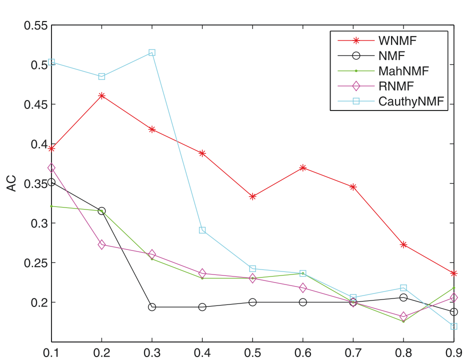
Figure 2 Recovered images from the ORL and YALE datasets corrupted by Salt and Pepper noise. For (a)–(h), the first row is sample images under the corrupted percentage p, and the last five rows are recovered images by weighted nonnegative matrix factorization (WNMF), NMF, MahNMF, RNMF and CauchyNMF.



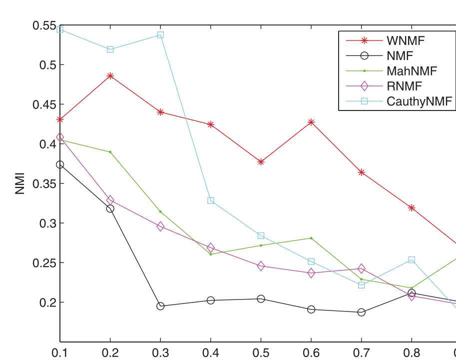
(a) Average ACs of the ORL dataset vs. the corrupted percentage



(b) Average NMIs of the ORL dataset vs. the corrupted percentage



(c) Average ACs of the YALE dataset vs. the corrupted percentage



(d) Average NMIs of the YALE dataset vs. the corrupted percentage

Figure 3 Evaluation on the ORL and YALE databases contaminated by Salt and Pepper noise.

- WNMF achieves relatively stable clustering results on the Salt and Pepper noise, that is to say, WNMF can hardly be affected by the outliers. When p becomes larger, WNMF still achieves a good performance.
- All the clustering results indicate that WNMF can learn a better subspace on the ORL dataset contaminated by Salt and Pepper noise.

6.3. Contiguous Occlusion

6.3.1. Visualization of recovered faces

Figure 4 and Table 3 present the recovery faces and the PSNRS of the ORL and YALE datasets contaminated by Contiguous Occlusion. According to experimental results, we observe that

- WNMF can achieve face recovery completely as the block size varies. CauchyNMF can recover some corrupted faces in the smaller block size (i.e., $b = 10$), but achieves worse performances in the end.
- As the block size increases, WNMF remains the highest PSNRs than other algorithms. CauchyNMF can only achieves satisfactory PSNRs when the block size is small enough.
- WNMF and CauchyNMF can handle Contiguous Occlusion. WNMF can completely handle Contiguous Occlusion, but CauchyNMF can remove Contiguous Occlusion when the corrupted region is very small.

Table 2 PSNRs on the ORL and YALE datasets contaminated by Salt and Pepper with different corrupted percentages from 10% to 25%.

