Neurocomputing 551 (2023) 126521

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Multi-view clustering via label-embedded regularized NMF with dual-graph constraints

Bin Li [a], Zhenqiu Shu [a,*,1], Yingbo Liu [b], Cunli Mao [a], Shengxiang Gao [a], Zhengtao Yu [a]

[a] School of Faculty of Information Engineering and Automation, Kunming University of Science and Technology, China
[b] Big Data Research Institute of Yunnan Economy and Society, Yunnan University of Finance and Economics, China

ARTICLE INFO

*Article history:*
Received 17 October 2022
Revised 11 May 2023
Accepted 30 June 2023
Available online 4 July 2023
Communicated by Zidong Wang

*Keywords:*
NMF
Multi-view clustering
Common representation
Semi-supervised
Graph iterative optimization

ABSTRACT

Nonnegative matrix factorization (NMF) methods have achieved remarkable performances in multi-view clustering due to their effectiveness and efficiency. To better obtain a low-dimensional common representation, the limited labels and the geometric structure of the multi-view data should be fully utilized in clustering. In this work, we introduce a novel multi-view learning approach, dubbed label-embedded regularized NMF with dual-graph constraints (LeNMF-DC), for clustering. Our proposed LeNMF-DC approach mainly utilizes matrix factorization to obtain a low-dimensional common representation of the multi-view data, in which the prior knowledge hidden in data can be fully explored. Specifically, we construct three graph regularization terms to preserve the manifold structure in the data, feature and label space, respectively. Moreover, we take advantage of the labels of the labeled samples without additional parameters. In addition, we develop an alternate iterative optimization scheme to solve the model of LeNMF-DC and then show its convergence rate. Compared with traditional multi-view clustering approaches, the labels of unlabeled samples in our proposed LeNMF-DC approach are assigned by the label constraint matrix rather than the clustering algorithm, and thus it avoids performance loss during the clustering. Experimental results on four benchmark datasets manifest that our LeNMF-DC approach can achieve superior performances than several state-of-the-art approaches in multi-view clustering.


## 1. Introduction

With the rapid development of electronic information technology, the ways of obtaining data are becoming more and more abundant. Therefore, we can obtain different representations of the same object from several different observations. In other words, multi-view data is generated by adopting different feature extractors. For example, multiple different features, such as HOG [1], LBP [2] and GIST [3], can be used to describe the same image sample. Generally, each view has its specific property and the semantic information contained between the views complements each other. Therefore, multi-view clustering approaches aim at fusing complementary information between different views to improve clustering performance. Recently, they have been widely applied to many fields, such as image classification [4,5], genetic diagnosis [6], disease diagnosis [7], speech processing [8,9] and data mining [10,11], etc.

Over the last several years, various multi-view learning approaches have been developed to explore underlying correlations and complementary information between different views. Zhang et al. [12] introduced a binary multi-view clustering (BMVC) model for large-scale multi-view data, which encodes multi-view data containing complementary information into Hamming space and then uses matrix factorization to obtain the cluster structures. Inspired by multi-kernel learning, Zeng et al. [13] put forward a multi-kernel collaborative clustering model. It considers that the common multi-kernel space conforms to global clustering. Xu et al. [14] proposed to embed the multiple discriminative subspace learning in multi-view clustering while adaptively reconciling these subspaces. Ye et al. [15] developed multi-view robust double-sided twin SVM (MvRDTSVM) to guarantee excellent classification performance and robustness of multi-view learning. Wang et al. [16] proposed to flexibly capture the underlying clustering structure of multi-view data and achieve a satisfactory inter-view agreement. Most of the aforementioned approaches, however, cannot fully integrate the information contained in complex multi-view data.

Recently, several matrix factorization approaches have been widely developed for multi-view clustering owing to their effec-

* Corresponding author.
*E-mail address:* shuzhenqiu@163.com (Z. Shu).
[1] Bin Li and Zhenqiu Shu have contributed equally to this work.

https://doi.org/10.1016/j.neucom.2023.126521
0925-2312/

tiveness and efficiency. Liu et al. [17] proposed a joint matrix factorization framework to fuse individual view clusters into one common cluster. Motivated by deep learning, Wei et al. [18] decomposed the multi-view data into a common subspace using deep matrix factorization. Tan et al. [19] proposed a co-regularized NMF approach with correlation constraints (CO-NMFCC) for multi-view clustering, which can generate more stable clustering results by exploiting consistent and complementary information between different views. Its advantages are usually easy to implement and inexpensive computational cost. However, these approaches cannot remove the influence of noises and outliers in clustering. In addition, the structure information of multi-view data cannot be fully utilized. To overcome these shortcomings, motivated by manifold learning [20], several NMF approaches with structure constraints have received increasing attention [21–25]. Cai et al. [21] proposed to preserve the manifold structure of the single-view data by imposing a graph regularization constraint. To improve the robustness of the algorithm, Du et al. [22] further proposed a graph-regularized multiple NMF approach using $L_{2,1}$-norm to measure the reconstruction error. Zhang et al. [23] utilized the graph-based regularized NMF approach to effectively fuse useful information contained in multi-view data. Rai et al. [24] introduced the NMF-based partial view clustering (PVC) approach. It is only applicable to two-view data but can be extended to $k$ partial-view scenario. Zong et al. [25] proposed a multi-manifold regularized NMF approach to exploit the local geometric structure of multi-view data. In order to utilize the label information of the data, some semi-supervised NMF methods [26–30] were proposed for multi-view learning. In references [26–28], they constructed the label constraint matrix and then incorporated it into the representation of each view. Liang et al. [29,30] attempted to improve the multi-view clustering performance by embedding the known labels into the consensus representation. Although the above semi-supervised NMF methods fully utilize the known labels among multi-view data, they seldom consider their manifold structure information.

To fully utilize the geometric structure information of both the multi-view data and the limited labels, we propose a novel multi-view clustering approach, namely label-embedded regularized NMF with dual-graph constraints (LeNMF-DC). In our LeNMF-DC approach, more prior knowledge, such as data manifold, feature manifold and label distribution, hidden in multi-view data are fully considered by using the regularization technology compared with traditional multi-view clustering approaches. Experimental results on four benchmark datasets demonstrate the advantage of our proposed LeNMF-DC approach in multi-view clustering. The main contributions of our work are summarized as follows:

(1) Our proposed LeNMF-DC approach constructs a label-embedded regularization term and then integrates it into the NMF framework. Therefore, the label distribution structure of multi-view data can be preserved in the new representation space.
(2) In LeNMF-DC, we adaptively learn the optimal combination of the graphs in the feature space and then construct the graph regularization term based the common representation. Thus, the feature manifold can be effectively preserved in low-dimensional feature representation. In addition, we linearly combine the graph regularizers of each view in the data space to capture the data manifold of multi-view data.
(3) Different from the aforementioned multi-view clustering approaches, our LeNMF-DC approach is able to assign the labels to the unlabeled samples by the label constraint matrix rather than the result of the clustering algorithm. Therefore, it can effectively reduce the performance loss during clustering procedure by using different clustering strategies.
(4) We develop an efficient alternate optimization scheme to solve the proposed model. Experimental results on several benchmark multi-view datasets demonstrate that our LeNMF-DC approach achieves better performances than other competitors in clustering.

The rest of the paper is organized as follows: In section 2, we review some related works of our proposed approach. Section 3 introduces the proposed approach in detail. Section 4 shows the experimental evaluations and analysis. Section 5 draws the conclusion of this work.

## 2. Related works

In this section, we briefly review the related modes of our proposed approach, such as NMF [31], constrained NMF (CNMF) [32] and dual graph regularized NMF (DNMF) [33]. To show the proposed model more clearly, Table 1 summarizes the notations used in this paper and their descriptions.

### 2.1. NMF and CNMF

Given a non-negative data matrix $X \in R^{m\times n}$, NMF aims at finding two non-negative matrices $U \in R^{m\times k}$ and $V \in R^{n\times k}$, and uses their product to approximate the data matrix $X$. Since $k \ll m$, the coefficient matrix $V$ can be regarded as a low-dimensional representation of the original data matrix. Mathematically, the NMF model is formulated as follows:

$$\min_{U,V}\left\|X - UV^T\right\|_F^2 \quad s.t. U \geqslant 0, V \geqslant 0, \tag{1}$$

where $\|.\|_F^2$ denotes the Frobenius norm.

To take advantage of the limited labels among data, several semi-supervised learning variants of NMF [34,35,10] were developed over the past few years. CNMF [32] was proposed to use the limited labels as a hard constraint to guide the matrix decomposition. Specifically, the indicator matrix $C$ is constructed from the known samples. $c_{ij} = 1$ denotes that the sample $x_i$ belongs to $j$th class, otherwise $c_{ij} = 0$. Then, CNMF constructs the label matrix $A$ based on the indicator matrix $C$ as follows:

$$A = \begin{pmatrix} C_{l\times c} & 0 \\ 0 & I_{n-l} \end{pmatrix}, \tag{2}$$

where $I_{n-l}$ is an identity matrix. The label constraint matrix $A$ is used to constrain the coefficient matrix $V$ by $V = AZ$, where $Z$ is the auxiliary matrix. It can be seen that if two data points $x_i$ and $x_j$ share the same label, their representation in low-dimensional space should be the same. Therefore, the objective function of CNMF can be expressed as follows:

$$O_F = \left\|X - UZ^TA^T\right\|_F^2. \tag{3}$$

### 2.2. DNMF

To explore the data manifold and the feature manifold, simultaneously, Shang et al.[33] proposed dual graph regularized NMF (DNMF) method for data representation. The geometric information of the data manifold and the feature manifold is considered in the matrix decomposition by constructing a nearest neighbor graph in the data space and the feature space, respectively. Therefore, the objective function of DNMF can be formulated as follows:

**Table 1**
Notations and descriptions.

| Notation | Description |
|---|---|
| $X \in \mathbb{R}^{m\times n}$ | The non-negative data matrix |
| $X^{(s)} \in \mathbb{R}^{m_s\times n}$ | The data matrix of the $s$-th view |
| $U \in \mathbb{R}^{m\times k}$ | The basis matrix in NMF |
| $U^{(s)} \in \mathbb{R}^{m_s\times k}$ | The basis matrix for the $s$-th view |
| $V \in \mathbb{R}^{n\times k}$ | The coefficient matrix in NMF |
| $A \in \mathbb{R}^{n\times c}$ | The label indicator matrix, $A=[A_l;A_u]$ $A_l \in \mathbb{R}^{l\times c}$, $A_u \in \mathbb{R}^{(n-l)\times c}$ |
| $Z \in \mathbb{R}^{c\times k}$ | The auxiliary matrix |
| $W^{V^{(s)}} \in \mathbb{R}^{n\times n}$ | The data graph matrix of the $s$-th view |
| $W^{U^{(s)}} \in \mathbb{R}^{m_s\times m_s}$ | The feature graph matrix of the $s$-th view |
| $D^{V^{(s)}} \in \mathbb{R}^{n\times n}$ | The diagonal matrix corresponding $W^{V^{(s)}}$ |
| $D^{U^{(s)}} \in \mathbb{R}^{m_s\times m_s}$ | The diagonal matrix corresponding $W^{U^{(s)}}$ |
| $L^V \in \mathbb{R}^{n\times n}$ | The data graph Laplacian matrix |
| $L^{U^{(s)}} \in \mathbb{R}^{m_s\times m_s}$ | The feature graph Laplacian matrix of the $s$-th view |
| $\tau_A^{(s)}$ | The weighting coefficient of $s$-th data graph Laplacian matrix |
| $\tau_V^{(s)}$ | The weighting coefficient of $s$-th feature graph Laplacian matrix |
| $n$ | The number of training samples |
| $l$ | The number of training samples with labels |
| $m_s$ | The feature dimensionality of the $s$-th view |
| $k$ | The feature dimensionality of the common space |
| $c$ | The number of data clusters |
| $\alpha, \lambda, \gamma$ | The trade-off regularization parameter |

$$\min_{U,V}\left\|X-UV^T\right\|_F^2+\lambda tr(V^TL^VV)+\mu tr(U^TL^UU),\quad \text{s.t. } U\geqslant 0, V\geqslant,0 \tag{4}$$

where $\lambda$ and $\mu$ are the nonnegative regularization parameters. $tr(U^TL^UU)$ and $tr(V^TL^VV)$ are the smoothness constraints imposed on the feature and the data, respectively. $L^U$ is the feature graph Laplacian matrix, and $L^V$ is the data graph Laplacian matrix, which are calculated by $L^U=D^U-W^U$ and $L^V=D^V-W^V$ respectively. Here, $D$ denotes the diagonal matrix and $W$ denotes the weight matrix of the graph. The 0–1 weighting scheme is usually used to construct $k$-nearest neighbor graph. Therefore, the affinity matrix of data graph is defined as follows:

$$W_{ij}^V=\begin{cases}1, & \text{if } X_{:,j}\in N_p(X_{:,i}),\\ 0, & \text{otherwise.}\end{cases} \tag{5}$$

where $N_p(X_{:,i})$ represents the nearest neighbor set of $X_{:,i}$.

