IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 5, 2019, accepted August 2, 2019, date of publication August 8, 2019, date of current version August 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933877

Adaptive Graph Regularization Discriminant Nonnegative Matrix Factorization for Data Representation

LIN ZHANG, ZHONGHUA LIU, LIN WANG, AND JIEXIN PU
Information Engineering College, Henan University of Science and Technology, Luoyang 471000, China

Corresponding author: Zhonghua Liu (lzhua_217@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1504610, in part by the National Key Research and Development Project under Grant 2016YFE0104600, and in part by the Natural Science Foundation of Henan Province under Grant 192102210130 and Grant 19B520008.

ABSTRACT Nonnegative matrix factorization, as a classical part-based representation method, has been widely used in pattern recognition, data mining and other fields. However, the traditional nonnegative matrix factorization directly factoring decomposes the original data, and the original data often contains a lot of redundancy and noise, which seriously affect the subsequent processing of the data. In this work, we propose an adaptive graph regularization discriminant nonnegative matrix factorization (AGDNMF) for image clustering. The AGDNMF algorithm makes full use of local structure information and a small amount of label information. In AGDNMF, the local structure information can be more accurate and the label information can prevent the points with the same label from being merged into one point. These two items are combined into the objective function of NMF. In addition, we provide the update rules for the corresponding optimization functions and prove its convergence. A large number of experiments on different data sets show that the proposed algorithm has good clustering performance.

INDEX TERMS Nonnegative matrix factorization, graph regularization, discriminative information, image clustering, data representation.

I. INTRODUCTION

In today's information age, we can get the information we want regardless of time or place. As a carrier of the information, data exist in many real applications. For example, we take photos, upload videos, and write text. High-dimensional data although brought us good life experience, but there are still many problems in the process of dealing with data. Therefore, how to extract effective information from high-dimensional data becomes very significant. In recent years, data representation plays an important role in pattern recognition and image processing [1]–[3]. A suitable data representation is helpful to reveal the potential information structure of the data very well, which is convenient for the next processing. In recent years, matrix factorization technology has attracted more and more attention as a popular technology for data representation [4]–[6].

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

The matrix factorization technique consists of finding two or more low-dimensional matrices, so that their products are good enough to approximate the original data matrix. At present, some popular matrix decomposition techniques are vector quantization (VQ) [7], singular value decomposition (SVD) [8], principal component analysis (PCA) [9], and non-negative matrix factorization (NMF) [10], and so on. Different from the VQ, SVD, and PCA methods, the two matrices obtained by NMF decomposition are all non-negative.

Nonnegative matrix factorization adds a nonnegative constraint to the matrix factorization, that is, all elements in the decomposed matrix are greater than or equal to zero. Because nonnegativity properties allow only additive combinations of original data, not subtractive combinations, they lead to a parts-based representation for NMF. This attribute is consistent with the physiological and psychological elements based on the representation of the parts of the human brain. As an ideal algorithm, NMF has shown advantages in image clustering [11]–[13], face recognition [14]–[16],

pattern recognition [17]–[19], and text clustering [20]–[22] and so on [23]–[25]. In recent years, some researchers put forward some new algorithms by adding additional constraints to NMF. By using the label information and the non-negative coefficient matrix to construct the regularization constraints, Babaee and Tsoukalas [26] propose discriminative NMF (DNMF). By adding sparse constraints to the decomposed base matrix, Li *et al.* [27] proposed local nonnegative matrix factorization (LNMF). Liu and Wu [28] presented constrained NMF (CNMF) by utilizing label information, in which the data points with the same labels will have the same representation. Guan and Tao [29] proposed manifold regularization discriminant nonnegative matrix decomposition (MD-NMF), which aims to add orthogonal constraints to the base matrix. Shang and Jiao [30] considering that the properties of the data are distributed according to the manifold, and propose a nonnegative matrix based on graph dual regularization.

In recent years, some studies have shown that more data information is likely to come from low-dimensional manifold structures embedded in high-dimensional space. In order to find potential manifold structures, many manifold learning algorithms are proposed, such as local linear embedding (LLE) [31], isometric mapping (ISOMAP) [32], and Laplacian feature mapping (LE) [33]. These algorithms all make use of local invariance and have been proved that the learning performance of such algorithms is significantly improved. Combined with local geometry structure, Cai *et al.* [34] proposed graph regularized nonnegative matrix factorization (GNMF). In GNMF, the data structure is encoded by the nearest neighbor graph.

The traditional methods of constructing graphs are often based on the original data, which often contains other influencing factors such as noise. This can result in a constructed neighbor graph that is not optimal. And in the real world, the original data usually contains a small amount of label information. Inspired by these conditions, this paper proposes a new algorithm called adaptive graph regularized discriminate nonnegative matrix factorization (AGDNMF). We construct a neighbor graph by adaptively assigning weights. The label information of data is simultaneously combined into our model to improve the recognition ability. The proposed algorithm not only captures local structure information better, but also makes full use of label information. The main contributions of this algorithm are as follow

- In our method, nonnegative matrix factorization, manifold learning and the label information are integrated into a unified framework.
- Adaptive to obtain the constructed neighbor graph. This can better capture local structural information and get better recognition performance.
- The introduction of this label constraint item can ensure that when there are a large number of labelled data points, the data points will not disappear and the clustering performance will be increased.
- In this paper, we propose an iterative scheme. The experimental results show that the algorithm has a good clustering effect.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the adaptive graph regularized discriminate non-negative matrix factorization in detail, and gives the convergence proof of the algorithm. In section 4, a lot of comparison experiments are conducted to prove the performance of the algorithm. The final section draws a conclusion.

II. THE RELATED WORKS

Before introducing our algorithm, let us briefly introduce some of the work that is closely related to this paper.

A. NONNEGATIVE MATRIX FACTORIZATION(NMF)

Suppose that there is a data matrix with n samples $X = [x_1, x_2, \ldots, x_n] \in R^{m\times n}$, and each column of X represents a face image. The dimension of each image is m-dimension. K represents the number of clusters. The purpose of NMF is to find two matrices $U \in R^{m\times k}$ and $V \in R^{n\times k}$ to represent the original data set as much as possible. We use the square of the Euclidean distance as the objective function to measure its similarity:

$$O = \left\| X - UV^T \right\|_F^2 \tag{1}$$

where $\|\cdot\|_F$ is a Frobenius matrix. The variables U and V are not convex together in this algorithm. Therefore, it is difficult to find the global minimum of this algorithm. In order to solve this problem, Lee and Seung proposed an iterative algorithm to find the local optimal solution.

