



Robust semi-supervised non-negative matrix factorization for binary subspace learning

Xiangguang Dai¹ · Keke Zhang¹ · Juntang Li² · Jiang Xiong¹ · Nian Zhang³ · Huaqing Li⁴

Received: 29 September 2020 / Accepted: 23 January 2021
© The Author(s) 2021

Abstract

Non-negative matrix factorization and its extensions were applied to various areas (i.e., dimensionality reduction, clustering, etc.). When the original data are corrupted by outliers and noise, most of non-negative matrix factorization methods cannot achieve robust factorization and learn a subspace with binary codes. This paper puts forward a robust semi-supervised non-negative matrix factorization method for binary subspace learning, called RSNMF, for image clustering. For better clustering performance on the dataset contaminated by outliers and noise, we propose a weighted constraint on the noise matrix and impose manifold learning into non-negative matrix factorization. Moreover, we utilize the discrete hashing learning method to constrain the learned subspace, which can achieve a binary subspace from the original data. Experimental results validate the robustness and effectiveness of RSNMF in binary subspace learning and image clustering on the face dataset corrupted by Salt and Pepper noise and Contiguous Occlusion.

Keywords Noise · Binary subspace learning · Graph regularization · Dimensionality reduction · Non-negative matrix factorization

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, and there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “Robust Semi-Supervised Non-negative Matrix Factorization for Binary Subspace Learning”.

✉ Huaqing Li
huaqingli@hotmail.com

Xiangguang Dai
daixiangguang@163.com

Keke Zhang
xk_zhang0924@126.com

Juntang Li
398340239@qq.com

Jiang Xiong
xjcq123@126.com

Nian Zhang
nzhang@udc.edu

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

Introduction

The processing and application of the original high-dimensional data are very challenging. To address this problem, many data dimensionality reduction methods were applied to solve image retrieval image indexing [1,2] and image classification [3]. To achieve a satisfactory subspace from dimensionality reduction, most of the studies mainly consider how to discover an effective low-dimensional representation from the original data. A common scheme is to dig out the geometrical structure information of the original data, which can lead to a more discriminative representation.

In the past decades, several dimensionality reduction techniques were presented such as principal components analysis (PCA) [4] and non-negative matrix factorization (NMF) [5], which can learn an effective subspace for classification

² Sate Grid Chongqing Yongchuan Electric Power Supply Branch, Chongqing, China

³ Department of Electrical and Computer Engineering, University of the District of Columbia, Washington, D.C. 20008, USA

⁴ College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China

and clustering. NMF decomposes an original non-negative matrix into two non-negative matrices, such that their product approximates to the original data matrix. The non-negative property is consistent with the humans perception, which is more meaningful in image representation. Due to the satisfactory performance of NMF, some extensions [6–19] were proposed and utilized to improve the clustering effect.

Traditional NMF is an unsupervised method and cannot be designed for clustering specially. To achieve the better clustering effect, some constraints (i.e., label propagation, manifold learning, pairwise constraint, etc.) were considered to constrain the subspace, which can learn a more effective parts-based representation. When the original data are heavily corrupted, NMF fails to achieve clustering. This is because its loss function is more sensitive to outliers and noise. Therefore, some researchers [6–12] proposed some other loss functions to try better matrix factorization. Gao et al. [9] presented a capped norm as the loss function and an outlier threshold to reduce outliers. However, there is no theory to adjust the threshold. Recently, Guan et al. [12] proposed the Truncated Cauchy loss (CauchyNMF) as the loss function and the three-sigma-rule to filter outliers. Although CauchyNMF learns a better subspace from the original data space contaminated by outliers and noise than other NMF methods, it leads to an unsatisfactory subspace when the outliers cannot follow the Gaussian distribution.

After dimensionality reduction by non-negative matrix factorization, the parts-based representation composed of real numbers will take much more time in clustering. Recently, data-dependent hashing methods [20–28] were put forward to learn the latent features of training data and achieved effective binary codes from different hash functions. It is obvious that a subspace composed of binary codes (−1 and 1 or 0 and 1) can reduce the clustering time. However, the traditional NMF cannot learn a parts-based representation composed of binary codes.

In this paper, based on data-dependent hashing methods, non-negative matrix factorization, and manifold learning, a novel dimensionality reduction method is presented to learn a subspace composed of binary codes from the original data space. Our achievements are as follows:

- A robust non-negative matrix factorization framework was proposed to remove outliers in the subspace. Moreover, the learned subspace composed of binary codes obeys the geometrical assumption of the original data.
- Our problem can be formulated as a mixed integer optimization problem. We transform it into several subproblems and elegantly solve these subproblems.
- Extensive experiments prove that our method can learn a subspace composed of binary codes from the dataset corrupted by Salt and Pepper noise and Contiguous Occlusion. Moreover, the clustering performance from

the subspace can demonstrate that our method can achieve the better clustering effect than other dimensionality reduction methods.