Similarly, the affinity matrix of feature graph is defined as follows

$$W_{ij}^U=\begin{cases}1, & \text{if } X_{:,j}^T\in N_p\left(X_{:,i}^T\right),\\ 0, & \text{otherwise.}\end{cases} \tag{6}$$

where $N_p(X_{:,i}^T)$ represents the nearest neighbor set of $X_{:,i}^T$.

### 2.3. Multi-view NMF

In multi-view clustering, NMF aim at learning the consensus representation from different views, which can fully exploit their correlations and complementary information. Liu et al. [17] roposed a multi-view clustering method via joint NMF (multiNMF), which learns the low-dimensional representation of each view separately and then seeks to obtain their common representation. However, it cannot directly learn the common representation from matrix decomposition. Therefore, Jia et al. [36] attempt to decompose the data matrix $\left\{X^{(s)}\right\}_{s=1}^S$ into a basis matrix set $\left\{U^{(s)}\right\}_{s=1}^S$ and a coefficient matrix $V$ as follows:

$$\min_{\left\{U^{(s)}\right\}_{s=1}^S,V}\sum_{s=1}^S\left\|X^{(s)}-U^{(s)}V^T\right\|_F^2,\quad \text{s.t. } U_{ik}^{(s)}\geqslant 0, V_{kj}\geqslant 0 \tag{7}$$

Here, one of the disadvantages in references [16,17] is that the manifold structure hidden in multi-view data in references [16,17] is neglected, and thus their performance improvement in clustering tasks is extremely limited. To solve this issues, Zhang et al. [23] introduced the feature graph regularizer into multiNMF for object recognition. Inspired by DNMF [33], Luo et al. [37] proposed the dual regularized multi-view NMF (DMvNMF) algorithm, which considers the data manifold and feature manifold of multi-view data. To further improve the clustering performance by fully utilizing the known labels, various semi-supervised NMF methods [38,39,30] were proposed for multi-view clustering. However, they cannot model the manifold structure of the known labels. To mitigate these issues, the proposed LeNMF-DC considers the geometric structure of data manifold, feature manifold and label distribution, simultaneously, and then incorporates them into the matrix decomposition. Therefore, it can more effectively capture the intrinsic feature information of multi-view data compared with traditional multi-view clustering methods.

## 3. The proposed method

In this section, we introduce our LeNMF-DC method in detail.

### 3.1. Motivation

To take full use of the supervised information, various variants of NMF were proposed for multi-view clustering over the past several years [33–35]. However, the aforementioned approaches have the following drawbacks: 1) They neglect the label distribution structure among the multi-view data; 2) The dual manifold structure embedded in multi-view data cannot be fully considered. To overcome these issues, we propose a novel semi-supervised learning approach, called LeNMF-DC, for multi-view clustering. In LeNMF-DC, the data graph regularization terms are used to explore the structure information of the data manifold. In addition, it adaptively learns the weights of the feature graphs constructed from different views, and then uses it to construct the graph regularizer in the feature space and the label space, respectively. Unlike most existing multi-view clustering approaches, our proposed LeNMF-DC approach is capable of assigning the labels for the unlabeled samples by the learned label constraint matrix instead of the clusting algorithm. Consequently, it effectively reduces the influence of the clustering algorithm on the final results. Fig. 1 shows the flowchart of the proposed LeNMF-DC approach.

### 3.2. Objective Function

Given a multi-view dataset $\left\{X^{(s)}\in R^{m_s\times n}\right\}_{s=1}^S$ consisting of $S$ views with $c$ category samples, the first $l$ samples are labeled and the rest $n-l$ samples are unlabeled. To fully utilize the limited label information, we construct the label indicator matrix $A$ as follows:

$$A=[A_l;A_u]=[a_1,\ldots,a_l,a_{l+1},\ldots,a_n]^T\in R^{n\times c}, \tag{8}$$

where $A_l$ and $A_u$ denote the label indicator matrix of the labeled samples and the unlabeled samples, respectively. Here, $a_{ij}=1$ denotes that $x_i$ is labeled with the $j$th class, otherwise $a_{ij}=0$. Inspired by [32,40], we integrate the label matrix $A$ into matrix decomposition as follows:

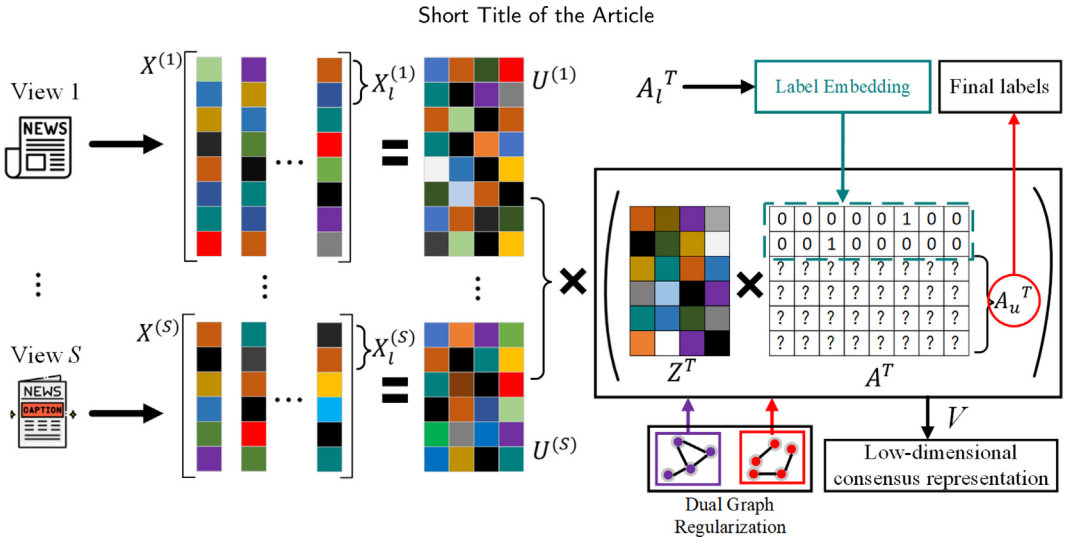


**Fig. 1.** The flowchart of the proposed LeNMF-DC approach.

$$V = AZ, \tag{9}$$

where $Z \in R^{c\times k}$ is an auxiliary matrix. It is worthing that the label indicator matrix $A$ in Eq. (9) is dynamically updated while the label indicator matrix in Eq. (2) is predefined and fixed. Therefore, our proposed LeNMF-DC method can directly learn the label matrix of the unlabeled samples.

To effectively discover the geometric structure information embedded into multi-view data, we construct two nearest graphs in the data and feature space, respectively. According to [41], the affinity matrix of the data graph is given as follows:

$$W_{ij}^{V^{(s)}} = \begin{cases} 1, & \text{if } x_j^{(s)} \in N_p\left(x_i^{(s)}\right) \text{or} x_i^{(s)} \in N_p\left(x_j^{(s)}\right) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $N_p(x_i^{(s)})$ represents the nearest neighbor set of $x_i^{(s)}$. The Laplacian matrix of the data graph is defined as $L^V = D^V - W^V$ and $D^V$ is a diagonal matrix calculated by $D_{ii}^V = \sum_j W_{ij}^V$. Therefore, the label structure distribution regularizer can be expressed as follows:

$$\begin{aligned} R_A &= \sum_{s=1}^{S} (\tau_A^{(s)})^{\mu_1} (\tfrac{1}{2} \sum_{i,j=1}^{n} \left\| a_i - a_j \right\|^2 W^{V^{(s)}}) \\ &= \sum_{s=1}^{S} (\tau_A^{(s)})^{\mu_1} ((\sum_{i=1}^{n} a_i a_i^T D_{ii}^{V^{(s)}}) - (\sum_{i,j=1}^{n} a_i a_i^T W_{ij}^{V^{(s)}})) \\ &= \sum_{s=1}^{S} (\tau_A^{(s)})^{\mu_1} (Tr(A^T D^{V^{(s)}} A) - Tr(A^T W^{V^{(s)}} A)) \\ &= Tr(A^T (\sum_{s=1}^{S} (\tau_A^{(s)})^{\mu_1} L^{V^{(s)}}) A) = Tr(A^T L^A A) \\ &\text{s.t.} \sum_{s=1}^{S} \tau_A^{(s)} = 1, \tau_A^{(s)} \geqslant 0 \end{aligned} \tag{11}$$

where $\tau_A^{(s)}$ denotes the weighting coefficient of $s$-th Laplacian matrix and $\mu_1$ is the parameter to control the weight distribution. Similarly, we can construct the manifold structure regularizer in the data space by minimizing the following problem:

$$\begin{aligned} R_V &= \sum_{s=1}^{S} (\tau_V^{(s)})^{\mu_2} (\tfrac{1}{2} \sum_{i,j=1}^{n} \tau_V^{(s)} \left\| Az_i - Az_j \right\|^2 W^{V^{(s)}}) \\ &= \sum_{s=1}^{S} (\tau_V^{(s)})^{\mu_2} (\sum_{i=1}^{n} Az_i z_i^T A^T D_{ii}^{V^{(s)}}) - \sum_{s=1}^{S} (\tau_V^{(s)})^{\mu_2} (\sum_{i,j=1}^{n} Az_i z_i^T A^T W_{ij}^{V^{(s)}}) \\ &= \sum_{s=1}^{S} (\tau_V^{(s)})^{\mu_2} Tr(Z^T A^T D^{V^{(s)}} AZ) - \sum_{s=1}^{S} (\tau_V^{(s)})^{\mu_2} Tr(Z^T A^T W^{V^{(s)}} AZ) \\ &= Tr(Z^T A^T (\sum_{s=1}^{S} (\tau_V^{(s)})^{\mu_2} L^{V^{(s)}}) AZ) \\ &= Tr(Z^T A^T L^V AZ) \\ &\text{s.t.} \sum_{s=1}^{S} \tau_V^{(s)} = 1, \tau_V^{(s)} \geqslant 0 \end{aligned} \tag{12}$$

where $\tau_V^{(s)}$ is the weighting coefficient of the $s$-th Laplacian graph and $\mu_2$ is the parameter to control the weights distribution. We can see that Eq. (12) can adaptively learn the optimal linear combination of the graphs from different views.

