

Received June 10, 2019, accepted June 18, 2019, date of publication June 24, 2019, date of current version July 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924520

Robust Adaptive Graph Regularized Non-Negative Matrix Factorization

XIANG HE, QI WANG[✉], (Senior Member, IEEE), AND XUELONG LI, (Fellow, IEEE)

Center for Optical Imagery Analysis and Learning (OPTIMAL), School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Qi Wang (crabwq@gmail.com)

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1107403, in part by the National Natural Science Foundation of China under Grant 61871470, Grant U1864204, and Grant 61773316, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, and in part by the Project of Special Zone for National Defense Science and Technology Innovation.

ABSTRACT Data clustering, which aims to divide the given samples into several different groups, has drawn much attention in recent years. As a powerful tool, non-negative matrix factorization (NMF) has been applied successfully in clustering tasks. However, there are still two main limitations. First, the original NMF treats equally both noisy and clean data, which leads to high sensitivity to noises and outliers. Second, the performance of graph-based NMFs highly depends on the input graph, that is, if a low-quality graph is constructed to regularize NMF, the clustering results will be bad. To address the above-mentioned problems, we propose a novel robust adaptive graph regularized non-negative matrix factorization (RAGNMF) for data clustering. To be specific, we develop a robust weighted NMF (RWNMF) that can assign small weights to noises and outliers and large weights to clean data. Thus, the robustness of NMF is improved. Moreover, in the process of matrix factorization, metric learning is combined to choose some discriminative features and compute more appropriate distances of samples. Then, an adaptive graph is learned to well regularize the NMF. The experimental results demonstrate that the proposed RAGNMF can achieve better clustering performance than most of the state-of-the-art methods.

INDEX TERMS Non-negative matrix factorization, data clustering, graph learning, robustness.

I. INTRODUCTION

Clustering is one of the most important tasks in the fields of data mining and machine learning [1], [2]. For the given data, the goal of clustering is to partition them into multiple different groups. In the past decades, plenty of clustering methods have been proposed, such as k-means [3], spectral clustering [4], support vector clustering [5], non-negative matrix factorization (NMF) [6]–[9]. Because of the powerful ability for clustering, NMF has been studied extensively for many years. NMF is proposed originally in the influential work [6]. Different from most matrix factorization methods, NMF guarantees the non-negativity on the resulting matrices. Therefore, NMF can acquire a parts-based representation because only additive (not subtractive) operation is allowed. Due to the advantage, NMF has been successfully applied in most fields, e.g., document clustering [10], face recognition [11], hyperspectral unmixing [12].

Although NMF can achieve some promising performance in many fields, it still suffers from two main problems.

The associate editor coordinating the review of this manuscript and approving it for publication was Walter Didimo.

First, the standard NMF is very sensitive to noises and outliers. Because NMF adopts the squared error function (i.e., Frobenius norm) to define the objective function, only a few outliers can dominate the whole loss and lead to bad factorization results. To improve the robustness, some robust NMF models have been developed. Hamza and Brady [13] proposed $L_1 + L_2$ function to measure the loss of NMF. But the solving algorithm is very time-consuming. Kong *et al.* [14] presented the robust $\ell_{2,1}$ -NMF, which used $\ell_{2,1}$ norm instead of Frobenius norm to measure the loss. Because of removing the square operation of each reconstruction error, $\ell_{2,1}$ -NMF decreases successfully the sensitivity to noises and outliers. Moreover, Li *et al.* [15] proposed to employ $\ell_{2,0.5}$ norm to minimize the loss. The robustness of NMF are further improved. However, the above $\ell_{2,0.5}$ -NMF and widely used $\ell_{2,1}$ -NMF are still influenced by some large outliers and produce poor reconstruction results.

The second problem is that the original NMF only focuses on the global data structures and ignore the local relationship [16], [17] between neighboring samples. This strongly limits the ability of clustering. Afterwards, many

graph regularized NMF methods [18]–[20] are developed to exploit both global and local data structures. GNMF [21], RMNMF [22] and HNMF [23] are several typical works of graph based NMF by incorporating graph Laplacian. However, they do not take full advantage of the intrinsic geometrical structures of the given data. Graph based NMFs first build a similarity graph [24], and then performs matrix factorization. So the performance of NMF is closely connected with the input graph. Besides, NMFAN [25] proposed to incorporate adaptive neighbors into NMF and achieved better clustering results. But all the above algorithms construct the similarity graph by the ordinary Euclidean distances between samples. If the feature space contains some noisy and irrelevant features [26], the Euclidean distance metric will be not accurate enough. Thus, the graph is also not appropriate to regularize NMF.

To better cope with the above limitations, we propose a robust weighted NMF (RWNMF) and a robust adaptive graph regularized NMF (RAGNMF). The main contributions of the paper are summarized as follows.

- 1) Some necessary weights are incorporated into NMF to significantly improve the robustness of NMF. The proposed RWNMF treats each data sample differently, and can automatically assign large weights for clean samples and small weights for noisy samples and outliers. Therefore, noises and outliers only have little influence to the reconstruction results.
- 2) The graph learning is integrated into NMF. Euclidean distances are not accurate enough to measure the real distances of data with noisy and irrelevant features. So we learn a metric matrix to compute the new distances by unsupervised metric learning. Thus an adaptive graph is learned by the new distances to better regularize NMF.
- 3) A uniform framework is developed to perform feature selection, adaptive graph construction and matrix factorization simultaneously. In the iterative process of the proposed RAGNMF, some discriminative features are selected to compute the more appropriate distances, which are used to learn adaptive graph. Finally, the robust weight NMF is carried out with the adaptive graph.

II. ROBUST WEIGHTED NMF

Let $X = [X_1, X_2, \dots, X_n] \in \mathbb{R}^{m \times n}$ be the given non-negative data matrix, and each column of X (i.e., X_i) represents a data sample. n and m are the number of samples and features, respectively. NMF aims to reconstruct the given data X by two non-negative matrices $U \in \mathbb{R}^{m \times k}$ and $V \in \mathbb{R}^{n \times k}$. U and V are also named as basis matrix and encoding matrix, respectively. The original NMF proposes to employ Frobenius norm to minimize the reconstruction errors, and it aims to solve the following problem

$$\min_{U \geq 0, V \geq 0} \|X - UV^T\|_F^2 = \sum_{i=1}^n \|(X - UV^T)_i\|_2^2. \quad (1)$$

However, the original NMF (1) is very sensitive to noises and outliers because it uses the squared error function to measure the loss. This leads to that only a few outliers with large errors will dominate easily the objective function. Afterwards, Kong *et al.* [14] presented $\ell_{2,1}$ -NMF to enhance the robustness of NMF. It removes the square operation of reconstruction errors by utilizing $\ell_{2,1}$ norm to replaces Frobenius norm. The robust $\ell_{2,1}$ -NMF is formulated as

$$\min_{U \geq 0, V \geq 0} \|X - UV^T\|_{2,1} = \sum_{i=1}^n \|(X - UV^T)_i\|_2. \quad (2)$$

However, when the given data contain some large noises or outliers, $\ell_{2,1}$ -NMF is still influenced and obtains poor reconstruction results. Only removing the square operation of the standard NMF (1) is not enough to avoid the negative effects from outliers.