Related works

Non-negative matrix factorization and its extensions

Supposed that any image can be represented by a vector $x_i \in R^m$ and the matrix $V = [x_1, \dots, x_n]$ denotes an original image space composed of n images. NMF can be utilized to discover two low-dimensional matrices $W \in R^{m \times r}$ and $H \in R^{r \times n}$, such that their product can best approximate to V , where r is a factorization rank and $r \ll \min\{m, n\}$. Generally, NMF can be mathematically formulated as follows:

$$\begin{aligned} \min_{W, H} \quad & \text{Loss}(V, WH) \\ \text{s.t.} \quad & W \geq 0, H \geq 0, \end{aligned} \quad (1)$$

where the function Loss is to measure the error between V and WH . The usual loss function can be L_1 norm, $L_{2,1}$ norm, or the Frobenius norm. Guan et al. [12] put forward a Truncated Cauchy loss to reduce outliers (CauchyNMF), which can be written by the following form:

$$\min_{W \geq 0, H \geq 0} F(W, H) = \sum_{i=1}^m \sum_{j=1}^n g\left(\frac{(V - WH)_{ij}}{\gamma}\right), \quad (2)$$

where $g(x) = \begin{cases} \ln(1+x), & 0 \leq x \leq \sigma \\ \ln(1+\sigma), & x > \sigma. \end{cases}$ The scale parameter σ can be computed by three-sigma-rule, and the truncation parameter γ can be achieved by the Nagy algorithm [12].

Most of NMF variants utilize various loss function to handle outliers, but these approaches cannot remove outliers in the subspace. To address this problem, a novel robust NMF framework was put forward as follows:

$$\begin{aligned} \min_{W, H, E} \quad & \text{Loss}(M, WH, E) + \lambda \Omega(E, W, H) \\ \text{s.t.} \quad & W \geq 0, H \geq 0, \end{aligned} \quad (3)$$

where M is the original data matrix contaminated by outliers, E is an error matrix, λ is a hyper-parameter, and Ω is the constraint term. Based on problem (3), Zhang et al. [11] proposed the following robust NMF problem:

$$\begin{aligned} \min_{W, H, E} \quad & \|M - WH - E\|_F^2 + \lambda \|E\|_M \\ \text{s.t.} \quad & W \geq 0, H \geq 0, \end{aligned} \quad (4)$$

where $\|E\|_M = \sum_{ij} |e_{ij}|$.

Data-dependent hashing methods

Assumed that an image sample is expressed by the vector $v \in R^m$ and the original data space is showed by the matrix $V = [v_1, \dots, v_n] \in R^{m \times n}$. Data-dependent hashing methods expect to find a binary code matrix $B \in \{-1, 1\}^{L \times n}$, which can maintain the semantic similarities of the data space. Usually, each column of B is L -bit codes for each v , where $L < m$.

To make full use of the label information of the original data, Shen et al. [25] put forwarded a supervised discrete hashing framework, which can generate binary codes with satisfactory linear classification. Supposed that an original data matrix $V = [v_1, \dots, v_n] \in R^{m \times n}$, a label matrix $Y \in \{0, 1\}^{c \times n}$, a binary code matrix $B \in \{0, 1\}^{L \times n}$, $W \in R^{L \times c}$, and $P \in R^{m \times L}$. SDH are able to sum up by:

$$\begin{aligned} \min_{W, B, P} \quad & \|Y - W^T B\|_F^2 + \lambda \|W\|_F^2 \\ & + \mu \|B - P^T \phi(V)\|_F^2 \\ \text{s.t.} \quad & B \in \{0, 1\}^{L \times n}, \end{aligned} \quad (5)$$

where $\phi(\cdot)$ is the RBF kernel mapping, μ and λ are penalty parameters. For the purpose of solving problem (5), we can optimize the following three subproblems:

$$\min_W \|Y - W^T B\|_F^2 + \lambda \|W\|_F^2 \quad (6)$$

and

$$\begin{aligned} \min_{W, B, P} \quad & \|Y - W^T B\|_F^2 + \mu \|B - P^T \phi(V)\|_F^2 \\ \text{s.t.} \quad & B \in \{0, 1\}^{L \times n} \end{aligned} \quad (7)$$

and

$$\|B - P^T \phi(V)\|_F^2, \quad (8)$$

until the stop condition is satisfied. Thus, we are able to obtain the local optimal solution of problem (5).