Then we construct a $p$-nearest neighbour graph in the feature space. The affinity matrix $W_{ij}^{U^{(s)}}$ of the feature graph is expressed as follows:

$$W_{ij}^{U^{(s)}} = \begin{cases} 1, & \text{if } x_j^{(s)^T} \in N_p\left(x_i^{(s)^T}\right) \text{or} x_i^{(s)^T} \in N_p\left(x_j^{(s)^T}\right) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $N_p(x_i^{(s)^T})$ represents the $p$-nearest neighbor set of $x_i^{(s)^T}$. The basis matrix of the $s$-view is expressed as $U_r^{(s)} = \left[u_1^{(s)}, u_2^{(s)}, \cdots, u_{m_s}^{(s)}\right]^T \in R^{m_s \times k}$. $D^{U^{(s)}}$ is a diagonal matrix and the feature graph Laplacian matrix is $L^{U^{(s)}} = D^{U^{(s)}} - W^{U^{(s)}}$. Therefore, the graph regularizer in the feature space is given as follows:

$$\begin{aligned} R_U^{(s)} &= \tfrac{1}{2} \sum_{i,j=1}^{m_s} \left\| u_i^{(s)} - u_j^{(s)} \right\|^2 W^{U^{(s)}} \\ &= \sum_{i=1}^{m_s} u_i^{(s)} u_i^{(s)^T} D_{ii}^{U^{(s)}} - \sum_{i,j=1}^{m_s} u_i^{(s)} u_i^{(s)^T} W_{ij}^{U^{(s)}} \\ &= Tr(U^{(s)^T} D^{U^{(s)}} U^{(s)}) - Tr(U^{(s)^T} W^{U^{(s)}} U^{(s)}) \\ &= Tr(U^{(s)^T} L^{U^{(s)}} U^{(s)}) \end{aligned} \tag{14}$$

The two graph regularizers in Eq. (12) and Eq. (14) are collectively referred to as dual graph constraints, which are designed to exploit the manifold structures of multi-view data in the data and feature spaces. To preserve the distribution structure of the feature, data and labels, the objective function of our proposed LeNMF-DC approach is written as follows:

$$\begin{aligned}&\min \sum_{s=1}^{S}\left\|X^{(s)}-U^{(s)}(AZ)^{T}\right\|_{F}^{2}+\sum_{s=1}^{S}\gamma tr(U^{(s)^{T}}L^{U^{(s)}}U^{(s)})\\&+\alpha tr(Z^{T}A^{T}L^{V}AZ)+\lambda tr(A^{T}L^{A}A)\\&\text{s.t. } U^{(s)}\geqslant 0, A\geqslant 0, Z\geqslant 0, L^{V}=\sum_{s=1}^{S}(\tau_{V}^{(s)})^{\mu_{2}}L^{V^{(s)}},\\&\sum_{s=1}^{S}\tau_{V}^{(s)}=1, L^{A}=\sum_{s=1}^{S}(\tau_{A}^{(s)})^{\mu_{1}}L^{V^{(s)}}, \sum_{s=1}^{S}\tau_{A}^{(s)}=1,\\&\tau_{V}^{(s)}\geqslant 0, \tau_{A}^{(s)}\geqslant 0\end{aligned} \tag{15}$$

where $\gamma, \alpha$ and $\lambda$ denote the regularization parameters.

### 3.3. Optimization

According to $\|X\|_{F}^{2}=tr(XX^{T})$, we develop an alternating iterative optimization scheme to solve the model (15). Therefore, the objective function (15) is rewritten as follows:

$$\begin{aligned}&O_{F}=\sum_{s=1}^{S}tr(X^{(s)}X^{(s)^{T}}+U^{(s)}(AZ)^{T}AZU^{(s)^{T}}-2X^{(s)}AZU^{(s)^{T}})\\&+\alpha tr(Z^{T}A^{T}L^{V}AZ)+\lambda tr(A^{T}L^{A}A)+\sum_{s=1}^{S}\gamma tr(U^{(s)^{T}}L^{U^{(s)}}U^{(s)})\\&\text{s.t. } L^{A}=\sum_{s=1}^{S}\tau_{A}^{(s)}L^{V^{(s)}}, L^{V}=\sum_{s=1}^{S}\tau_{V}^{(s)}L^{V^{(s)}}\\&\sum_{s=1}^{S}\tau_{V}^{(s)}=1, \sum_{s=1}^{S}\tau_{A}^{(s)}=1\\&U^{(s)}\geqslant 0, A\geqslant 0, Z\geqslant 0, \tau_{V}^{(s)}\geqslant 0, \tau_{A}^{(s)}\geqslant 0\end{aligned} \tag{16}$$

Let $\sigma, \varrho$ and $\omega$ be the Lagrange multipliers for the constraints $U^{(s)}\geqslant 0, Z\geqslant 0$ and $A\geqslant 0$, respectively. Therefore, the Lagrange function $L$ for problem (14) is formulated as follows:

$$L=O_{F}+tr(\sigma U^{(s)^{T}})+tr(\varrho Z^{T})+tr(\omega A^{T}), \tag{17}$$

(1) **Updating** $U^{(s)}$: By calculating the partial derivative of Eq. (15) w.r.t. $U^{(s)}$, we have

$$\frac{\partial L}{\partial U^{(s)}}=-2X^{(s)}AZ+2U^{(s)}(AZ)^{T}AZ+2\gamma L^{U^{(s)}}U^{(s)}+\sigma, \tag{18}$$

According to Karush–Kuhn–Tucker (KKT) condition $\sigma_{ij}^{(s)}u_{ij}^{(s)}=0$, the formula (16) is rewritten as follows:

$$(-X^{(s)}AZ+U^{(s)}(AZ)^{T}AZ+\gamma L^{U^{(s)}}U^{(s)})_{ij}(u^{(s)})_{ij}=0, \tag{19}$$

Therefore, we can get the updating rule of $U^{(s)}$ as follows:

$$u_{ij}^{(s)}\leftarrow u_{ij}^{(s)}\frac{(X^{(s)}AZ+\gamma W^{U^{(s)}}U^{(s)})_{ij}}{(U^{(s)}(AZ)^{T}AZ+\gamma D^{U^{(s)}}U^{(s)})_{ij}}, \tag{20}$$

(2) **Updating** $A$: We take the partial derivative of Eq. (15) w.r.t. $A$, and get the following formula:

$$\begin{aligned}\frac{\partial L}{\partial A}=&-2X^{(s)^{T}}U^{(s)}Z^{T}+2AZU^{(s)^{T}}U^{(s)}Z^{T}+2\lambda L^{A}A+2\alpha L^{V}AZZ^{T}\\&+\omega,\end{aligned} \tag{21}$$

Similarly, according to KKT condition $\omega_{ij}a_{ij}=0$, the problem (19) is reformulated as the following form:

$$\left(-X^{(s)^{T}}U^{(s)}Z^{T}+AZU^{(s)^{T}}U^{(s)}Z^{T}+\lambda L^{A}A+\alpha L^{V}AZZ^{T}\right)_{ij}a_{ij}=0, \tag{22}$$

Therefore, we can get the following updating rules:

$$a_{ij}\leftarrow a_{ij}\frac{(X^{(s)^{T}}U^{(s)}Z^{T}+\lambda W^{V}A+\alpha W^{V}AZZ^{T})_{ij}}{(AZU^{(s)^{T}}U^{(s)}Z^{T}+\lambda D^{V}A+\alpha D^{V}AZZ^{T})_{ij}}, \tag{23}$$

(3) **Updating** $Z$: By taking the partial derivative of Eq. (15) w.r.t. $Z$ and then setting it to zero, we have

$$\frac{\partial L}{\partial Z}=-2A^{T}X^{(s)^{T}}U^{(s)}+2A^{T}AZU^{(s)^{T}}U^{(s)}+2\alpha A^{T}L^{V}AZ+\varrho, \tag{24}$$

Since KKT condition $\varrho_{ij}z_{ij}=0$, we get the following equation:

$$(-A^{T}X^{(s)^{T}}U^{(s)}+A^{T}AZU^{(s)^{T}}U^{(s)}+\alpha A^{T}L^{V}AZ)_{ij}z_{ij}=0, \tag{25}$$

Since $L^{V}=D^{V}-W^{V}$, we derive the updating rule as follows:

$$z_{ij}\leftarrow z_{ij}\frac{(A^{T}X^{(s)^{T}}U^{(s)}+\alpha A^{T}W^{V}AZ)_{ij}}{(A^{T}AZU^{(s)^{T}}U^{(s)}+\alpha A^{T}W^{V}AZ)_{ij}}, \tag{26}$$

(4) **Updating** $\tau_{A}^{(s)}$: Let us denote $P_{A}^{(s)}=Tr(A^{T}L^{A^{(s)}}A)$. The objective function (17) w.r.t. $\tau_{A}^{(s)}$ is given as follows:

$$\begin{aligned}&\min_{\tau_{A}^{(s)}}\sum_{s=1}^{S}(\tau_{A}^{(s)})^{\mu_{1}}P_{A}^{(s)}\\&\text{s.t. } \sum_{s=1}^{S}\tau_{A}^{(s)}=1, \tau_{A}^{(s)}\geqslant 0\end{aligned} \tag{27}$$

The Lagrange function of Eq. (27) is rewritten as follows:

$$\min_{\tau_{A}^{(s)}}\sum_{s=1}^{S}(\tau_{A}^{(s)})^{\mu_{1}}P_{A}^{(s)}-\psi_{1}(\sum_{s=1}^{S}\tau_{A}^{(s)}-1), \tag{28}$$

where $\psi_{1}$ is the Lagrange multiplier. By setting the derivative of Eq. (28) w.r.t. $\tau_{A}^{(s)}$ to zero, we have

$$\tau_{A}^{(s)}=(\frac{\psi_{1}}{\mu_{1}P_{A}^{(s)}})^{\frac{1}{\mu_{1}-1}}, \tag{29}$$

We replace $\tau_{A}^{(s)}$ into $\sum_{s=1}^{S}\tau_{A}^{(s)}=1$ in Eq. (29) and then obtain

$$\tau_{A}^{(s)}=\frac{(\mu_{1}P_{A}^{(s)})^{\frac{1}{1-\mu_{1}}}}{\sum_{s=1}^{S}\mu_{1}P_{A}^{(s)}}, \tag{30}$$

(5) **Updating** $\tau_{V}^{(s)}$: Let us denote $P_{V}^{(s)}=Tr(Z^{T}A^{T}L^{V^{(s)}}AZ)$. The objective function (17) w.r.t. $\tau_{V}^{(s)}$ is written as follows:

$$\begin{aligned}&\min_{\tau_{s}}\sum_{s=1}^{S}(\tau_{V}^{(s)})^{\mu_{2}}P_{V}^{(s)}\\&\text{s.t. } \sum_{s=1}^{S}\tau_{V}^{(s)}=1, \tau_{V}^{(s)}\geqslant 0\end{aligned} \tag{31}$$

The Lagrange function of Eq. (31) is written as follows:

$$\min_{\tau_{V}^{(s)}}\sum_{s=1}^{S}(\tau_{V}^{(s)})^{\mu_{2}}P_{V}^{(s)}-\psi_{2}(\sum_{s=1}^{S}\tau_{V}^{(s)}-1), \tag{32}$$

where $\psi_{2}$ is the Lagrange multiplier. Similarly, by setting the derivative of Eq. (32) w.r.t. $\tau_{V}^{(s)}$ to zero, we obtain

$$\tau_V^{(s)} = \left(\frac{\psi_2}{\mu_2 P_V^{(s)}}\right)^{\frac{1}{\mu_2 - 1}}, \tag{33}$$

We substitute $\sum_{s=1}^{S} \tau_V^{(s)} = 1$ for $\tau_V^{(s)}$ in Eq. (33) and then have

$$\tau_V^{(s)} = \frac{\left(\frac{\mu_2 P_V^{(s)}}{S}\right)^{\frac{1}{1-\mu_2}}}{\sum_{s=1}^{S} \mu_2 P_V^{(s)}}, \tag{34}$$

The optimization procedure of our LeNMF-DC algorithm is summarized in Algorithm 1.