We can analyze the sensitivity of the original NMF (1) from another perspective. The total loss is the sum of all individual reconstruction loss, and each sample is treated equally. But the given data X usually has some noises [27] and outliers. It is more appropriate to incorporate some necessary weights to measure the importance for each sample. In other words, the robust NMF is supposed to treat each sample with different confidence. We can naturally implement the idea by minimizing the weighted sum of each reconstruction error $\|(X - UV^T)_i\|_2^2$. Thus, we propose the following robust weighted NMF (RWNMF) model

$$\begin{aligned} & \min_{U, V, W} \sum_i W_i * \|(X - UV^T)_i\|_2^2 + \alpha \|W\|_F^2 \\ \iff & \min_{U, V, W} \operatorname{Tr}(X - UV^T)W(X - UV^T)^T + \alpha \|W\|_F^2, \\ & \text{s.t. } U \geq 0, \quad V \geq 0, \quad W_i \geq 0, \quad \sum_{i=1}^n W_i = C_1, \end{aligned} \quad (3)$$

where C_1 is a positive constant, and α is a regularization parameter to balance the weight distribution. W is a diagonal weight matrix with diagonal element W_i , which represents the weight of the i -th sample. Specifically, if removing the second regularization term or setting α as 0, we will obtain the trivial solution, i.e., only one weight is assigned to C_1 and other weights are 0. On the contrary, if $\alpha \rightarrow \infty$, all weights will be assigned to the same values, and then RWNMF is equivalent to the original NMF.

III. ROBUST ADAPTIVE GRAPH REGULARIZED NMF

One important property is that NMF has non-negative constraints on U and V . This helps NMF to obtain a parts-based representation [21] of the given data X . So NMF can provide a superior physiological and psychological interpretation for non-negative data, including documents, face images and etc. However, by observing (1) we can get that NMF only focuses on the global data structures of X and ignores the local relationship between data. So many graph based NMF models have been developed to further capture the local

geometrical structures. Among these methods, GNMF [21] is a well-known effective method, which incorporates graph Laplacian into NMF to exploit the data correlation of neighboring samples. To be specific, the objective function of GNMF can be seen as follows

$$\min_{U \geq 0, V \geq 0} \|X - UV^T\|_F^2 + \lambda \text{Tr}(V^T L_S V), \quad (4)$$

where $L_S = D_S - S$ is called graph Laplacian. S is data similarity matrix, and D_S is degree matrix, which is a diagonal matrix with $D_i = \sum_j^n S_{ij}$ as the diagonal element. The local data structures are integrated by the added second term, which enforces that the new representation of data samples will also be close to each other if the samples are close in the original geometrical space.

However, many graph regularized NMFs do not build a discriminative graph to capture the intrinsic data structures. On one hand, most graph construction methods need to compute the distances between all pairs of data samples. The graph will be low-quality if the computed distances are not accurate enough. On the other hand, once the graph is constructed, it will keep fixed in the subsequent steps. So the clustering performance of NMF will be limited if the input graph is not optimal. Therefore, it is very necessary to build a high-quality graph [28].

In most cases, similarity matrix S is computed by the following Gaussian kernel function

$$S_{ij} = e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}} = e^{-\frac{(x_i - x_j)^T(x_i - x_j)}{2\sigma^2}}, \quad (5)$$

where S_{ij} stands for the similarity between i -th and j -th samples, and σ is Gaussian kernel parameter to control the values of similarity. However, the traditional Euclidean distance is not appropriate to measure the real distances between samples because the original feature space usually contains some noisy and irrelevant features. Naturally, some discriminative features should have higher weights to compute the distances. Therefore, it is desirable to learn a new metric matrix M to obtain more accurate distances. By incorporating matrix M , the new distances are computed by

$$\|x_i - x_j\|_M^2 = (x_i - x_j)^T M (x_i - x_j), \quad (6)$$

where M is a diagonal matrix for simplification. In fact, the above distance metric (6) is weighted Euclidean distance. The more appropriate similarity can be obtained by replacing $\|x_i - x_j\|_2^2$ with (6).

In the reconstruction problem of NMF, we incorporate the metric matrix M to reconstruct the given data matrix X . Therefore, by integrating M into GNMF (4), we get the following adaptive graph regularized NMF model

$$\begin{aligned} \min_{U, V, M} & \quad \text{Tr}(X - UV^T)^T M (X - UV^T) \\ & + \lambda \text{Tr}(V^T L_S V) + \beta \|M\|_F^2, \\ \text{s.t. } & U \geq 0, \quad V \geq 0, \quad M_i \geq 0, \quad \sum_{i=1}^m M_i = C_2, \end{aligned} \quad (7)$$

where C_2 is another positive constant, and β is a regularization parameter. The third term can bring the same benefit as $\alpha \|W\|_F^2$ in problem (3). M_i is the i -th diagonal element of M . Note that when $\beta \rightarrow \infty$, model (7) is equivalent to GNMF. Actually, model (7) can be viewed as an unsupervised metric learning [29] method to learn the metric matrix M . Besides, it is also a feature selection [30] method because the learned M can measure the importance of each feature and select the discriminative features. By combining the robust weighted NMF (3), we propose the following robust adaptive graph regularized NMF (RAGNMF) model

$$\begin{aligned} \min_{U, V, W, M} & \quad \text{Tr}[M(X - UV^T)W(X - UV^T)^T] \\ & + \lambda \text{Tr}(V^T L_S V) + \alpha \|W\|_F^2 + \beta \|M\|_F^2, \\ \text{s.t. } & U \geq 0, \quad V \geq 0, \quad W_i \geq 0, \quad M_i \geq 0, \\ & \sum_{i=1}^n W_i = C_1, \quad \sum_{i=1}^m M_i = C_2, \\ & S_{i,j} = e^{-\frac{\|x_i - x_j\|_M^2}{2\sigma^2}} = e^{-\frac{(x_i - x_j)^T M (x_i - x_j)}{2\sigma^2}}. \end{aligned} \quad (8)$$

Note that when $\alpha \rightarrow \infty$ and $\beta \rightarrow \infty$, the proposed RAGNMF is equivalent to GNMF. Moreover, the solution of the proposed RAGNMF is uncertain like GNMF [21]. Suppose that U and V are the optimal solutions of RAGNMF, and then for any positive diagonal matrix Λ , $U\Lambda$ and $V\Lambda^{-1}$ will have the same value in the first term of RAGNMF (8). To eliminate the uncertainty, we also adopt the same strategy (Euclidean length of each column in U is set to 1, and V is accordingly adjusted so that UV^T does not change) as GNMF [21].