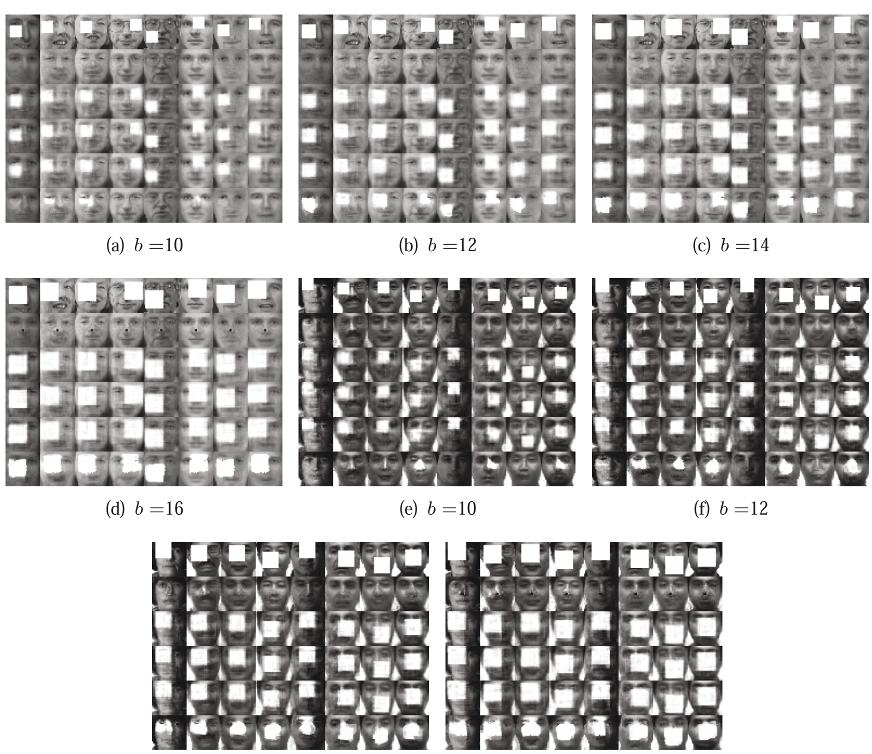
	ORL				YALE			
p(%)	10	15	20	25	10	15	20	25
WNMF	25.28	25.17	25.05	24.89	20.28	20.38	20.17	19.92
NMF	20.83	19.44	18.38	17.55	17.82	16.41	15.31	14.50
MahNMF	22.82	21.68	20.79	19.98	17.73	16.34	15.31	14.38
RNMF	23.47	22.32	21.16	20.04	20.26	19.19	18.17	16.96
CauchyNMF	23.76	23.8	23.93	24.09	18.67	18.65	18.62	18.30

Note: WNMF, weighted nonnegative matrix factorization; NMF, nonnegative matrix factorization; PSNR, peak signal-to-noise ratio.

Table 3 PSNRs on the ORL and YALE datasets contaminated by Contiguous Occlusion with different block sizes from 10 to 16.

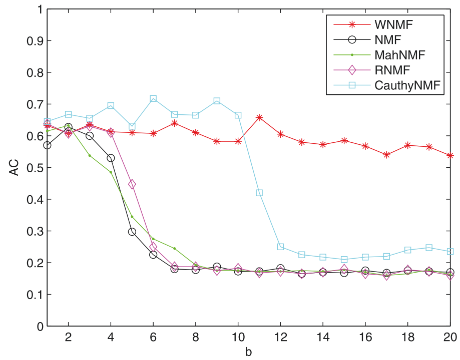
	ORL				YALE			
b	10	12	14	16	10	12	14	16
WNMF	23.83	23.54	23.15	21.79	20.09	19.68	19.18	18.48
NMF	15.5	14.19	13.03	11.97	12.79	11.44	10.26	9.195
MahNMF	15.55	14.23	13.03	11.95	12.78	11.43	10.23	9.203
RNMF	15.53	14.2	13.03	11.97	12.8	11.45	10.27	9.192
CauchyNMF	20.93	15.29	13.42	12.09	16.84	13.54	10.88	9.489

Note: WNMF, weighted nonnegative matrix factorization; NMF, nonnegative matrix factorization; PSNR, peak signal-to-noise ratio.

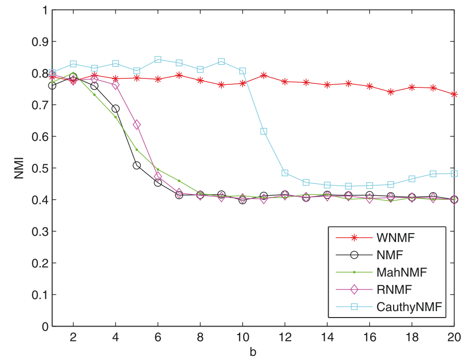


(a) $b = 10$ (b) $b = 12$ (c) $b = 14$ (d) $b = 16$ (e) $b = 10$ (f) $b = 12$ (g) $b = 14$ (h) $b = 16$