**Algorithm 1**: Our LeNMF-DC algorithm

**Require:** Given a multi-view data $\left\{X^{(s)} \in R^{m_s \times n}\right\}_{s=1}^{S} = \{x_1^{(s)}, x_2^{(s)}, \cdots, x_n^{(s)}\}$, the first $l$ data points are labeled with $c$ class and the remaining $n - l$ data points are unlabeled; the dimension of low-dimensional representation $k$; the parameters $\lambda, \alpha\ \gamma, \mu_1$ and $\mu_2$; the maximum iteration times $T$;

**Ensure:** The index of the largest element in each row of $A$ as the label.

1: **Initialize:**
2: Construct the weight matrix $W^{(V)}$ by Eq. (10).
3: Construct the weight matrix $W^{(U^{(s)})}$ by Eq. (13).
4: Initialize the weight $\tau_A^{(s)} = \frac{1}{S}$ and $\tau_V^{(s)} = \frac{1}{S}$ for each view;
5: Initialize $U^{(s)}, A_u$ and $Z$.
6: **while** $t < T$ **do**
7: Update $U^{(s)}$ using Eq. (20);
8: Update $A$ using Eq. (23);
9: Update $Z$ using Eq. (26);
10: Update $\tau_A^{(s)}$ using Eq. (30);
11: Update $\tau_V^{(s)}$ using Eq. (34);
12: **end while**

### 3.4. Computational Complexity

The main computational cost of the LeNMF-DC approach includes the construction of the weight matrix and the iterative optimization procedure. The computational complexity is denoted using the big $O$ notation. $c$ represents the number of clusters, $k$ stands for the desired reduced dimension. We take $O(n^2m + nm^2)$ computational cost to construct the matrices $W^{U^{(s)}}$ and $W^{V^{(s)}}$. For each round of updates, the time complexity of updating $U^{(s)}$ is $O(2mck + 2m^2k + 3mk)$, the time complexity of updating $A$ is $O(4cnk + 2n^2c + 3nc)$ and the time complexity of updating $Z$ is $O(2mkc + 2c^2k + 4ck)$. In addition, we need $O(n^2ck)$ and $O(m^2nk)$ to update $\tau_A^{(s)}$ and $\tau_A^{(s)}$, respectively. Assuming that our proposed model converges after $t$ iterations, the total computational complexity is $O(t(n^2m + nm^2 + S(4mkc + 2m^2k + 2n^2c + 2c^2k + m^2nk + n^2ck)))$. Since $c$ and $k$ are generally smaller than $m$ and $n$, respectively, the overall computational complexity of our proposed optimization scheme is $O(t(n^2m + nm^2 + S(m^2(2k + nk) + n^2(2c + ck))))$.

## 4. Experiments

In this section, we conducted a series of experiments on four benchmark multi-view datasets to verify the effectiveness of our proposed LeNMF-DC approach.

### 4.1. Datasets and Evaluation Metrics

In our work, the Handwritten, COIL20, BBC and BBCSport datasets were used to the experimental dataset. The relevant properties of these datasets can be briefly summarized as follows.

**Handwritten 2 Sources** [2] [42]: This handwritten digits (0–9) dataset is collected from two sources (MNIST handwritten digits and USPS handwritten digits) with a total of 2000 samples.

**COIL20** [3] [43]: It consists of 1400 images of 20 classes. For each image, the 1024-dimensional intensity, 3304 -dimensional LBP and 6750-dimensional Gabor feature are extracted to form a multi-view dataset.

**BBC** [4] [44]: It consists of 685 documents related to five news topics (business, entertainment, politics, sports, technology) from the BBC News Corpus. In this dataset, each document is described by four different views.

**BBCSport** [5] [44]: This dataset consists of 544 documents linked to five themes (athletics, cricket, football, rugby and tennis) collected from the BBC Sport website. In this dataset, each document is described by two views.

To fairly evaluate the performances of different algorithms, we employed six most well-known evaluation metrics including Accuracy (ACC)[45], Normalized Mutual Information (NMI)[46], Purity [47], F-score, Precision and Recall. Their detailed definition are given as follows:

**(1) ACC:** It can be calculated by the following formula:

$$\mathrm{ACC} = \frac{\sum_{1}^{n} \partial(g_i, map(l_i))}{n}, \tag{35}$$

where $l_i$ is the output label of the model, $g_i$ is the true label of the sample and $n$ is the number of samples. $map(\cdot)$ is the optimal mapping function. The delta function $\partial(\cdot) = 1$ if $g_i = map(l_i)$, otherwise $\partial(\cdot) = 0$.

**(2) NMI:** The mutual information (MI) can be calculated by the following formula:

$$\mathrm{MI}(L, G) = \sum_{l_i \in C, g_i \in G} p(l_i, g_i) log_2 \frac{p(l_i, g_i)}{p(l_i) \cdot p(g_i)}, \tag{36}$$

where $p(\cdot)$ is the probability and $p(\cdot, \cdot)$ is the joint probability. The definition of NMI is given as follows:

$$\mathrm{NMI}(L, G) = \frac{\mathrm{MI}(L, G)}{max(H(L), H(G))}, \tag{37}$$

where $H(\cdot)$ denotes the entropy function. $L$ is the label set predicted by the model and $G$ is the true label set of the samples. In general, the more similar the two categories are, the greater the NMI value is.

**(3) Purity:** In clustering, the definition of purity is given as follows:

$$\mathrm{Purity}(\Omega, C) = \frac{1}{N} \sum_{k} \max_{j} |w_k \cap c_j|, \tag{38}$$

where $N$ is the total number of samples. $\Omega = \{w_1, w_2, \cdots, w_K\}$ denotes cluster division and $C = \{c_1, c_2, \cdots, c_j\}$ denotes the real class division. It is easy to know that the value of purity ranges from 0 to 1. Generally, the closer the value of purity is to 1, the better the clustering result is.

**(4) Precision, Recall and F-score:** We first introduce four concepts: **(a)** TP: It indicates that similar samples are assigned to the

[2] **https://cs.nyu.edu/roweis/data.html**
[3] **https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php**
[4] **http://mlg.ucd.ie/datasets/bbc.html**
[5] **http://mlg.ucd.ie/datasets/bbc.html**

same category; **(b)** TN: It means that different type samples are assigned to different categories; **(c)** FP: It indicates that different type samples are classified into the same category; **(d)** FN: It indicates that similar samples are classified into different categories. Therefore, Precision can be defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{39}$$

It is obvious that Precision is the percentage of true positives in the retrieved results. Recall can be computed as follows:

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{40}$$

and F-score can be calculate as follows:

$$\text{F} - \text{score} = (1 + \tau^2)\frac{\text{Precision} \cdot \text{Recall}}{\tau^2(\text{Precision} + \text{Recall})}, \tag{41}$$

where the parameter $\tau$ is used to balance the weight of Precision and Recall.

### 4.2. Baselines

We compared the proposed LeNMF-DC approach with other state-of-the-art competitors. These algorithms were mainly divided into three categories: (1) Single-view clustering algorithms: NMF [31] and graph regularized NMF (GNMF [21])); (2) Unsupervised learning multi-view clustering algorithms: multi-view clustering via joint NMF (MultiNMF [17]), multi-view clustering via deep matrix factorization (DMVC [48]), auto-weighted multi-view clustering via deep matrix decomposition (AWMVC [49]), NMF with co-orthogonal constraints (NMF-CC [50]), pseudo-label guided collective matrix factorization (PLCMF [51]); (3) Semi-supervised learning multi-view clustering algorithms: partially shared latent factor (PSLF [38]), partially shared deep matrix factorization (PSDMF [39]), graph-regularized partially shared NMF (GPSNMF [30]). These algorithms are described in detail as follows:

- **NMF**: This is the basic NMF approach for single-view data representation.
- **GNMF**: It imposes a graph regularization constraint on the NMF for single-view data representation.
- **MultiNMF**: This multi-view approach learns a consensus matrix from the multi-view data via a soft regularization term for clustering.
- **DMVC**: This approach adopts deep matrix factorization approach to represent the multi-view data.
- **AWMVC**: It exposes the hierarchical semantics of input data in a hierarchical manner and automatically assigns the weight for each view.
- **NMFCC**: It imposes the co-orthogonal constraints to deal with influence of orthogonality of basis matrix and internal vector representing matrix.
- **PSLF**: It uses for the shared latent representation learning which composed of specific latent factors in each view, and thus the consistency and complementary among views are fully explored. According to different final labels generated by different clustering approaches, PSLF can be extended into two different applications: (1) PSLF-$k$: the cluster labels generated by $k$-means; (2) PSLF-$w$: the cluster labels obtained by $Y = W^T V$, where $W$ denotse the regression coefficient matrix and $V$ denotes the latent factor.
- **PLCMF**: It performs the clustering on each view to generate the pseudo-labels and then guides the matrix factorization.
- **PSDMF**: This approach incorporates partially common structures into deep matrix factorization and combines with graph regularizer and semi-supervised regression models. Therefore, it can remove the influence of irrelevant information.
- **GPSNMF**: It encodes the geometric structure information using the affinity graph constructed from each view. Similar to PSLF, GPSNMF can also be divided into GPSNMF-$k$ and GPSNMF-$w$ according to the different clustering algorithms.

### 4.3. Clustering Results

In this subsection, we conducted some experiments to evaluate the effectiveness of the proposed approach on the Hadwritten2-sources, COIL20, BBC and BBCSport datasets. Among these baseline algorithms, both NMF and GNMF belong to single-view algorithms and the rest are multi-view learning approaches. For single-view algorithms, the best performance was different views is used as the final result. For multi-view approaches, we run them ten times and recorded the average values as the final results. In addition, we reported the clustering performances of the semi-supervised learning algorithms when the available labeled samples were set to 5%, 10% and 15%, respectively. In particular, the best performances under different label ratios were displayed in bold with different colors.

The clustering results of our LeNMF-DC approach and all baseline comparison approaches on the Hadwritten2sources, COIL20, BBC and BBCSports datasets are shown in Table 2–6, respectively. As can be seen from Table 2, when the label ratio on the Hadwritten2sources dataset is set to 10%, compared with PSDMF, the proposed LeNMF-DC method improves 8.37%, 6.85%, 5.11%, 7.55% and 2.29% on ACC, Purity, F-score, Precision and Recall, respectively. Besides, when the label ratio on the COIL20 dataset is set to 5%, our proposed method can achieve 3.28%, 1.51%, 3.28%, 2.76%, 2.4% and 3.12% improvement on ACC, NMI, Purity, F-score, Precision and Recall in comparison to GPSNMF-w, respectively. In addition, it can be seen that our proposed method outperforms other competitors on both the BBC and BBCSport datasets in most cases. Specifically, when the label ratio on the BBC dataset is set to 15%, our method can improve ACC, NMI, Purity, F-score, Precision and Recall by 2.59%, 4.06%, 3.63%, 3.86%, 14.82%, and 3.13% compared with the PSDMF method, respectively. Furthermore, we can find that our proposed method can achieve the best performances on the BBCSport dataset in most cases regardless of the proportion of labeled samples. Moreover, it can be found that our proposed method achieves better performances than other unsupervised multi-view clustering methods on different datasets.

According to the experimental results on four datasets, we can get the following observations:

(1) It is noted that the clustering performances of the semi-supervised learning approaches, such as PLSF, PSDMF and GPSNMF and LeNMF-DC, can be improved as the number of labeled samples increases. It indicates the labels among the multi-view data can effectively improve the performances in clustering. This phenomenon is consistent with our understanding of semi-supervised learning theory.
(2) It can be seen that the proposed LeNMF-DC approach achieves very competitive performances compared with other competitors on four datasets. This is because our proposed LeNMF-DC approach take more prior knowledge hidden in multi-view data (e.g. data manifold, feature manifold, label distribution) into account than other approaches. It demonstrates the effectiveness of our proposed LeNMF-DC approach in multi-view clustering.