IV. OPTIMIZATION

In this section, we will derive an efficient optimization algorithm to solve the proposed RAGNMF (8). Because RWNMF is a special case of RAGNMF, it also can be solved effectively. To be specific, the proposed objective function (8) with four variables is solved by alternative updating strategy, i.e., when the problem is optimized with respect to one variable, the other variables are fixed, and then repeat the procedure until converge. The detailed iterative updating method is shown as follows.

Update U and V : fix W and M , and then U , V can be solved as follows

$$\begin{aligned} \min_{U \geq 0, V \geq 0} & \quad \text{Tr}[M(X - UV^T)W(X - UV^T)^T] \\ & + \lambda \text{Tr}(V^T L_S V). \end{aligned} \quad (9)$$

Let A_{ik} and B_{jk} be the Lagrange multipliers for U_{ik} and V_{jk} respectively. So the Lagrangian function can be obtained as

$$\begin{aligned} \min_{U \geq 0, V \geq 0} & \quad \text{Tr}[M(X - UV^T)W(X - UV^T)^T] \\ & + \text{Tr}(AU^T) + \text{Tr}(BV^T) \\ & + \lambda \text{Tr}(V^T L_S V). \end{aligned} \quad (10)$$

Then we can get the following equations by making the partial derivatives of the above problem (10) to 0

$$2MUV^TWV - 2MXWV + A = 0, \quad (11)$$

$$2WVU^TMU + 2\lambda L_S V - 2WX^T MU + B = 0. \quad (12)$$

According to KKT conditions [31], we have

$$(MUV^TWV)_{ir}U_{ir} - (MXWV)_{ir}U_{ir} = 0, \quad (13)$$

$$(WVU^TMU + \lambda L_S V)_{jr}V_{jr} - (WX^T MU)_{jr}V_{jr} = 0. \quad (14)$$

The above equations produce the following updating rules

$$U_{ir} \leftarrow U_{ir} \frac{(MXWV)_{ir}}{(MUV^TWV)_{ir}}, \quad (15)$$

$$V_{jr} \leftarrow V_{jr} \frac{(WX^T MU)_{jr}}{(WVU^TMU + \lambda L_S V)_{jr}}. \quad (16)$$

Update W: fix U , V and M , and then W can be solved as follows

$$\begin{aligned} \min_W \operatorname{Tr}[E^M W] + \alpha \|W\|_F^2 \\ \text{s.t. } W_i \geq 0, \quad \sum_{i=1}^n W_i = C_1, \end{aligned} \quad (17)$$

where $E^M = (X - UV^T)^T M (X - UV^T)$. The above problem (17) can be converted to

$$\begin{aligned} \min_W \sum_{i=1}^n (W_i + \frac{1}{2\alpha} E_i^M)^2, \\ \text{s.t. } W_i \geq 0, \quad \sum_{i=1}^n W_i = C_1, \end{aligned} \quad (18)$$

where E_i^M is the diagonal element of E^M . The problem (18) is a quadratic programming (QP) problem, and has a closed form solution via the iterative algorithm [32].

Update M: fix U , V and W , and then M can be solved as follows

$$\begin{aligned} \min_M \operatorname{Tr}[E^W M] + \beta \|M\|_F^2, \\ \text{s.t. } M_i \geq 0, \quad \sum_{i=1}^m M_i = C_2, \end{aligned} \quad (19)$$

where $E^W = (X - UV^T)W(X - UV^T)^T$. The above problem (19) is similar to the problem (17) of updating W , and also can be solved by the iterative algorithm [32].

Finally, we summarize the detailed optimization steps of the proposed RWNMF and RAGNMF in Algorithm 1 and 2 respectively.

V. EXPERIMENTS

In this section, we conduct extensive experiments to verify the effectiveness of the proposed RWNMF and RAGNMF. Some synthetic data is firstly generated to prove the robustness of RWNMF, and then eight benchmark datasets are used to justify the clustering performance of RAGNMF. Note that in the following experiments, we set C_1 as n , the number of samples of X , and set C_2 as m , the number of features of X .

Algorithm 1 Robust Weighted NMF (RWNMF)

Input: Non-negative data X , the number of clusters k .

Parameter: Regularization parameter α .

Output: Basis matrix U , encoding matrix V , the learned weighted matrix W .

Initialize: weighted matrix W is set as identity matrix initially.

While not converged **do**

- 1) Update U according to $U_{ir} \leftarrow U_{ir} \frac{(XWV)_{ir}}{(UV^TWV)_{ir}}$.
- 2) Update V according to $V_{jr} \leftarrow V_{jr} \frac{(WX^T U)_{jr}}{(WVU^TU)_{jr}}$.
- 3) Update W by solving (18) (matrix M is set as identity matrix).

End while

Algorithm 2 Robust Adaptive Graph Regularized NMF (RAGNMF)

Input: Non-negative data X , the number of clusters k , the input graph S .

Parameter: Regularization parameters λ, α, β .

Output: Basis matrix U , encoding matrix V , the learned weighted matrix W and metric matrix M .

Initialize: weighted matrix W and metric matrix M are set as identity matrix initially.

While not converged **do**

- 1) Update graph similarity matrix S according to (5), (6).
- 2) Update U according to (15).
- 3) Update V according to (16).
- 4) Update W by solving (18).
- 5) Update M by solving (19).

End while

A. ROBUSTNESS OF RWNMF

1) EXPERIMENTS ON SYNTHETIC DATA

For verifying the robustness of the proposed RWNMF, some state-of-the-art robust NMF methods are considered as comparison algorithms. They are listed as follows in detail.

- 1) **the original NMF** [33]: using Frobenius norm to measure the reconstruction errors.
- 2) **$\ell_{2,1}$ -NMF** [14]: using $\ell_{2,1}$ norm to measure the reconstruction errors.
- 3) **$\ell_{2,0.5}$ -NMF** [15]: using $\ell_{2,0.5}$ norm to measure the reconstruction errors.
- 4) **the proposed RWNMF**: using weighted Frobenius norm to measure the reconstruction errors.