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{\left(UV^TV\right)_{ik}} \tag{2}$$

$$v_{jk} \leftarrow v_{jk} \frac{(X^TU)_{jk}}{\left(VU^TU\right)_{jk}} \tag{3}$$

B. GRAPH REGULARIZED NONNEGATIVE MATRIX FACTORIZATION(GNMF)

By introducing manifold regularized entries in the decomposition of the basic nonnegative matrix, Cai *et al.* [34] propose a graph of regularized nonnegative matrix decomposition. GNMF can keep the local structure information very well. The objective function of the GNMF is as follows:

$$\begin{aligned} &\min O = \left\| X - UV^T \right\|_F^2 + \lambda Tr\left(V^TLV\right) \\ &s.t.\ U \geq 0, \quad S \geq 0 \end{aligned} \tag{4}$$

where $Tr(\bullet)$ is the trace of the matrix. L is a Laplacian matrix.

This objective function is a convex function for variables U and V, so we find the local minimum by iteratively updating the rules.

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{\left(UV^TV\right)_{ik}} \tag{5}$$

$$v_{jk} \leftarrow v_{jk} \frac{\left(X^T U + \lambda WV\right)_{jk}}{\left(VU^T U + \lambda DV\right)_{jk}} \tag{6}$$

C. CONSTRAINED NONNEGATIVE MATRIX FACTORIZATION (CNMF)

Both nonnegative matrix factorization and graph regularization nonnegative matrix factorization are unsupervised learning algorithms, and they do not use label information. Liu and Wu [28] introduce label information as a constraint, so that sample points with the same label are reduced in dimension and mapped into the same class. Assume that the first l samples in the original data have labels and n-l samples have no labels. This algorithm introduces a label matrix C. When X_i is labelled with j-th class, $C_{ij} = 1$, otherwise $C_{ij} = 0$. Based on this we obtain a label constraint matrix A:

$$A = \begin{pmatrix} C_{l\times k} & 0 \\ 0 & I_{n-l} \end{pmatrix} \tag{7}$$

where I_{n-l} is an identity matrix.

The objective function of this algorithm can be expressed as:

$$O = \left\| X - U\,(AZ)^T \right\|_F^2 \tag{8}$$

The iterative update rules for this algorithm are as follows:

$$u_{ik} \leftarrow u_{ik} \frac{(XAZ)_{ik}}{\left(A^T AZU^T U\right)_{ik}} \tag{9}$$

$$z \leftarrow z \frac{\left(A^T X^T U\right)}{\left(A^T AZU^T U\right)} \tag{10}$$

III. ADAPTIVE GRAPH REGULARIZATION DISCRIMINANT NONNEGATIVE MATRIX FACTORIZATION (AGDNMF)

With the research on manifold learning in recent years, the local structure can represent data better than global structure. The GNMF algorithm preserves the local structural information by constructing neighbor graphs, but the construction of traditional neighbor graphs has the following two disadvantages: First, it is sensitive to the value of the parameters. Second, the weight of the construct depends on the original data. Therefore, the obtained neighbor graph is not optimal. At the same time, the weight matrix does not change after it is generated. Although CNMF considers the label information of the data, it doesn't use the manifold structure information. Based on these problems, we propose a semi-supervised adaptive graph regularization discriminant nonnegative matrix factorization (AGDNMF). We build weight maps by adaptively and add label information as constraints. AGDNMF can perform graph construction and nonnegative matrix factorization simultaneously and satisfies points with the same label in low-dimensional space without merging into one point.

A. AGDNMF MODEL

We firstly define a similarity matrix S, the probability between any two points v_i and v_j is s_{ij}.The greater the similarity between the two samples is, the closer the distance between the two points is. We can determine S according to the following.

$$\min \sum_{i,j}^{n} \left\| V_i - V_j \right\|_2^2 s_{ij}$$
$$s.t.\ \forall i, \quad s_i^T 1 = 1,\ 0 \le s_{ij} \le 1 \tag{11}$$

where s_i is the i-th column vector in similarity matrix S.

In spectral analysis, $Ls = D - \frac{S^T+S}{2}$ is called a Laplacian matrix. D is the degree matrix, its i-th diagonal element is $\sum_j \frac{s_{ij}+s_{ji}}{2}$. Given $F = [f_1, \ldots, f_2] \in R^{n\times k}$, we can get classical spectral clustering:

$$\sum_{i,j} \left\| f_i - f_j \right\|_2^2 s_{ij} = 2Tr\left(F^T LsF\right) \tag{12}$$

The neighbor graph in the ideal state wants the similarity matrix to contain k connected components, while the traditional similarity matrix generally has only one connected component. So we use an important property of Laplace to solve this problem.

Theorem 1: In the Laplacian matrix, the number of zero in the eigenvalue is equal to the number of connected regions of the graph. According to this theorem, we add this constraint to Equation (11), we have:

$$\min \sum_{i,j} \left\| v_i - v_j \right\|_2^2 s_{ij}$$
$$s.t.\ \forall i, \quad s_i^T = 1,\ 0 \le s_{ij} \le 1,\ rank\,(Ls) = n - k \tag{13}$$

This problem is difficult to solve because this constraint is dependent on the similarity matrix S. To solve this problem, we can define the i-th minimum eigenvalue of the Ls matrix as $\sigma_i\,(Ls)$. Since Ls is a positive semi-definite matrix, we can get $\sigma_i\,(Ls) \ge 0$. At this point, $rank\,(Ls) = n - k$ can be expressed as $\sum_{i=1}^{c} \sigma_i\,(Ls) = 0$. According to Ky Fan's theorem (Fan 1949), we can get

$$\sum_{i=1}^{k} \sigma_i\,(Ls) = \min_{F\in R^{n\times k}, F^T F = I} Tr\left(F^T LsF\right) \tag{14}$$

Equation (13) can be rewritten as

$$\min \sum_{i,j} \left\| v_i - v_j \right\|_2^2 s_{ij} + 2\lambda Tr\left(F^T LsF\right)$$
$$s.t. \forall i, \quad s_i^T = 1,\ 0 \le s_{ij} \le 1,\ F \in R^{n\times k}, F^T F = 1 \tag{15}$$

In order to get better clustering results, a little label information contained in the data is used. We introduce matrix $Q \in R^{c\times n}$ as follows:

$$Q_{ij} = \begin{cases} 1 & if\ the\ sample\ j\ is\ labeled \\ & and\ is\ from\ class\ i \\ 0 & otherwise \end{cases} \tag{16}$$

where c is the class number of the observed data. Suppose we have $n = 6$ samples, $n_l = 4$ of which are labelled as $c_1 = 1$,