**Table 2**
Clustering performance on the Handwritten2Sources dataset.

| Method | label ratio | ACC(%) | NMI(%) | Purity(%) | F-score(%) | Precision(%) | Recall(%) |
|---|---|---|---|---|---|---|---|
| NMF | - | 57.20±3.14 | 48.93±2.88 | 57.20±2.71 | 42.43±1.58 | 41.28±2.01 | 43.65±2.24 |
| GNMF | - | 53.10±3.51 | 51.12±1.96 | 58.15±2.45 | 42.44±2.25 | 40.22±2.63 | 44.45±3.10 |
| Multi-NMF | - | 45.76±2.54 | 37.67±3.20 | 59.76±2.93 | 43.15±1.17 | 37.22±4.08 | 51.94±3.25 |
| DMVC | - | 50.49±2.56 | 55.29±2.08 | 46.79±2.31 | 43.38±1.59 | 41.39±2.92 | 45.58±1.97 |
| AWMVC | - | 78.33±1.24 | 77.58±1.62 | 70.49±1.14 | 73.58± 1.38 | 70.12±1.20 | 77.54±1.36 |
| NMF-CC | - | 66.20±2.51 | 54.31±1.69 | 66.20±2.50 | 48.93±2.14 | 47.35±1.65 | 50.63±1.76 |
| PLCMF | - | 73.50±0.88 | 63.79±1.04 | 74.75±0.47 | 62.42±1.21 | 59.21±0.69 | 66.00±1.11 |
| PLSF-*k* | 5% | 62.39±2.41 | 62.04±3.02 | 66.92±2.67 | 53.78±2.58 | 50.36±1.24 | 57.78±3.07 |
| | 10% | 60.51±1.24 | 59.45±2.27 | 64.52±3.36 | 51.35±2.52 | 48.43±3.16 | 54.74±4.12 |
| | 15% | 60.83±3.16 | 60.43±2.83 | 65.40±2.96 | 52.33±3.14 | 49.05±3.57 | 56.26±2.62 |
| PLSF-*w* | 5% | 74.06±1.21 | 56.77±1.38 | 74.06±2.07 | 56.73±2.43 | 55.74±1.85 | 57.76±1.97 |
| | 10% | 81.54±1.36 | 66.56±2.31 | 81.34±1.96 | 66.89±2.58 | 66.36±3.14 | 67.44±2.37 |
| | 15% | 84.86±2.14 | 71.50±2.52 | 84.86±1.57 | 72.52±1.85 | 72.22±2.27 | 72.82±2.62 |
| PSDMF | 5% | 81.14±6.47 | **82.62±6.95** | 82.75±7.64 | **76.00±6.18** | 72.40±4.28 | **80.17±8.20** |
| | 10% | 82.55±7.53 | ***83.88±6.88*** | 84.07±5.73 | 77.92±7.84 | 75.10±6.88 | 81.12±6.52 |
| | 15% | 83.19±9.24 | **84.32±5.76** | 84.86±5.77 | 78.96±4.79 | 76.39±7.43 | 81.90±6.80 |
| GPSNMF-*k* | 5% | 70.05±2.12 | 62.81±1.69 | 71.42±2.39 | 58.45±2.64 | 57.91±4.36 | 59.00±2.58 |
| | 10% | 75.78±2.25 | 68.84±4.72 | 72.58±4.16 | 63.25±3.96 | 62.53±4.27 | 65.07±4.41 |
| | 15% | 80.59±2.53 | 72.80±2.86 | 80.59±3.47 | 68.67±2.59 | 66.99±4.68 | 70.44±2.49 |
| GPSNMF-*w* | 5% | 73.32±2.24 | 55.27±3.26 | 73.32±3.51 | 55.81±3.36 | 54.88±2.96 | 56.77±3.49 |
| | 10% | 83.61±2.86 | 69.12±3.57 | 83.61±3.88 | 70.68±2.66 | 70.49±4.13 | 70.87±3.58 |
| | 15% | 90.59±2.37 | 80.78±3.25 | 90.59±3.07 | 82.16±3.61 | 81.98±2.76 | 82.35±2.50 |
| Ours | 5% | **85.78±2.51** | 76.16±2.69 | **85.78±3.61** | 75.05±2.52 | **74.24±1.96** | 75.90±2.88 |
| | 10% | ***90.92±2.15*** | 82.82±2.01 | ***90.92±1.58*** | ***83.03±1.64*** | ***82.65±2.25*** | ***83.41±1.34*** |
| | 15% | **91.72±2.14** | 84.21±2.30 | **91.72±1.83** | **84.43±2.68** | **84.05±2.66** | **84.42±2.74** |

**Table 3**
Clustering performance on the COIL20 dataset.

| Method | label ratio | ACC(%) | NMI(%) | Purity(%) | F-score(%) | Precision(%) | Recall(%) |
|---|---|---|---|---|---|---|---|
| NMF | - | 62.01±1.35 | 73.35±2.56 | 65.97±1.96 | 55.17±2.41 | 48.36±3.74 | 64.22±2.93 |
| GNMF | - | 80.21±3.61 | 93.20±2.17 | 85.07±2.15 | 82.03±2.10 | 77.03±3.07 | 87.74±5.63 |
| Multi-NMF | - | 62.51±2.54 | 75.66±3.20 | 64.32±2.93 | 56.12±1.17 | 46.37±4.08 | 70.28±3.25 |
| DMVC | - | 81.04±3.26 | 92.31±2.53 | 81.53±1.58 | 79.28±1.93 | 72.32±3.42 | 88.96±1.97 |
| AWMVC | - | 61.04±2.31 | 75.27±2.11 | 64.67±3.62 | 66.05± 1.83 | 83.38±3.09 | 54.68±1.48 |
| NMF-CC | - | 78.92±5.12 | 89.17±1.53 | 80.33±3.07 | 75.68±3.36 | 71.04±3.76 | 82.47±1.69 |
| PLCMF | - | 67.64±0.69 | 78.10±1.53 | 72.12±0.12 | 63.62±1.29 | 59.64±2.61 | 67.59±0.83 |
| PLSF-*k* | 5% | 65.60±2.69 | 79.79±4.74 | 70.62±2.76 | 58.32±8.74 | 50.35±7.59 | 69.81±4.27 |
| | 10% | 68.75±6.91 | 82.23±3.19 | 74.85±4.63 | 61.31±7.05 | 52.63±5.36 | 73.42±5.61 |
| | 15% | 71.73±2.97 | 78.97±2.35 | 73.69±5.27 | 58.63±3.81 | 50.61±5.01 | 69.67±6.30 |
| PLSF-*w* | 5% | 71.78±4.76 | 77.28± 7.38 | 73.68±4.09 | 62.46±3.29 | 59.49±7.72 | 65.75±8.69 |
| | 10% | 79.94±3.71 | 82.47±5.13 | 79.94±1.63 | 70.37±3.64 | 65.94±4.09 | 75.44±5.68 |
| | 15% | 83.25±3.39 | 84.50±3.17 | 83.91±2.93 | 75.84±3.02 | 75.50±2.87 | 76.19±2.33 |
| PSDMF | 5% | 64.85±6.47 | 50.77±6.95 | 68.39±7.64 | 54.47±6.18 | 40.56±4.28 | 66.04±8.20 |
| | 10% | 67.72±7.53 | 55.39±6.88 | 70.59±5.73 | 58.75±7.84 | 46.09±6.88 | 68.39±6.52 |
| | 15% | 70.66±9.24 | 58.46±5.76 | 74.27±5.77 | 74.35±4.79 | 52.59±7.43 | 74.85±6.80 |
| GPSNMF-*k* | 5% | 79.90±2.69 | 89.32±5.37 | 81.51±7.50 | 78.45±6.82 | 76.24±5.94 | 80.79±7.16 |
| | 10% | 81.64±2.57 | 90.35±5.68 | 83.72±4.37 | 80.97±6.62 | 79.18±4.59 | 82.84±4.29 |
| | 15% | 84.89±5.89 | 91.18±2.67 | 87.09±4.19 | 82.80±7.53 | 81.26±6.93 | 84.41±1.93 |
| GPSNMF-*w* | 5% | 92.03±2.38 | 94.57±4.13 | 92.03±6.10 | 89.17±3.97 | 87.79±5.92 | 90.61±8.49 |
| | 10% | 97.38±2.54 | ***97.74±6.15*** | 97.38±3.59 | 95.73±5.88 | 95.39±4.72 | 96.08±3.91 |
| | 15% | 98.77±2.93 | 98.73±5.81 | 98.77±4.17 | 97.78±3.61 | 97.63±2.59 | 97.93±4.13 |
| Ours | 5% | **95.31±5.37** | **96.08±2.90** | **95.31±5.41** | **91.93±2.63** | **90.19±5.17** | **93.73±2.06** |
| | 10% | ***97.85±4.93*** | 97.61±4.20 | ***97.85±4.73*** | ***96.02±2.16*** | ***95.73±3.55*** | ***96.32±3.43*** |
| | 15% | **99.11±3.92** | **98.88±3.18** | **99.11±3.01** | **98.28±3.69** | **98.22±1.57** | **98.35±3.28** |

(3) Both PLSF and GPSNMF adopt different clustering strategies on four datasets to obtain the final labels. It can be found from the experimental results that PLSF and GPSNMF obtains differences performances with different clustering strategies. Since our proposed LeNMF-DC approach can assign the labels of the unlabeled samples according to the label indicator matrix, it effectively avoids the performance loss owing to the clustering algorithm.
(4) It is easy to know that GNMF, PSDMF and GPSNMF employ the graph regularizer to preserve the manifold structural information of the original data. However, our proposed LeNMF-DC approach constructs the dual graph regularization term in the data and feature space to explore the feature manifold and the data manifold. Therefore, our proposed LeNMF-DC approach can achieve better performances than these approaches.
(5) We can see that our proposed LeNMF-DC approach achieves better performances compared with other semi-supervised learning approaches. The possible reason is that our proposed LeNMF-DC approach adopts an automatic weighting strategy to construct the label regularizer and thus preserves the label distribution of the labeled samples in low-dimensional feature space.

### *4.4. Parameter Sensitivity Analysis and Ablation Experiments*

It can be seen that our proposed LeNMF-DC approach includes the following parameters: the desired reduced dimension $k$, the

**Table 4**
Clustering performance on the BBC dataset.

| Method | label ratio | ACC(%) | NMI(%) | Purity(%) | F-score(%) | Precision(%) | Recall(%) |
|---|---|---|---|---|---|---|---|
| NMF | - | 52.57±1.64 | 36.08±0.87 | 40.86±0.35 | 43.07±2.69 | 48.91±5.49 | 62.52±3.01 |
| GNMF | - | 62.62±4.08 | 44.06±1.28 | 53.57±0.09 | 43.07±6.25 | 52.25±2.14 | 57.85±3.72 |
| Multi-NMF | - | 46.13±1.08 | 24.45±0.87 | 46.28±1.53 | 43.03±0.96 | 42.95±2.60 | 43.27±0.53 |
| DMVC | - | 48.18±1.63 | 38.91±0.48 | 46.06±5.39 | 25.67±0.31 | 38.46±0.57 | 40.97±1.29 |
| AWMVC | - | 57.34±3.68 | 39.53±2.57 | 68.85±1.79 | 48.96±2.64 | 40.62±1.72 | 64.23±2.66 |
| NMF-CC | - | 67.36±1.05 | 43.99±0.96 | 65.87±0.37 | 52.48±1.04 | 54.27±0.63 | 60.89±1.82 |
| PLCMF | - | 63.36±0.54 | 39.95±0.08 | 63.36±0.19 | 58.59±0.58 | 47.06±0.62 | 75.60±1.06 |
| PLSF-*k* | 5% | 66.21±6.39 | 44.19±5.47 | 66.21±5.27 | 50.47±8.36 | 38.55±6.58 | 73.04±7.19 |
| | 10% | 69.64±2.53 | 42.29±1.82 | 69.64±1.96 | 58.80±6.19 | 48.43±3.75 | 74.82±5.38 |
| | 15% | 77.15±2.81 | 54.36±4.83 | 77.15±1.96 | 64.89±3.66 | 55.60±4.88 | 77.91±3.63 |
| PLSF-*w* | 5% | 73.58±1.58 | 49.23±2.37 | 73.58±2.13 | 70.97±3.82 | 58.08±1.92 | **73.66±1.24** |
| | 10% | 76.32±4.17 | 54.21±2.63 | 75.32±1.95 | 62.59±3.17 | 54.16±1.67 | 74.14±1.05 |
| | 15% | 80.76±2.39 | 56.55±0.96 | 80.76±1.53 | 70.13±2.57 | 63.64±3.16 | 78.11±2.81 |
| PSDMF | 5% | 66.21±6.39 | 44.19±5.47 | 66.21±5.27 | 50.47±8.36 | 38.55±6.58 | 73.04±7.19 |
| | 10% | 69.64±5.49 | 53.67±6.68 | 74.21±6.29 | 62.37±7.96 | 58.76±9.18 | 76.39±7.53 |
| | 15% | 81.62±5.69 | 64.31±5.39 | 82.24±8.57 | 75.69±5.13 | 68.57±6.85 | 72.93±5.73 |
| GPSNMF | 5% | 75.73±3.24 | 55.07±1.77 | 75.73±2.15 | 62.05±3.20 | 51.42±1.59 | 78.22±2.31 |
| | 10% | 78.90±1.82 | ***64.21±1.03*** | 80.52±1.52 | ***71.38±2.03*** | 64.18±3.64 | ***80.41±3.51*** |
| | 15% | 80.24±1.27 | 64.88±2.51 | 80.24±1.09 | 68.70±2.59 | 69.95±3.67 | 67.50±1.52 |
| GPSNMF-*w* | 5% | 77.11±2.01 | 53.73±1.65 | 77.11±3.58 | 62.13±2.67 | 55.43±3.18 | 70.69±2.49 |
| | 10% | 81.82±2.68 | 58.62±3.24 | 81.82±1.98 | 70.57±2.62 | 64.51±4.06 | 77.88±2.73 |
| | 15% | 80.07±1.64 | 56.54±5.34 | 80.07±5.50 | 68.38±2.63 | 60.51±2.49 | **78.60±3.04** |
| Ours | 5% | **84.95±4.02** | **61.90±5.47** | **84.95±3.68** | **73.86±2.90** | **75.94±2.51** | 71.90±2.47 |
| | 10% | ***81.98±3.68*** | 60.79±3.32 | ***81.98±4.09*** | 70.44±3.91 | ***74.04±3.73*** | 67.19±4.21 |
| | 15% | **84.21±2.57** | **68.37±3.26** | **85.87±1.96** | **79.55±5.07** | **83.39±3.92** | 76.06±3.28 |