In response to this problem (6), we have:

$$W = (BB^T + \lambda)^{-1}BY^T. \quad (9)$$

In response to this problem (7), we make the following assumptions:

- z^T is the l th row of B and B' is the matrix of B not including z .
- q^T is the l th row of $WY + \mu P^T \phi(V)$ and Q' is the matrix of Q not including q .
- v^T is the l th row of W and W' is the matrix of W not including v .

We can safely come to the conclusion that:

$$z = \text{sign}(q - B'^T W'v). \quad (10)$$

With regard to problem (8), the solution is:

$$P = (\phi(V)\phi(V)^T)^{-1}\phi(V)B^T. \quad (11)$$

Consequently, it is easy to search the local optimal solution of the problem (5) using (9), (10), and (11).

Problem formulation

When the original data are destroyed by outliers and noise, existing NMF methods have the following shortcomings: (1) They are unable to learn an effective and powerful subspace from the original data space. (2) The parts-based representation is unable to retain the geometrical structure information of the original data. (3) These NMF methods are unable to learn a subspace with binary codes.

For problem (4), Zhang et al. [11] assumed that the outliers of the error matrix E are very sparse. Yet, the abnormal position of the original data is ignored. Supposed that we become aware of some outlier locations. For an image space $M \in R^{m \times n}$, a weighted graph S marks the location of outliers by the following equation:

$$S_{ij} = \begin{cases} 0, & \text{if } M_{ij} \text{ is an outlier,} \\ 1, & \text{otherwise.} \end{cases} \quad (12)$$

Hence, we rewrite the constraint on E as follows:

$$\|E \otimes S\|_F^2. \quad (13)$$

To obtain the geometric information in subspace, manifold regularization can be used to establish the relation between the original data space and the subspace. Therefore, a common used method called manifold learning [29] is as follows:

$$\text{tr}(H(D - U)H^T), \quad (14)$$

where tr is the trace of a matrix, $U_{jl} = e^{-\frac{\|x_j - x_l\|^2}{\sigma}}$ and $D_{ii} = \sum_j W_{ij}$.

In summary, combining (13), (14), and (4) results in our robust semi-supervised non-negative matrix factorization for binary subspace learning (RSNMF). Given a non-negative data matrix $V \in R^{m \times n}$ and a factorization r , one hopes to achieve a code matrix $B \in \{-1, +1\}^{r \times n}$ from V . Our proposed robust semi-supervised non-negative matrix factorization (RSNMF) can be utilized to learn binary codes for clustering. There are three properties as follows:

- The learned subspace can remove outliers and noise similar to (4).
- The subspace composed of binary codes can be learned from the data space similar to (5). (5) is a supervised problem; therefore, we delete the first two terms of (5) to be an unsupervised problem.
- The low-dimensional space composed of binary codes should remain similarity or dissimilarity of the original data space similar to (14).

Combining (4), (5), and (14), our problem can be formulated as follows:

$$\begin{aligned}
 & \min_{W, H, E, B, P} F(W, H, E, B, P) \\
 & = \|M - WH - E\|_F^2 + \lambda \|E \otimes S\|_F^2 \\
 & \quad + \gamma \text{tr}(H(D - U)H^T) \\
 & \quad + \gamma \text{tr}(B(D - U)B^T) + \alpha \|B - PH\|_F^2 \\
 & \text{s.t. } W \geq 0, H \geq 0, B \in \{-1, 1\}.
 \end{aligned} \tag{15}$$

where λ , γ , and α are hyper-parameters.

Optimization scheme

Problem (15) is a non-convex optimization problem. Thus, it is unable to search the global optimal solution. A generic framework for solving problem (15) is the “block-coordinate-descent” method (BCD) [30], in which one block variables are solved in order under relevant constraints and the remaining variables remain fixed. Thus, problem (15) can be converted into several convex problems and solve them in turn until convergence. For (15), we have five block variables W , H , B , P , and E . Thus, BCD is able to optimize the five matrices in turn. Supposed that the k th solution of problem (15) has been realized. The $k + 1$ th solution can be searched by:

$$\begin{aligned}
 E^{k+1} = \arg \min_E & \|M - W^k H^k - E\|_F^2 \\
 & + \lambda \|E \otimes S\|_F^2
 \end{aligned} \tag{16}$$

and

$$\begin{aligned}
 W^{k+1} = \arg \min_W & \|M - WH^k - E^{k+1}\|_F^2 \\
 \text{s.t. } W & \geq 0
 \end{aligned} \tag{17}$$

