Applied Intelligence
https://doi.org/10.1007/s10489-022-03881-x

Dual local learning regularized NMF with sparse and orthogonal constraints

Zhenqiu Shu[1] · Furong Zuo[2] · Wenli Wu[2] · Congzhe You[2]

Accepted: 7 June 2022

Abstract
Non-negative matrix factorization (NMF) has shown remarkable competitiveness in the past few years. To fully exploit various known prior knowledge hidden in data, this paper proposes a dual local learning regularized NMF with sparse and orthogonal constraints (DLLNMF-SO) algorithm. DLLNMF-SO constructs two local learning regularizers to consider the geometric structure and discriminative information embedded in data and feature space, respectively. Besides, it makes full use of sparse self-representation information by adding the $l_{2,1}$-norm constraint. Meanwhile, the orthogonal constraint is imposed on the basis vectors to preserve the correspondence between samples and basic vectors. We give an efficient iterative updating scheme for the optimization problem of DLLNMF-SO and provides its convergence guarantee. We demonstrate that our proposed approach outperforms other competitors by conducting serval experiments on three benchmark datasets.

Keywords NMF · Sparse · Orthogonal · Geometric structure · Discriminative information

1 Introduction

It is a widespread fundamental topic about the representation of high-dimensional data. For decades, data representation methods play a significant role in handling high-dimensional data in real-world applications including image classification [1], document clustering [2], face recognition [3], etc. They aim at exploring the semantic information among data via the low-dimensional vectors. During the past decades, some classic matrix factorization-based data representation methods, such as PCA [4], LDA [5], NMF [6], manifold learning [7], have proposed in various practical applications.

NMF is a well-known data representation approach because of its powerful interpretability in psychological and physiological. Its purpose tries to provide the product of two low rank factors matrices to best approximate the original data matrix. By imposing the nonnegative constraint, NMF allows only additive, not subtractive operations and thus is considered a parts-based representation approach. Therefore, it shows a more reasonable explanation for the real-world data. Up to now, some variants of NMF have a wide range of practical applications, such as multi-view clustering [8], cross-modal retrieval [9], tumor classification [10], hyperspectral unmixing [11], Fault detection [12].

In recent years, some studies seek to improve the representation power of NMF by fully utilizing the known prior knowledge [13–15]. The work [13] presented a graph regularized NMF (GNMF) approach, which constructs an affinity graph to approximately encode geometric manifold information embedded in data. A local learning regularizer ties to predict the label of the sample by adopting its adjacent data set [13]. By exploiting the manifold geometric in both the data and feature space, a dual graph regularizer is proposed and then integrated into the model of NMF [14]. To fully consider the high-order information among data, Wang et al. [15] further introduced a dual hyper-graph regularized NMF approach, and then applied it to tumor classification. Shu et al. [16] presented a rank-constrained NMF approach, which imposes the rank constraint on the learned graph. Thus, it obtains a connected graph regularizer that is beneficial to the cluster analysis. To explore the geometric structure information embedded in samples effectively, various variants of NMF [16, 17] have been proposed and achieved more superior performances in real-world applications.

✉ Zhenqiu Shu
shuzhenqiu@163.com

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

[2] School of Computer Engineering, Jiangsu University of Technology, Changzhou 231001, China

Published online: 02 August 2022

The above-mentioned algorithms are unsupervised learning methods and thus ignore the class label information in the limited labeled samples. To alleviate this problem, several semi-supervised and supervised learning approaches were developed. Liu et al. [19] used the known partial label to constrain the coefficient matrix by constructing an auxiliary matrix. Therefore, it maps the partially labeled points from the same category to the same subspace. To explore more discriminative representation from the data, Li et al. [20] proposed to explicitly pursue the block-diagonal structure of the new representation. Trigeorgis et al. [21] developed a Deep semi-NMF (DNMF) method to learn the hidden attribute representations, whose weakly supervised extension was also proposed by adding the graph regularizer. The work [22] presented a supervised learning based NMF approach, which not only integrates the nonnegative and orthogonal constraints but also considers discriminative information among data. Ma et al. [23] developed a discriminative subspace matrix factorization approach for multi-view clustering. It not only considers the pseudo-information and the intrinsic geometry structure, but also uses the $l_{2,1}$-norm to improve the robustness.

In our work, a dual local learning regularized NMF with sparse and orthogonal constraints (DLLNMF-SO) is proposed for data representation, which not only considers the discriminative information and local geometric structure hidden in dual space using two local learning regularizers, but also simultaneously imposes the sparse and orthogonal constraints during matrix decomposition. The optimization strategy of the proposed method is provided and its convergence proofs are given. Experiments performed on several image datasets show that the proposed DLLNMF-SO algorithm yields superior results in clustering.

The remainder of this paper is organized as follows. In Section 2, we present the related methods. In Section 3, we introduce our proposed DLLNMF-SO method. Section 4 reports the experiments and its performances. In Section 5, we draw a conclusion.

2 Related works

2.1 NMF

Given a non-negative data matrix, $X = [x_1, x_2, \cdots x_n] \in R^{m \times n}$, NMF tries to decompose the original data matrix X into two low-rank nonnegative matrices $U = [u_{ik}] \in R^{m \times k}$ and $V = [v_{jk}] \in R^{n \times k}$. Thus, NMF aims to minimum the optimization problem as follows:

$$O_{NMF} = \left\|X - UV^T\right\|^2 = \sum_{i,j}\left(x_{ij} - \sum_{k=1}^{K} u_{ik}v_{jk}\right)^2 s.t. U \geq 0, V \geq 0, \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm. By adopting the multiplicative iterative method [9], Eq. (1) can be solved according to the following the updating rules:

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

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

2.2 GNMF

GNMF aims at constructing an affinity graph to describe the manifold structure embedded in high-dimensional ambient space [10]. By applying the regularization technology, GNMF seeks to minimize the following optimization problem:

$$O_{GNMF} = \left\|X - UV^T\right\|_F^2 + \lambda Tr\left(V^TLV\right) s.t. U \geq 0, V \geq 0., \tag{4}$$

where $Tr(\cdot)$ denotes the trace of a matrix and λ is the regularization parameter. W is the affinity matrix and D stands for the diagonal matrix, $D_{jj} = \sum_l W_{jl}$. $L = D - W$ is a graph Laplacian matrix.

Obviously, model (4) is a nonconvex problem with respect to U and V together. Fortunately, we can transform it into a convex problem by optimizing only one variable while fixing others variables. Thus, the alternating updating strategy is employed to achieve a local optimal solution. Therefore, the model (4) can derive the following equations:

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

$$v_{ij} \leftarrow v_{ij}\frac{\left(X^TU + \lambda WV\right)_{ij}}{\left(VU^TU + \lambda DV\right)_{ij}}. \tag{6}$$

2.3 LLNMF

The motivation of the local learning method is that the label of each sample is estimated by its nearest neighbor samples [13]. Therefore, it can discover the local geometric structure and explore discriminative information of high-dimensional samples. Therefore, LLNMF tries to minimum the following problem:

$$O_{DLLNMF} = \left\|X - UV^T\right\|_F^2 + \delta Tr\left(V^TL_vV\right) s.t. U \geq 0, V \geq 0.. \tag{7}$$

In Eq. (7), the first term denotes the reconstruction term, and the second term is the local learning regularizer. $L = (M - I)^T(M - I)$, where M is defined as:

$$M_{ij} = \begin{cases} \alpha_{ij}, \text{if} x_i \in N(x_i) \\ 0, \text{otherwise} \end{cases}. \tag{8}$$

Eq. (7) can derive the updating rules as follows:

$$u_{ij} \leftarrow u_{ij} \sqrt{\frac{(XV)_{ij}}{(UV^T V)_{ij}}}, \tag{9}$$

$$v_{ij} \leftarrow v_{ij} \sqrt{\frac{(X^T U + \delta L^- V)_{ij}}{(VU^T U + \alpha L^+ V)_{ij}}}, \tag{10}$$

where $L^+_{v_{ij}} = \left(|L_{v_{ij}}| + L_{v_{ij}}\right)/2$ and $L^-_{v_{ij}} = \left(|L_{v_{ij}}| - L_{v_{ij}}\right)/2$.

