BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 05/15 pp55–74
DOI: 10.26599/BDMA.2023.9020004
Volume 7, Number 1, March 2024

Discriminatively Constrained Semi-Supervised Multi-View Nonnegative Matrix Factorization with Graph Regularization

Guosheng Cui, Ye Li*, Jianzhong Li, and Jianping Fan

Abstract: Nonnegative Matrix Factorization (NMF) is one of the most popular feature learning technologies in the field of machine learning and pattern recognition. It has been widely used and studied in the multi-view clustering tasks because of its effectiveness. This study proposes a general semi-supervised multi-view nonnegative matrix factorization algorithm. This algorithm incorporates discriminative and geometric information on data to learn a better-fused representation, and adopts a feature normalizing strategy to align the different views. Two specific implementations of this algorithm are developed to validate the effectiveness of the proposed framework: Graph regularization based Discriminatively Constrained Multi-View Nonnegative Matrix Factorization (GDCMVNMF) and Extended Multi-View Constrained Nonnegative Matrix Factorization (ExMVCNMF). The intrinsic connection between these two specific implementations is discussed, and the optimization based on multiply update rules is presented. Experiments on six datasets show that the effectiveness of GDCMVNMF and ExMVCNMF outperforms several representative unsupervised and semi-supervised multi-view NMF approaches.

Key words: multi-view; semi-supervised clustering; discriminative information; geometric information; feature normalizing strategy

1 Introduction

Clustering is a very important unsupervised model in machine learning[1] and pattern recognition[2] because of its ability to capture the latent structure of the data. Generally, the data mostly used nowadays are not only of high dimension[3] but also derived from multiple sources[4] because of the rapid development of information technology. This is usually referred to as views or modalities in the literature[5]. These multiple views generally contain complementary and interaction information[6]; however, the primary issue is how to effectively fuse the information obtained from different views in the learning procedure[7]. In recent years, many algorithms have been proposed to handle multi-view clustering tasks; however, due to the potential of deep neural networks, various deep learning based multi-view clustering approaches, such as deep matrix factorization[8] and auto-encoder-based methods[9], have been presented.

Traditional machine learning based multi-view clustering algorithms can be grouped into three

• Guosheng Cui and Ye Li are with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen 518055, China. E-mail: gs.cui@siat.ac.cn; ye.li@siat.ac.cn.
• Jianping Fan is with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with University of Chinese Academy of Sciences, Beijing 100049, China. E-mail: jp.fan@siat.ac.cn.
• Jianzhong Li is with School of Computer Science and Control Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. E-mail: lijzh@siat.ac.cn.
* To whom correspondence should be addressed.
Manuscript received: 2022-11-24; revised: 2023-03-06; accepted: 2023-04-04

categories: (1) k-means-based methods[10]; (2) spectral graph based methods[11]; and (3) Nonnegative Matrix Factorization (NMF) based methods[12]. Multi-view clustering methods based on k-means clustering are usually implemented in the feature space, and do not use the geometric structure and discriminative information presented in the multi-view data[13]. Spectral graph based multi-view clustering methods have been extended from single-view spectral clustering to multi-view configuration; however, such methods necessitate building a similarity matrix for each class, resulting in high computational costs[14]. NMF[15], a powerful feature extracting technique, has been widely used in multi-view clustering tasks. In a Multi-View Nonnegative Matrix Factorization (MVNMF) framework, each point can be represented with an efficient low-dimensional feature vector.

The common objective function of traditional MVNMF can be expressed as follows:

$$\sum_{v} \|\boldsymbol{X}^v - \boldsymbol{S}^v (\boldsymbol{G}^v)^{\mathrm{T}}\|_{\mathrm{F}}^2 \tag{1}$$

where $\boldsymbol{X}^v$, $\boldsymbol{S}^v$, and $\boldsymbol{G}^v$ are the data matrix, the basis matrix, and the coefficient matrix of the v-th view, respectively, and $\|\cdot\|_{\mathrm{F}}^2$ is the Frobenius norm. "T" denotes the transposition. The key problem is to design an efficient fusing strategy to integrate multiple view information into one compact representation. MultiNMF[16] is the first to attempt to learn a consensus representation by minimizing the objective function of MVNMF with a centroid co-regularization term, which is defined as: $\|\boldsymbol{G}^v \boldsymbol{P}^v - \boldsymbol{G}_c\|$. Here, $\boldsymbol{G}_c$ is a common consensus matrix and $\boldsymbol{P}^v$ is a diagonal matrix. By introducing $\boldsymbol{P}^v$, the feature scale of different views is normalized to be similar; thus, multiple views can be aligned effectively. In conjunction with centroid co-regularization, many variants methods are developed consequently[17−19]. Wang et al.[20] developed a multi-view clustering method based on NMF and pairwise measurements (namely MPMNMF). In this model, a pair-wise co-regularization is introduced, which can be defined using Euclidean distance or kernel as $\sum_{1\leqslant v,\, w\leqslant V,\, v\neq w} \|\boldsymbol{G}^v - \boldsymbol{G}^w\|_{\mathrm{F}}^2$ (MPMNMF_1) or $\sum_{1\leqslant v,\, w\leqslant V,\, v\neq w} \|\boldsymbol{G}^v (\boldsymbol{G}^v)^{\mathrm{T}} - \boldsymbol{G}^w (\boldsymbol{G}^w)^{\mathrm{T}}\|_{\mathrm{F}}^2$ (MPMNMF_2). The features from different views are pushed close to each other using pairwise co-regularization, and alignment is acquired. In Ref. [21], a method named Uniform Distribution NMF (UDNMF) based on nonnegative matrix tri-factorization[22] was proposed. This method factorizes each view into three matrices, i.e., basis matrix, shared embedding matrix, and coefficient matrix, and the column summation of the product of basis matrix and shared embedding matrix is constrained to be 1. Centroid co-regularization $\|\boldsymbol{G}^v - \boldsymbol{G}_c\|$ is used to align the multiple views. The above mentioned methods all attempt to align the multiple views to fuse information. However, Wang et al.[23] attempted a different approach and presented a Locality-Preserved Diverse NMF (LP-DiNMF) method. In this study, they introduced a diverse term $\sum_{v\neq w} \|\boldsymbol{G}^v \odot \boldsymbol{G}^w\|_1$. By minimizing it, the heterogeneity of the different views is encouraged, and more comprehensive information is expected. In Ref. [24], a multi-view clustering method named robust Neighboring constraint NMF (rNNMF) was proposed. This method handles the noise and outliers among the views by defining a reconstruction term and a neighbor-structure-preserving term using the $\ell_{2,1}$-norm. Recently, some "deep" models have also been proposed to tackle the multi-view clustering problem[25]. Inspired by deep semi-nonnegative matrix factorization[26], Zhao et al.[27] proposed a multi-view deep semi-nonnegative matrix factorization. This method uses adaptive weights for different views. In Ref. [28], an auto-weighted deep matrix factorization was presented to tackle multi-view clustering task. A shared coefficient matrix was introduced in their study to fuse the information of the multiple views, and an auto-weighted strategy was adopted to balance different views. Also, in terms of diversity[23], Luong et al.[29] developed a method named Orthogonal Diverse Deep NMF (ODD-NMF). In this method, a diverse term $\sum_{v\neq w} \mathrm{tr}\,(\boldsymbol{G}_m^v (\boldsymbol{G}_m^w)^{\mathrm{T}})$ (m denotes the m-th layer) was designed to boost the diverse information of data.