and

$$\begin{aligned}
 H^{k+1} \\
 = \arg \min_H & \|M - W^{k+1}H - E^{k+1}\|_F^2 \\
 & + \alpha \|B - PH\|_F^2 + \gamma \text{tr}(H(D - U)H^T) \\
 \text{s.t. } H & \geq 0,
 \end{aligned} \tag{18}$$

and

$$P^{k+1} = \arg \min_P \|B^{k+1} - PH^{k+1}\|_F^2 \tag{19}$$

and

$$\begin{aligned}
 B^{k+1} \\
 = \arg \min_B & (\alpha \|B^{k+1} - P^{k+1}H^{k+1}\|_F^2 \\
 & + \gamma \text{tr}(B(D - U)B^T)) \\
 \text{s.t. } B & \in \{-1, 1\}^{r \times n}.
 \end{aligned} \tag{20}$$

It is easy to get the solution of problems (4) and (4) as follows:

$$e_{ij} \leftarrow \frac{m_{ij} - (WH)_{ij}}{1 + \lambda s_{ij}}. \tag{21}$$

$$w_{il} \leftarrow w_{il} \frac{(MH^T)_{il} - (EH^T)_{il}}{(WHH^T)_{il}}. \tag{22}$$

For (18), we can utilize Nesterov’s optimal method [31] to solve it. To save space, we do not introduce this algorithm. For problem (19), we have the optimal solution:

$$P = BH^T(HH^T)^{-1} \tag{23}$$

or

$$P^T = (HH^T)^{-1}HB^T, P = (P^T)^T. \tag{24}$$

Using a function RRC in [25], Eq. (24) can be realized. Therefore, the solution of (19) can be realized by:

$$\begin{aligned}
 P & = \text{RRC}(H^T, B^T, 0) \\
 P & = P^T.
 \end{aligned} \tag{25}$$

For problem (20), using the discrete cyclic coordinated descent method, we can achieve its local optimal solution. First, problem (20) can be converted into the following form:

$$F(B) = (\alpha \|B - PH\|_F^2 + \beta \text{tr}(BKC B^T)), \tag{26}$$

where $K = D - U$ and $C \in R^{n \times n}$ is an identity matrix. Second, some assumptions are made as follows:

- z is the k th column of B and B' is the matrix of B excluding z . Thus, $B = [z \ B']$.
- c is the k th column of C and C' is the matrix of C excluding c . Thus, $C = [c \ C']$.
- k^T is the k th row of K and K' is the matrix of K excluding k . Thus, $K = \begin{bmatrix} k^T \\ K' \end{bmatrix}$.
- $Q = PH$. q is the k th column of Q and Q' is the matrix of B excluding q . Thus, $Q = [q \ Q']$.



Thirdly, B can be learned column by column. The first term of problem (26) is rewritten as follows:

$$\begin{aligned}
 & \|B - PH\|_F^2 \\
 &= \|z - q\|_F^2 + \|B' - Q'\|_F^2 \\
 &= \text{tr}((z - q)^T(z - q)) + \text{const} \\
 &= \text{tr}(z^T z - z^T q - q^T z + q^T q) + \text{const} \\
 &= -2\text{tr}(q^T z) + \text{const}.
 \end{aligned} \quad (27)$$

Similar to the second term of problem (26), we are able to conclude that:

$$\begin{aligned}
 & \text{tr}(BKC B^T) \\
 &= \text{tr}((zk^T + B'K')(cz^T + C'B'^T)) \\
 &= \text{tr}(zk^T cz^T + zk^T C'B'^T + B'K'cz^T \\
 &\quad + B'K'C'B'^T) \\
 &= \text{tr}(zk^T C'B'^T) + \text{tr}(B'K'cz^T) + \text{const} \\
 &= 2\text{tr}(k^T C'B'^T z) + \text{const}.
 \end{aligned} \quad (28)$$

Therefore, problem (20) can be rewritten by:

$$\begin{aligned}
 \min_z \quad & \beta \text{tr}(k^T C'B'^T z) - \alpha \text{tr}(q^T z) \\
 \text{s.t.} \quad & z \in \{-1, 1\}^r.
 \end{aligned} \quad (29)$$

This problem brings about the following optimal solution:

$$z = \text{sign}(\alpha q - \beta B' C'^T k). \quad (30)$$

Obviously, each z can be calculated from the pre-learned B' in advance. Hence, we can implement B before each z is updated. In [25], it is recommended (30) be used to learn the binary code matrix B in t times, where $t = 5$.