2.4 DGNMF

Motivated by GNMF, DGNMF is further proposed by constructing two graphs to model the feature and data manifold, respectively. Thus, DNMF aims at minimizing the following problem:

$$J_{DNMF} = \|X - UV^T\|_F^2 + \alpha Tr(U^T L_U U) + \beta Tr(V^T L_V V) \quad s.t. U \geq 0, V \geq 0. \tag{11}$$

where α and β represent the parameters of two graph regularizers, respectively. Similarly, Eq. (11) can be updated by the following rules:

$$u_{ij} \leftarrow u_{ij} \frac{(XV + \beta W^U U)_{ij}}{(UV^T V + \beta D^U U)_{ij}}, \tag{12}$$

$$v_{ij} \leftarrow v_{ij} \frac{(X^T U + \alpha W^V V)_{ij}}{(VU^T U + \alpha D^V V)_{ij}}. \tag{13}$$

3 The proposed method

3.1 Motivation

In the past few years, many matrix factorization methods have proposed for dealing with high-dimensional data. However, they cannot take fully use of the prior knowledge hidden in data, such as local structure information, sparseness and orthogonality. To solve these issues, we propose a novel data representation algorithm, called dual local learning regularized NMF with sparse and orthogonal constraints (DLLNMF-SO). In DLLNMF-SO, the prior knowledge embedded in data can been fully utilized. Specifically, a dual local learning regularizer is constructed to consider the local structure information in data and feature space. In addition, the sparse and orthogonal constraints are imposed on the basis matrix, and thus the robustness of the proposed DLLNMF-SO algorithm can be greatly enhanced. Therefore, our proposed DLLNMF-SO method can effectively capture the intrinsic structural information by considering the prior knowledge of high-dimensional data. Figure 1 shows the framework of the proposed DLLNMF-SO method.

3.2 Dual local learning regularizer

Given a data matrix $X = [x_1, x_2, \cdots x_n] \in R^{m \times n}$, x_i represents a data point.$N(x_i)$ denotes nearest neighbor sample set of x_i and n_i represents the size of $N(x_i)$. We construct a function $f_i^l(x)$ ($1 \leq l \leq c$) to predict the label of $\{x_j\}_{x_j \in N(x_i)}$ by its neighborhood points. The superscript l of $f_i^l(x)$ means that it is assigned to the l-th cluster, and its subscript i implies that it uses the neighborhood of x_i for training. The predictor $f_i^l(x)$ can be fitted using the following minimum problem:

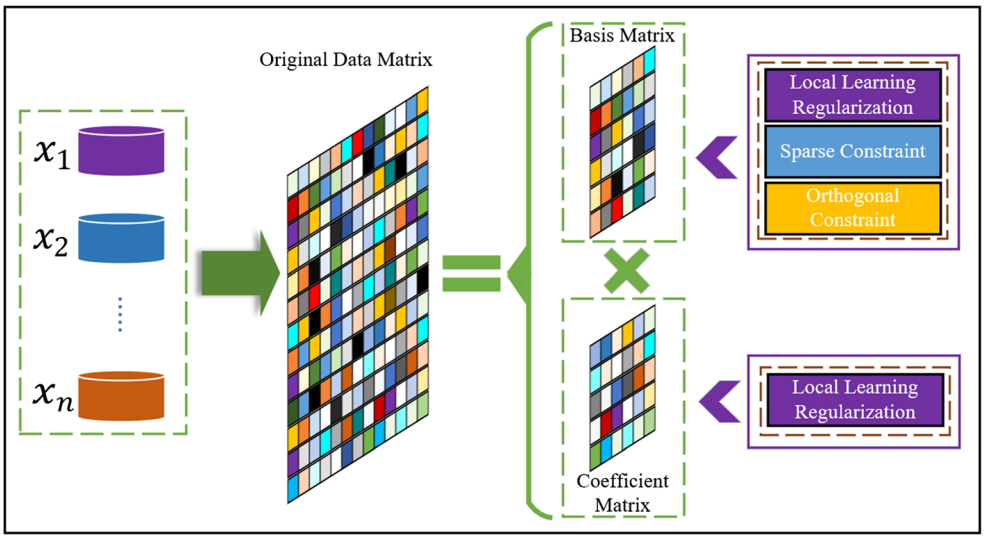


Fig. 1 The framework of the proposed DLLNMF-SO method

$$J_i^l = \frac{1}{n_i} \sum_{x_j \in N(x_i)} \left(f_i^l(x_i) - v_j^l \right)^2 + \gamma_i \left\| f_i^l \right\|_I^2, \tag{14}$$

where v_i^l is the l-th clustering assignment of x_i and γ_i is a regularization parameter. The first term is used to calculate the sum of the error between the sample and its neighborhood samples, and the second term seeks to smooth the function f_i^l. In Eq. (14), we simply set $\gamma_i = \gamma$ and $n_i = k$, where $1 \le i \le m$.

After giving the definition of a positive definite kernel function K: $\chi \times \chi \to R$, the predictor function $f_i^l(\mathrm{x}_i)$ can be expressed as follows:

$$f_i^l = \sum_{j=1}^{n_i} \beta_{ij}^l K(\mathrm{x}_i, \mathrm{x}_j), \tag{15}$$

where β_{ij}^l is the expansion coefficient. Thus, the Eq. (15) is further reformulated as follows:

$$f_i^l = K_i \beta_i^l, \tag{16}$$

where $f_i^l \in R^{n_i}$ denotes the vector $\left[f_i^l(x_j) \right]^T$. $K_i \in R^{n_i \times n_i}$ stands for the kernel matrix constructed on the neighborhood of x_i and $\beta_i^l = \left[\beta_{i1}, \cdots, \beta_{in_i} \right]^T \in R^{n_i \times 1}$ denotes the expansion coefficient vector. By putting formula (16) into formula (14), we can derive the following loss function:

$$J_i^l = \frac{1}{n_i} \left\| K_i \beta_i^l - v_j^l \right\|^2 + \gamma \left(\beta_i^l \right)^T K_i \beta_i^l, \tag{17}$$

where $v_i^l \in R^{n_i}$ denotes the vector $\left[v_i^l \right]^T$.

By calculating the partial derivatives of J_i^l with respect to β_i^l and then setting it to zero, that is, $\frac{\partial J_i^l}{\partial \beta_i^l} = 0$, the solution of Eq. (17) is given as follows:

$$\beta_i^l = (K_i + n_i \gamma I)^{-1} v_i^l, \tag{18}$$

where I denotes a $n_i \times n_i$ identity matrix. By putting Eq. (18) into Eq. (15), we have

$$f_i^l(x_i) = K_i^T (K_i + n\gamma I)^{-1} v_i^l = \alpha_i^T v_i^l, \tag{19}$$

where $k_i \in R^{n_i}$ stands for the vector $[k(x_i, x_j)]^T$ and $\alpha_i = [\alpha_{i1}, \cdots, \alpha_{in_i}]^T \in R^{n_i \times 1}$.

After constructing the predictor (19), the local regularization technology aims to minimize the sum of the error as

$$J = \sum_{l=1}^{m} \sum_{i=1}^{n} \left\| f_i^l(x_i) - v_i^l \right\|^2. \tag{20}$$

where $v^l = \left[v_1^l, \cdots, v_n^l \right]$ is the l-th row of V. By substitute Eq. (19) into Eq. (20), we have

$$\begin{aligned} J &= \sum_{l=1}^{m} \sum_{i=1}^{n} \left\| k_i^T (K^u i + n_i \gamma I)^{-1} v_i^l - v_i^l \right\|_F^2 \\ &= \sum_{i=1}^{m} \left\| S v^l - v^l \right\|^2 \\ &= Tr\left(V^T (S-I)(S-I) V \right) \\ &= Tr\left(V^T L_v V \right) \end{aligned}, \tag{21}$$

where $L_v = (S - I)(S - I)$ and $S \in R^{n \times n}$ can be defined as follows:

$$S_{ij} = \begin{cases} \alpha_{ij}, \text{if} \mathrm{x}_j \in M(\mathrm{x}_i) \\ 0, \text{otherwise} \end{cases}. \tag{22}$$

Therefore, Eq. (21) is named as local learning regularizer in the feature space.