Recently, some semi-supervised multi-view clustering methods have been proposed. In Ref. [30], Jiang et al. presented a unified latent factor learning method, in which a regression term is introduced to fit the partially labeled data points. However, Liu et al.[31] developed a partially shared NMF, which can separately model common and private information of data. In Refs. [32], Liang et al. expanded the work in Ref. [31] by incorporating a graph regularization term. In Refs. [31] and [32], the authors tried to use both distinct and shared information to improve the clustering performance. However, deciding the dimensions of the distinct and shared parts of the coefficient matrix for these two methods is difficult. Many approaches based on constrained NMF, which is a semi-supervised NMF model designed for single-view data, have also been developed. Wang et al.[33] proposed an Adaptive Multi-View semi-supervised NMF (AMVNMF) based on Constrained NMF (CNMF), a semi-supervised NMF method designed for single-view data. This method adopts $\ell_{2,1}$-norm to measure the reconstruction loss to make it more robust to the outliers. A centroid co-regularization is used to align the multiple views. Cai et al.[34] developed a semi-supervised MVNMF approach based on CNMF with sparseness constraint (namely MVCNMF), which factorizes each view in the CNMF framework and aligns multiple views using Euclidean distance based pairwise co-regularization, similar to the approach presented in Ref. [20]. The $\ell_{2,1}$-norm regularization is imposed on the auxiliary matrix in each view to select features. In Ref. [35], a similar method named Multi-View Orthonormality-CNMF (namely MVOCNMF) was proposed. MVOCNMF differs from MVCNMF in that it replaces the $\ell_{2,1}$-norm regularization with an orthonormality constraint, which is imposed on the auxiliary matrix in each view to normalize the feature scale. Wang et al.[36] developed a semi-supervised multi-view clustering model based on anchor graph, in which the anchors are constructed using label information. In Ref. [37], Nie et al. presented an auto-weighted multi-view learning method, that can adaptively model the intrinsic structure of data. Based on the work of Nie et al.[37], Liang et al.[38] proposed a label propagation based NMF. In this model an intrinsic structure of data was constructed as in Ref. [37], which helps the label propagation use the limited labeled data points. Additionally, this method adopts the $\ell_{2,1}$-norm to measure the reconstruction loss. In Ref. [39], Zhao et al. developed a deep semi-supervised NMF model. In this model, two graphs are constructed to discover the discriminant information of data, where the affinity graph ensures intra-class compactness and the penalty graph ensures inter-class distinctness. Recently, Chen et al.[40] presented a deep semi-supervised multi-view clustering method based on the autoencoder framework with pairwise constraint. The pairwise constraint was used to encode the partial label information, and a clustering layer was constructed to produce clustering results.

From the above review, it can be said that for semi-supervised multi-view clustering methods, the key point is to discover the discriminant information of data. To guarantee the learned feature more discriminative, the essential idea is to ensure the intra-class compactness and inter-class distinctiveness. Many semi-supervised MVNMFs have adopted CNMF to use the partial label information of data. However, they can only guarantee the intra-class compactness of data because of the intrinsic property of CNMF. From the above references, many methods have also adopted $\ell_{2,1}$-norm to measure the reconstruction loss, claiming that it is more robust to outliers. However, the $\ell_{2,1}$-norm and Frobenius-norm (i.e., $\ell_{2,2}$-norm) have not been compared under the same model configuration. In addressing the above issues, this study presents a novel discriminatively constrained semi-supervised MVNMF with a feature alignment strategy. Two specific implementations of this model are introduced, i.e., Graph regularization based Discriminatively Constrained Multi-View Nonnegative Matrix Factorization (GDCMVNMF) and Extended Multi-View CNMF (ExMVCNMF). In GDCMVNMF, a more general $\ell_{2,p}$-norm is used to measure the reconstruction loss of data, which helps us compare the influence of $\ell_{2,p}$-norm when p is set to different values, such as 0.5, 1, and 2. Additionally, the innner connection of GDCMVNMF and ExMVCNMF is revealed. ExMVCNMF retains the CNMF property of mapping the data points within the same class into the same feature vector to ensure the intra-class compactness of data. The influence of intra-class compactness on the model can also be revealed by comparing GDCMVNMF and ExMVCNMF. The contributions of this study are summarized as follows:

- A general discriminatively constrained semi-supervised MVNMF with a feature alignment strategy is proposed in this study. Two specific implementations of this model, i.e., GDCMVNMF and ExMVCNMF, are introduced, and their corresponding optimizing strategies are also presented.
- The inner connection of GDCMVNMF and ExMVCNMF is revealed, so the influence of the intrinsic property of CNMF, i.e., the intra-class compactness of data, on the mining of the discriminative information of data, can be discussed.
- A more general $\ell_{2,p}$-norm is used in GDCMVNMF to measure data reconstruction loss.