Some two-dimensional synthetic data points are used to validate the robustness of the proposed RWNMF. As shown in Fig. 1, 20 2-D points are generated in red, and obviously the top 6 points are outliers. After running the proposed RWNMF and other comparison methods (k is set as 1), we draw the reconstructed 1-D points of the original 2-D data. These reconstruction points of different algorithms have different shapes and colors. It is easy to observe from Fig. 1

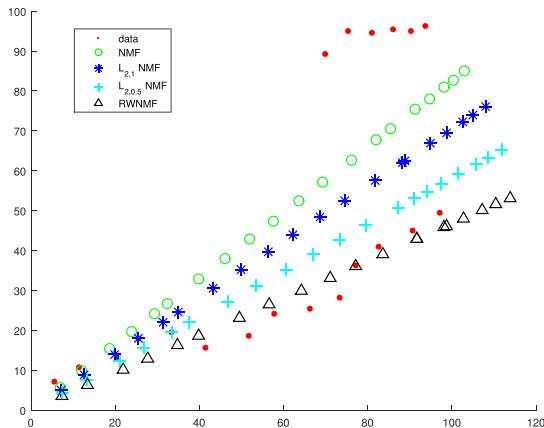


FIGURE 1. Reconstruction results of four methods on synthetic data.

TABLE 1. Weights of 6 outliers for all methods.

| Outliers Id | 1 | 2 | 3 | 4 | 5 | 6 |
|---------------------|--------|--------|--------|--------|--------|--------|
| NMF | 1 | 1 | 1 | 1 | 1 | 1 |
| $\ell_{2,1}$ -NMF | 0.0305 | 0.0292 | 0.0351 | 0.0326 | 0.0402 | 0.0390 |
| $\ell_{2,0.5}$ -NMF | 0.0018 | 0.0017 | 0.0020 | 0.0019 | 0.0022 | 0.0021 |
| RWNMF | 1.2e-5 | 1.2e-5 | 1.3e-5 | 1.3e-5 | 1.4e-5 | 1.3e-5 |

that the robustness is “NMF < $\ell_{2,1}$ -NMF < $\ell_{2,0.5}$ -NMF < RWNMF”. Specifically, the reconstruction points of the standard NMF strongly deviate the normal 2-D data, which indicates that the original NMF is strongly influenced by the top 6 outliers. The reason is that NMF uses the square of errors to minimize the loss, which leads to that only a few outliers or noises can easily dominate the loss function. Moreover, $\ell_{2,1}$ -NMF removes the square operation of reconstruction errors by considering $\ell_{2,1}$ norm instead of Frobenius norm to define the objective function. So $\ell_{2,1}$ -NMF has better robustness than the original NMF. $\ell_{2,0.5}$ -NMF further improves the robustness of NMF by replacing the square operation with square root operation, which decreases the adverse effects of noises and outliers.

However, the above robust NMF models, $\ell_{2,1}$ -NMF and $\ell_{2,0.5}$ -NMF, can still not avoid the negative influence of noises and outliers. Obviously, the black triangle points in Fig. 1 accurately reconstruct the given 2-D points and are not influenced by the 6 outliers. Therefore, the proposed RWNMF is the most robust method of all comparison algorithms. This is because the proposed RWNMF can automatically find the outliers and set their weights as almost 0. Besides, The fact is also proven by Table 1, which shows the weight values of six outliers. It can be observed that all weights of the original NMF are 1, because NMF treats equally all data points. The weights of $\ell_{2,1}$ -NMF, $\ell_{2,0.5}$ -NMF and the proposed RWNMF are lower than that of the original NMF. So their robustness is better than NMF. Furthermore, the proposed RWNMF has the smallest weights of six outliers (close to 0). Therefore, RWNMF can prevent the reconstruction results from being influenced by noises and outliers and acquire the best robustness.



FIGURE 2. Some noisy face images of ORL dataset. The upper face images is polluted by 4 × 4 block noises, and the lower face images is polluted by 8 × 8 block noises.

Note that in the experimental results Fig. 1 and Table 1, the value of parameter α is 50. The similar robust results can be obtained when α changes from 20 to 300. If the value of α is too small or too large, the performance of RWNMF will be bad. For instance, when $\alpha = 1$, the last 16 points are considered as outliers and have the weights of 0. When $\alpha = 1000$, all points have almost the same weights of 1 and the reconstruction results are influenced by outliers like the original NMF.

2) EXPERIMENTS ON REAL-WORLD DATA

To further verify the robustness and effectiveness of the proposed RWNMF, we also use real-world data to carry out some experiments. A face recognition dataset ORL is employed to perform the experiments. The detailed descriptions of ORL are in section V-B.2. We resize all images to 23×28 and randomly add some noises to four images of each individual. All noises are square blocks with different sizes. As shown in Fig. 2, it exhibits some noisy face images of ORL dataset. Some block noises cover the key face parts, e.g., eyes, noses, and mouths.

In the following experiments, the sizes of block noises are set as 2×2 , 4×4 , 6×6 , and 8×8 . Using ACC and NMI (their detailed descriptions are in section V-B.1) as the evaluation metrics, the clustering results are reported in Table 2. Note that the value of α is set as 10. It is similar to the experiments on synthetic data that α can not be too small or too large. It can be seen easily from Table 2 that our proposed RWNMF can always obtain the best clustering results in the ORL dataset with noises of different sizes. The fact suggests that the proposed RWNMF has the best robustness than the original NMF, and two other improved NMFs. Moreover, another interesting thing is that when the sizes of noises are small (e.g., 2×2), ACC of the proposed RWNMF only improves less than 1% comparing with the original NMF. When the added noises have the larger sizes (e.g., 8×8), ACC of the proposed RWNMF improves more than 4%. This indicates that the proposed RWNMF can effectively decrease the negative influence of large noises and improve the robustness of the original NMF.

TABLE 2. Experimental results of ORL face dataset with some noises of different sizes.

| Size | Metric | NMF | $\ell_{2,1}$ -NMF | $\ell_{2,0.5}$ -NMF | RWNMF |
|------|--------|--------|-------------------|---------------------|---------------|
| 2×2 | ACC | 0.6727 | 0.6730 | 0.6742 | 0.6821 |
| | NMI | 0.8266 | 0.8309 | 0.8314 | 0.8336 |
| 4×4 | ACC | 0.6520 | 0.6560 | 0.6534 | 0.6674 |
| | NMI | 0.8089 | 0.8080 | 0.8035 | 0.8189 |
| 6×6 | ACC | 0.5495 | 0.5617 | 0.5724 | 0.5863 |
| | NMI | 0.7386 | 0.7479 | 0.7543 | 0.7578 |
| 8×8 | ACC | 0.3888 | 0.3915 | 0.4152 | 0.4298 |
| | NMI | 0.6130 | 0.6101 | 0.6258 | 0.6332 |

B. CLUSTERING PERFORMANCE OF RAGNMF

1) EVALUATION METRICS

In the subsection, two commonly used evaluation indexes of measuring the clustering performance are introduced. They are given as below in detail.