$c_2 = 3, c_3 = c_4 = 2$. According to this matrix Q can be expressed as

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{17}$$

We introduce this introduced matrix as a constraint into NMF.

$$O_l = \left\| Q - AV_l^T \right\|^2 \tag{18}$$

where $V_l = [v_1, v_2, \cdots, v_{n_l}, 0, 0, \cdots 0]^T \in R^{n\times k}$. Matrix $A \in R^{c\times k}$ is an auxiliary matrix and its value can be negative, its purpose is to get better Q to get the best value. We apply the above constraints to the NMF, and finally we get the objective function of AGDNMF as follows.

$$\begin{aligned} \min O = & \left\| X - UV^T \right\|_F^2 + \alpha \sum_{i,j} \left\| v_i - v_j \right\|_2^2 s_{ij} \\ & + 2\lambda Tr\left(F^T L s F\right) + \gamma \left\| Q - AV_l^T \right\|^2 \\ s.t.\ & \mathrm{U} \geq 0, \quad \mathrm{V} \geq 0, \ \forall i, \ s_i^T = 1, \ 0 \leq s_{ij} \leq 1, \\ & F \in R^{n\times c}, \quad F^T F = 1 \end{aligned} \tag{19}$$

B. THE UPDATE RULES OF AGDNMF

In the objective function Equation (19), the variables (U, V, S, F, A) are not simultaneously convex, so it is difficult to directly find the global minimum. Below we propose an iterative update algorithm to get the local optimal solution of the objective function O.

1) FIX U, V, A AND F, UPDATE S

With fixed U, V, A and F, the Equation (19) can be rewritten as

$$\begin{aligned} & \min \alpha \sum_{i,j} \left\| v_i - v_j \right\|_2^2 s_{ij} + 2\lambda Tr\left(F^T L s F\right) \\ & s.t.\ \forall i, \quad s_i^T = 1, \ 0 \leq s_{ij} \leq 1 \end{aligned} \tag{20}$$

According to Equation (12), we get

$$\begin{aligned} & \min \alpha \sum_{i,j} \left\| v_i - v_j \right\|_2^2 s_{ij} + \lambda \sum_{i,j} \left\| f_i - f_j \right\|_2^2 s_{ij} \\ & s.t.\ \forall i, \quad s_i^T = 1, \ 0 \leq s_{ij} \leq 1 \end{aligned} \tag{21}$$

We define the matrix M, let $m_{ij} = \left\| v_i - v_j \right\|_2^2$. And matrix N, let $n_{ij} = \left\| f_i - f_j \right\|_2^2$. We use the vector $d_{ij} = m_{ij} + \lambda n_{ij}$ to represent $d_i \in R^{n\times 1}$. We can rewrite Equation (21) as follows.

$$\begin{aligned} & \min_{s_i} \left\| s_i + \frac{1}{2\sigma} d_i \right\|_2^2 \\ & s.t.\ s_i^T = 1, \quad 0 \leq s_{ij} \leq 1 \end{aligned} \tag{22}$$

The variable σ is determined by an adaptive number of neighbors, and this variable changes during each iteration.

2) FIX U, V, A AND S, UPDATE F

With fixed U, V, A and S, the Equation (19) is transformed into

$$\min_{F\in R^{n\times c}, F^T F = 1} Tr\left(F^T L s F\right) \tag{23}$$

The optimal solution of F is formed from the eigenvectors of the k minimum eigenvalues in the Ls.

3) FIX S AND F, UPDATE U, V AND A

When S and F are fixed, the third term in Equation (19) can be considered as a constant. Equation (19) can be can be rewritten as follows.

$$\begin{aligned} O = & Tr((X - UV^T)(X - UV^T)^T) + \lambda Tr(V^T L V) \\ & + \gamma Tr((Q - AV_l^T)(Q - AV_l^T)^T) \\ = & Tr(XX^T) - 2Tr(XVU^T) + Tr(UV^T VU^T) \\ & + \lambda Tr(V^T L V) + \gamma Tr(QQ^T) - 2\gamma Tr(QV_l A^T) \\ & + \gamma Tr(AV_l^T V_l A^T) \end{aligned} \tag{24}$$

We introduce the Lagrange multipliers φ_{ik} and ϕ_{jk}, constrain $u_{ik} \geq 0$ and $v_{jk} \geq 0$. The corresponding Lagrangian function Γ can be written as

$$\begin{aligned} \Gamma = & -2Tr(XVU^T) + Tr(UV^T VU^T) + Tr(\varphi U) \\ & + Tr(\psi V) + \lambda Tr(V^T L V) - 2\gamma Tr(QV_l A^T) \\ & + \gamma Tr(AV_l^T V_l A^T) \end{aligned} \tag{25}$$

The partial derivative of the function Γ relative to U, V and A is

$$\begin{aligned} \frac{\partial \Gamma}{\partial U} &= -2XV + 2UV^T V + \varphi \\ \frac{\partial \Gamma}{\partial V} &= -2X^T U + 2VU^T U + 2\lambda LV \\ &\quad - 2\gamma Q^T A + 2\gamma V_l A^T A + \psi \\ \frac{\partial \Gamma}{\partial A} &= -2QV_l + 2AV_l^T V_l \end{aligned} \tag{26}$$

We use the KKT condition to solve the equations $\frac{\partial \Gamma}{\partial U} = 0$, $\frac{\partial \Gamma}{\partial V} = 0$, $\frac{\partial \Gamma}{\partial A} = 0$. We can get

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{(UV^T V)_{ik}} \tag{27}$$

$$v_{jk} \leftarrow v_{jk} \frac{(X^T U + \lambda WV + \gamma(V_l A^T A)^- + \gamma(Q^T A)^+)_{jk}}{(VU^T U + \lambda DV + \gamma(V_l A^T A)^+ + \gamma(Q^T A)^-)_{jk}} \tag{28}$$

$$A \leftarrow QV_l (V_l^T V_l)^{-1} \tag{29}$$

For any matrix W, W^+ and W^- are respectively defined: $W^+ = (|W| + W)/2$, and $W^- = (|W| - W)/2$. We introduce the following theorem to get the local optimal solution of convergence.

Theorem 2: The objective function in Equation (19) is non-increasing under the update rules in Equation (27). The objective function is invariant under this update rule if and only if U, V and A are at a stable point. This theorem ensures that when U, V and A converge to a point, a local optimal solution can be obtained.