This can help to explore the effect of p value on the model under the same configuration.

- The effectiveness of the proposed methods is verified by comparing with several recently proposed representative unsupervised and semi-supervised methods on six datasets.

The rest of the paper is organized as follows: Section 2 briefly reviews some related works to the proposed methods. Section 3 presents the proposed Semi-Supervised Multi-View Nonnegative Matrix Factorization (S^{2}MVNMF) along with its first implementation and detailed optimization algorithm. Section 4 describes another implementation of S^{2}MVNMF along with its corresponding optimization procedure and computation complex analysis. Section 5 presents extensive experiments conducted to verify the effectiveness of the proposed methods. Finally, Section 6 concludes the study.

In Table 1, some relative notations used in the study are summarized for clarity.

Table 1 Important notations in the paper.

Notation	Description
n	Number of samples
C	Number of classes
l	Number of labeled samples
V	Number of views
m^v	Dimension of features for original view v
d^v	Dimension of low-dimensional feature for view v
$\boldsymbol{e}$	All-one column vector
$x_i^v \in \mathbf{R}_+^{m^v}$	Sample of original view v
$g_i^v \in \mathbf{R}_+^{d^v}$	i-th column vector of $\boldsymbol{G}^v$
$\boldsymbol{I}$	Identity matrix
$\boldsymbol{X}^v$	Data matrix for view v
$\boldsymbol{S}^v$	Basis matrix for view v
$\boldsymbol{G}^v$	Coefficient matrix for view v
$\boldsymbol{G}_c$	Consistent coefficient matrix to align the coefficient matrices learned from different views
$\boldsymbol{L}^v$	Graph laplacian matrix for view v
$\boldsymbol{W}^v$	Similarity matrix for view v
$\boldsymbol{D}^v$	Degree matrix for view v
$\boldsymbol{A}_{\text{lc}}$	Label constraint matrix defined in Ref. [41]
$\boldsymbol{Z}$	Auxiliary matrix defined in Ref. [41]
$\boldsymbol{Z}^v$	Auxiliary matrix for view v
$\boldsymbol{Z}_c$	Consistent auxiliary matrix to align the auxiliary matrices of different views
$\lVert\cdot\rVert_{\text{F}}$	Frobenius norm of a matrix
$\lVert\cdot\rVert_{2,1}$	$\ell_{2,1}$ norm of a matrix
tr $(\cdot)$	Trace operator of a matrix
$\odot$	Element-wise product operator

2 Related Work

2.1 CNMF

CNMF[41] is a semi-supervised NMF method that uses label information as additional constraints to improve the discriminating power of the resulting matrix decomposition. The specific formulation of CNMF is as follows:

$$O_{\text{CNMF}} = \lVert \boldsymbol{X} - \boldsymbol{S}(\boldsymbol{A}_{\text{lc}}\boldsymbol{Z})^{\text{T}} \rVert_{\text{F}}^2 \tag{2}$$

where $\boldsymbol{X} = [x_1, x_2, \ldots, x_i, \ldots, x_n]$ is a data matrix, $\boldsymbol{S}$ is a basis matrix, $\boldsymbol{Z}$ is an auxiliary matrix, and $\boldsymbol{A}_{\text{lc}}$ is a label constraint matrix, which is as follows:

$$\boldsymbol{A}_{\text{lc}} = \begin{pmatrix} \boldsymbol{H}_{l\times C} & 0 \\ 0 & \boldsymbol{I}_{n-l} \end{pmatrix} \tag{3}$$

where $\boldsymbol{H}$ is label matrix of the labeled data points and $\boldsymbol{I}_{n-l}$ is an identity matrix of dimension $n-l$. Here, l is the number of labeled data points, n is the number of total data points, and C is the number of classes. When $l=5$, x_1 and x_2 belong to Class I, x_3 and x_4 belong to Class II, and x_5 belongs to Class III. Matrix $\boldsymbol{A}_{\text{lc}}$ can be represented as follows:

$$\boldsymbol{A}_{\text{lc}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \boldsymbol{I}_{n-5} \end{pmatrix} \tag{4}$$

The coefficient matrix $\boldsymbol{G}$ in the original NMF is represented as $\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}$ in CNMF. Considering the definition of the label constraint matrix $\boldsymbol{A}_{\text{lc}}$, if x_i and x_j are in the same class, then $g_i = g_j$, where g_i and g_j are low dimensional representations of the i-th and j-th data points, respectively. In sum, for the labeled data points, the data points with the same label are mapped to the same low-dimensional vector.

2.2 Robust structured nonnegative matrix factorization

Robust Structured Nonnegative Matrix Factorization (RSNMF)[42] is another semi-supervised NMF method that attempts to learn representation using a block-diagonal structure. The objective function of RSNMF is as follows:

$$O_{\text{RSNMF}} = \lVert \boldsymbol{X} - \boldsymbol{S}\boldsymbol{G}^{\text{T}} \rVert_{2,p}^p + \mu \lVert \boldsymbol{I}_{\text{block}} \odot \boldsymbol{G} \rVert_{\text{F}}^2,$$
$$\text{s.t., } \boldsymbol{S}, \boldsymbol{G} \geqslant 0,\ \sum_{i=1}^{d} \boldsymbol{G}_{hi} = 1, \forall h \tag{5}$$