Algorithm 1 RSNMF

Input: $V \in R_+^{m \times n}$, $W \in R_+^{m \times r}$, $H \in R_+^{r \times n}$, $E \in R_+^{m \times n}$, $iter$, $B \in \{-1, +1\}^{r \times n}$, $P \in R^{r \times r}$, λ , γ , α

Output: B

Initialize D and U (14)

for $k=0$ to $iter$ **do**

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

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

Computing H by [31]

Computing P by (25)

Computing B by (30)

end for

Experimental results

Experiment setup

We compare our proposed method (RSNMF) with NMF [5], RNMF_L1 [9], PCA [4], and CauchyNMF [12] on the clustering performances of two datasets (i.e., ORL and YALE) contaminated by Salt and Pepper noise and Contiguous Occlusion.

ORL contains a set of the frontal face images, and these images were established from 1992 to 1994 by Cambridge University. There are 40 various persons, and each person includes 10 images. Each image were taken at different facial expressions, times, and so on. The format of each image is PGM with the size of 92×122 pixels. We scale down each image to 32×32 pixels. YALE was constructed by the Center for Computational Vision and Control of Yale University. There are 165 images of 15 persons, and each person contains 11 pictures. Each image was taken by different facial expressions or configurations with the size of 100×100 pixels. Some example images from ORL and YALE are presented in Fig. 1.

To verify the clustering ability on the corrupted data, we propose two corruptions including Salt and Pepper noise and Contiguous Occlusion. Salt and Pepper noise is utilized to change a portion of pixel values to be 0–255. The corrupted percentage of pixels is from 0.05 to 0.8 with the step size



(a) Some sample images from ORL



(b) Some sample images from YALE

Fig. 1 Sample images from ORL and YALE

0.05. Contiguous Occlusion randomly corrupts a block of each image and the pixels of the block is filled with 255. The corrupted block size is proposed to be 1 to 24 with the step size 2. Supposed that $r = 32$, $\alpha = 0.1$, $\gamma = 1e^{-5}$, $\lambda = 100$, and $iter = 200$. We propose Accuracy (AC) and Normalized Mutual Information (NMI) [32] to validate the clustering effect of each algorithm.

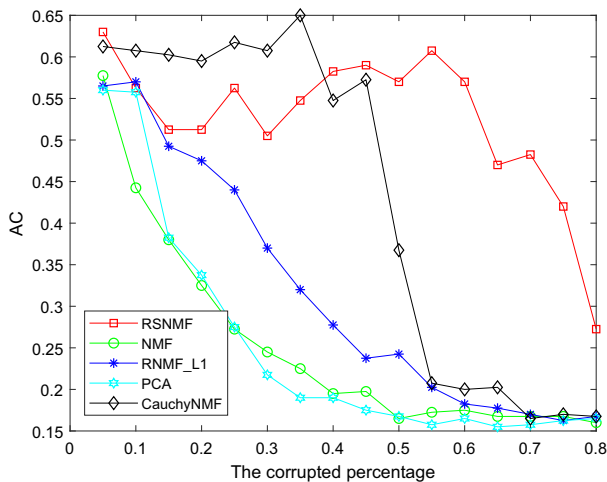
Salt and Pepper noise

Figure 2 presents the clustering performance on ORL and YALE when the two datasets are contaminated by Salt and Pepper noise. From these figures, we observe that:

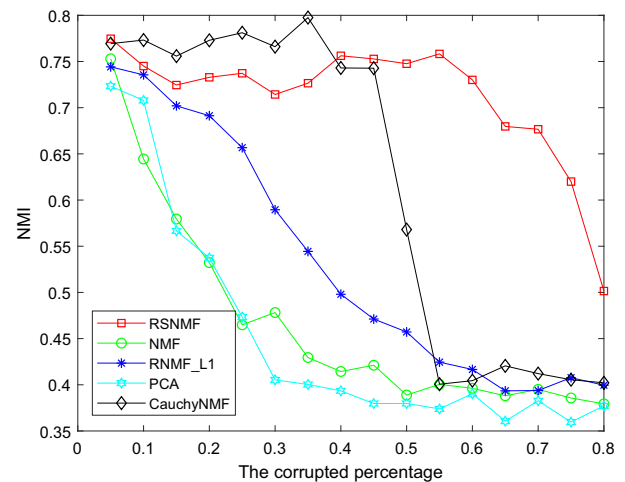
- For ORL, the best AC and NMI obtained by CauchyNMF are 0.65 and 0.8, respectively. For YALE, the best AC

and NMI achieved by CauchyNMF are 0.48 and 0.52, respectively.