**Table 5**
Clustering performance on the BBCSport dataset.

| Method | label ratio | ACC(%) | NMI(%) | Purity(%) | F-score(%) | Precision(%) | Recall(%) |
|---|---|---|---|---|---|---|---|
| NMF | - | 57.77±0.69 | 45.06±0.91 | 44.87±1.25 | 48.44±2.63 | 42.95±1.64 | 43.27±3.07 |
| GNMF | - | 54.69±3.25 | 56.77±1.97 | 56.36±3.19 | 56.36±1.59 | 56.63±0.93 | 58.84±3.61 |
| Multi-NMF | - | 53.74±2.56 | 32.71±2.14 | 54.72±1.91 | 54.87±2.80 | 52.25±0.27 | 57.85±1.95 |
| DMVC | - | 41.47±1.57 | 23.35±1.94 | 49.15±2.01 | 36.87±3.72 | 38.24±2.49 | 43.16±1.84 |
| AWMVC | - | 65.33±1.87 | 41.52±2.51 | 62.62±2.26 | 55.67±0.93 | 49.83±4.57 | 61.75±1.10 |
| NMF-CC | - | 56.71±4.28 | 41.94±1.63 | 63.07±4.72 | 49.23±5.01 | 45.78±2.69 | 53.68±1.54 |
| PLCMF | - | 74.22±0.52 | 55.98±1.94 | 79.48±0.36 | 63.87±0.61 | 67.76±0.28 | 60.58±1.81 |
| PLSF-*k* | 5% | 61.70±2.12 | 42.24±3.63 | 64.22±2.58 | 50.82±1.69 | 37.55±4.79 | 78.62±2.83 |
| | 10% | 68.16±5.36 | 48.55±5.27 | 68.16±3.61 | 51.18±4.15 | 38.51±2.69 | 76.32±1.87 |
| | 15% | 81.65±2.47 | 53.39±3.62 | 75.54±1.94 | 63.57±2.59 | 61.80±3.27 | 65.44±4.30 |
| PLSF-*w* | 5% | 80.46±5.13 | 59.51±3.62 | 80.46±1.83 | 66.76±2.47 | 60.24±3.06 | 74.86±4.26 |
| | 10% | 87.96±6.07 | 71.70±3.82 | 87.96±2.96 | 78.91±4.60 | 74.35±2.17 | 84.08±5.99 |
| | 15% | 90.91±2.17 | 78.87±4.23 | 90.91±1.90 | 81.59±5.30 | 76.89±1.33 | 86.90±4.29 |
| PSDMF | 5% | 78.29±5.87 | 60.67±6.43 | 89.68±7.76 | 70.58±8.62 | 68.74±5.49 | 72.87±6.37 |
| | 10% | 85.58±8.65 | 65.92±7.59 | 82.64±4.96 | 73.79±7.54 | 68.53±6.68 | 76.26±5.76 |
| | 15% | 86.84±6.58 | 68.76±7.52 | 83.71±5.89 | 76.56±4.29 | 74.91±7.56 | 78.83±8.55 |
| GPSNMF-*k* | 5% | 81.24±2.67 | 68.21±3.06 | 82.40±2.25 | 68.15±1.86 | 56.60±1.93 | **85.64±2.17** |
| | 10% | 84.90±3.51 | 72.46±3.24 | 84.90±2.76 | 77.49±1.68 | 78.00±1.09 | 77.00±3.51 |
| | 15% | 88.31±2.37 | 82.27±1.54 | 88.31±4.06 | 81.49±2.51 | 71.69±3.32 | **94.38±2.49** |
| GPSNMF-*w* | 5% | 81.82±2.47 | 62.15±4.18 | 81.82±1.83 | 71.20±3.29 | 64.66±3.64 | 79.22±3.14 |
| | 10% | 86.73±3.07 | 68.88±2.49 | 86.73±4.16 | 76.82±4.02 | 70.55±3.64 | 84.32±5.07 |
| | 15% | 88.37±2.87 | 72.37±3.29 | 88.37±3.66 | 80.73±2.13 | 76.46±3.00 | 85.52±5.13 |
| Ours | 5% | **89.98±1.30** | **77.30±0.68** | **90.22±0.27** | **85.35±1.15** | **85.57±0.68** | 85.17±0.57 |
| | 10% | ***94.08±0.28*** | ***82.12±0.58*** | ***94.08±2.21*** | ***88.04±1.84*** | ***89.23±0.41*** | ***86.89±2.07*** |
| | 15% | **95.04±0.47** | **84.77±1.54** | **95.04±1.29** | **89.42±2.39** | **91.11±2.51** | 87.80±0.96 |

**Table 6**
Ablation experiment results.

| methods | HW2sources | | | COIL20 | | | BBC | | | BBCSport | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | NMI | Purity | ACC | NMI | Purity | ACC | NMI | Purity | ACC | NMI | Purity |
| LeNMF-DC,$\lambda$=0 | 0.8665 | 0.7651 | 0.8665 | 0.9355 | 0.9456 | 0.9355 | 0.8170 | 0.5576 | 0.8170 | 0.9271 | 0.8190 | 0.9271 |
| LeNMF-DC,$\alpha$=0 | 0.8499 | 0.7456 | 0.8499 | 0.9056 | 0.9225 | 0.9056 | 0.8595 | 0.6298 | 0.8595 | 0.9191 | 0.8066 | 0.9191 |
| LeNMF-DC,$\gamma$=0 | 0.6932 | 0.6043 | 0.6934 | 0.4503 | 0.5139 | 0.4539 | 0.6893 | 0.4129 | 0.6893 | 0.7746 | 0.5311 | 0.7746 |
| LeNMF-DC,$\mu_1$=0 | 0.8281 | 0.7201 | 0.8281 | 0.9538 | 0.9563 | 0.9538 | 0.5945 | 0.3781 | 0.6366 | 0.9261 | 0.7994 | 0.9261 |
| LeNMF-DC,$\mu_2$=0 | 0.8312 | 0.7149 | 0.8312 | 0.9122 | 0.9144 | 0.9122 | 0.7494 | 0.5345 | 0.7642 | 0.9281 | 0.7948 | 0.9281 |
| LeNMF-DC | 0.9092 | 0.8282 | 0.9092 | 0.9577 | 0.9581 | 0.9577 | 0.8198 | 0.6079 | 0.8198 | 0.9408 | 0.8212 | 0.9408 |

regularization parameters $\lambda, \alpha$ and $\gamma$, the weight distribution parameter $\mu_1$ and $\mu_2$. In this paper, we empirically set the values of $\mu_1$ and $\mu_2$ to 0.01 and their sensitivity analysis was ignored due to space limitation. Therefore, the sensitivity analysis of the rest parameters $k, \lambda, \alpha$ and $\gamma$ were given in this subsection. We evaluated the sensitivity of a parameter while fixing other parameters.

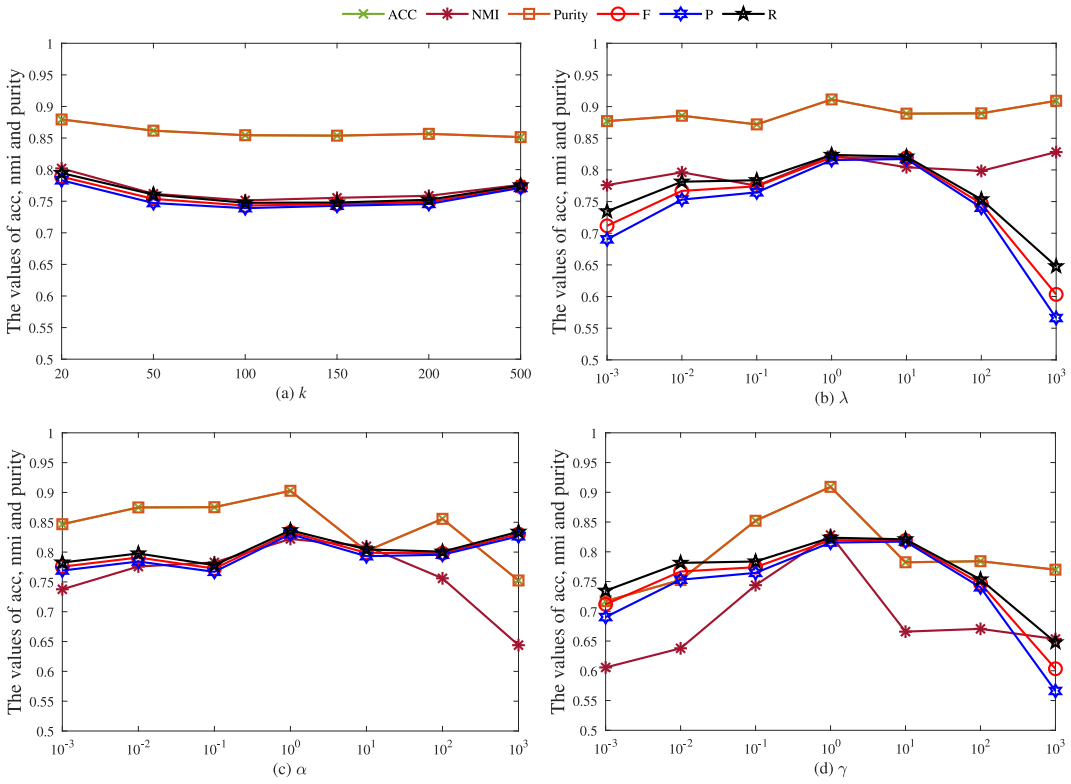


**Fig. 2.** 10 bpParameter sensitivity of LeNMF-DC on the 3Sources dataset: (a) varying k while setting $\lambda = 1$, $\alpha = 0.1$ and $\gamma = 0.1$, (b) varying $\lambda$ while setting $k = 20$, $\alpha = 0.1$ and $\gamma = 0.1$, (c) varying $\alpha$ while setting $k = 20$, $\lambda = 1$ and $\gamma = 0.1$, (d) varying $\gamma$ while setting $k = 20$, $\alpha = 1$ and $\gamma = 0.1$.