C. CONVERGENCE PROOF

In this section we present the proof of Theorem 2. We can see that the iterative formula for U in the objective function is consistent with the NMF. In NMF algorithm, the author has proved that the objective function of NMF is non-increasing under this iterative formula. For this we only need to prove that the objective function of AGDNMF does not increase under the iteration formulas (28). We introduce a helper function that uses the expectation maximization algorithm to prove theorem 2.

Lemma 1: Assuming that there is a function G, when the condition $G\left(v, v'\right) \geq F\left(v\right), G\left(v, v\right) = F\left(v\right)$ is satisfied, the function f does not grow under the following rules.

$$v^{t+1} = \arg\min_{v} G\left(v, v^{t}\right) \tag{30}$$

$F\left(v^{t+1}\right) = F\left(v^{t}\right)$ is established only if the local minimum value of $G\left(v, v^{t}\right)$ is X^{t}. The local optimal value F, which converges to $v_{\min} = \arg\min_{v} F\left(v\right)$, can be obtained by iterative Equation (30). Therefore, we prove that the iteration of V is the same under update rules (28) (30) when constructing an appropriate helper function. $F_{v_{ab}}$ represents elements in v that are only related to v_{ab}, F', F'' are the first derivative and second derivative of the v_{ab}. F' and F'' are defined as follows.

$$\begin{aligned} F'_{ab}(V_{ab}) = &(-2X^{T}U + 2VU^{T}U + 2\lambda LV \\ &- 2\gamma Q^{T}A + 2\gamma V_{l}A^{T}A)_{ab} \end{aligned} \tag{31}$$

$$F''_{ab}(V_{ab}) = \begin{cases} 2(U^{T}U + \lambda L + \gamma(A^{T}A))_{bb} & if\ a \leq N_{l} \\ 2((U^{T}U)_{bb} + (\lambda L)_{aa}) & otherwise \end{cases} \tag{32}$$

Lemma 2: An auxiliary function $F_{v_{ab}}\left(v\right)$ for v_{ab} is defined as follows.

$$\begin{aligned} G(v_{ab}, v^{t}_{ab} = &F_{ab}(v^{t}_{ab}) + F'_{ab}(v^{t}_{ab})(v_{ab} - v^{t}_{ab}) \\ &+ \frac{1}{v^{t}_{ab}}[V^{t}U^{T}U + \lambda DV + \gamma(V^{t}_{l}A^{T}A)^{+} \\ &+ \gamma(Q^{T}A)^{-}]_{ab}(v_{ab} - v^{t}_{ab})^{2} \end{aligned} \tag{33}$$

Proof: Obviously, if $G\left(v, v\right) = F_{v_{ab}}\left(v\right)$, we only need to prove $G\left(v, v^{t}_{ab}\right) \geq F_{v_{ab}}\left(v\right)$. The Taylor expansion of $F_{v_{ab}}\left(v\right)$ at v^{t}_{ab} is as follows.

$$\begin{aligned} F_{ab}(v) = &F_{ab}(v^{t}_{ab}) + F'_{ab}(v^{t}_{ab})(v_{ab} - v^{t}_{ab}) \\ &+ \frac{1}{2}F''_{ab}(v^{t}_{ab})(v_{ab} - v^{t}_{ab})^{2} \end{aligned} \tag{34}$$

We prove that $G\left(v, v^{t}_{ab}\right) \geq F_{v_{ab}}\left(v\right)$, we only need to prove that the following inequality is true.

$$\begin{aligned} &\frac{1}{V^{t}_{ab}}[V^{t}U^{T}U + \lambda DV^{t} + \gamma(V^{t}_{l}A^{T}A)^{+} + \gamma(Q^{T}A)^{-}]_{ab} \\ &\geq \frac{1}{2}F''_{ab}(V^{t}_{ab}) \end{aligned} \tag{35}$$

TABLE 1. The description of data sets.

Dataset	Sample number (N)	Dimensionality (M)	Class number (C)
Yale	165	2000	15
ORL	400	644	40
COIL20	1440	1024	20
HD	10000	256	10
Jaffe	213	4096	10
AR	700	2000	50

Combining Equation (31) to (34), Equation (35) can be rewritten as

$$\begin{cases} \frac{1}{V^{t}_{ab}}[V^{t}U^{T}U + \lambda DV + \gamma(V^{t}_{l}A^{T}A)^{+} + \gamma(Q^{T}A)^{-}]_{ab} \\ \geq (U^{T}U + \gamma(A^{T}A))_{bb} + (\lambda L)_{aa} \\ \qquad if\ a \leq N_{l} \\ \frac{1}{V^{t}_{ab}}[V^{t}U^{T}U + \lambda DV + \gamma(V^{t}_{l}A^{T}A)^{+} + \gamma(Q^{T}A)^{-}]_{ab} \\ \geq (U^{T}U)_{bb} + (\lambda L)_{aa} \\ \qquad otherwise \end{cases} \tag{36}$$

We compare the inequalities on both sides.

$$\begin{aligned} (V^{t}U^{T}U)_{ab} &= \sum_{h=1}^{K} v^{(t)}_{ah}(U^{T}U)_{hb} \\ &= V^{t}_{ab}(U^{T}U)_{bb} + \sum_{h=1,h\neq b}^{K} V^{t}_{ah}(U^{T}U)_{hb} \\ &\Rightarrow \\ \frac{(V^{t}U^{T}U)_{ab}}{V^{t}_{ab}} &\geq (U^{T}U)_{bb} \\ \lambda(DV)_{ab} &= \lambda\sum_{h=1}^{N} D_{ah}v^{t}_{hb} \\ &= \lambda D_{aa}V^{t}_{ab} + \lambda\sum_{h=1,h\neq a}^{N} D_{ah}v^{t}_{hb} \\ \lambda D_{aa}V^{t}_{ab} &\geq \lambda(D - W)_{aa}V^{t}_{ab} = \lambda L_{aa}V^{t}_{ab} \\ &\Rightarrow \\ \frac{\lambda(DV)_{ab}}{V^{t}_{ab}} &\geq \lambda L_{aa} \end{aligned} \tag{37, 38}$$

It can be seen from the above that the inequality holds, so $G\left(v, v^{t}_{ab}\right) \geq F_{v_{ab}}\left(v\right)$ is established and Lemma 2 is true. Since $G\left(v, v^{t}\right)$ is a helper function of $F_{v_{ab}}\left(v\right)$, it can be concluded that $F_{v_{ab}}\left(v\right)$ is non-incremental under the update rule (28).

Regarding the convergence of the A update rule, since this update is obtained by deriving A for the Lagrangian function, but A is not subject to this constraint. In fact, from the perspective of the convexity of A, each iteration is equivalent to minimizing A. Thus, Theorem 2 is true for (29).