Clustering Accuracy (ACC) finds the one-to-one relationship between true classes and clustering results, and obtains the data samples that each cluster has from the corresponding class. The detailed definition can be seen as follows

$$ACC = \frac{\sum_{i=1}^n \delta(\text{map}(r_i), l_i)}{n}, \quad (20)$$

where r_i represents the clustering result of x_i , and l_i stands for the true label of data x_i . n is the number of the whole samples. $\text{map}(r_i)$ denotes the optimal map function, and it can acquire a permutation to best match the clustering labels with the true labels. Besides, $\delta(a, b)$ represents the delta function which equals 1 if $a = b$, and equals 0 otherwise [14], [22].

Normalized Mutual Information (NMI) is another widely used quantitative metric to measure the quality of clustering. To be specific, it is defined as follows.

$$NMI = \frac{\sum_{i=1}^c \sum_{j=1}^c n_{i,j} \log \frac{n_{i,j}}{n_i \hat{n}_j}}{\sqrt{(\sum_{i=1}^c n_i \log \frac{n_i}{n})(\sum_{j=1}^c \hat{n}_j \log \frac{\hat{n}_j}{n})}}, \quad (21)$$

where c represents the number of classes, n_i denotes the number of data belonging to the cluster C_i ($1 \leq i \leq c$), \hat{n}_j is the number of data included in the class L_j ($1 \leq j \leq c$), and $n_{i,j}$ stands for the number of overlapped data between class L_j and cluster C_i [14], [22].

2) DATASET DESCRIPTIONS

We perform the data clustering experiments on eight public available datasets, which include four image datasets (ORL, Yale, USPS, and JAFFE), and four non-image datasets (Iris, Ecoli, Vote, and Control) from UCI Machine Learning Repository [34]. Among the image datasets, ORL, Yale, JAFFE are face datasets, and USPS is handwritten digit dataset. The detailed statistics of the above eight datasets are summarized in Table 3.

TABLE 3. Descriptions of eight datasets.

| Datasets | Samples (n) | Dimensions (m) | Class (c) |
|----------|-----------------|--------------------|---------------|
| ORL | 400 | 168 | 40 |
| Yale | 165 | 256 | 15 |
| USPS | 300 | 256 | 10 |
| JAFFE | 213 | 256 | 10 |
| Iris | 150 | 4 | 3 |
| Ecoli | 336 | 7 | 8 |
| Vote | 435 | 16 | 2 |
| Control | 600 | 60 | 6 |

3) COMPARISON METHODS

To prove the effectiveness of the proposed RAGNMF for clustering, we employ other eight clustering methods as the comparison algorithms, which are

- 1) **k-means** [35]: the widely used clustering algorithm.
- 2) **NMF** [33]: the standard NMF using Frobenius norm to define the loss function.
- 3) **$\ell_{2,1}$ -NMF** [14]: a robust NMF model using $\ell_{2,1}$ norm to define the loss function.
- 4) **$\ell_{2,0.5}$ -NMF** [15]: another robust NMF model using $\ell_{2,0.5}$ norm to define the loss function.
- 5) **RWNMF**: the proposed robust NMF model using adaptive weights to define the loss function.
- 6) **GNMF** [21]: the classical graph-based NMF combining the local data structures with NMF.
- 7) **RMNMF** [22]: an improved graph-based NMF which further employs $\ell_{2,1}$ norm to improve the robustness.
- 8) **NMFAN** [25]: another improved graph-based NMF which uses adaptive neighbors to better explore manifold structures.

4) EXPERIMENTAL SETTINGS

For all eight datasets, their values are normalized to [0,1] firstly. For GNMF and RMNMF, the Gaussian kernel parameter σ of graph Laplacian is selected in the range of {0.1, 0.2, 0.5, 0.8, 1, 1.5, 2, 4}. For GNMF, RMNMF and NMFAN, we set the regularization parameters by searching the grid {0.001, 0.01, 0.1, 1, 10, 100}. For k-means, NMF, $\ell_{2,1}$ -NMF and $\ell_{2,0.5}$ -NMF, no parameters need to be tuned.

For the proposed RAGNMF, the Gaussian kernel parameter σ is set as the same value of GNMF. Parameter λ is selected by searching from {0.001, 0.01, 0.1, 1, 10, 100}, and parameters α and β are set by searching the grid {1, 5, 10, 20, 50, 100}.

Because one or more parameters of some comparing algorithms need to be tuned, we run all methods with different parameter settings and report the best clustering results. Besides, we set k as the true label number of all datasets for each method. Note that k-means is used as post-processing method for NMF, $\ell_{2,1}$ -NMF, $\ell_{2,0.5}$ -NMF, GNMF, NMFAN, and the proposed RWNMF, RAGNMF to obtain the final clustering results. The detailed clustering steps are as follows. For the above NMF-based methods, they all can obtain an

TABLE 4. Experimental results measured by ACC of nine clustering methods.

| Datasets | k-means | NMF | $\ell_{2,1}$ -NMF | $\ell_{2,0.5}$ -NMF | RWNMF | GNMF | RMNMF | NMFAN | RAGNMF |
|----------|---------|--------|-------------------|---------------------|--------|--------|--------|--------|---------------|
| ORL | 0.6645 | 0.6802 | 0.6877 | 0.6880 | 0.6898 | 0.6842 | 0.6942 | 0.6986 | 0.7058 |
| Yale | 0.5148 | 0.5185 | 0.5381 | 0.5354 | 0.5367 | 0.5230 | 0.5124 | 0.5267 | 0.5486 |
| USPS | 0.6785 | 0.6313 | 0.6487 | 0.6497 | 0.6523 | 0.6563 | 0.6356 | 0.6765 | 0.6820 |
| JAFFE | 0.8333 | 0.8704 | 0.8948 | 0.9052 | 0.9126 | 0.9089 | 0.8932 | 0.9185 | 0.9342 |
| Iris | 0.8933 | 0.8013 | 0.8238 | 0.8279 | 0.8356 | 0.9507 | 0.9435 | 0.9535 | 0.9707 |
| Ecoli | 0.5747 | 0.5658 | 0.5458 | 0.5512 | 0.5647 | 0.7211 | 0.7158 | 0.7235 | 0.7476 |
| Vote | 0.8387 | 0.8018 | 0.8046 | 0.8053 | 0.8096 | 0.8023 | 0.8215 | 0.8483 | 0.8630 |
| Control | 0.4357 | 0.4120 | 0.4228 | 0.4236 | 0.4275 | 0.5732 | 0.5576 | 0.5835 | 0.5886 |

TABLE 5. Experimental results measured by NMI of nine clustering methods.