We also construct a similar regularizer in data space by adopting the positive definite kernel function:

$$Q_i^l = \frac{1}{m_i} \sum_{x_j' \in M(x_i')} \left(f_i'^l - u_j^l \right) + \eta \left\| f_i'^l \right\|_F^2, \tag{23}$$

where $M\left(x_i'\right)$ denotes the neighborhood set of x_i', and m_i is the size of $M\left(x_i'\right)$. u_j^l indicates that x_j' is assigned to the l-th class, and η denotes the regularization parameter. $\left\| f_i'^l \right\|_F^2$ aims at smoothing the local geometric structure embedded in data. For simplicity, we set $m_1 = m_2 = \cdots = m_n = k$.

Similarly, the local learning regularization term in the data space can be given as

$$\begin{aligned} Q &= \sum_{l=1}^{m} \sum_{i=1}^{n} \left\| k_{ui}^T \left(K_{ui} + n_i \eta I' \right)^{-1} u_i^l - u_i^l \right\|_F^2 \\ &= \sum_{i=1}^{m} \left\| P v^l - v^l \right\|^2 \\ &= Tr\left(U^T \left(S - I' \right) \left(S - I' \right) U \right) \\ &= Tr\left(U^T L_u U \right) \end{aligned}, \tag{24}$$

where $u^l = \left[u_1^l, \cdots, u_m^l \right]^T \in R^m$ denotes the l-th row of U, $L_u = (P - I)(P - I)$ and $P \in R^{m \times m}$ is given as

$$P_{ij} = \begin{cases} \alpha_{ij}', \text{if} \mathrm{x}_j' \in M\left(\mathrm{x}_i' \right) \\ 0, \text{otherwise} \end{cases}, \tag{25}$$

3.3 Sparse constraint

In the past decade, sparse representation plays an important role in some real-world fields [23, 24]. Several studies have shown that $l_{2,1}$-norm based sparse constraint is insensitive to noise and outliers, and thus an effectively improve the robustness of algorithm. Therefore, the basis matrix U based on $l_{2,1}$-norm metric can be represented as

$$\|U\|_{2,1} = \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{n} |U_{i,j}|^2}. \tag{26}$$

Obviously, Eq. (26) not only guarantees the sparsity of rows in matrix U, but also ensures that the discriminant features are selected in the representation space.

3.4 Orthogonal constraint

To ensure the uniqueness of the solution, the orthogonal constraint has been widely applied to data representation [26]. By enforcing the orthogonal constraint on the basis matrix U, the degree of freedom during matrix decomposition is eliminated and each image is accurately corresponding to the basic vector. Thus, we can get the orthogonal constraint as follows:

$$U^T U = I. \tag{27}$$

3.5 Objective function

To fully consider more prior knowledge among data, we can give the model of the proposed DLLNMF-SO method as follows:

$$\begin{aligned} O_{DLLNMF}\text{-SO} &= \left\|X - UV^T\right\|_F^2 + \alpha Tr\left(V^T L_V V\right) \\ &+ \beta \mathrm{TrTr}\left(U^T L_U U\right) \\ &+ \theta \|U\|_{2,1} s.t. U \geq 0, V \geq 0, U^T U = I., \end{aligned} \tag{28}$$

where α, β and θ denote the nonnegative parameters. The first term is the approximation error. The second and third terms denote the local learning regularizers in feature and data space, respectively. The last term is the sparse constraint imposed on the basis matrix.

3.6 Optimization

It is clear to see that the problem (28) is a nonconvex problem for both U and V together. Therefore, we cannot obtain a global minimum of Eq. (28). Fortunately, the iterative updating strategy is used to optimize the proposed model and a local minimum of Eq. (28) can be achieved. Generally, one variable can be updated while other variables are fixed.

Using matrix property, the model (28) can be further reformulated as follows:

$$\begin{aligned} O_{\mathrm{DLLNMF-SO}} &= Tr\left(\left(X - UV^T\right)\left(X - UV^T\right)\right) \\ &+ \alpha Tr\left(V^T L_v V\right) + \beta Tr\left(U^T L_u U\right) + \theta \|U\|_{2,1} \\ &= Tr\left(XX^T\right) - 2Tr\left(XVU^T\right) + Tr\left(UV^T VU^T\right). \\ &+ \alpha Tr\left(V^T L_v V\right) + \beta Tr\left(U^T L_u U\right) + \theta \|U\|_{2,1} \\ &\quad s.t. U \geq 0, V \geq 0, U^T U = I. \end{aligned} \tag{29}$$

Let Ψ_{ij} and Φ_{ij}be the corresponding Lagrange multipliers for the constraint on $U_{ij} \geq 0$ and $V_{ij} \geq 0$, respectively. Thus, the Lagrange function of Eq. (29) is given as follows:

$$\begin{aligned} L = Tr\left(XX^T\right) - 2Tr\left(XVU^T\right) + Tr\left(UV^T VU^T\right) + \alpha Tr\left(V^T L_v V\right) + \beta Tr\left(U^T L_u U\right) \\ + Tr\left(\Psi V^T\right) + Tr\left(\Phi U^T\right) + \gamma Tr\left(U^T U - I\right) + \theta Tr\left(U^T Q U\right) \end{aligned}, \tag{30}$$

where $Q = [q_{ij}] \in R^{m \times m}$ is a diagonal matrix.

(1) Update Q:

The i-th diagonal element q_{ii} of the diagonal matrix Q is given as follows:isis

$$q_{ii} = \frac{1}{2\|U_i\|_2}. \tag{31}$$

To solve the overflow problem, a small constant ε is added into the matrix Q. Therefore, the formula (31) can be rewritten as follows:

$$q_{ii} = \frac{1}{2max\left(\|U_i\|_2, \varepsilon\right)}. \tag{32}$$

(2) Update U:

The partial derivatives of L with respect to U is given as follows:

$$\frac{\partial L}{\partial U} = -2XV + 2UV^T V + 2\beta L_u U + 2\theta QU + 2\gamma U + \Phi. \tag{33}$$

Using the KKT condition $\Phi_{ij} U_{ij} = 0$ [27], the formula (33) becomes

$$\left(-2XV + 2UV^T V + 2\beta L_u U + 2\theta QU + 2\gamma U\right)_{ij} U_{ij} = 0. \tag{34}$$

Introduce

$$L_u = L_u^+ - L_u^-, \tag{35}$$

where $L_{uij}^+ = \left(\left|L_{uij}\right| + L_{uij}\right)/2$ and $L_{uij}^- = \left(\left|L_{uij}\right| - L_{uij}\right)/2$.

Substitute $L_u = L_u^+ - L_u^-$ into the Eq. (34), we have

$$\left(-2XV + 2UV^TV + 2\beta L_u^+ U - 2\beta L_u^- U + 2\theta QU + 2\gamma U\right)_{ij} U_{ij} = 0. \tag{36}$$

Eq. (36) leads to the following updating rule:

$$u_{ij} \leftarrow u_{ij} \frac{(XV + \beta L_u{}^- U)_{ij}}{\left(UV^TV + \beta L_u{}^+ U + \lambda U + \theta QU\right)_{ij}}. \tag{37}$$

(3) Update V:

By calculating the partial derivatives of L with respect to V, we have

$$\frac{\partial L}{\partial V} = -2X^TU + 2VU^TU + 2\alpha L_v V + \Psi. \tag{38}$$

Similarly, we get the following equation for v_{ij}:

$$\left(-2X^TU + 2VU^TU + 2\alpha L_v V\right)_{ij} V_{ij} = 0. \tag{39}$$

Introduce

$$L_v = L_v^+ - L_v^-, \tag{40}$$

where $L_{v_{ij}}^+ = \left(|L_{v_{ij}}| + L_{v_{ij}}\right)/2$ and $L_{u_{ij}}^- = \left(|L_{u_{ij}}| - L_{u_{ij}}\right)/2$.

Substitute $L_v = L_v^+ - L_v^-$ into the Eq. (39), we have

$$\left(-2X^TU + 2VU^TU + 2\alpha L_v^+ U - 2\alpha L_v^- U\right)_{ij} V_{ij} = 0. \tag{41}$$

Eq. (41) can derive the following updating rule:

$$v_{ij} \leftarrow v_{ij} \frac{\left(X^TU + \alpha L_v{}^- V\right)_{ij}}{\left(VU^TU + \alpha L_v{}^+ V\right)_{ij}}. \tag{42}$$

Algorithm1 DLLNMF-SO

Input: Data set X, parameters α, β, θ and λ.

1. Construct two regularizers by solving Eq. (21) and Eq. (24);
2. **while** not converged **do**

Update Q by solving (32);
Update U by solving (37);
Update V by solving (42);

end while

Output: U and V.