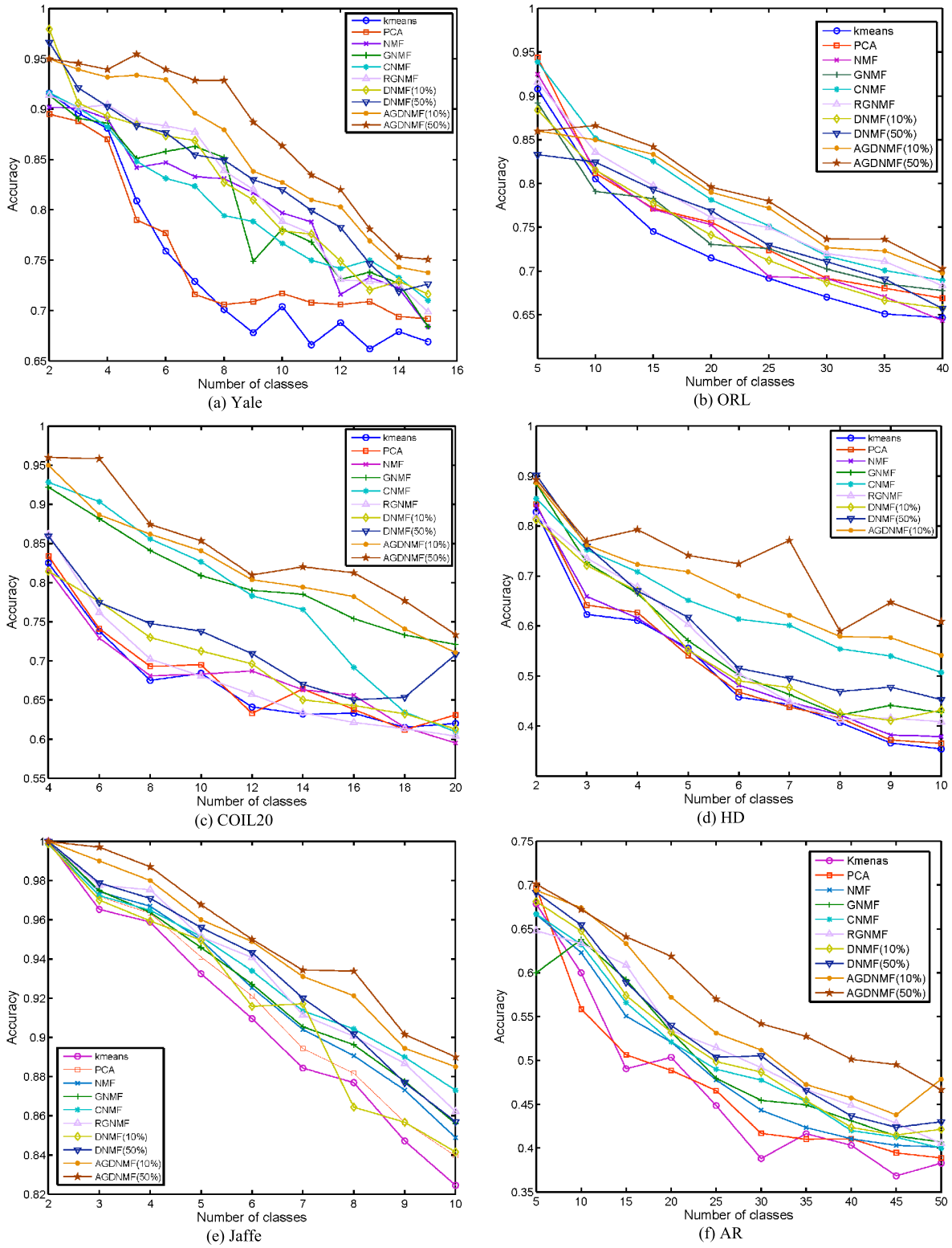
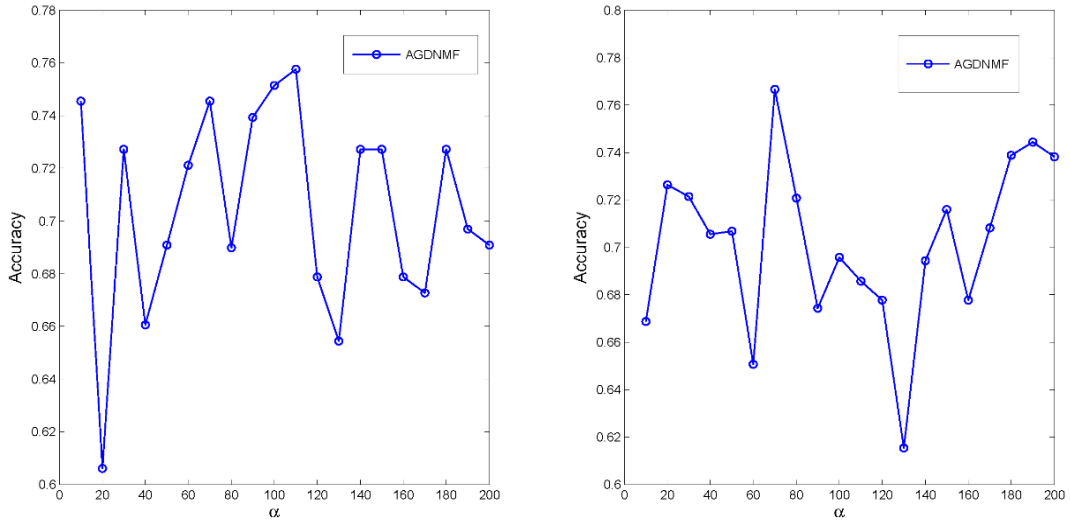


FIGURE 1. Clustering results of each algorithm on four data sets. (a) Yale, (b) ORL, (c) COIL20, (d) HD, (e) JAFFE, (F) AR.

D. COMPUTATIONAL COMPLEXITY ANALYSIS

In this paper, the objective function proposed is obtained through continuous iteration. So we discuss the computational complexity of AGDNMF.

We use O to represent the computational complexity of the algorithm. The computational complexity of the algorithm is described by three arithmetic operations. Each iteration update based on AGDNMF requires addition



(a) Yale (b) COIL20

FIGURE 2. The clustering results on AGNMF on two databases versus parameter α. (a) Yale, (b) COIL20.

operation

$$2MNK + 2(M+N)K^2 + N(p+3)K - \frac{2}{3}K^3 + (3N_l + S)K^2 + 4N_lK + \frac{1}{2}K(K-l) + M$$

multiplication operation

$$2MNK + 2(M+N)K^2 + (M+N)K + N(\text{p+1})K - \frac{2}{3}Kn^3 + (3N_l + S)K^2 + 6N_lK + \frac{1}{2}K(K-l) + M$$

division operation

$$(M+N)K + K$$

The overall complexity of AGDNMF is

$$O(MNK)$$

IV. EXPERIMENTS

In this section, a lot of image clustering experiments with AGDNMF algorithm have been implemented on six common databases. And some related algorithms are also carried out for comparison.