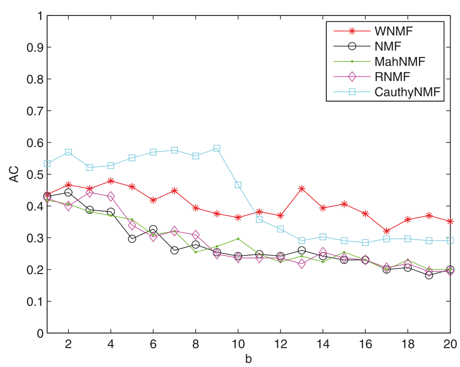
Figure 4 Recovered images from the ORL and YALE datasets corrupted by Contiguous Occlusion. For (a)–(h), the first row is sample images under the corrupted block size b, and the last five rows are recovered images by weighted nonnegative matrix factorization (WNMF), NMF, MahNMF, RNMF and CauchyNMF.



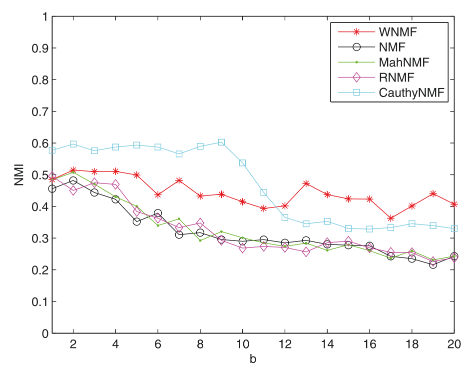
(a) Average ACs of the ORL dataset vs. the corrupted block size



(b) Average NMIs of the ORL dataset vs. the corrupted block size



(c) Average ACs of the YALE dataset vs. the corrupted block size



(d) Average NMIs of the YALE dataset vs. the corrupted block size

Figure 5 Evaluation on the ORL and YALE databases contaminated by Contiguous Occlusion.

6.3.2. Clustering

According to Figure 5, we can conclude that

- WNMF is more robust to remove a large number of outliers of Contiguous Occlusion, which denotes that WNMF can learn a more robust representation from the ORL and YALE datasets corrupted by Contiguous Occlusion.
- NMF, MahNMF and RNMF cannot handle Contiguous Occlusion, which indicates that they cannot achieve a robust subspace for clustering.
- CauthyNMF achieve excellent clustering results in the beginning, however, they perform unstable as the ORL and YALE datasets are contaminated by serious corruptions.

6.4. Convergence Study

The update rules (18), (21) and (22) for optimizing WNMF are iterative. These rules are proved to be convergent. In this subsection, we investigate whether these rules can be convergent. Figure 6 presents the convergence curves of WNMF on the ORL and YALE dataset corrupted by Salt and Peeper noise and Contiguous Occlusion. For each figure, the x-axis is the iteration number, and the y-axis denotes the objective value defined in (23). Suppose that $p = 0.5$, $b = 12$, $r = 50$ and the maximum iteration number is 500. It is obvious that the iterative rules for WNMF have a fast convergence.

7. CONCLUSION

This paper proposed an effective weighted NMF model to handle outliers and noise. The advantages of the proposed framework are as follows: (1) WNMF is more effective and robust to handle Salt and Pepper noise and Contiguous Occlusion. (2) WNMF can achieve a cleaner data space and a smaller factorization error when the ORL and YALE datasets are contaminated by Salt and Pepper noise and Contiguous Occlusion. (3) WNMF can learn a more robust low-dimensional presentation for clustering when the ORL and YALE datasets are contaminated with heavy corruptions.

CONFLICT OF INTEREST

We would like to submit the enclosed manuscript entitled "Weighted Nonnegative Matrix Factorization for Image Inpainting and Clustering " which we wish to be considered for publication in "International Journal of Computational Intelligence Systems". No conflict of interest exits in the submission of this manuscript, and this manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in