| Datasets | k-means | NMF | $\ell_{2,1}$ -NMF | $\ell_{2,0.5}$ -NMF | RWNMF | GNMF | RMNMF | NMFAN | RAGNMF |
|----------|---------|--------|-------------------|---------------------|--------|--------|--------|--------|---------------|
| ORL | 0.8328 | 0.8347 | 0.8413 | 0.8419 | 0.8434 | 0.8433 | 0.8418 | 0.8458 | 0.8521 |
| Yale | 0.5789 | 0.5541 | 0.5673 | 0.5661 | 0.5726 | 0.5521 | 0.5486 | 0.5776 | 0.5891 |
| USPS | 0.6176 | 0.5914 | 0.5936 | 0.5924 | 0.5942 | 0.6071 | 0.5983 | 0.6123 | 0.6387 |
| JAFFE | 0.8632 | 0.8795 | 0.8872 | 0.8957 | 0.9015 | 0.9175 | 0.9045 | 0.9203 | 0.9368 |
| Iris | 0.7515 | 0.6381 | 0.6453 | 0.6524 | 0.6689 | 0.8420 | 0.8352 | 0.8515 | 0.8942 |
| Ecoli | 0.5402 | 0.5048 | 0.4567 | 0.4615 | 0.4831 | 0.5757 | 0.5682 | 0.5712 | 0.5907 |
| Vote | 0.3738 | 0.3065 | 0.3064 | 0.3078 | 0.3165 | 0.3078 | 0.3568 | 0.3921 | 0.4254 |
| Control | 0.6521 | 0.6003 | 0.6163 | 0.6165 | 0.6179 | 0.7132 | 0.6987 | 0.7186 | 0.7203 |

encoding matrix $V \in \mathbb{R}^{n \times k}$. Considering each row of V as a new representation of each data sample, we use k-means to get the clustering results like [14], [21], [25]. Furthermore, we repeat all algorithms ten times and show the average clustering results.

5) EXPERIMENTAL RESULTS AND ANALYSES

After carrying out each method under their best parameter setups, we report the experimental results of eight datasets in Table 4 and 5. They are shown by the above aforementioned evaluation metrics ACC and NMI. By observing carefully the clustering results in Table 4 and 5, we have the following detailed experimental analyses.

- 1) These graph based NMF methods (GNMF, RMNMF, NMFAN and the proposed RAGNMF) achieve the better clustering performance than k-means and the original NMF in most datasets. This is because graph based NMFs utilize similarity graph to exploit the local geometrical structures of the given data. Thus global and local structures are jointly learned for NMF to improve significantly the clustering performance.
- 2) The clustering results of $\ell_{2,1}$ -NMF, $\ell_{2,0.5}$ -NMF, RMNMF, and the proposed RWNMF, RAGNMF are better than the standard NMF. Among these algorithms, $\ell_{2,1}$ -NMF and RMNMF use $\ell_{2,1}$ norm to improve the robustness of NMF, while the proposed RWNMF and RAGNMF employ the weighted Frobenius norm to decrease the sensitivity to noises and outliers. This suggests that robust NMF can improve the clustering performance of NMF. Besides, we find that the proposed RWNMF outperforms $\ell_{2,1}$ -NMF and $\ell_{2,0.5}$ -NMF by further observation. The main reason is that RWNMF can avoid the adverse influence of large noises and outliers.

- 3) It can be clearly observed that NMFAN and the proposed RAGNMF show some improvement compared to other graph regularized NMFs, i.e., GNMF and RMNMF. Although all the above four methods employ the graph Laplacian to explore the local data structures, GNMF and RMNMF only use the input graph, which is fixed in the whole process of matrix factorization. The initial graph may be not appropriate enough to regularize NMF and will restrict the clustering performance of NMF. Different from GNMF and RMNMF, NMFAN and the proposed RAGNMF learn an adaptive graph to well regularize NMF and achieve the better clustering results.
- 4) One can be noticed that the proposed RAGNMF achieves the best clustering results in all eight datasets. The fact suggests that the proposed RAGNMF can bring respectable improvement for NMF and is a superior clustering method. The main advantages are summarized as follows. First, the proposed RAGNMF is a robust NMF because the large noises and outliers have very small weights to dominate the whole loss. Second, the proposed RAGNMF also performs graph learning and uses an adaptive graph to exploit the local nonlinear information. Thus both global and local geometrical structures are fully exploited and utilized.

6) PARAMETER TUNING

In the subsection, we will show the clustering accuracy of the proposed RAGNMF under different parameter settings. Three important parameters (i.e., λ , α and β) need to be tuned for the proposed RAGNMF. Specifically, λ , α and β balance graph learning term, adaptive weights term and metric learning term respectively. When tuning the parameter λ , we set other parameters as their optimal values. The same

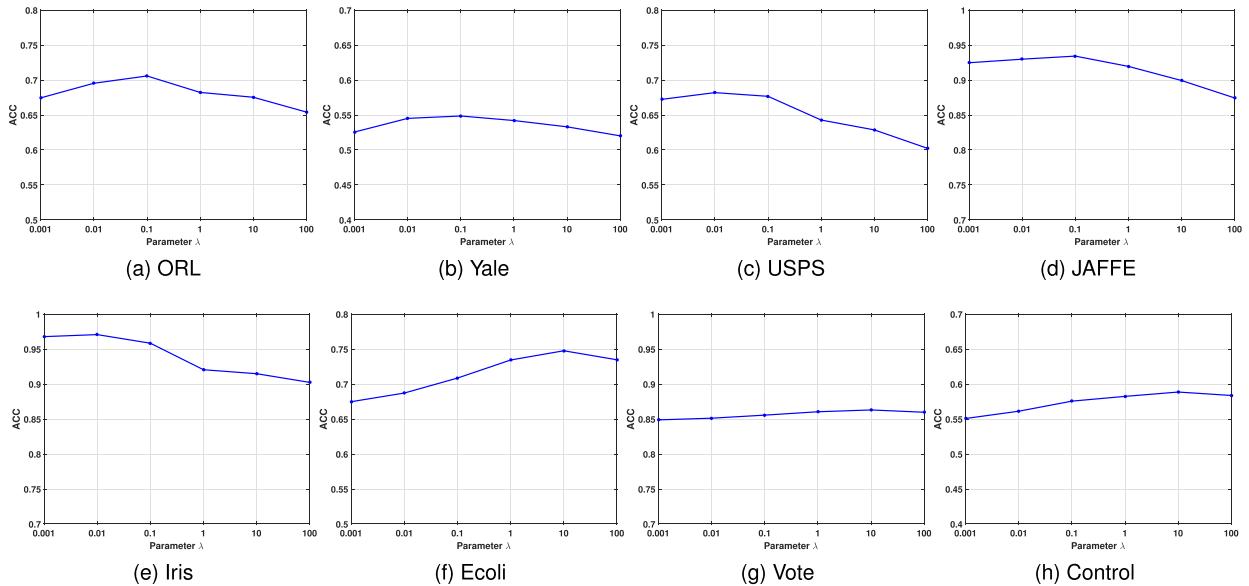


FIGURE 3. ACC of the proposed RAGNMF with respect to the parameter λ in eight datasets.