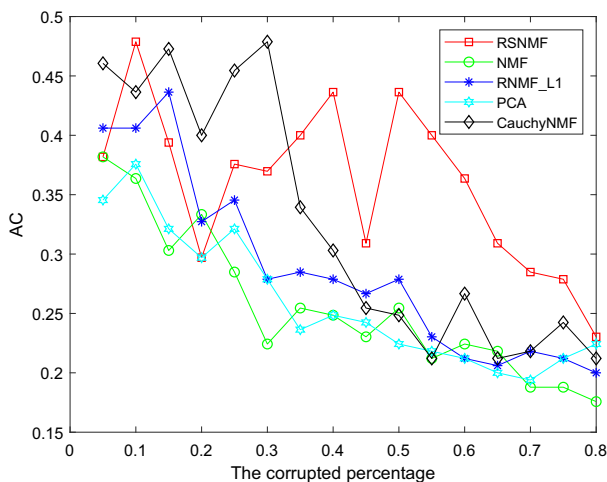
- RSNMF and CauchyNMF have the better ACs and NMIs than other methods. CauchyNMF performs better than RSNMF in the beginning; however, it becomes worse when the corrupted percentage varies.
- For the smaller corrupted percentage, all methods can not only learn a satisfactory subspace, but also achieve excellent clustering results. When ORL is heavily corrupted (i.e., the corrupted percentage greater than 0.4), only RSNMF remains the satisfactory clustering results.