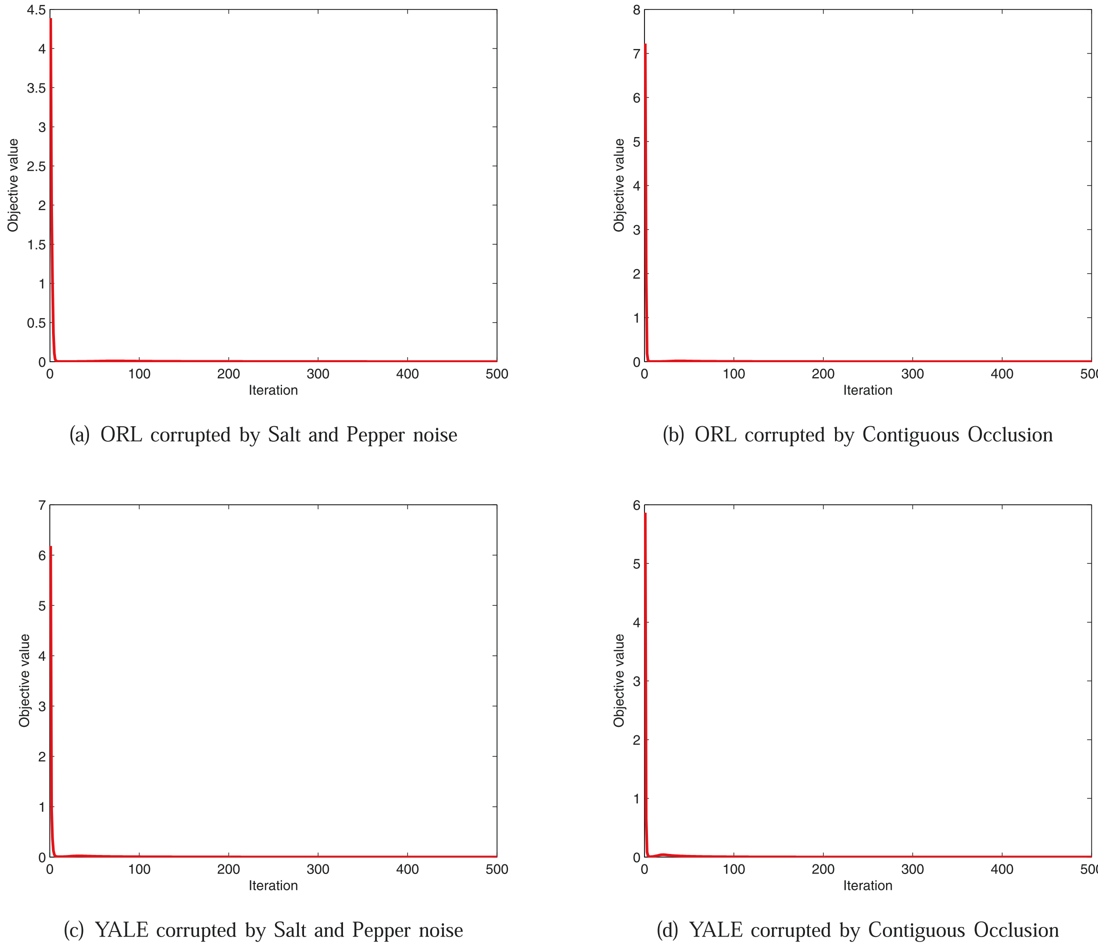


Figure 6 Convergence curves on ORL and YALE corrupted by Salt and Peeper noise and Contiguous Occlusion.

part. All the authors listed have approved the manuscript that is enclosed.

AUTHORS' CONTRIBUTIONS

KeKe Zhang and Nian Zhang proposed a weighted non-negative matrix factorization framework and designed the related algorithm. Jiang Xiong tested the algorithm and demonstrated the robustness and effectiveness of the proposed algorithm. Xiangguang Dai wrote and revised the manuscript.

ACKNOWLEDGMENTS

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 and National Science Foundation (NSF) grant #1505509 and DoD grant #W911NF1810475.