where μ is a balancing parameter, d is the dimension of latent feature, and $\odot$ is an element-wise product operator. Here, $\|\cdot\|_{2,p}$ denotes the $\ell_{2,p}$-norm of a matrix, for a matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_{2,p}=\left(\sum_j\left(\sum_i \boldsymbol{A}_{ij}^2\right)^{p/2}\right)^{1/p}$. When $0<p<1$, the $\ell_{2,p}$-norm can produce more robust solutions compared to the $\ell_{2,1}$-norm; when $p=1$, the $\ell_{2,p}$-norm reduces to $\ell_{2,1}$-norm; when $p=2$, the $\ell_{2,p}$-norm is equivalent to the Frobenius norm. $\boldsymbol{I}_{\text{block}}=[\bar{\boldsymbol{I}};\,\hat{\boldsymbol{0}}]\in\mathbf{R}^{n\times m}$ is an indicator matrix, and $\hat{\boldsymbol{0}}\in\mathbf{R}^{(n-l)\times d}$ is the zero matrix corresponding to the unlabeled samples. Here, $\bar{\boldsymbol{I}}\in\mathbf{R}^{l\times d}$ is defined as labeled samples and is expressed as follows:

$$\bar{\boldsymbol{I}}=\begin{pmatrix}\bar{\boldsymbol{0}}_1 & \boldsymbol{1} & \ldots & \boldsymbol{1}\\ \boldsymbol{1} & \bar{\boldsymbol{0}}_2 & \ldots & \boldsymbol{1}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{1} & \boldsymbol{1} & \ldots & \bar{\boldsymbol{0}}_C\end{pmatrix} \tag{6}$$

where $\bar{\boldsymbol{0}}_c\in\mathbf{R}^{n_c\times m_s}(c=1,2,\ldots,C)$ is a zero matrix for the c-th class. Here, n_c is the number of labeled samples in the c-th class, and m_s is the dimension of each subspace. By minimizing the second term in Eq. (5), the coefficient matrix learned by this method, especially for the labeled samples, is restricted to block diagonal structure, and intra-class distinction is enlarged.

2.3 Representative semi-supervised MVNMF based on CNMF

2.3.1 AMVNMF

AMVNMF[33] extends CNMF to handle multi-view clustering task for the first time. The objective function of AMVNMF is as follows:

$$\begin{aligned}O_{\text{AMVNMF}}=\sum_{v=1}^{V}(&\|\boldsymbol{X}^v-\boldsymbol{S}^v(\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v)^{\text{T}}\|_{2,1}+\\ &(\alpha^v)^{\kappa}\,\|\,\boldsymbol{Z}^v\boldsymbol{P}^v-\boldsymbol{Z}_c\|_{\text{F}}^2),\\ &\text{s.t.},\ \sum_{v=1}^{V}\alpha^v=1,\alpha^v\geqslant 0\end{aligned} \tag{7}$$

where $\|\cdot\|_{2,1}$ denotes $\ell_{2,1}$-norm of a matrix. For matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_{2,1}=\sum_j\sqrt{\sum_i \boldsymbol{A}_{ij}^2}$. Here, $(\alpha^v)^{\kappa}$ is a weighting parameter, with κ as a hyper-parameter, which can be adaptively changed during optimization. The diagonal matrix $\boldsymbol{P}^v$ is introduced to normalize $\boldsymbol{S}^v$ using $\boldsymbol{S}^v(\boldsymbol{P}^v)^{-1}$. In this manner, the scale of $\boldsymbol{Z}^v$ is constrained to be within the same range[16]. Here, $\boldsymbol{Z}_c$ is a consistent auxiliary matrix shared across multiple views, and $\boldsymbol{P}^v=\text{Diag}\left(\sum_{i=1}^{m^v}\boldsymbol{S}_{i,1}^v,\sum_{i=1}^{m^v}\boldsymbol{S}_{i,2}^v,\ \ldots,\ \sum_{i=1}^{m^v}\boldsymbol{S}_{i,d^v}^v\right)$.

The reconstruction term of AMVNMF is measured using the $\ell_{2,1}$-norm to enhance the robustness of this method against noises and outliers.

2.3.2 MVCNMF

MVCNMF[34] is a recently proposed semi-supervised NMF method whose objective function is as follows:

$$\begin{aligned}O_{\text{MVCNMF}}=\sum_{v=1}^{V}\Big\{&\theta_v\,\|\boldsymbol{X}^v-\boldsymbol{S}^v(\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v)^{\text{T}}\|_{\text{F}}^2+\\ &\sum_{s=1}^{V}\theta_{vs}\,\|\boldsymbol{Z}^v-\boldsymbol{Z}^s\|_{\text{F}}^2+\lambda\|\boldsymbol{Z}^v\|_{2,1}\Big\},\\ &\text{s.t.},\ \boldsymbol{S}^v,\boldsymbol{Z}^v\geqslant 0\end{aligned} \tag{8}$$

where θ_v, θ_{vs}, and λ are the balancing parameters for the corresponding terms. Although the subscripts of θ_v and θ_{vs} vary with different views, they are shared across multiple views according to Ref. [34], i.e., $\theta_1=\theta_2=\ldots=\theta_V$, and $\theta_{11}=\theta_{12}=\ldots=\theta_{VV}$. In contrast to AMVNMF, which introduces a consistent auxiliary matrix $\boldsymbol{Z}_c$ to align different views, MVCNMF attempts to align different views by enforcing them to be similar to each other, as can be seen from the second term in Eq (8). The $\ell_{2,1}$-norm is also imposed on $\boldsymbol{Z}^v$ to select efficient and robust features.

2.3.3 MVOCNMF

MVOCNMF[35] is another recently proposed semi-supervised NMF method. The objective function of MVOCNMF is as follows:

$$\begin{aligned}O_{\text{MVOCNMF}}=\sum_{v=1}^{V}\Big\{&\theta_v\,\|\boldsymbol{X}^v-\boldsymbol{S}^v(\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v)^{\text{T}}\|_{\text{F}}^2+\\ &\sum_{s=1}^{V}\frac{1}{2}\theta_{vs}\,\|\boldsymbol{Z}^v-\boldsymbol{Z}^s\|_{\text{F}}^2+\lambda\,\|\boldsymbol{B}\odot(\boldsymbol{Z}^v(\boldsymbol{Z}^v)^{\text{T}})-\boldsymbol{I}\|_{\text{F}}^2\Big\},\\ &\text{s.t.},\ \boldsymbol{S}^v,\boldsymbol{Z}^v\geqslant 0\end{aligned} \tag{9}$$

where $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{B}$ is a selected constraint matrix which is defined as

$$\boldsymbol{B}_{ij}=\begin{cases}1, & i=j \text{ or } 1\leqslant i,j\leqslant C;\\ 0, & \text{others}\end{cases} \tag{10}$$

The last term in Eq (9) is applied to impose the normalization constraint on z_i^v (the i-th row of z^v), i.e., $z_i^v\,(z_i^v)^{\text{T}}=1$, $1\leqslant i\leqslant n-l+C$. This constraint aims to restrict the features of different views to the same scale.