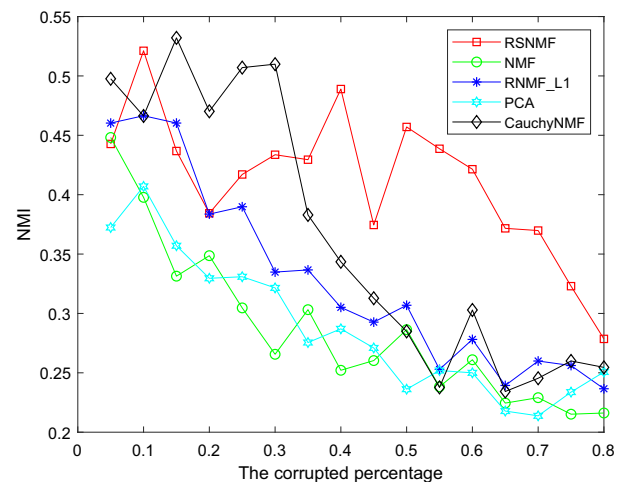
(a) Clustering ACs on ORL



(b) Clustering NMIs on ORL

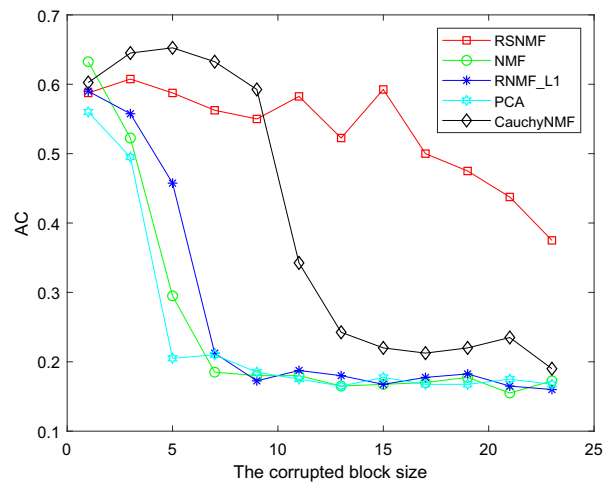


(c) Clustering ACs on YALE

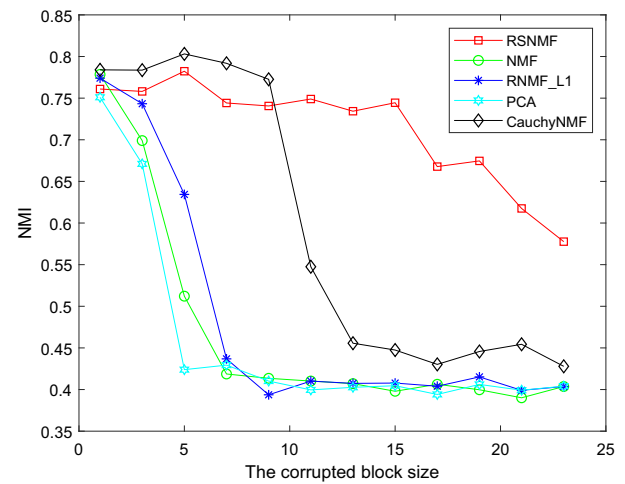


(d) Clustering NMIs on YALE

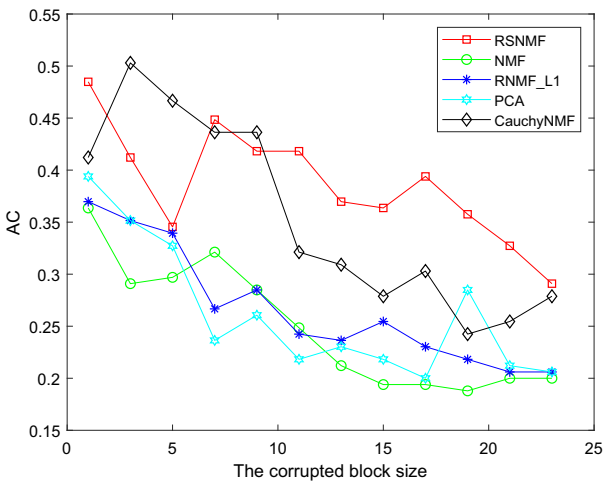
Fig. 2 Clustering ACs and NMIs on ORL and YALE when the corrupted percentage varies from 0.05 to 0.8 with the step size 0.05



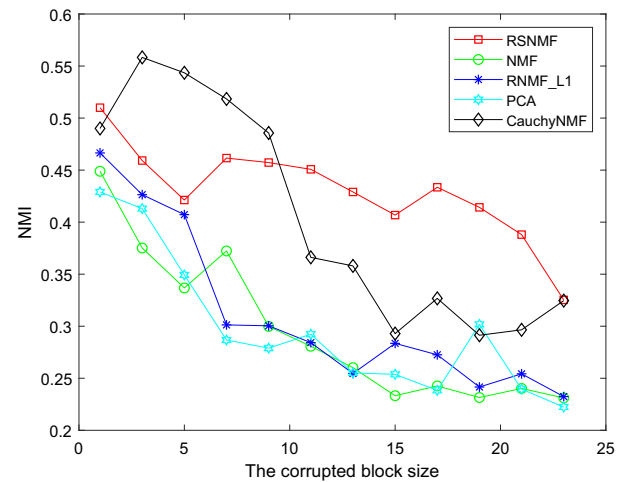
(a) Clustering ACs on ORL



(b) Clustering NMIs on ORL



(c) Clustering ACs on YALE



(d) Clustering NMIs on YALE

Fig. 3 Clustering ACs and NMIs on ORL and YALE when the corrupted block size varies from 1 to 24 with the step size 2

Contiguous occlusion

Figure 3 presents the clustering performance on ORL and YALE when the two datasets are contaminated by Contiguous Occlusion. From these figures, there are some interesting points as follows:

- NMF, PCA, and RNMF_L1 achieve satisfactory clustering results when the block size is very small.
- CauchyNMF achieves the best AC and NMI with the smaller block size (i.e., the block size less than 6). As the block size is greater than 10, CauchyNMF performs worse rapidly.

- Although RSNMF performs not better than CauchyNMF in the beginning, it remains stable clustering results as the block size increases.

Conclusion

This paper presented a robust semi-supervised non-negative matrix factorization for binary subspace learning (RSNMF) to handle Salt and Pepper noise and Contiguous Occlusion. The clustering performances demonstrate that our method achieves the following advantages. First, RSNMF can learn a more effective and discriminative parts-based representation composed of binary codes from ORL and YALE corrupted by Salt and Pepper noise and Contiguous Occlusion. Second,

RSNMF is more robust to outliers than the existing dimensionality reduction methods.