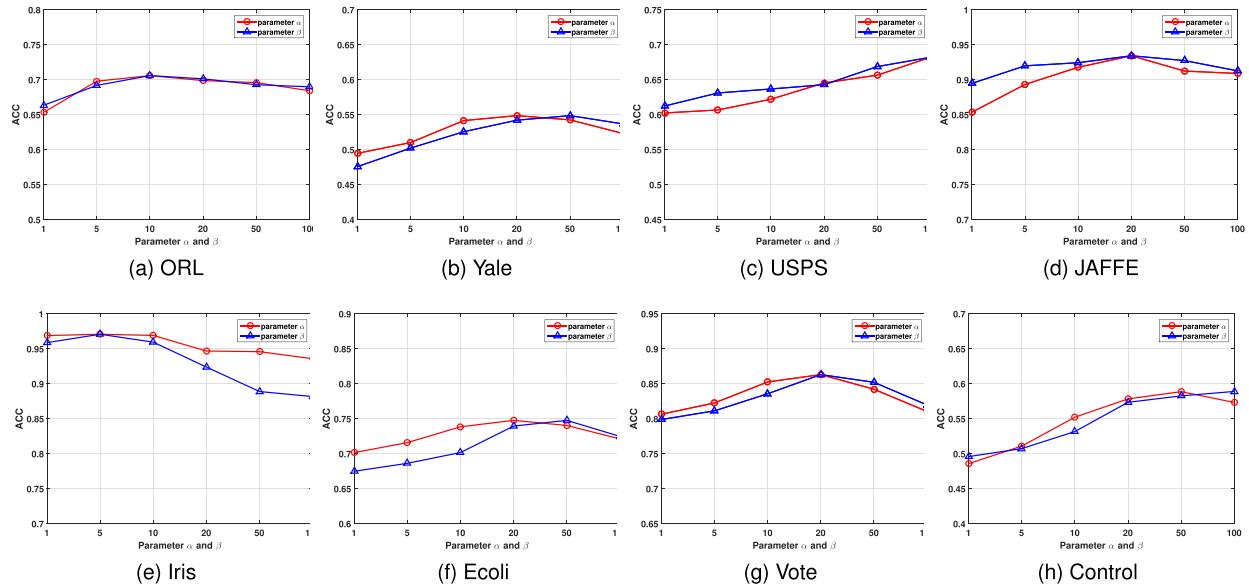


FIGURE 4. ACC of the proposed RAGNMF with respect to the parameters α and β in eight datasets.

setting is also for parameters α and β . Fig. 3 displays the ACC of RAGNMF when λ changes from $\{0.001, 0.01, 0.1, 1, 10, 100\}$. It is easy to see that the proposed RAGNMF is relatively stable to λ . As shown in Fig. 4, it exhibits the ACC of RAGNMF with respect to the parameters α and β in eight datasets. The proposed RAGNMF has poor performance only when the values of α and β are too large (i.e., Iris dataset) or too small (i.e., Control dataset). In a word, RAGNMF is only a little sensitive to α and β .

7) CONVERGENCE ANALYSES

Here we will make some convergence analyses of the proposed RAGNMF. The optimization process of solving RAGNMF is by updating four variables iteratively.

When updating U and V , the convergence is guaranteed by GNMF [21] and the subproblem has nonincreasing objective values. When updating W and M , Huang *et al.* [32] have proven that the global minimum solution can be obtained. Therefore, the proposed RAGNMF has nonincreasing objective values and converges to local optimal solution. Besides, some experiments are performed to verify the convergence of RAGNMF. As shown in Fig. 5, the proposed RAGNMF has good convergence performance in all eight datasets.

8) TIME COMPLEXITY

In this subsection, some experiments are conducted to validate the efficiency of RAGNMF. All experiments are performed in MATLAB R2015a in a Windows 7 machine

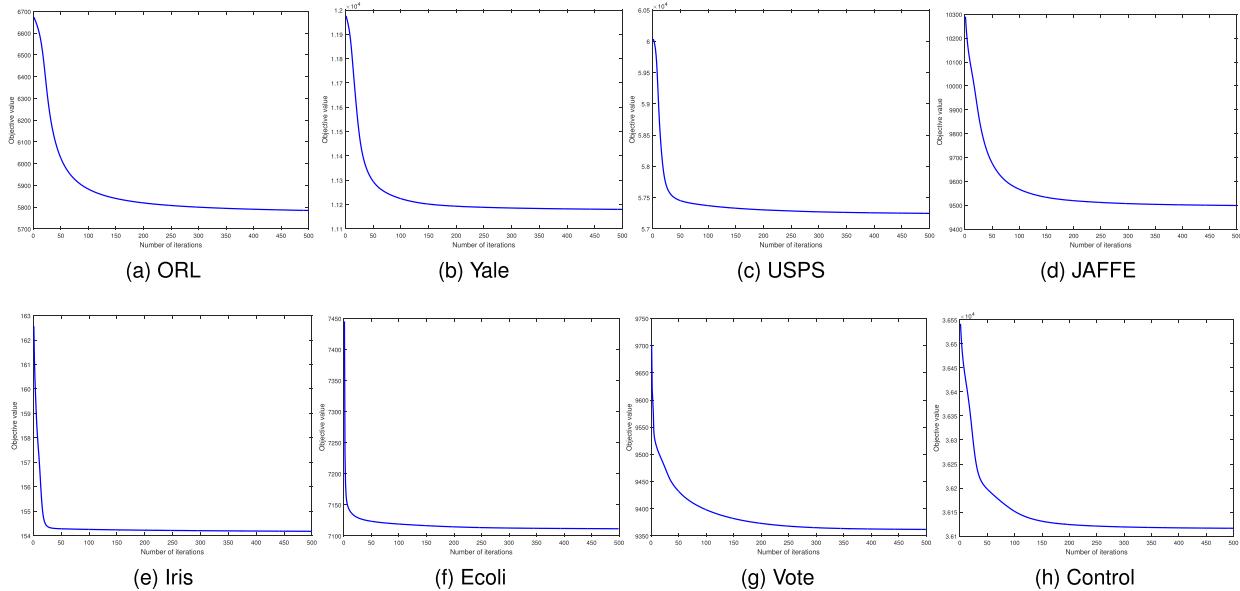


FIGURE 5. Convergence curves of the proposed RAGNMF in eight datasets.