The abovementioned recently proposed semi-supervised multi-view NMF methods have the following drawbacks. First, they all focus on how to scale the features of different views and combine them.

However, no attempt is made to extract the intrinsic geometric structure information from the data, including labeled samples and unlabeled samples. Second, because they are all based on the original CNMF method due to the definition of label constraint matrix A_{lc}, the samples from the same class are mapped into the same vector (see Fig. 1). However, the distinction between these different class vectors is not ensured, which harms the identifiability of the features learned by the model. To address these issues, the following sections propose that the geometric structure and discriminative information on the data should be simultaneously considered to improve the quality of the features. Additionally, a feature normalizing strategy different from the abovementioned methods is introduced in this study to effectively align the features of multiple views.

3 Proposed S²MVNMF

3.1 GDCMVNMF

For a semi-supervised algorithm, it is important to simultaneously consider the label information and geometrical structure information on the data. This way, information from labeled and unlabeled samples is explored to enhance the learning algorithm's performance. Based on this idea, a unified framework for S²MVNMF can be expressed as follows:

$$O_{\mathrm{S^2MVNMF}} = \sum_{v=1}^{V}\Big\{D\left(\boldsymbol{X}^v \| \boldsymbol{S}^v (\boldsymbol{G}^v)^{\mathrm{T}}\right) + \Omega_l\left(\boldsymbol{G}_l^v\right) + \Omega_g\left(\boldsymbol{G}^v\right) + \Omega_a\left(\boldsymbol{G}^v\right)\Big\},\quad \text{s.t., } \boldsymbol{S}^v, \boldsymbol{G}^v \geqslant 0 \tag{11}$$

where $D\left(\boldsymbol{X}^v \| \boldsymbol{S}^v (\boldsymbol{G}^v)^{\mathrm{T}}\right)$ is the measurement to quantify

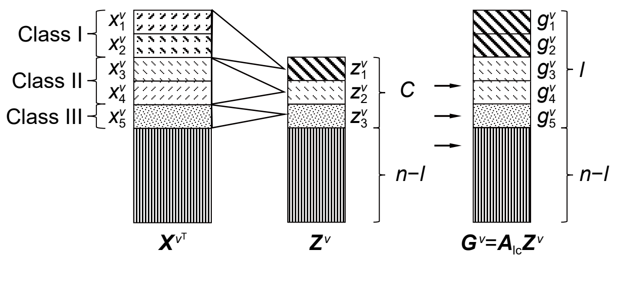


Fig. 1 Correspondence of X^v, Z^v and $A_{\mathrm{lc}}Z^v$ for the labeled samples in Refs. [33–35]. Samples with the same labels, such as x_1^v and x_2^v, are mapped into the same vectors, i.e., $g_1^v = g_2^v = z_1^v$. Similarly, for x_3^v and x_4^v, $g_3^v = g_4^v = z_2^v$; for x_5^v, $g_5^v = z_3^v$. Additionally, the distinction of these vectors is not ensured in the learning procedure (take the example in Section 2.1).

the distance between $\boldsymbol{X}^v$ and $\boldsymbol{S}^v (\boldsymbol{G}^v)^{\mathrm{T}}$. $\Omega_l\left(\boldsymbol{G}_l^v\right)$ is the term defined to explore the label information provided by the data, $\boldsymbol{G}_l^v$ is the feature corresponding to the labeled samples, $\Omega_g(\cdot)$ is the regularizer to use the geometric information on the data, and $\Omega_a(\cdot)$ is a term to align the features from multiple views.

In the above framework, the $\ell_{2,p}$-norm based reconstruction cost is adopted for each view to enhancing the robustness of the model to noises or outliers. Thus, the first term in Eq. (1) is as follows:

$$D\left(\boldsymbol{X}^v \| \boldsymbol{S}^v (\boldsymbol{G}^v)^{\mathrm{T}}\right) = \|\boldsymbol{X}^v - \boldsymbol{S}^v (\boldsymbol{G}^v)^{\mathrm{T}}\|_{2,p}^{p} \tag{12}$$

The block-diagonal structure constraint is imposed on the coefficient matrix, such as RSNMF, to use the discriminative information provided by the labeled samples. Specific definition of $\Omega_l\left(\boldsymbol{G}_l^v\right)$ is as follows:

$$\Omega_l\left(\boldsymbol{G}_l^v\right) = \|\boldsymbol{I}_{\mathrm{block}} \odot \boldsymbol{G}^v\|_{\mathrm{F}}^2 \tag{13}$$

For the third term $\Omega_g(\boldsymbol{G}^v)$, a graph regularizer is constructed for convenience to encode the geometric information on the entire data, including labeled and unlabeled samples, and is defined as follows:

$$\Omega_g(\boldsymbol{G}^v) = \sum_{i=1}^{n}\sum_{j=1}^{n} \|g_i^v - g_j^v\|_2^2\, \boldsymbol{W}_{ij}^v = \mathrm{tr}\left((\boldsymbol{G}^v)^{\mathrm{T}} \boldsymbol{L}^v \boldsymbol{G}^v\right) \tag{14}$$

where $\boldsymbol{L}^v = \boldsymbol{D}^v - \boldsymbol{W}^v$, $\boldsymbol{D}_{ii}^v = \sum_j \boldsymbol{W}_{ij}^v$ (or $\boldsymbol{D}_{ii}^v = \sum_j \boldsymbol{W}_{ji}^v$). $\boldsymbol{W}^v$ is defined as follows:

$$\boldsymbol{W}_{ij}^v = \begin{cases} \mathrm{e}^{-\|g_i^v - g_j^v\|_2^2/2\delta^2}, & \text{if } g_i^v \in N_k(g_j^v) \text{ or } g_j^v \in N_k(g_i^v); \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

where δ is a predefined parameter. In this study, δ is fixed as 1 for simplicity, and $N_k(g_i^v)$ consists of k Nearest Neighbors (k-NN) of g_i^v. For the final term $\Omega_a(\boldsymbol{G}^v)$, a shared common consensus matrix $\boldsymbol{G}_c$ is introduced to align the multiple views. Then, $\Omega_a(\boldsymbol{G}^v)$ is given as follows:

$$\Omega_a(\boldsymbol{G}^v) = \|\boldsymbol{G}^v - \boldsymbol{G}_c\|_{\mathrm{F}}^2 \tag{16}$$