A. DATA DESCRIPTION

Six image data sets are used to validate our method, which is described in detail as follows (also see Table 1).

1) YALE FACE DATABASE

The Yale face database contains 165 grayscale images of 15 people, each with 11 images. The image size of each picture is 80 × 100. In this experiment we manually adjusted the image to a size of 40 × 50.

2) ORL FACE DATABASE

The Yale face database contains 400 grayscale images of 40 people, each with 10 images. The image size of each picture is 92 × 112. In this experiment we manually adjusted the image to a size of 23 × 28.

3) COIL20 DATASET

The COIL20 dataset contains 1440 grayscale images of 20 objects, each with 10 images. The image size of each picture is 64 × 64. In this experiment we manually adjusted the image to a size of 32 × 32.

4) HANDWRITTEN DIGITS DATASET

The handwritten digits database contains 10000 images from 0 to 9 totally, each containing 1000 images. The image size of each picture is 16 × 16.

5) JAFFE FACE DATABASE

The Jaffe face database contains 213 grayscale images of 10 objects. The size of each picture is 256 × 256. In this experiment, we manually adjusted the image to a size of 64 × 64.

6) AR FACE DATABASE

The AR face database contains over 4000 different face images. We select 25 male and 25 female for testing. In this experiment, we manually adjusted the image to a size of 32 × 32.

B. COMPARISON METHOD

In order to demonstrate that our AGDNMF method improves the clustering performance, we compare it with the following algorithm, such as other K-means clustering method [35], PCA [9], non-negative matrix factorization NMF [10], graph

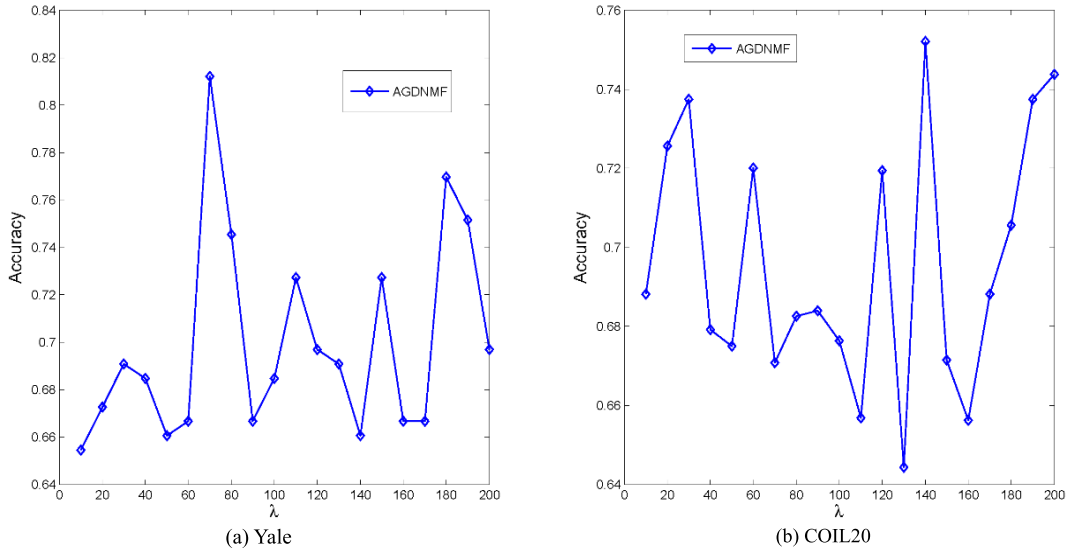


FIGURE 3. The clustering results on AGNMF on two databases versus parameter λ. (a) Yale, (b) COIL20.

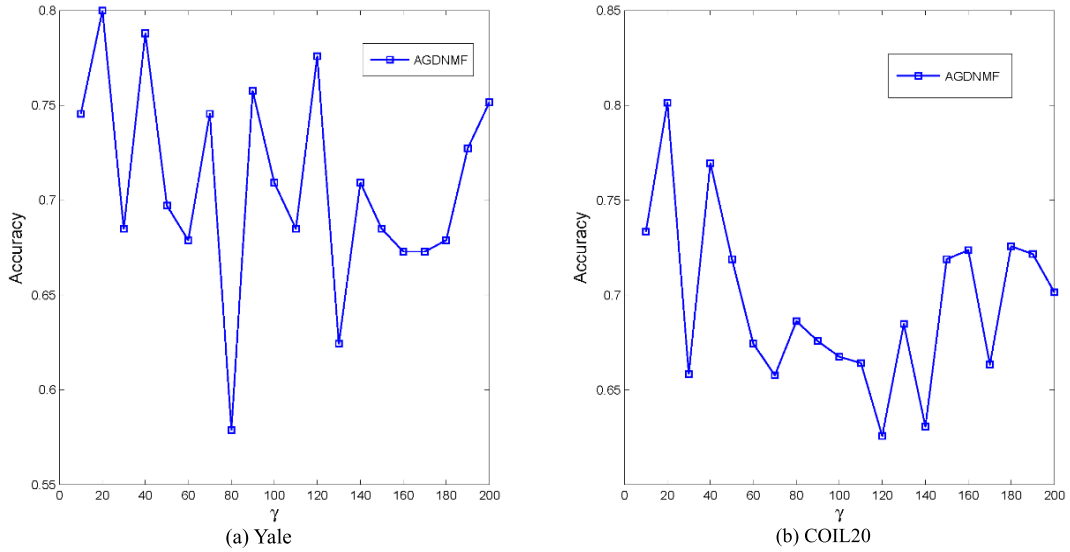


FIGURE 4. The clustering results on AGNMF on two databases versus parameter γ. (a) Yale, (b) COIL20.

regularized nonnegative matrix factorization (GNMF) [34], constrained nonnegative matrix factorization (CNMF) [28], robust graph regularized nonnegative matrix factorization (RGNMF) [36], discriminative nonnegative matrix factorization (DNMF) [26].