3.7 Convergence analysis

Definition 1 $G(x, x')$ is an auxiliary function for $F(x)$, if it satisfies the conditions $G(x, x') \geq F(x)$ and $G(x, x) = F(x)$.

Lemma 1 If $G(x, x')$ is an auxiliary function of $F(x)$, the function $F(x)$ is non-increasing under the following update rule:

$$x^{t+1} = argmin_x G(x, x^t). \tag{43}$$

Proof. $F(x^{t+1}) \leq G(x^{t+1}, x^t) \leq G(x^t, x^t) = F(x^t)$.

The model of DLLNMF-SO is further reformulated as follows:

$$F(V) = \left\|X - U^TV\right\|_F^2 + \beta tr\left(U^TL_uU\right) + \alpha tr\left(V^TL_vV\right) + \theta\|U\|_{2,1} \quad s.t. U^TU = I \tag{44}$$

F_{vij} represents the part of model (44) that is only related to the element V_{ij} in V. Therefore, we can write the first-order and second-order differential of formula (44) as follows:

$$F'_{v_{ij}} = \left[\frac{\partial F}{\partial V}\right]_{ij} = \left[-2X^TU + 2VU^TU + 2\alpha L_v V\right]_{ij}, \tag{45}$$

$$F''_{v_{ij}} = 2\left[U^TU\right]_{ij} + 2\alpha[L_v]_{ij}. \tag{46}$$

The following function

$$G\left(v_{ij}, v_{ij}^t\right) = F_{v_{ij}}\left(v_{ij}^t\right) + F'_{v_{ij}}\left(v_{ij}^t\right)\left(v_{ij} - v_{ij}^t\right) + \frac{\left[VU^TU + \alpha L_v^+ V\right]_{ij}}{V_{ij}^t}\left(v_{ij} - v_{ij}^t\right)^2. \tag{47}$$

is the auxiliary function of variable v_{ij} in Eq. (44).

Proof: We can see $G\left(v_{ij}, v_{ij}\right) = F_{v_{ij}}\left(v_{ij}\right)$. Thus, we only need to prove that $G\left(v_{ij}, v_{ij}^t\right) \geq F_{v_{ij}}\left(v_{ij}\right)$. The Taylor series expansion of the function $F_{ij}(v_{ij})$ is given as follows:

$$F_{vij}\left(V_{ij}\right) = F_{vij}\left(v_{ij}^t\right) + F'_{vij}\left(v_{ij}^t\right)\left(v_{ij} - v_{ij}^t\right) + \left\{\left[U^TU\right]_{ij} + \alpha[L_v]_{ij}\right\}\left(v_{ij} - v_{ij}^t\right)^2. \tag{48}$$

By observing Eq. (47) and Eq. (48), it is easy to find that $G\left(v_{ij}, v_{ij}^t\right) \geq F_{v_{ij}}\left(v_{ij}\right)$ is equivalent to

$$\frac{\left[VU^TU + \alpha L_v^+ V\right]_{ij}}{V_{ij}^t} \geq \left[U^TU\right]_{ij} + \alpha[L_v]_{ij}.$$

We can get

$$\left[VU^TU\right]_{ij} = \sum_l^K V_{il}^t\left[U^TU\right]_{lj} \geq V_{ij}^t\left[U^TU\right]_{jj},$$

and

$$\alpha L_v^+ V = \alpha\sum_{l=1}^M L_{v_{il}}^+ v_{il}^t \geq \alpha L_v v_{il}^t \geq \alpha\left(L_{v_{il}}^+ - L_{v_{il}}^-\right)v_{il}^t = \alpha L_{v_{ii}} v_{ij}^t.$$

Consequently, $G\left(v_{ij}, v_{ij}^t\right) \geq F_{v_{ij}}\left(v_{ij}\right)$ is hold.

Fig. 2 Face images from the FERET database

Theorem 1 The model (28) is non-increasing by using the updating rules (37) or (42).

Proof: Putting Eq. (47) into Eq. (43), we can obtain that

$$v_{ij}^{t+1} = v_{ij}^{t} - v_{ij}^{t} \frac{F'_{vij}\left(v_{ij}^{t}\right)}{2\left(VU^{T}U + \alpha L_{v}^{+}V\right)_{ij}} = v_{ij}^{t} \frac{\left(X^{T}U + \alpha L_{v}^{-}V\right)_{ij}}{\left(VU^{T}U + \alpha L_{v}^{+}V\right)_{ij}} \tag{49}$$

Since Eq. (47) is an auxiliary function of F_{ij}, the convergence of Eq. (28) can be proved using Eq. (37). Similarly, the proposed model (28) is non-increasing by adopting Eq. (42) and its convergence is easy to prove.

3.8 Computational complexity of DLLNMF-SO

The computational complexity of DLLNMF-SO is discussed in this subsection. Firstly, the proposed DLLNMF-SO algorithm needs $O(n^3 + m^3)$ to construct a dual local learning regularizer by adopting the kernel regression method, where n and m denote the number and the dimensionality of image sample, respectively. Secondly, the proposed solution scheme needs $O(\boldsymbol{tnmc})$ to iteratively update the U, V and Q, where t and c denote the iteration times and the number of clusters, respectively. In summary, the overall computational complexity of our proposed method is $O(\boldsymbol{tnmc} + n^3 + m^3)$.

4 Experiments

To investigate our proposed DLLNMF-SO approach, we conducted the clustering experiments on three real image datasets including FERET, Yale B, and ORL databases. To verify the effectiveness of DLLNMF-SO, several well-known representative clustering methods, such as k-means (KM), NMF, GNMF, CNMF, LLNMF, SODNMF [28] and DGNMF [21], are used as the comparison algorithms. In addition, we adopt accuracy (AC) and normalized mutual information (NMI) to fairly compare the results of various approached in clustering, and their detailed definitions can be found in [12].

4.1 FERET face database

The first experiment is carried out on the FERET face database, which includes 1400 face images captured from 200

Table 1 The clustering performances on the FERET database

AC								
P	KM	NMF	GNMF	LLNMF	DGNMF	SODNMF	DNMF	DLLNMF-SO
100	0.287±0.016	0.260±0.011	0.224±0.013	0.220±0.007	0.217±0.012	0.218±0.016	0.302±0.016	**0.306±0.003**
120	0.271±0.016	0.255±0.014	0.208±0.006	0.213±0.008	0.202±0.004	0.214±0.009	**0.288±0.004**	0.284±0.013
140	0.262±0.006	0.248±0.007	0.202±0.005	0.208±0.008	0.203±0.004	0.211±0.007	0.295±0.003	**0.301±0.010**
160	0.258±0.003	0.236±0.009	0.203±0.006	0.204±0.001	0.199±0.002	0.200±0.010	0.290±0.005	**0.317±0.012**
180	0.250±0.007	0.232±0.007	0.197±0.005	0.198±0.001	0.195±0.004	0.197±0.012	0.293±0.005	**0.299±0.005**
200	0.253±0.007	0.230±0.005	0.197±0.003	0.196±0.004	0.190±0.005	0.198±0.018	0.284±0.004	**0.295±0.007**
Avg	0.264±0.009	0.244±0.009	0.205±0.006	0.207±0.005	0.201±0.005	0.206±0.012	0.292±0.006	**0.300±0.008**
NMI								
P	KM	NMF	GNMF	LLNMF	DGNMF	SODNMF	DNMF	DLLNMF-SO
100	0.618±0.016	0.598±0.012	0.592±0.009	0.596±0.004	0.587±0.012	0.438±0.032	0.594±0.011	**0.626±0.005**
120	0.609±0.006	0.605±0.010	0.599±0.004	0.601±0.003	0.594±0.006	0.431±0.038	0.596±0.009	**0.614±0.015**
140	0.622±0.005	0.614±0.007	0.608±0.006	0.611±0.003	0.610±0.002	0.436±0.034	0.614±0.003	**0.639±0.009**
160	0.623±0.004	0.614±0.005	0.619±0.005	0.620±0.003	0.617±0.002	0.418±0.043	0.620±0.004	**0.635±0.006**
180	0.629±0.008	0.618±0.001	0.626±0.001	0.623±0.004	0.623±0.002	0.397±0.034	0.634±0.003	**0.647±0.002**
200	0.637±0.004	0.624±0.005	0.633±0.002	0.629±0.003	0.630±0.004	0.391±0.020	**0.642±0.003**	0.639±0.007
Avg	0.624±0.007	0.612±0.006	0.612±0.004	0.613±0.003	0.610±0.005	0.416±0.034	0.617 **±0.006**	**0.633±0.007**

The boldface indicates the best performance

Fig. 3 Face samples from the Yale B database

persons. Each face image was resized and reshaped by a 1024-dimensional vector. Some samples chosen from the FERET database are shown in Fig. 2.