Thus, summarizing the terms mentioned above, the final objective function of the proposed method can be described as follows:

$$O_{\mathrm{GDCMVNMF}} = \sum_{v=1}^{V}\Big\{\|\boldsymbol{X}^v - \boldsymbol{S}^v \boldsymbol{G}^{v\mathrm{T}}\|_{2,p}^{p} + \alpha\, \|\boldsymbol{I}_{\mathrm{block}} \odot \boldsymbol{G}^v\|_{\mathrm{F}}^2 + \beta \cdot \mathrm{tr}\left((\boldsymbol{G}^v)^{\mathrm{T}} \boldsymbol{L}^v \boldsymbol{G}^v\right) + \gamma\, \|\boldsymbol{G}^v - \boldsymbol{G}_c\|_{\mathrm{F}}^2\Big\},\quad \text{s.t., } \boldsymbol{S}^v, \boldsymbol{G}^v, \boldsymbol{G}_c \geqslant 0 \tag{17}$$

where α, β, and γ are the balancing parameters for corresponding terms.

In contrast to the recently proposed semi-supervised multi-view NMF methods mentioned above, inter-class discriminative information among labeled samples and geometric structure information are simultaneously considered. Therefore, the proposed algorithm in this study is termed GDCMVNMF.

One critical challenge in handling multi-view tasks is to align multiple features to effectively fuse information on different views. Therefore, to make this possible, the scales of multiple features must be restricted to be comparable. Based on this, the column vectors of $\boldsymbol{S}^v$ are constrained as $\|\boldsymbol{S}^v_{\cdot j}\|_2 = 1$; however, directly optimizing the above objective function extremely complicates the optimization problem significantly. An alternative scheme to the direct strategy is to compensate the norms of the basis matrix into the coefficient matrix. The objective function of GDCMVNMF in Eq. (17) can then be rewritten as follows:

$$O_{\text{GDCMVNMF}} = \sum_{v=1}^{V} \Big\{ \|\boldsymbol{X}^v - \boldsymbol{S}^v(\boldsymbol{G}^v)^{\mathrm{T}}\|_{2,p}^{p} + \alpha\, \|\boldsymbol{I}_{\text{block}} \odot \boldsymbol{G}^v \boldsymbol{P}^v\|_{\mathrm{F}}^2 + \beta \cdot \operatorname{tr}\left((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{G}^v\boldsymbol{P}^v\right) + \gamma\, \|\boldsymbol{G}^v\boldsymbol{P}^v - \boldsymbol{G}_c\|_{\mathrm{F}}^2 \Big\}$$
$$\text{s.t., } \boldsymbol{S}^v, \boldsymbol{G}^v, \boldsymbol{G}_c \geqslant 0 \tag{18}$$

where $\boldsymbol{P}^v$ is defined as

$$\boldsymbol{P}^v = \operatorname{Diag}\left(\sqrt{\sum_{i=1}^{m^v}\left(\boldsymbol{S}^v_{i,1}\right)^2}, \sqrt{\sum_{i=1}^{m^v}\left(\boldsymbol{S}^v_{i,2}\right)^2}, \ldots, \sqrt{\sum_{i=1}^{m^v}\left(\boldsymbol{S}^v_{i,d^v}\right)^2}\right) \tag{19}$$

In the following sections, the optimization algorithms of GDCMVNMF are introduced in detail. For a specific view v, the objective function defined in Eq. (18) is non-convex for both $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$; however, it is convex for $\boldsymbol{S}^v$ or $\boldsymbol{G}^v$ when either is fixed. Section 3.2 presents an iterative multiplicative updating procedure to solve the above problem.

3.2 Optimization algorithm of GDCMVNMF

To minimize Eq. (18), the minimizing problem is divided into several manageable subproblems.

For the v-th view, the other views are not involved in the optimization of $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$. Letting $\boldsymbol{Q}^v = \boldsymbol{X}^v - \boldsymbol{S}^v(\boldsymbol{G}^v)^{\mathrm{T}}$, the minimizing problem in Eq. (18) can be written as follows:

$$\min_{\boldsymbol{S}^v, \boldsymbol{G}^v, \boldsymbol{G}_c \geqslant 0} \|\boldsymbol{Q}^v\|_{2,p}^p + \alpha\, \|\boldsymbol{I}_{\text{block}} \odot \boldsymbol{G}^v\boldsymbol{P}^v\|_{\mathrm{F}}^2 + \beta \cdot \operatorname{tr}\left((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{G}^v\boldsymbol{P}^v\right) + \gamma\, \|\boldsymbol{G}^v\boldsymbol{P}^v - \boldsymbol{G}_c\|_{\mathrm{F}}^2 \tag{20}$$

3.2.1 Fixing G_c and G^v, updating S^v

When $\boldsymbol{G}_c$ and $\boldsymbol{G}^v$ are fixed, Eq. (20) can be rewritten as follows:

$$\min_{\boldsymbol{S}^v \geqslant 0} \|\boldsymbol{Q}^v\|_{2,p}^p + \alpha \cdot \operatorname{tr}\left((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{I}_{\text{block}} \odot \boldsymbol{G}^v)^{\mathrm{T}}(\boldsymbol{I}_{\text{block}} \odot \boldsymbol{G}^v)\boldsymbol{P}^v\right) + \beta \cdot \operatorname{tr}\left((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{G}^v\boldsymbol{P}^v\right) + \gamma \cdot \operatorname{tr}\left((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{G}^v\boldsymbol{P}^v - 2\boldsymbol{G}_c^{\mathrm{T}}\boldsymbol{G}^v\boldsymbol{P}^v\right) \tag{21}$$