REFERENCES

[1] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature. 401 (1999), 788–791.
[2] G. Casalino, N. Del Buon, C. Mencar, Subtractive clustering for seeding non-negative matrix factorizations, Inf. Sci. 257 (2014), 369–387.
[3] W. Wu, Y. Jia, S. Kwong, J. Hou, Pairwise constraint propagation-induced symmetric nonnegative matrix factorization, IEEE Trans. Neural Netw. Learn. Syst. 29 (2018), 6348–6361.
[4] H. Li, K. Li, J. An, W. Zhang, K. Li, An efficient manifold regularized sparse non-negative matrix factorization model for large-scale recommender systems on GPUs, Inf. Sci. 496 (2019), 464–484.
[5] X. Liu, W. Wang, D. He, P. Jiao, D. Jin, C.V. Cannistraci, Semi-supervised community detection based on non-negative matrix factorization with node popularity, Inf. Sci. 381 (2017), 304–321.
[6] X. Peng, D. Chen, D. Xu, Hyperplane-based nonnegative matrix factorization with label information, Inf. Sci. 493 (2019), 1–9.
[7] A.B. Hamza, D.J. Brady, Reconstruction of reflectance spectra using robust nonnegative matrix factorization, IEEE Trans. Signal Process. 54 (2006), 3637–3642.
[8] D. Kong, C. Ding, H. Huang, Robust nonnegative matrix factorization using L21-norm, in Proceedings of the 20th ACM

international conference on Information and knowledge management, Glasgow, Scotland, 2011, pp. 673–682.
[9] N. Guan, D. Tao, Z. Luo, J. Shawetaylor, MahNMF: Manhattan non-negative matrix factorization, arXiv:1207.3438v1, 2012, http://arxiv.org/abs/1207.3438v1
[10] H. Gao, F. Nie, W. Cai, H. Huang, Robust capped norm nonnegative matrix factorization, in ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 2015, pp. 871–880.
[11] L. Du, X. Li, Y. Shen, Robust nonnegative matrix factorization via half-quadratic minimization, in IEEE International Conference on Data Mining, Brussels, Belgium, 2012, pp. 201–210.
[12] L. Zhang, Z. Chen, M. Zheng, X. He, Robust non-negative matrix factorization, Front. Electr. Electron. Eng. China. 6 (2015), 192–200.
[13] N. Guan, T. Liu, Y. Zhang, D. Tao, L. Davis, Truncated cauchy non-negative matrix factorization, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2018), 246–259.
[14] Q. Gu, J. Zhou, Local learning regularized nonnegative matrix factorization, in Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), Morgan Kaufmann Publishers Inc, Pasadena, CA, USA, 2009.
[15] C. Peng, Z. Kang, C. Chen, Q. Cheng, Nonnegative matrix factorization with local similarity learning, arXiv preprint arXiv:1907.04150, 2019. https://arxiv.org/abs/1907.04150
[16] Z. Kang, H. Pan, S. Hoi, Z. Xu, Robust graph learning from noisy data, IEEE Trans. Cybern. 50 (2019), 1833–1843.
[17] G. Chen, C. Xu, J. Wang, *et al.*, Robust non-negative matrix factorization for link prediction in complex networks using manifold regularization and sparse learning, Physica A Stat. Mech. Appl. 539 (2020), 122882.
[18] Z. Guo, Y. Zhang, A sparse corruption non-negative matrix factorization method and application in face image processing and recognition, Measurement. 136 (2019), 429–437.
[19] L. Zhao, L. Tao, C. Zou, Unsupervised learning of phonemes of whispered speech in a noisy environment based on convolutive non-negative matrix factorization, Inf. Sci. 257 (2014), 115–126.
[20] N. Gillis, Robustness analysis of hottopixx, a linear programming model for factoring nonnegative matrices, SIAM J. Matrix Anal. Appl. 34 (2013), 1189–1212.
[21] Y. Lu, Z. Lai, Y. Xu, J. You, X. Li, C. Yuan, Projective robust nonnegative factorization, Inf. Sci. 364 (2016), 16–32.
[22] Z. Li, J. Tang, X. He, Robust structured nonnegative matrix factorization for image representation, IEEE Trans. Neural Netw. Learn. Syst. 29 (2018), 1947–1960.
[23] D. Cai, X. He, J. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011), 1548–1560.
[24] D. Cai, X. He, J. Han, Document clustering using locality preserving indexing, IEEE Trans. Knowl. Data Eng. 17 (2005), 1624–1637.