In each time, P categories face images were randomly sampled from the FERET database to verify the proposed method. For a fair comparison, we ran all algorithms ten times and recorded their average performances. The clustering performances of all methods are summarized in Table 1. Noting that the average results of the proposed DLLNMF-SO approach are superior to other competitors with P varying from 100 to 200. This is because our proposed method not only takes full of the geometric structure and the discriminative ability of data by incorporating the dual regularizer into the model of NMF, but also improves the representation power by imposing the orthogonal and sparse constraints.

4.2 Yale B face database

The cropped Yale B face database includes 2414 face samples of 38 persons with different poses and illumination conditions. The size of all images was normalized to 64 × 64 pixels. Some samples from the Yale B database are shown in Fig. 3.

Similarly, we evaluated the effectiveness of our proposed DLLNMF-SO method by randomly sampling P categories images from the Yale B database. We ran each method ten times and reported its average values. The detailed clustering performances of different algorithms on the Yale B database are shown in Table 2. By observing the data from the Table 2, we can easily see that the results of both GNMF and DGNMF outperform k-means and NMF. This is because the clustering performances

Table 2 The clustering performances on the Yale B database

AC								
P	KM	NMF	GNMF	LLNMF	DGNMF	SODNMF	DNMF	DLLNMF-SO
17	0.133±0.022	0.188±0.027	0.319±0.050	0.305±0.017	0.402±0.038	0.299±0.027	0.267±0.030	**0.411±0.021**
20	0.118±0.020	0.168±0.006	0.257±0.067	0.306±0.032	0.370±0.028	0.312±0.040	0.253±0.015	**0.414±0.018**
23	0.128±0.009	0.188±0.008	0.278±0.020	0.295±0.010	0.387±0.042	0.286±0.027	0.238±0.024	**0.391±0.044**
26	0.114±0.012	0.178±0.012	0.242±0.052	0.292±0.019	0.362±0.029	0.291±0.037	0.234±0.032	**0.391±0.010**
29	0.114±0.011	0.175±0.013	0.218±0.032	0.291±0.012	0.338±0.032	0.269±0.016	0.213±0.017	**0.382±0.012**
32	0.109±0.009	0.184±0.003	0.246±0.041	0.287±0.013	0.358±0.007	0.268±0.016	0.212±0.019	**0.401±0.012**
35	0.106±0.013	0.166±0.006	0.210±0.055	0.276±0.020	0.342±0.012	0.259±0.018	0.202±0.013	**0.377±0.020**
38	0.109±0.002	0.171±0.011	0.179±0.056	0.285±0.015	0.349±0.015	0.259±0.026	0.218±0.019	**0.355±0.011**
Avg	0.116±0.012	0.177±0.010	0.244±0.046	0.292±0.017	0.364±0.026	0.280±0.026	0.230±0.021	**0.390±0.019**
NMI								
P	KM	NMF	GNMF	LLNMF	DGNMF	SODNMF	DNMF	DLLNMF-SO
17	0.090±0.027	0.197±0.042	0.369±0.078	0.326±0.014	0.486±0.068	0.342±0.020	0.285±0.027	**0.487±0.005**
20	0.078±0.032	0.193±0.021	0.296±0.070	0.337±0.020	0.452±0.037	0.363±0.038	0.281±0.032	**0.466±0.010**
23	0.121±0.012	0.261±0.007	0.338±0.024	0.327±0.009	0.474±0.036	0.364±0.033	0.291±0.026	**0.479±0.016**
26	0.111±0.025	0.258±0.019	0.314±0.068	0.352±0.014	0.466±0.040	0.372±0.043	0.305±0.026	**0.471±0.023**
29	0.122±0.023	0.263±0.012	0.299±0.035	0.357±0.010	0.453±0.027	0.358±0.015	0.301±0.019	**0.469±0.010**
32	0.143±0.015	0.290±0.018	0.346±0.047	0.365±0.010	0.479±0.019	0.373±0.018	0.310±0.020	**0.502±0.011**
35	0.135±0.017	0.285±0.019	0.306±0.070	0.360±0.015	0.474±0.008	0.361±0.019	0.309±0.014	**0.481±0.007**
38	0.153±0.006	0.299±0.017	0.278±0.081	0.372±0.018	0.477±0.003	0.372±0.020	0.334±0.019	**0.485±0.009**
Avg	0.119±0.020	0.256±0.019	0.318±0.059	0.349±0.014	0.470±0.029	0.363±0.026	0.302±0.023	**0.480±0.011**

The boldface indicates the best performance

Fig. 4 Face sampled from the ORL database

can be greatly improved by fully considering the intrinsic manifold structure of samples. Besides, the results of the proposed DLLNMF-SO algorithm are superior than both SODNMF and DNMF on Yale B database. This is because the proposed DLLNMF-SO method not only captures the discriminative information by constructing a dual local learning regularizer, but also imposes the sparse and orthogonal constraints on the basic matrix to improve the representation ability of the algorithm.

4.3 ORL face database

The ORL database includes 40 different themes for each with ten different images. The light and facial expressions of images were taken under various times or different environments. All images were cropped into 32 × 32 pixels and each pixel 256 levels of gray. Figure 4 shows some images used in the experiment.

Similarly, we employed the same experimental scheme as the previous two subsections. The clustering performances of nine algorithms on the ORL database are reported in Table 3. It is clear to see that DLLNMF-SO obtains the best results among all the comparison algorithms. This is because our proposed DLLNMF-SO method discovers more known prior knowledge embedded in the data than other state-of-the-art methods. Specifically, in DLLNMF-SO, the intrinsic geometric structure and discriminative information hidden in samples in dual space are fully utilized by using the dual local learning regularizers, and the sparse and orthogonal constraints are imposed on the basis matrix to improve the discriminative ability and obtain a powerful low-dimensional representation.

Table 3 The clustering performances on the ORL database

AC								
P	KM	NMF	GNMF	LLNMF	DGNMF	SODNMF	DNMF	DLLNMF-SO
20	0.581±0.194	0.572±0.030	0.523±0.029	0.336±0.031	0.509±0.060	**0.616±0.067**	0.407±0.048	0.589±0.067
23	0.524±0.050	0.521±0.060	0.505±0.065	0.317±0.022	0.500±0.075	0.586±0.075	0.407±0.049	**0.596±0.044**
26	0.560±0.040	0.557±0.066	0.526±0.024	0.304±0.019	0.516±0.027	0.591±0.050	0.435±0.031	**0.592±0.039**
29	0.558±0.052	0.524±0.044	0.534±0.047	0.304±0.014	0.515±0.025	0.592±0.035	0.446±0.037	**0.602±0.036**
32	0.523±0.010	0.515±0.026	0.507±0.029	0.295±0.019	0.508±0.033	0.578±0.041	0.467±0.028	**0.583±0.014**
35	0.540±0.045	0.509±0.027	0.491±0.039	0.280±0.02	0.468±0.039	0.560±0.040	0.449±0.026	**0.588±0.027**
38	0.525±0.019	0.502±0.026	0.501±0.015	0.293±0.024	0.485±0.012	**0.576±0.033**	0.469±0.031	0.571±0.017
40	0.542±0.022	0.510±0.035	0.493±0.016	0.293±0.009	0.475±0.030	0.568±0.045	0.469±0.032	**0.575±0.022**
Avg	0.544±0.054	0.526±0.039	0.510±0.033	0.303±0.019	0.497±0.038	0.583±0.048	0.444±0.035	0.587±0.033
NMI								
P	KM	NMF	GNMF	LLNMF	DGNMF	SODNMF	DNMF	DLLNMF-SO
20	0.701±0.020	0.695±0.031	0.644±0.022	0.737±0.031	0.635±0.042	0.729±0.050	0.501±0.043	**0.757±0.057**
23	0.685±0.034	0.695±0.038	0.648±0.052	0.472±0.023	0.641±0.069	0.719±0.055	0.519±0.044	**0.726±0.037**
26	0.713±0.026	0.707±0.036	0.66±0.0128	0.485±0.017	0.665±0.020	0.736±0.036	0.560±0.049	**0.738±0.0314**
29	0.720±0.035	0.694±0.023	0.696±0.034	0.494±0.013	0.686±0.023	0.733±0.025	0.592±0.035	**0.750±0.018**
32	0.697±0.007	0.692±0.015	0.670±0.023	0.491±0.021	0.670±0.026	0.736±0.030	0.634±0.026	**0.737±0.017**
35	0.714±0.023	0.696±0.018	0.675±0.038	0.494±0.023	0.657±0.035	0.731±0.024	0.618±0.030	**0.748±0.018**
38	0.705±0.012	0.699±0.020	0.678±0.006	0.507±0.017	0.669±0.015	**0.746±0.019**	0.658±0.019	0.740±0.009
40	0.718±0.009	0.707±0.016	0.686±0.007	0.520±0.009	0.680±0.013	0.744±0.033	0.657±0.022	**0.746±0.012**
Avg	0.718±0.021	0.698±0.025	0.671±0.024	0.525±0.019	0.662±0.030	0.734±0.034	0.592±0.034	0.742±0.025