Let

$$\boldsymbol{U}_1^v = \left[(\boldsymbol{I}_{\text{block}} \odot \boldsymbol{G}^v)^{\mathrm{T}}(\boldsymbol{I}_{\text{block}} \odot \boldsymbol{G}^v)\right] \odot \boldsymbol{I},$$
$$\boldsymbol{U}_2^v = \left[(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{G}^v\right] \odot \boldsymbol{I} = \boldsymbol{U}_2^{v+} - \boldsymbol{U}_2^{v-},$$
$$\boldsymbol{U}_3^v = \left[(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{G}^v\right] \odot \boldsymbol{I} \tag{22}$$

where $\boldsymbol{U}_2^{v+} = [(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{D}^v\boldsymbol{G}^v] \odot \boldsymbol{I}$ and $\boldsymbol{U}_2^{v-} = [(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{W}^v\boldsymbol{G}^v] \odot \boldsymbol{I}$.

Then, Formula (21) with $\boldsymbol{P}^v$ defined in Eq. (19) can be deformed as follows:

$$\min_{\boldsymbol{S}^v \geqslant 0} \|\boldsymbol{Q}^v\|_{2,p}^p + \alpha \cdot \operatorname{tr}\left(\boldsymbol{S}^v\boldsymbol{U}_1^v(\boldsymbol{S}^v)^{\mathrm{T}}\right) + \beta \cdot \operatorname{tr}\left(\boldsymbol{S}^v\boldsymbol{U}_2^v(\boldsymbol{S}^v)^{\mathrm{T}}\right) + \gamma \cdot \operatorname{tr}\left(\boldsymbol{S}^v\boldsymbol{U}_3^v(\boldsymbol{S}^v)^{\mathrm{T}} - 2\boldsymbol{G}_c^{\mathrm{T}}\boldsymbol{G}^v\boldsymbol{P}^v\right) \tag{23}$$

For the constraint $\boldsymbol{S}^v = [\boldsymbol{S}^v_{ih}] \geqslant 0$, the Lagrangian multiplier $\varXi = [\xi_{ih}]$ is introduced. Then, the Lagrangian function $\mathcal{L}$ of Formula (23) is obtained as follows:

$$\mathcal{L} = \|\boldsymbol{Q}^v\|_{2,p}^p + \alpha \cdot \operatorname{tr}\left(\boldsymbol{S}^v\boldsymbol{U}_1^v(\boldsymbol{S}^v)^{\mathrm{T}}\right) + \beta \cdot \operatorname{tr}\left(\boldsymbol{S}^v\boldsymbol{U}_2^v(\boldsymbol{S}^v)^{\mathrm{T}}\right) + \gamma \cdot \operatorname{tr}\left(\boldsymbol{S}^v\boldsymbol{U}_3^v(\boldsymbol{S}^v)^{\mathrm{T}} - 2\boldsymbol{G}_c^{\mathrm{T}}\boldsymbol{G}^v\boldsymbol{P}^v\right) + \operatorname{tr}\left(\varXi(\boldsymbol{S}^v)^{\mathrm{T}}\right) \tag{24}$$

Then, the partial derivative of Eq. (24) with respect to $\boldsymbol{S}^v$ can be expressed as follows:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{S}^v} = -2\boldsymbol{Q}^v\boldsymbol{E}^v\boldsymbol{G}^v + 2\alpha\boldsymbol{S}^v\boldsymbol{U}_1^v + 2\beta\boldsymbol{S}^v\boldsymbol{U}_2^v + 2\gamma\boldsymbol{S}^v\boldsymbol{U}_3^v - 2\gamma\boldsymbol{S}^v(\boldsymbol{P}^v)^{-1}\boldsymbol{U}_4^v + \varXi \tag{25}$$

where $\boldsymbol{E}^v$ is a diagonal matrix with $\boldsymbol{E}^v_{ii} = p/[2\,\|\boldsymbol{Q}^v_{\cdot i}\|_2^{2-p}]$ and $\boldsymbol{U}_4^v = [\boldsymbol{G}_c^{\mathrm{T}}\boldsymbol{G}^v] \odot \boldsymbol{I}$.

Setting the above expression to zero and using the Karush-Kuhn-Tucker (KKT) condition[43] of $\xi_{ih}\boldsymbol{S}^v_{ih} = 0$, then the following update rule is obtained:

$$\boldsymbol{S}^v_{ih} = \boldsymbol{S}^v_{ih}\frac{\left(\boldsymbol{X}^v\boldsymbol{E}^v\boldsymbol{G}^v + \beta\boldsymbol{S}^v\boldsymbol{U}_2^{v-} + \gamma\boldsymbol{S}^v(\boldsymbol{P}^v)^{-1}\boldsymbol{U}_4^v\right)_{ih}}{\left(\boldsymbol{S}^v(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{E}^v\boldsymbol{G}^v + \alpha\boldsymbol{S}^v\boldsymbol{U}_1^v + \beta\boldsymbol{S}^v\boldsymbol{U}_2^{v+} + \gamma\boldsymbol{S}^v\boldsymbol{U}_3^v\right)_{ih}} \tag{26}$$

3.2.2 Fixing G_c and S^v, updating G^v

After updating $\boldsymbol{S}^v$, $\boldsymbol{P}^v$ (defined in Eq. (19)) is used to normalize the columns of $\boldsymbol{S}^v$, and the norm is compensated to $\boldsymbol{G}^v$, that is,

$$\boldsymbol{S}^v \Leftarrow \boldsymbol{S}^v(\boldsymbol{P}^v)^{-1}, \boldsymbol{G}^v \Leftarrow \boldsymbol{G}^v\boldsymbol{P}^v \tag{27}$$