C. EXPERIMENT SETUP

In order for each algorithm to achieve the best results, the parameters in each algorithm are appropriate. All experiments are repeated 20 times on each dataset, we take the average of 20 test results as the final result. For all experimental results, we adopted accuracy (AC) as a unified evaluation standard. We use this precision standard to measure the clustering performance of each algorithm. Suppose a data set contains n images. For the i-th data point, l_i and t_i represent the obtained cluster label and real label, respectively.

$$AC = \frac{\sum_{i=1}^{N} \delta(t_i, map(l_i))}{N} \tag{39}$$

where $\delta(\boldsymbol{x}, \boldsymbol{y})$ is the measurement function. If $\boldsymbol{x} = \boldsymbol{y}$, the value of $\delta(\boldsymbol{x}, \boldsymbol{y})$ is 1, otherwise it is 0. $map(\boldsymbol{l_i})$ is a mapping function that maps the clustering labels of each data point to the real label. We can get the optimal mapping result by Kuhn-Munkres algorithm [37]. Fig. 1 show the clustering results in Yale, ORL, COIL20, HD, Jaffe, AR.

D. EXPERIMENTAL RESULTS

Fig. 1 shows the clustering resulting of several methods on four image databases. From Fig.2, we can conclude that we can get the following conclusions. First of all, we can see that

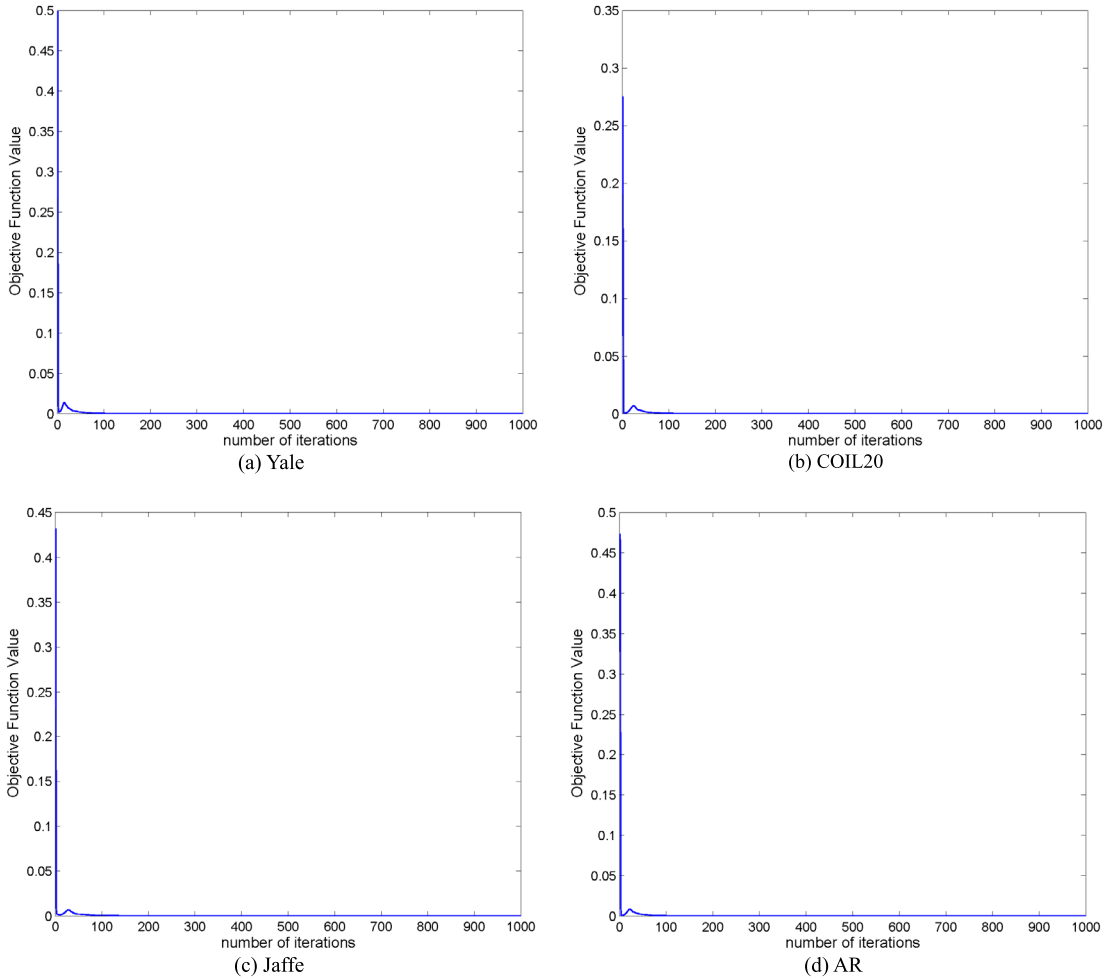


FIGURE 5. The convergence of GDNMF on four databases. (a) Yale, (b) ORL, (c) Jaffe, (d) AR.

the algorithm based on NMF is superior to other algorithms, which shows that part-based learning can better retain data information and get better results. Second, the graph-based approach performs better than other algorithms, indicating that the local structure information contains a lot of hidden information. Third, our algorithm is superior to other algorithms that do not use label information, which shows that label information is very important to image clustering. Finally, our proposed AGDNMF algorithm is superior to other algorithms in most cases. This shows that this algorithm can make good use of local structure information and label information to obtain better data representation.

E. PARAMETERS SELECT

The algorithm we proposed contains three parameters α, λ and γ. We experimented on two of the databases to verify the effect of these three parameters on the algorithm. We take all the classifications of each database for testing. When testing one parameter, the other two parameters are fixed. Figure 2, Figure 3 and Figure 4 show the clustering results when the three parameters are different. We can see that changing the value of the parameters, the results of different databases are not the same.

F. CONVERGENCE STUDY

The updating rules of this algorithm are obtained through iteration. In order to prove its convergence, experiments were carried out on four databases. Figure 5 shows the convergence curve on Yale, ORL, Jaffe and AR databases. We can find that AGDNMF converges at a very fast speed and can converge iteratively within 100 iterations. This also verified in our statement of the computational complexity, although this algorithm needs a lot of iterations, but the whole process is very fast.

V. CONCLUSION

In this paper, we propose a novel semi-supervised adaptive graph regularization discriminant nonnegative matrix factorization (AGDNMF). Neighbor graphs are constructed adaptively in the algorithm to obtain better local structural information to promote the image clustering accuracy. At the same time, the label information is employed as a

constraint, to avoid the data points with the same label are not merged into the same point, and the clustering performance is improved. We simultaneously propose an update rule method to solve AGDNMF. A large number of experiments have proved that AGDNMF algorithm has better performance than other algorithms.