TABLE 6. Comparison of running time in JAFFE dataset.

| Methods | ACC | NMI | Time(s) |
|---------------------|--------|--------|---------|
| k-means | 0.8333 | 0.8632 | 0.11 |
| NMF | 0.8704 | 0.8795 | 0.23 |
| $\ell_{2,1}$ -NMF | 0.8948 | 0.8872 | 0.62 |
| $\ell_{2,0.5}$ -NMF | 0.9052 | 0.8957 | 0.64 |
| RWNMF | 0.9126 | 0.9015 | 0.94 |
| GNMF | 0.9089 | 0.9175 | 0.29 |
| RMNMF | 0.8932 | 0.9045 | 2.67 |
| NMFAN | 0.9185 | 0.9203 | 2.43 |
| RAGNMF | 0.9342 | 0.9368 | 2.13 |

with Intel Core i7-3770 3.40 GHz CPU and 32-GB RAM. JAFFE dataset is chosen as the experimental data. As shown in Table 6, it shows the clustering results and running times for all methods. From the table, we can observe that 1) k-means is the fastest method but it has the worst clustering results. 2) NMF, $\ell_{2,1}$ -NMF, $\ell_{2,0.5}$ -NMF, RWNMF and GNMF have low time complexity but their clustering results are not good enough. 3) RMNMF and NMFAN have promising performance but the running time is relatively high. 4) the proposed RAGNMF can achieve the best results and also has acceptable time complexity.

VI. CONCLUSION

In this paper, we propose two NMF-based methods, i.e., robust weighted NMF (RWNMF) and robust adaptive graph regularized NMF (RAGNMF). Most existing methods of NMF usually suffer from two main problems. First, they are not robust enough to handle with noises and outliers. Second, for graph based NMFs, they do not fully exploit the local structures by the fixed input graph. To cope with the above problems, weighted Frobenius norm is presented to define the reconstruction loss function of NMF, and then large noises

and outliers have very small influence to the reconstruction results. Thus the robustness of NMF is improved significantly. Then RWNMF is proposed with the idea. Besides, metric learning is incorporated into NMF to learn a metric matrix, which can be used to compute the more accurate distances between samples. The graph is then updated iteratively with the new distances. Finally, an adaptive graph is learned to better guide NMF. It is worth noting that the process also can extra finish the task of feature selection. Then RAGNMF is developed for data clustering by combining the RWNMF. Two efficient and elegant optimization algorithms are also presented to solve the proposed RWNMF and RAGNMF. Extensive experiments on a synthetic dataset and eight public benchmark datasets have demonstrated the robustness of RWNMF and the superior clustering ability of RAGNMF, respectively.

REFERENCES

- [1] W. Zhou, C. Wu, Y. Yi, and G. Luo, "Structure preserving non-negative feature self-representation for unsupervised feature selection," *IEEE Access*, vol. 5, pp. 8792–8803, 2017.
- [2] Z. Yang, Y. Zhang, W. Yan, Y. Xiang, and S. Xie, "A fast non-smooth nonnegative matrix factorization for learning sparse representation," *IEEE Access*, vol. 4, pp. 5161–5168, 2016.
- [3] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Oakland, CA, USA, 1967, vol. 1, no. 14, pp. 281–297.
- [4] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [5] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support vector clustering," *J. Mach. Learn. Res.*, vol. 2, pp. 125–137, Mar. 2002.
- [6] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788, 1999.
- [7] J. Bobadilla, R. Bojorquez, A. H. Esteban, and R. Hurtado, "Recommender systems clustering using Bayesian non negative matrix factorization," *IEEE Access*, vol. 6, pp. 3549–3564, 2018.
- [8] X. Luo, M. Zhou, M. Shang, S. Li, and Y. Xia, "A novel approach to extracting non-negative latent factors from non-negative big sparse matrices," *IEEE Access*, vol. 4, pp. 2649–2655, 2016.

- [9] W. Zhu and Y. Yan, "Joint linear regression and nonnegative matrix factorization based on self-organized graph for image clustering and classification," *IEEE Access*, vol. 6, pp. 38820–38834, 2018.
- [10] Z. Xiong, Y. Zang, X. Jiang, and X. Hu, "Document clustering with an augmented nonnegative matrix factorization model," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2014, pp. 348–359.
- [11] W.-S. Chen, Y. Li, B. Pan, and C. Xu, "Block kernel nonnegative matrix factorization and its application to face recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3446–3452.
- [12] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016.
- [13] A. B. Hamza and D. J. Brady, "Reconstruction of reflectance spectra using robust nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3637–3642, Sep. 2006.
- [14] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using L21-norm," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 673–682.
- [15] Z. Li, J. Tang, and X. He, "Robust structured nonnegative matrix factorization for image representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1947–1960, May 2018.
- [16] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 977–986.
- [17] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.
- [18] L. Zhang, Q. Zhang, B. Du, D. Tao, and J. You, "Robust manifold matrix factorization for joint clustering and feature extraction," in *Proc. AAAI*, 2017, pp. 1662–1668.
- [19] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3840–3853, May 2017.
- [20] C. Peng, Z. Kang, Y. Hu, J. Cheng, and Q. Cheng, "Robust graph regularized nonnegative matrix factorization for clustering," *ACM Trans. Knowl. Discovery Data*, vol. 11, no. 3, p. 33, 2017.
- [21] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
- [22] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust manifold nonnegative matrix factorization," *ACM Trans. Knowl. Discovery Data*, vol. 8, no. 3, p. 11, 2014.
- [23] K. Zeng, J. Yu, C. Li, J. You, and T. Jin, "Image clustering by hyper-graph regularized non-negative matrix factorization," *Neurocomputing*, vol. 138, pp. 209–217, Aug. 2014.
- [24] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1969–1976.
- [25] S. Huang, Z. Xu, and F. Wang, "Nonnegative matrix factorization with adaptive neighbors," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 486–493.
- [26] J. J.-Y. Wang, J. Z. Huang, Y. Sun, and X. Gao, "Feature selection and multi-kernel learning for adaptive graph regularized nonnegative matrix factorization," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1278–1286, 2015.
- [27] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.*, vol. 11, no. 3, p. 258, 2019.
- [28] L. Zhang, Q. Zhang, B. Du, J. You, and D. Tao, "Adaptive manifold regularized matrix factorization for data clustering," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 3399–3405.
- [29] J. Yu, D. Tao, J. Li, and J. Cheng, "Semantic preserving distance metric learning and applications," *Inf. Sci.*, vol. 281, pp. 674–686, Oct. 2014.
- [30] J.-Y. Wang, I. Almasri, and X. Gao, "Adaptive graph regularized nonnegative matrix factorization via feature selection," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 963–966.
- [31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [32] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015.
- [33] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.
- [34] D. Dheeru and E. Karra Taniskidou. (2017). *UCI Machine Learning Repository*. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [35] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.



XIANG HE received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2017, where he is currently pursuing the M.S. degree in computer science with the Center for Optical Imagery Analysis and Learning (OPTIMAL). His research interests include machine learning and computer vision.



QI WANG (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

XUELONG LI (M'02–SM'07–F'12) is currently a Full Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, Shaanxi, China.