The boldface indicates the best performance

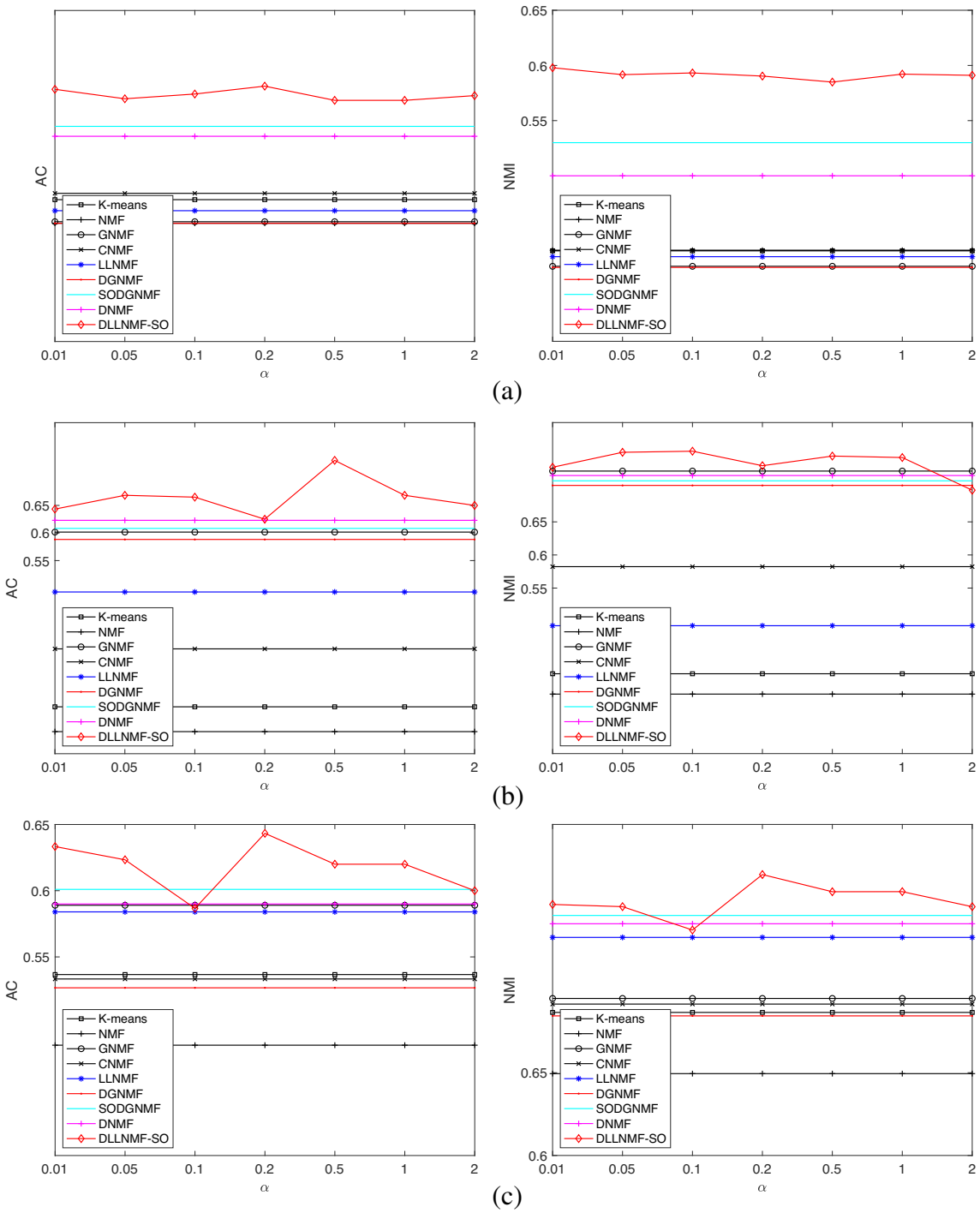


Fig. 5 The performances varied with parameter α

4.4 Parameter analysis

The model of our proposed DLLNMF-SO method contains four essential parameters:α, β, θ and λ. In above three experiments, we set α=0.1, β=2000, θ =50 and λ =40,000. In this section, we conducted some experiments on the FERET, Yale B, and ORL databases to mainly discuss the sensitivity of these parameters. Specifically, we varied only one parameter and fixed other parameters. Figures 5 and 6 plot the clustering results of seven algorithms with the different values of the parameters α and β on three datasets. The parameters α and β aims to control

the contributions of the local structure information of the data in the feature and data space. It can be seen that the performances of our proposed DLLNMF-SO can achieve the best performances in most cases. In addition, Figs. 7 and 8 show the clustering performances of all methods by varying θ and λ. It can be observed that the performance of our proposed DLLNMF-SO algorithm dramatically decreases if the values of the parameters θ and λ are too large or too small. Therefore, we can see that each regularization term of our DLLNMF-SO method plays a role in improving the clustering performances, which also demonstrates the motivations of imposing the constraints.