Acknowledgements 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 KJQN201901203), Natural Science Foundation of Chongqing (Grant No. cstc2019jcyj-bshX0101), Foundation of Chongqing Three Gorges University, and National Science Foundation (NSF) Grant #2011927 and DoD Grant #W911NF1810475.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Datta R, Joshi D, Jia LI, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age, ACM. Comput Surv 40(2):35–94
- Chen L, Xu D, Tsang IW, Li X (2014) Spectral embedded hashing for scalable image retrieval. IEEE Trans Cybern 44(7):1180–1190
- Banerjee B, Bovolo F, Bhattacharya A, Bruzzone L, Chaudhuri S, Mohan BK (2015) A new self-training-based unsupervised satellite image classification technique using cluster ensemble strategy. IEEE Geosci Remote Sens Lett 12(4):741–745
- Turk M, Pentland A (1991) Eigenfaces for recognition. J Cognit Neurosci 3(1):71–86
- Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791
- Hamza AB, Brady DJ (2006) Reconstruction of reflectance spectra using robust nonnegative matrix factorization. IEEE Trans Signal Process 54(9):37637–3642
- Kong D, Ding C, Huang H (2011) robust nonnegative matrix factorization using L21-norm. In: Proceedings of the 20th ACM international conference on information and knowledge management, pp 673–682
- Guan N, Tao D, Luo Z, Shawetaylor J (2012) MahNMF: manhattan non-negative matrix factorization. J Mach Learn Res. [arXiv:1207.3438v1](https://arxiv.org/abs/1207.3438v1)
- Gao H, Nie F, Cai W, Huang H (2015) Robust capped norm non-negative matrix factorization. In: ACM international on conference on information and knowledge management, pp 871–880
- Du L, Li X, Shen Y (2012) Robust nonnegative matrix factorization via half-quadratic minimization. In: IEEE international conference on data mining, pp 201–210
- Zhang L, Chen Z, Zheng M, He X (2015) Robust non-negative matrix factorization. Front Electr Electron Eng China 6(2):192–200
- Guan N, Liu T, Zhang Y, Tao D, Davis L (2018) Truncated Cauchy non-negative matrix factorization. IEEE Trans Pattern Anal Mach Intell 41(1):246–259
- Casalino G, Del Buon N, Mencar C (2014) Subtractive clustering for seeding non-negative matrix factorizations. Inf Sci 257:369–387
- Wu W, Jia Y, Kwong S, Hou J (2018) Pairwise constraint propagation-induced symmetric nonnegative matrix factorization. IEEE Trans Neural Netw Learn Syst 29(12):6348–6361
- Li H, Li K, An J, Zhang W, Li K (2018) An efficient manifold regularized sparse non-negative matrix factorization model for large-scale recommender systems on GPUs. Inf Sci. <https://doi.org/10.1016/j.ins.2018.07.060>
- Liu X, Wang W, He D, Jiao P, Jin D, Cannistraci CV (2017) Semi-supervised community detection based on non-negative matrix factorization with node popularity. Inf Sci 381:304–321
- Peng X, Chen D, Xu D (2019) Hyperplane-based nonnegative matrix factorization with label information. Inf Sci 493:1–9
- Kang Z, Pan H, Hoi S, Xu Z (2019) Robust graph learning from noisy data. IEEE Trans Cybern 50:1833–1843
- Li Z, Tang J, He X (2018) Robust structured nonnegative matrix factorization for image representation. IEEE Trans Neural Netw Learn Syst 29(5):1947–1960
- Weiss Y, Antonio T, Rob F (2009) Spectral hashing. In: Advances in neural information processing systems, pp 1753–1760. <https://proceedings.neurips.cc/paper/2008/file/d58072be2820e8682c0a27c0518e805e-Paper.pdf>
- Gong Y, Lazebnik S, Gordo A et al (2012) Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval. IEEE Trans Pattern Anal Mach Intell 32(12):2916–2929
- Shen F, Shen C, Shi Q, et al. (2013) Inductive hashing on manifolds. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1562–1569
- Wang J, Kumar S, Chang S (2012) Semi-supervised hashing for large-scale search. IEEE Trans Pattern Anal Mach Intell 34(12):2393–2406
- Kulis B, Darrell T (2009) Learning to hash with binary reconstructive embeddings. In: Advances in neural information processing systems, 22 pp 1042–1050. <https://proceedings.neurips.cc/paper/2009/file/6602294be910b1e3c4571bd98c4d5484-Paper.pdf>
- Shen F, Shen C, Liu W, et al. (2015) Supervised discrete hashing. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 37–45
- Gui J, Liu T, Sun Z et al (2018) Fast supervised discrete hashing. IEEE Trans Pattern Anal Mach Intell 40(2):490–496
- Liu W, Wang J, Ji R et al (2012) Supervised hashing with kernels. In: 2012 IEEE conference on computer vision and pattern recognition, pp 2074–2081
- Lin G, Shen C, Van den Hengel A (2015) Supervised hashing using graph cuts and boosted decision trees. IEEE Trans Pattern Anal Mach Intell 32(11):2317–2331
- Cai D, He X, Han J, Huang TS (2011) Graph regularized non-negative matrix factorization for data representation. IEEE Trans Pattern Anal Mach Intell 33(8):1548–1560
- Tseng P (2001) Convergence of a block coordinate descent method for nondifferentiable minimization. J Optim Theory Appl 109(3):475–494
- Guan N, Tao D, Luo Z, Yuan B (2012) NeNMF: an optimal gradient method for nonnegative matrix factorization. IEEE Trans Signal Process 60(6):2882–2898
- Cai D, He X, Han J (2005) Document clustering using locality preserving indexing. IEEE Trans Knowl Data Eng 17(12):1624–1637

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.