REFERENCES

[1] V. Kaburlasos and G. Papakostas, "Learning distributions of image features by interactive fuzzy lattice reasoning in pattern recognition applications," *IEEE Comput. Intell. Mag.*, vol. 10, no. 3, pp. 42–51, Aug. 2015.

[2] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[3] X. Li, X. Shen, Z. Shu, Q. Ye, and C. Zhao, "Graph regularized multilayer concept factorization for data representation," *Neurocomputing*, vol. 238, pp. 139–151, May 2017.

[4] Z. Huang, H. Zhu, J. T. Zhou, and X. Peng, "Multiple marginal Fisher analysis," *IEEE Trans. Ind. Informat.*, vol. 66, no. 12, pp. 9798–9807, Dec. 2018.

[5] X. Peng, Z. Yu, Z. Yi, and H. Tang, "Constructing the L2-graph for robust subspace learning and subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1053–1066, Apr. 2016.

[6] S. Zhang, Y. Xia, and X. Li, "Diversified recommendation algorithm for hybrid label based on matrix factorization," in *Proc. IEEE 3rd Int. Conf. Cloud Comput. Big Data Anal.*, Apr. 2018, pp. 39–44.

[7] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer, 1992.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2000.

[9] I. Jolliffe, *Principle Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[11] 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.

[12] Q. Liao and Q. Zhang, "Local coordinate based graph-regularized NMF for image representation," *Signal Process.*, vol. 124, pp. 103–114, Jul. 2016.

[13] W. Zhu, Y. Yan, and Y. Peng, "Topological structure regularized nonnegative matrix factorization for image clustering," *Neural Comput. Appl.*, pp. 1–19, Jun. 2018. doi: 10.1007/s00521-018-3572-4.

[14] W.-S. Chen, Y. Zhao, B. Pan, and B. Chen, "Supervised kernel nonnegative matrix factorization for face recognition," *Neurocomputing*, vol. 205, pp. 165–181, Sep. 2016.

[15] S. Wang, J. Peng, and W. Liu, "Discriminative separable nonnegative matrix factorization by structured sparse regularization," *Signal Process.*, vol. 120, pp. 620–626, Mar. 2016.

[16] Z. Wu, Z. Yu, J. Yuan, and J. Zhang, "A twice face recognition algorithm," *Soft Comput.*, vol. 20, no. 3, pp. 1007–1019, 2016.

[17] M. Lu, X.-J. Zhao, L. Zhang, and F.-Z. Li, "Semi-supervised concept factorization for document clustering," *Inf. Sci.*, vol. 331, pp. 86–98, Feb. 2016.

[18] N. B. Erichson, A. Mendible, S. Wihlborn, and J. N. Kutz, "Randomized nonnegative matrix factorization," *Pattern Recognit. Lett.*, vol. 104, pp. 1–7, Mar. 2018.

[19] R. Shang, Z. Zhang, L. Jiao, W. Wang, and S. Yang, "Global discriminative-based nonnegative spectral clustering," *Pattern Recognit.*, vol. 55, pp. 172–182, Jul. 2016.

[20] J. Fan and J. Wang, "A collective neurodynamic optimization approach to nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2344–2356, Oct. 2017.

[21] C.-X. Jin and Q.-C. Bai, "Text clustering algorithm based on the graph structures of semantic word co-occurrence," in *Proc. Int. Conf. Inf. Syst. Artif. Intell.*, Jun. 2017, pp. 497–502.

[22] L. M. Abualigah, A. T. Khader, and M. A. Al-Betar, "Multi-objectives-based text clustering technique using K-mean algorithm," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Inf. Technol.*, Jul. 2016, pp. 1–6.

[23] M. T. Belachew and N. Del Buono, "A dynamical system approach for continuous nonnegative matrix factorization," *Medit. J. Math.*, vol. 14, no. 1, p. 14, Feb. 2017.

[24] X. Dai, C. Li, X. He, and C. Li, "Nonnegative matrix factorization algorithms based on the inertial projection neural network," *Neural Comput. Appl.*, pp. 1–15, Jan. 2018. doi: 10.1007/s00521-017-3337-5.

[25] M. Fakhry, P. Svaizer, and M. Omologo, "Audio source separation in reverberant environments using β-divergence-based nonnegative factorization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 7, pp. 1462–1476, Jul. 2017.

[26] M. Babaee, S. Tsoukalas, M. Babaee, G. Rigoll, and M. Datcu, "Discriminative nonnegative matrix factorization for dimensionality reduction," *Neurocomputing*, vol. 173, pp. 212–223, Jan. 2016.

[27] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. 207–212.

[28] H. Liu and Z. Wu, "Non-negative matrix factorization with constraints," in *Proc. 24th AAAI Conf. Artif. Intell.*, 2010, pp. 1–6.

[29] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.

[30] F. Shang, L. C. Jiao, and F. Wang, "Graph dual regularization non-negative matrix factorization for co-clustering," *Pattern Recognit.*, vol. 45, no. 6, pp. 2237–2250, 2012.

[31] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[32] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[33] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 585–591.

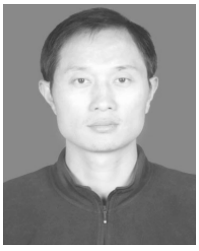
[34] D. Cai, X. He, J. Han, and T. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[35] X. Wang and Y. Bai, "The global Minmax k-means algorithm," *SpringerPlus*, vol. 5, no. 1, p. 1665, 2016.

[36] 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, 2017 Art. no. 33.

[37] L. Lovász and M. Plummer, *Matching Theory*. Budapest, Hungary: Akadémiai Kiadó, 1986.

LIN ZHANG is currently pursuing the M.S. degree with the Information Engineering College, Henan University of Science and Technology, China. Her current research interests include pattern recognition and machine learning.



ZHONGHUA LIU received the B.S. degree in computer engineering from the First Aeronautical Institute of the Air Force, in 1998, the M.S. degree in computer software and theory from Xihua University, in 2005, and the Ph.D. degree in pattern recognition and intelligent systems from the Nanjing University of Science and Technology, in 2011. He is currently an Associate Professor with the Information Engineering College, Henan University of Science and Technology. His current research interests include pattern recognition, face recognition, image processing, and scene matching.

LIN WANG received the Ph.D. degree in image and signal processing from INSA de lyon, France, in 2016. She is currently a Lecturer with the Information Engineering College, Henan University of Science and Technology, Luoyang, China. Her current research interests include image processing and computer vision.

JIEXIN PU is currently a professor with the Information Engineering College, Henan University of Science and Technology. His research interests include computer application technology, pattern recognition and computer networks, and information security.