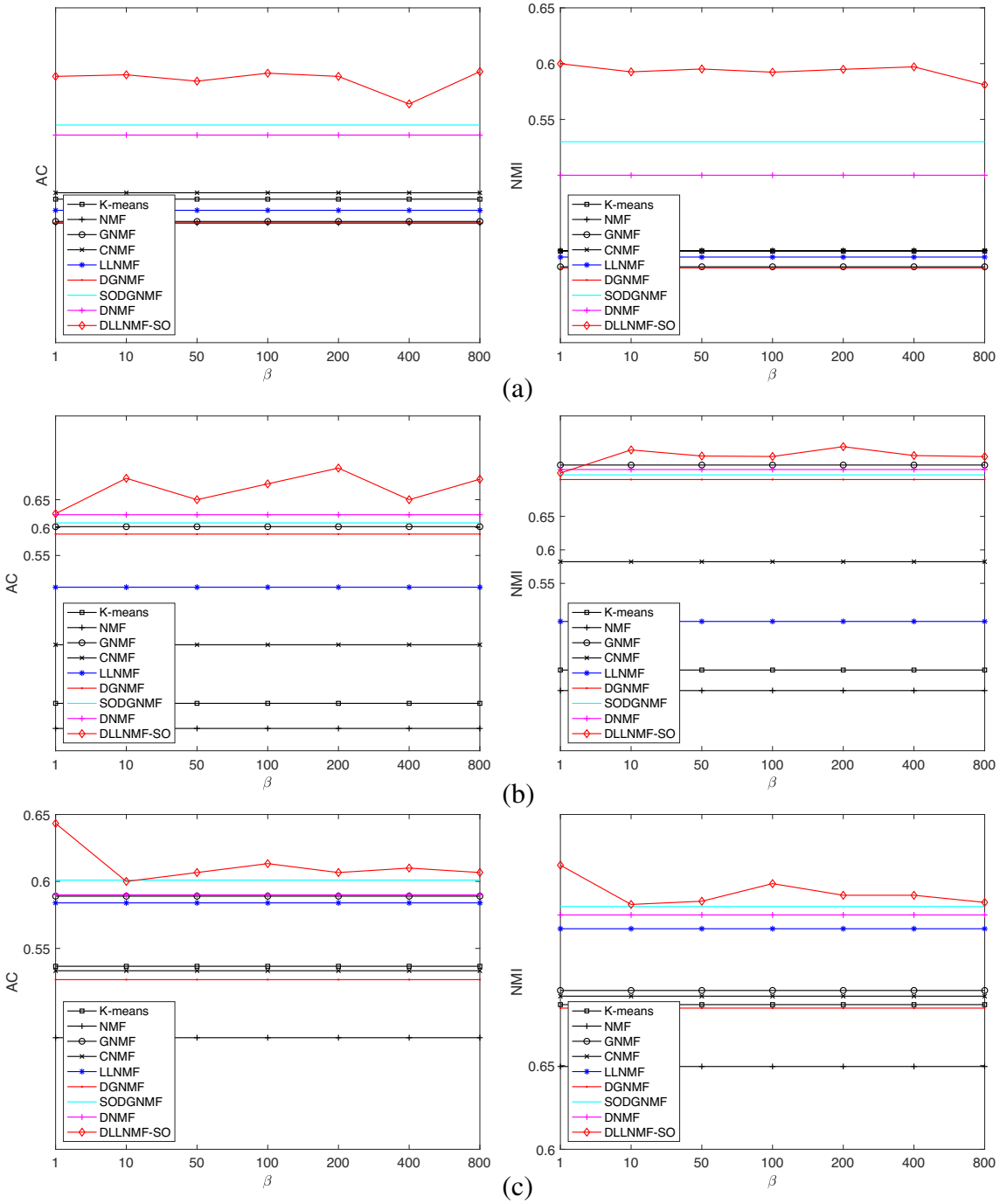


Fig. 6 The performances varied with parameter β

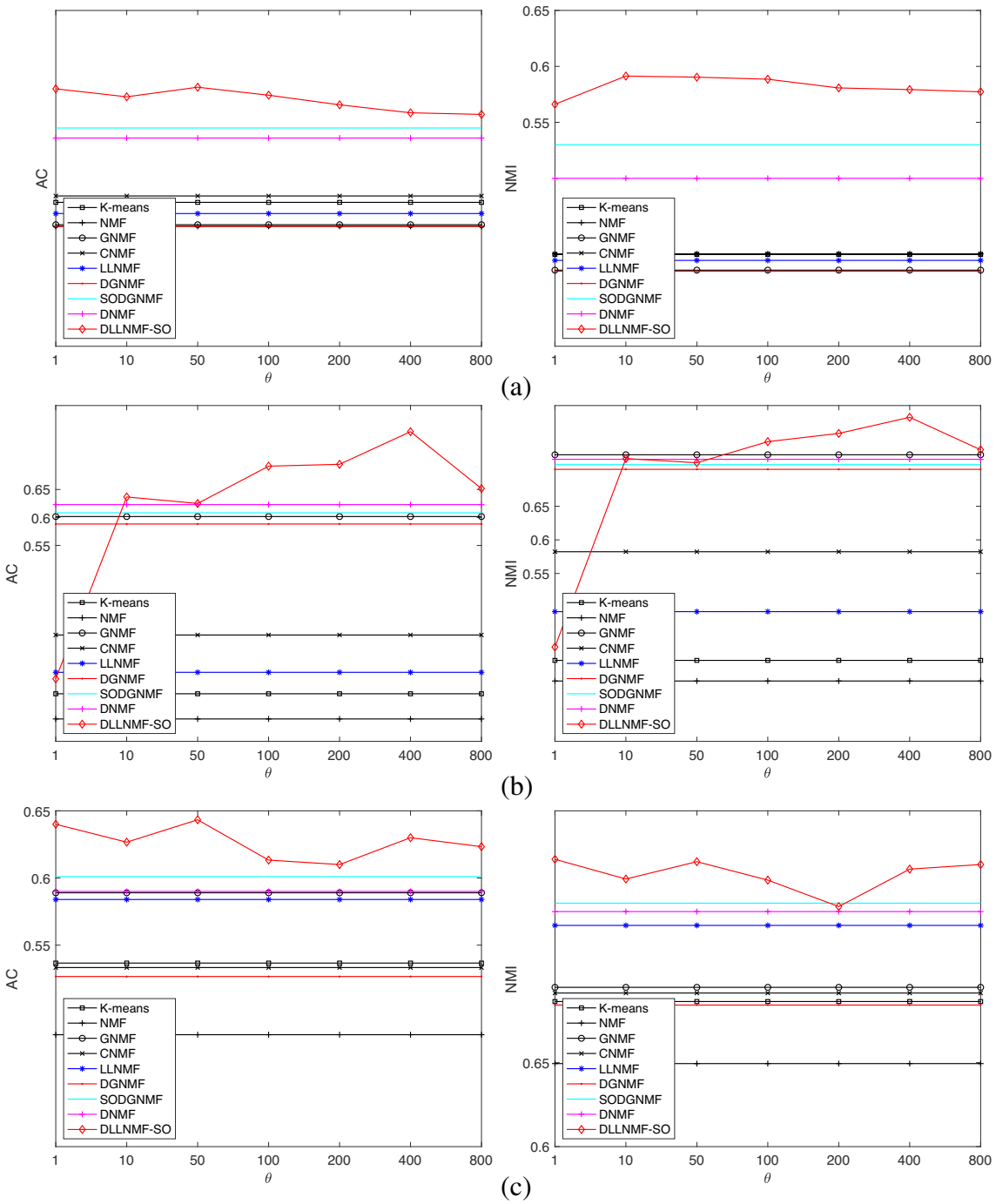


Fig. 7 The performances varied with parameter θ

5 Conclusion

In this paper, we have introduced a novel data representation method, called dual local learning regularized NMF with sparse and orthogonal constraints (DLLNMF-SO), in which the local geometric structure and discriminative information in both the data and feature space can be not only fully considered using a dual regularizer, but also the discriminative ability is greatly improved by imposing the sparse and orthogonal constraints. The iterative updating rules of the proposed