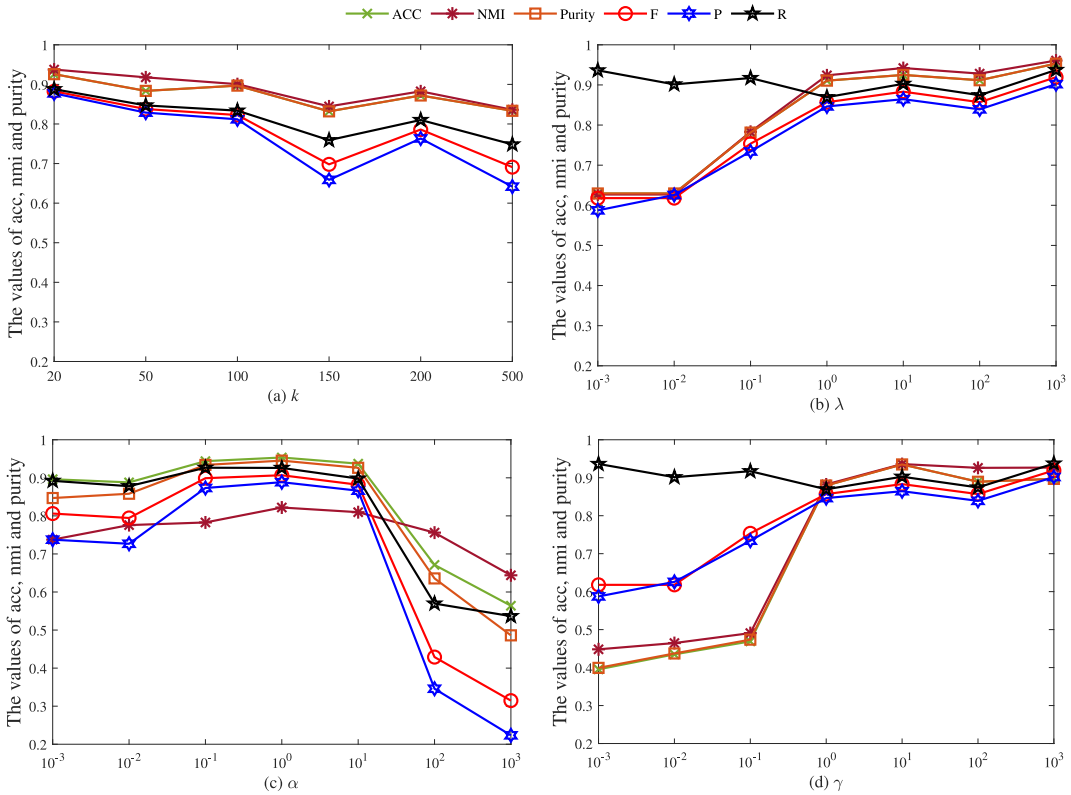


**Fig. 3.** Parameter sensitivity of LeNMF-DC on the 3Sources dataset: (a) varying $k$ while setting $\lambda = 1$, $\alpha = 0.1$ and $\gamma = 0.1$, (b) varying $\lambda$ while setting $k = 20$, $\alpha = 0.1$ and $\gamma = 0.1$, (c) varying $\alpha$ while setting $k = 20$, $\lambda = 1$ and $\gamma = 0.1$, (d) varying $\gamma$ while setting $k = 20$, $\alpha = 1$ and $\gamma = 0.1$.

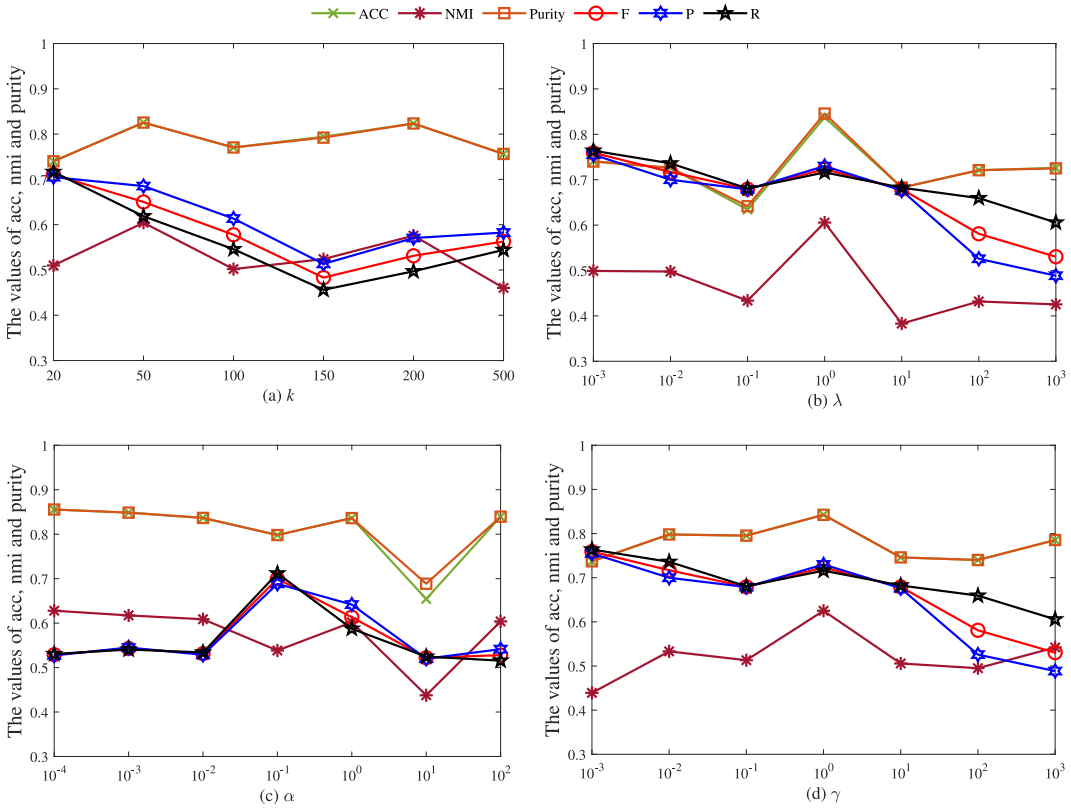


**Fig. 4.** Parameter sensitivity of LeNMF-DC on the BBC dataset: (a) varying $k$ while setting $\lambda = 1$, $\alpha = 1$ and $\gamma = 1$, (b) varying $\lambda$ while setting $k = 50$, $\alpha = 0.01$ and $\gamma = 1$, (c) varying $\alpha$ while setting $k = 50$, $\lambda = 1$ and $\gamma = 1$, (d) varying $\gamma$ while setting $k = 50$, $\alpha = 1$ and $\gamma = 0.01$.

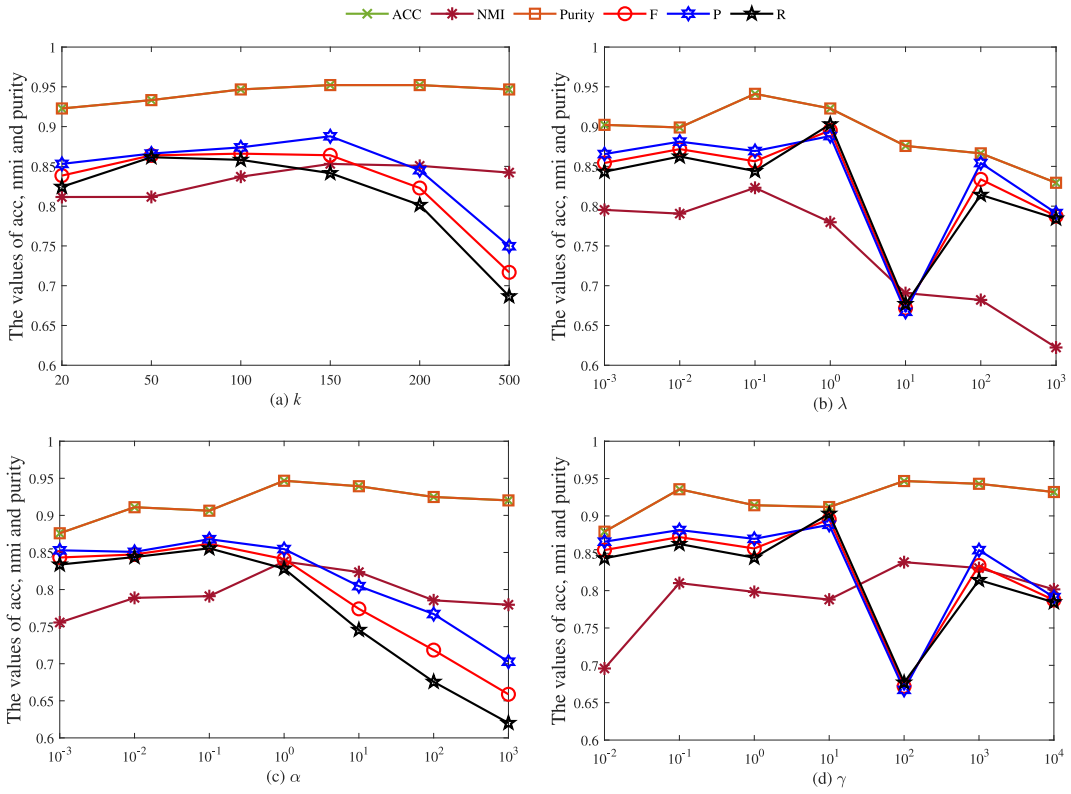


**Fig. 5.** Parameter sensitivity of DLENMF on the BBCSport dataset: (a) varying $k$ while setting $\lambda = 0.1$, $\alpha = 1$ and $\gamma = 1$, (b) varying $\lambda$ while setting $k = 50$, $\alpha = 1$ and $\gamma = 1$, (c) varying $\alpha$ while setting $k = 50$, $\lambda = 0.1$ and $\gamma = 1$, (d) varying $\gamma$ while setting $k = 50$, $\alpha = 0.1$ and $\gamma = 1$.

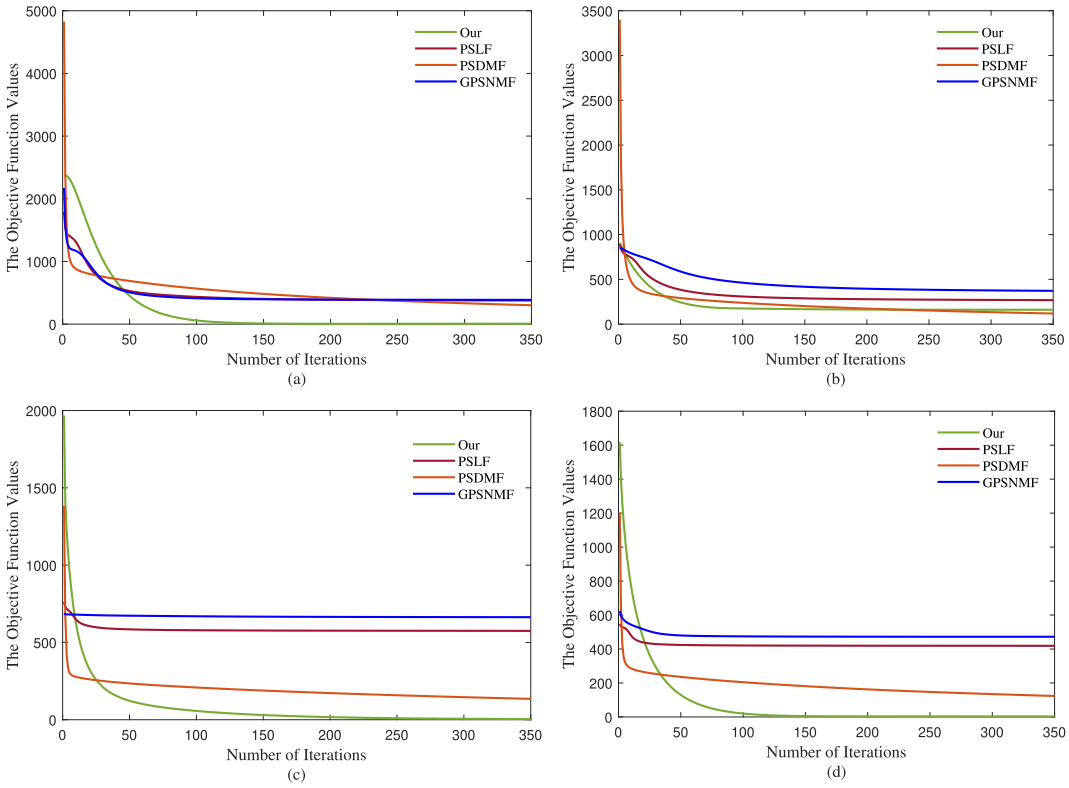


**Fig. 6.** Convergence curves of our proposed LeNMF-DC approach on different datasets: (a) HW2sources; (b) COIL20; (c) BBC; (d) BBCSport.

Specifically, the parameters $\lambda, \alpha$ and $\gamma$ were tuned in the range of {0.001, 0.01, 0.1, 1, 10, 100, 1000}, {0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000} and {0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000}, respectively. Figs. 2–5 show the clustering performances of our proposed approach on the Handwritten, COIL20, BBC and BBCSport datasets, respectively. It can be seen that the proposed LeNMF-DC model can achieve stable performances in a larger range of the parameters.

In addition, we conducted ablation experiments to evaluate the components of our proposed method. Specifically, the influence of a certain component on the model performance is tested by setting the regularization parameter of the corresponding component to zero. For example, when testing the influence of the data graph regularization item on the model performance, the parameter $\alpha$ is set to zero. Besides, we carried out ablation experiments with 10% of available labeled data, and then reported the ACC, NMI and Purity values in Table 5.