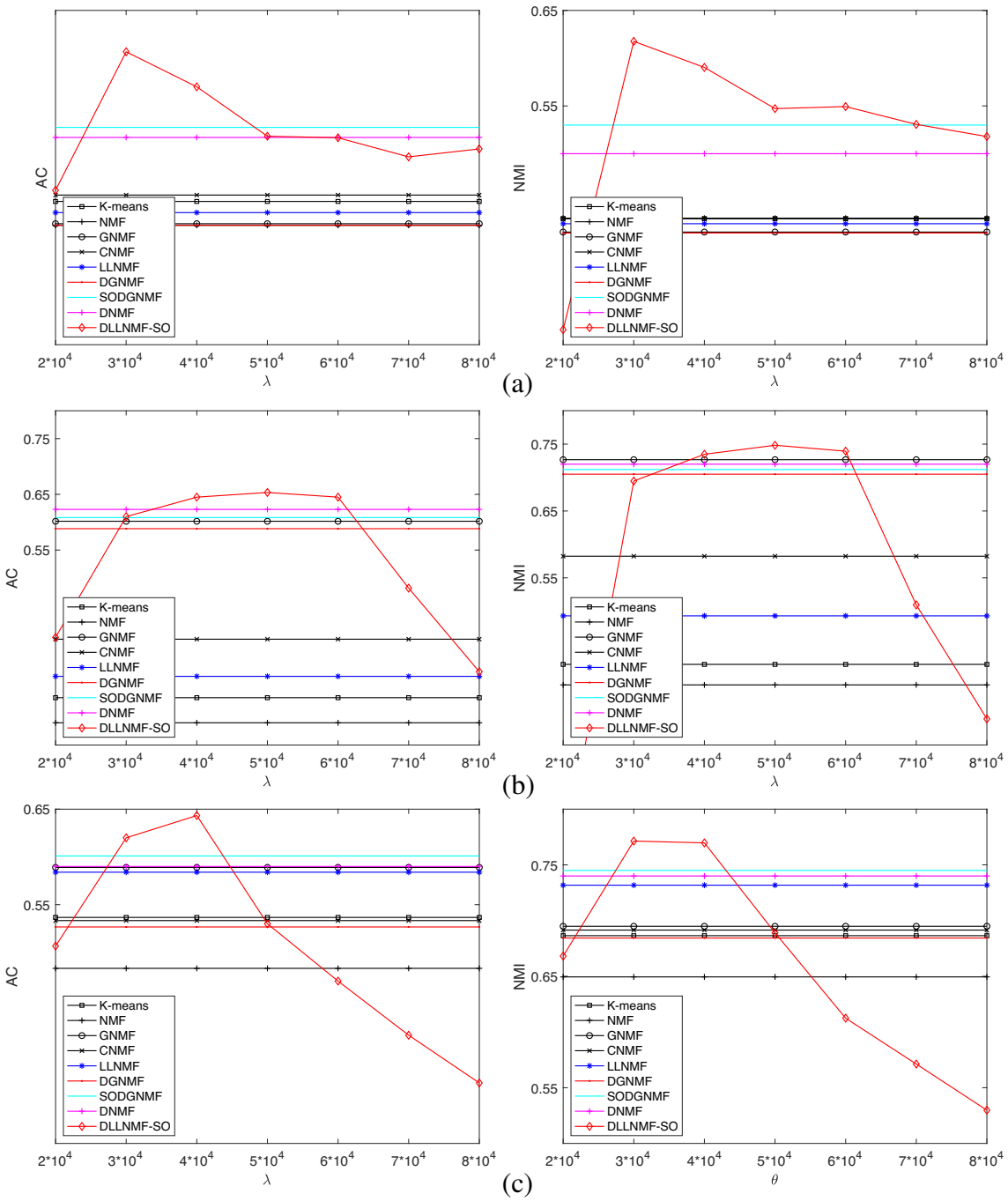


Fig. 8 The performances varied with parameter λ

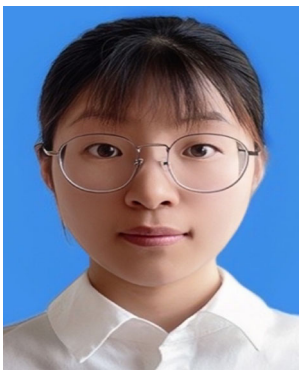
DLLNMF-SO method are given and its convergence proofs are proved. Numerical experiments on three benchmark databases have confirmed that the proposed DLLNMF-SO algorithm consistently outperforms other competitors in clustering.

Acknowledgements This work was supported by the National Natural Science Foundation of China [Grant No. 61603159, 61902160, U21B2027, 62162033], Yunnan Fundamental Research Projects [Grant No. 202101BE070001–056, 202101AT070438], Yunnan Provincial Major Science and Technology Special Plan Projects [Grant No. 202002AD080001, 202103AA080015].

References

1. Shu Z, Weng Z, Yu Z et al (2022) Correntropy-based dual graph regularized nonnegative matrix factorization with Lp smoothness for data representation. Appl Intell 52(7):7653–7669
2. Zhang Z, Zhang Y, Liu G et al (2020) Joint label prediction based semi-supervised adaptive concept factorization for robust data representation. IEEE Trans Knowl Data Eng 32(5):952–970
3. Shu Z, Wu X, Fan H et al (2017) Parameter-less auto-weighted multiple graph regularized nonnegative matrix factorization for data representation. Knowl-Based Syst 131:105–112
4. Jolliffe I (1989) Principal component analysis. Springer-Verlag, New York, NY, USA, pp 41–64
5. Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs fisherfaces: recognition using class specific linear projection. IEEE Trans Patt Anal Mach Intell 19(7):711–720
6. Lee DD, Seung HS (2000) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791
7. Yan S, Xu D, Zhang B (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29(1):40–51
8. Cai H, Liu B, Xiao Y (2020) Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization. Inf Sci 536(10):171–184
9. Zhang D, Wu X-J (2020) Scalable discrete matrix factorization and semantic autoencoder for cross-media retrieval. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2020.3032017
10. Jiao C, Gao Y, Yu N et al (2020) Hyper-graph regularized constrained NMF for selecting differentially expressed genes and tumor classification. IEEE J Biomed Health Inform 99:3002–3011
11. Lu X, Dong L, Yuan Y (2020) Subspace clustering constrained sparse NMF for hyperspectral unmixing. IEEE Trans Geosci Remote Sens 58(5):3007–3019. https://doi.org/10.1109/TGRS.2019.2946751
12. Xiu X, Fan J, Yang Y et al (2021) Fault detection using structured joint sparse nonnegative matrix factorization. IEEE Trans Instrum Meas 70:1–11
13. Cai D, He X, Han J et al (2011) Graph regularized nonnegative matrix factorization for data representation. IEEE Trans Pattern Anal Mach Intell 33(8):1548–1560
13. Gu Q, Zhou J (2009) Local learning regularized nonnegative matrix factorization. Twenty-first Int Joint Conf Artif Intell: 1044–1051
14. Shang F, Jiao LC, Wang F (2012) Graph dual regularization non-negative matrix factorization for co-clustering. Pattern Recogn 45(6):2237–2250
15. Wang C, Yu N, Wu M et al (2019) Dual hyper-graph regularized supervised NMF for selecting differentially expressed genes and tumor classification. IEEE Trans Comput Biol Bioinform 24(10): 3002–3011
16. Shu Z, Wu X, You C et al (2020) Rank-constrained nonnegative matrix factorization for data representation. Inf Sci 528:133–146
17. Shu Z, Zhou J, Huang P et al (2016) Local and global regularized sparse coding for data representation. Neurocomputing 198(29): 188–197
18. Shu Z, Sun Y, Tang J et al (2022) Adaptive graph regularized deep semi-nonnegative matrix factorization for data representation. Neural Process Lett. https://doi.org/10.1007/s11063-022-10882-x
19. Liu H, Wu Z, Cai D, Huang TS (2012) Constrained nonnegative matrix factorization for image representation. IEEE Trans Pattern Anal Mach Intell 34(7):1299–1311
20. Li Z, Tang J (2018) Robust structured nonnegative matrix factorization for image representation. IEEE Trans Neural Networks Learn Syst 29(5):1947–1960
21. Trigeorgis G, Bousmalis K et al (2017) A deep matrix factorization method for learning attribute representations. IEEE Trans Pattern Anal Mach Intell 39(3):1692–1700
22. Lu Y, Lai Z, Xu Y, Li X et al (2017) Nonnegative discriminant matrix factorization. IEEE Trans Circ Syst Video Technol 27(7): 1392–1405
23. Ma J, Zhang Y, Zhang L (2021) Discriminative subspace matrix factorization for multi-view data clustering. Patt Recogn. https://doi.org/10.1016/j.patcog.2020.107676
24. Li X, Zhang Y, Ge Z et al (2021) Adaptive nonnegative sparse representation for hyperspectral image super-resolution. IEEE J Selected Topics Appl Earth Observ Remote Sensing 14:4267–4283
25. Chen J, Yang S, Wang Z et al (2021) Efficient sparse representation for learning with high-dimensional data. IEEE Trans Neural Networks Learn Syst. https://doi.org/10.1109/TNNLS.2021.3119278
26. Ding C, Li T, Pen W et al (2006) Orthogonal nonnegative matrix tri-factorizations for clustering. ACM SIGKDD Int Conf Knowledge Discovery Data Mining:126–135
27. Shang R, Zhang Z, Jiao L et al (2016) Self-representation based dual-graph regularized feature selection clustering. Neurocomputing 171:1242–1253
28. Meng Y, Shang R, Jiao L, et al. Dual-graph regularized non-negative matrix factorization with sparse and orthogonal constraints Eng Appl Artif Intell, 2018, 69: 24–35

Zhenqiu Shu received the Ph.D. degree in computer applications at Nanjing University of Science and Technology. In February 2021, he joined the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, where he is currently an associate professor. Before joining in Kunming University of Science and Technology University, he had been a postdoctoral in Jiangnan University for four years. His research interests include image processing, computer vision and machine learning.



Furong Zuo is currently pursuing toward bachelor degree at School of Computer Engineering, Jiangsu University of Technology, Changzhou, China. His current research interests include pattern recognition, machine learning, and computer vision.

Wenli Wu is currently pursuing toward bachelor degree at School of Computer Engineering, Jiangsu University of Technology, Changzhou, China. His current research interests include pattern recognition, machine learning, and computer vision.

Congzhe You received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2016. He is currently a lecturer with School of Computer Engineering, Jiangsu University of Technology. His research interests include pattern recognition, machine learning, and computer vision.