### *4.5. Convergence Study*

This subsection experimentally evaluated the convergence of the optimization scheme on four datasets. Fig. 6 shows the trend of the values of the objective function (13) as the number of iterations increases. It can be seen that our proposed LeNMF-DC approach can converge within 120, 70, 220 and 120 iterations on the four datasets, respectively. Therefore, it shows the efficiency of our developed optimization strategy.

## 5. Conclusions

In this work, we introduce a novel multi-view learning approach, namely LeNMF-DC, which fully utilizes the limited labels and the geometric structure of multi-view data. By introducing a label auxiliary matrix without additional parameters, we construct the label regularization term to preserve the distribution structure of the labels. Moreover, we can directly learn the labels of the unlabeled samples instead of the clustering algorithm, and thus reduce the performance loss during the clustering procedure. Besides, we take full advantage of the manifold structure of the multi-view data in feature and data spaces by constructing a dual graph regularizer. Extensive experiments on four benchmark datasets demonstrate the advantage of the proposed LeNMF-DC approach in multi-view clustering.

## CRediT authorship contribution statement

**Bin Li:** Writing - original draft, Data curation. **Zhenqiu Shu:** Conceptualization, Writing - review & editing, Supervision. **Yingbo Liu:** Writing - review & editing. **Cunli Mao:** Writing - review & editing. **Shengxiang Gao:** Writing - review & editing. **Zhengtao Yu:** Funding acquisition.

## Data availability

The authors are unable or have chosen not to specify which data has been used.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China [Grant No. 61603159, 62162033, U21B2027, 6226070477], Yunnan Provincial Major Science and Technology

Special Plan Projects [Grant No. 202002AD080001, 202103AA080015], Yunnan Foundation Research Projects [Grant No. 202201AT070154, 202101BE070001-056].

## References


[1] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on pattern analysis and machine intelligence 24 (7) (2002) 971–987.

[2] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1, Ieee, 2005, pp. 886–893.

[3] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2) (2004) 91–110.

[4] L. Zhang, L. Zhang, B. Du, J. You, D. Tao, Hyperspectral image unsupervised classification by robust manifold matrix factorization, Information Sciences 485 (2019) 154–169.

[5] D. Guillamet, J. Vitria, B. Schiele, Introducing a weighted non-negative matrix factorization for image classification, Pattern Recognition Letters 24 (14) (2003) 2447–2454.

[6] B. Li, P. l. Zhang, H. Tian, S. Mi, D. Liu, and G. Ren, "A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox," Expert Systems with Applications, vol. 38, no. 8, pp. 10000–10009, 2011.

[7] Y. Gao, G. Church, Improving molecular cancer class discovery through sparse non-negative matrix factorization, Bioinformatics 21 (21) (2005) 3970–3975.

[8] S. Leglaive, L. Girin, R. Horaud, Semi-supervised multichannel speech enhancement with variational autoencoders and non-negative matrix factorization, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 101–105.

[9] P.D. O'grady and B.A. Pearlmutter, "Discovering speech phones using convolutive non-negative matrix factorisation with a sparseness constraint," Neurocomputing, vol. 72, no. 1-3, pp. 88–101, 2008.

[10] Z. Shu, Y. Sun, J. Tang, C. You, Adaptive graph regularized deep semi-nonnegative matrix factorization for data representation, Neural Processing Letters (2022) 1–19.

[11] Z. Shu, Z. Weng, Z. Yu, C. You, Z. Liu, S. Tang, X. Wu, Correntropy-based dual graph regularized nonnegative matrix factorization with lp smoothness for data representation, Applied Intelligence 52 (7) (2022) 7653–7669.

[12] Z. Zhang, L. Liu, F. Shen, H.T. Shen, L. Shao, Binary multi-view clustering, IEEE transactions on pattern analysis and machine intelligence 41 (7) (2018) 1774–1782.

[13] S. Zeng, X. Wang, H. Cui, C. Zheng, D. Feng, A unified collaborative multikernel fuzzy clustering for multiview data, IEEE Transactions on Fuzzy Systems 26 (3) (2017) 1671–1687.

[14] J. Xu, J. Han, and F. Nie, "Discriminatively embedded k-means for multi-view clustering," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5356–5364, 2016.

[15] Q. Ye, P. Huang, Z. Zhang, Y. Zheng, L. Fu, and W. Yang, "Multiview learning with robust double-sided twin svm," IEEE Transactions on Cybernetics, 2021.

[16] Y. Wang, L. Wu, X. Lin, J. Gao, Multiview spectral clustering via structured low-rank matrix factorization, IEEE transactions on neural networks and learning systems 29 (10) (2018) 4833–4843.

[17] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in Proceedings of the 2013 SIAM international conference on data mining, pp. 252–260, SIAM, 2013.

[18] S. Wei, J. Wang, G. Yu, C. Domeniconi, and X. Zhang, "Multi-view multiple clusterings using deep matrix factorization," in Proceedings of the AAAI conference on artificial intelligence, vol. 34, pp. 6348–6355, 2020.

[19] Y. Tan, W. Ou, F. Long, P. Wang, Y. Xue, Multi-view clustering via co-regularized nonnegative matrix factorization with correlation constraint, in: 2016 7th International Conference on Cloud Computing and Big Data (CCBD), IEEE, 2016, pp. 1–6.

[20] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, Journal of machine learning research 7 (11) (2006).

[21] D. Cai, X. He, J. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE transactions on pattern analysis and machine intelligence 33 (8) (2010) 1548–1560.

[22] G. Du, L. Zhou, L. Wang, Q. Xiao, H. Chen, Multi-view clustering via nonnegative matrix factorization with l 21 norm, in: Fuzzy Systems and Data Mining V, IOS Press, 2019, pp. 363–370.

[23] X. Zhang, H. Gao, G. Li, J. Zhao, J. Huo, J. Yin, Y. Liu, L. Zheng, Multi-view clustering based on graph-regularized nonnegative matrix factorization for object recognition, Information Sciences 432 (2018) 463–478.

[24] N. Rai, S. Negi, S. Chaudhury, O. Deshmukh, Partial multi-view clustering using graph regularized nmf, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 2192–2197.

[25] L. Zong, X. Zhang, L. Zhao, H. Yu, Q. Zhao, Multi-view clustering via multi-manifold regularized non-negative matrix factorization, Neural Networks 88 (2017) 74–89.

[26] H. Cai, B. Liu, Y. Xiao, L. Lin, Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization, Knowledge-Based Systems 182 (2019) .

[27] H. Cai, B. Liu, Y. Xiao, L. Lin, Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization, Information Sciences 536 (2020) 171–184.

[28] N. Liang, Z. Yang, Z. Li, S. Xie, Label prediction based constrained non-negative matrix factorization for semi-supervised multi-view classification, Neurocomputing 512 (2022) 443–455.

[29] N. Liang, Z. Yang, Z. Li, S. Xie, Co-consensus semi-supervised multi-view learning with orthogonal non-negative matrix factorization, Information Processing & Management 59 (5) (2022) .

[30] N. Liang, Z. Yang, Z. Li, S. Xie, C.-Y. Su, Semi-supervised multi-view clustering with graph-regularized partially shared non-negative matrix factorization, Knowledge-Based Systems 190 (2020) .

[31] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.

[32] H. Liu, Z. Wu, X. Li, D. Cai, T.S. Huang, Constrained nonnegative matrix factorization for image representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (7) (2011) 1299–1311.

[33] F. Shang, L. Jiao, F. Wang, Graph dual regularization non-negative matrix factorization for co-clustering, Pattern Recognition 45 (6) (2012) 2237–2250.

[34] Z. Shu, C. Zhao, P. Huang, Local regularization concept factorization and its semi-supervised extension for image representation, Neurocomputing 158 (2015) 1–12.

[35] Z. Shu, Y. Zhang, P. Li, C. You, Z. Liu, H. Fan, X. Wu, Dual local learning regularized nonnegative matrix factorization and its semi-supervised extension for clustering, Neural Computing and Applications 33 (2021) 6213–6231.

[36] Y. Jia, M. Salzmann, T. Darrell, Factorized latent spaces with structured sparsity, Advances in neural information processing systems 23 (2010).

[37] P. Luo, J. Peng, Z. Guan, J. Fan, Dual regularized multi-view non-negative matrix factorization for clustering, Neurocomputing 294 (2018) 1–11.

[38] J. Liu, Y. Jiang, Z. Li, Z.-H. Zhou, H. Lu, Partially shared latent factor learning with multiview data, IEEE transactions on neural networks and learning systems 26 (6) (2014) 1233–1246.

[39] H. Huang, N. Liang, W. Yan, Z. Yang, Z. Lit, W. Sun, Partially shared semi-supervised deep matrix factorization with multi-view data, in: 2020 International Conference on Data Mining Workshops (ICDMW), IEEE, 2020, pp. 564–570.

[40] W. Wu, S. Kwong, J. Hou, Y. Jia, H.H.S. Ip, Simultaneous dimensionality reduction and classification via dual embedding regularized nonnegative matrix factorization, IEEE Transactions on Image Processing 28 (8) (2019) 3836–3847.

[41] J. Wen, Z. Zhang, Z. Zhang, L. Fei, M. Wang, Generalized incomplete multiview clustering with flexible locality structure diffusion, IEEE transactions on cybernetics 51 (1) (2020) 101–114.

[42] H. Wang, Y. Yang, B. Liu, H. Fujita, A study of graph-based system for multi-view clustering, Knowledge-Based Systems 163 (2019) 1009–1019.

[43] S.A. Nene, S.K. Nayar, H. Murase, et al., "Columbia object image library (coil-20)," 1996.

[44] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in Proceedings of the 23rd international conference on Machine learning, pp. 377–384, 2006.

[45] H.W. Kuhn, The hungarian method for the assignment problem, Naval research logistics quarterly 2 (1–2) (1955) 83–97.

[46] A. Strehl, J. Ghosh, Cluster ensembles—a knowledge reuse framework for combining multiple partitions, Journal of machine learning research vol. 3, no. Dec (2002) 583–617.

[47] C.D. Manning, An introduction to information retrieval, Cambridge University Press, 2009.

[48] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," in Thirty-first AAAI conference on artificial intelligence, 2017.

[49] S. Huang, Z. Kang, Z. Xu, Auto-weighted multi-view clustering via deep matrix decomposition, Pattern Recognition 97 (2020) .

[50] N. Liang, Z. Yang, Z. Li, W. Sun, S. Xie, Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints, Knowledge-Based Systems 194 (2020) .

[51] D. Wang, S. Han, Q. Wang, L. He, Y. Tian, and X. Gao, "Pseudo-label guided collective matrix factorization for multiview clustering," IEEE Transactions on Cybernetics, 2021.


**Bin Li** is currently pursuing the master's degree at the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. His research interests mainly focus on multi-view clustering, deep multi-view and deep clustering.

**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. His research interests include image processing, computer vision and machine learning.

**Yingbo Liu** is an Associate Professor at Yunnan University of Finance and Economics. He is the Director of the Cloud Computing Center of Big Data Research Institute of Yunnan Economy and Society. He was a postdoctoral researcher at Yunnan Academy of Scientific & Technical Information (2015–2018) and a Research Fellow at SCSE of NTU, Singapore (20182019). He obtained his Ph.D. degree from the University of Chinese Academy of Sciences (2015), M.S. degree (2011) in Computer Science and Application and B.S. degree (2008) in Automation from Kunming University of Science and Technology. His current research interests include big data, statistic analysis, and machine learning application.

**Cunli Mao** received the Ph.D. degree in computer application technology from the Kunming University of Science and Technology, Kunming, China, in 2014. He is currently a Professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. His current research interests include natural language processing and machine learning.

**Shengxiang Gao** received the Ph.D. degree in computer application technology from the Kunming University of Science and Technology, Kunming, China, in 2016. She is currently an associate professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. Her current research interests include natural language processing and information retrieval.

**Zhengtao Yu** received the Ph.D. degree in computer application technology from the Beijing Institute of Technology, Beijing, China, in 2005. He is currently a Professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. His current research interests include natural language processing, information retrieval, and machine learning.