When $\boldsymbol{G}_c$ and $\boldsymbol{S}^v$ are fixed, the minimizing problem in Formula (20) is reduced to

$$\min_{\boldsymbol{G}^v\geqslant 0} \|\boldsymbol{Q}^v\|_{2,p}^p + \alpha\cdot \operatorname{tr}((\boldsymbol{I}_{\text{block}}\odot\boldsymbol{G}^v)^{\mathrm{T}}(\boldsymbol{I}_{\text{block}}\odot\boldsymbol{G}^v))+ \beta\cdot\operatorname{tr}((\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{G}^v)+\gamma\cdot\operatorname{tr}((\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{G}^v-2\boldsymbol{G}_c^{\mathrm{T}}\boldsymbol{G}^v) \tag{28}$$

For the constraint $\boldsymbol{G}^v=[G^v_{jh}]\geqslant 0$, the Lagrangian multiplier $\Psi=[\psi_{jh}]$ is introduced. Then, the Lagrangian function $\mathcal{L}$ of Formula (28) is obtained as follows:

$$\mathcal{L} = \|\boldsymbol{Q}^v\|_{2,p}^p + \alpha\cdot \operatorname{tr}((\boldsymbol{I}_{\text{block}}\odot\boldsymbol{G}^v)^{\mathrm{T}}(\boldsymbol{I}_{\text{block}}\odot\boldsymbol{G}^v))+ \beta\cdot\operatorname{tr}((\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{G}^v)+\gamma\cdot\operatorname{tr}((\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{G}^v-2\boldsymbol{G}_c^{\mathrm{T}}\boldsymbol{G}^v)+\operatorname{tr}(\Psi(\boldsymbol{G}^v)^{\mathrm{T}}) \tag{29}$$

Then, the partial derivative of Eq. (29) with respect to $\boldsymbol{G}^v$ is expressed as follows:

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{G}^v} = -2\boldsymbol{E}^v(\boldsymbol{Q}^v)^{\mathrm{T}}\boldsymbol{S}^v + 2\alpha\boldsymbol{I}_{\text{block}}\odot\boldsymbol{G}^v + 2\beta\boldsymbol{D}^v\boldsymbol{G}^v - 2\beta\boldsymbol{W}^v\boldsymbol{G}^v + 2\gamma\boldsymbol{G}^v - 2\gamma\boldsymbol{G}_c + \Psi \tag{30}$$

Similarly, letting $\frac{\partial\mathcal{L}}{\partial\boldsymbol{G}^v}=0$ and using KKT condition of $\psi_{jh}G^v_{jh}=0$, the following update rule can be obtained:

$$G^v_{jh} = G^v_{jh}\frac{((\boldsymbol{E}^v)^{\mathrm{T}}(\boldsymbol{X}^v)^{\mathrm{T}}\boldsymbol{S}^v+\beta\boldsymbol{W}^v\boldsymbol{G}^v+\gamma\boldsymbol{G}_c)_{jh}}{((\boldsymbol{E}^v)^{\mathrm{T}}\boldsymbol{G}^v(\boldsymbol{S}^v)^{\mathrm{T}}\boldsymbol{S}^v+\alpha\boldsymbol{I}_{\text{block}}\odot\boldsymbol{G}^v+\beta\boldsymbol{D}^v\boldsymbol{G}^v+\gamma\boldsymbol{G}^v)_{jh}} \tag{31}$$

3.2.3 Fixing $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$, updating $\boldsymbol{G}_c$

The partial derivative of Eq. (18) with respect to $\boldsymbol{G}_c$ is as follows (in each iteration, $\boldsymbol{S}^v$ is normalized):

$$\frac{\partial O_{\text{GDCMVNMF}}}{\partial\boldsymbol{G}_c} = \frac{\partial\sum_{v=1}^{V}\gamma\|\boldsymbol{G}^v-\boldsymbol{G}_c\|_{\mathrm{F}}^2}{\partial\boldsymbol{G}_c} = \sum_{v=1}^{V}[-2\gamma\boldsymbol{G}^v+2\gamma\boldsymbol{G}_c]=0 \tag{32}$$

Then, the exact solution for $\boldsymbol{G}_c$ is

$$\boldsymbol{G}_c = \frac{\sum_{v=1}^{V}\boldsymbol{G}^v}{V}\geqslant 0 \tag{33}$$

Algorithm 1 summarizes the optimizing scheme of GDCMVNMF.

3.3 Computational complexity analysis of GDCMVNMF

In this section, the computational complexity of GDCMVNMF is analyzed and expressed in big O notation[2]. For a specific view v in one iteration, updating $\boldsymbol{S}^v$ requires the calculation of $\boldsymbol{S}^v(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{E}^v\boldsymbol{G}^v$, $\boldsymbol{X}^v\boldsymbol{E}^v\boldsymbol{G}^v$, $\boldsymbol{S}^v\boldsymbol{U}^v_1$, $\boldsymbol{S}^v\boldsymbol{U}^v_2$, $\boldsymbol{S}^v\boldsymbol{U}^v_3$, and $\boldsymbol{S}^v(\boldsymbol{P}^v)^{-1}\boldsymbol{U}^v_4$; furthermore, the cost of $\boldsymbol{S}^v(\boldsymbol{G}^v)^{\mathrm{T}}\boldsymbol{E}^v\boldsymbol{G}^v$ and $\boldsymbol{X}^v\boldsymbol{E}^v\boldsymbol{G}^v$ is $O((d^v)^2(n+m^v))$ and $O(d^vn+m^vnd^v)$, respectively. Total cost of $\boldsymbol{S}^v\boldsymbol{U}^v_1$, $\boldsymbol{S}^v\boldsymbol{U}^v_2$, $\boldsymbol{S}^v\boldsymbol{U}^v_3$, and $\boldsymbol{S}^v(\boldsymbol{P}^v)^{-1}\boldsymbol{U}^v_4$ is $O(m^vd^v+Kd^vn+(d^v)^2n)$, where K is the number of k-NNs in the graph; thus, the cost for updating $\boldsymbol{S}^v$ is $O(m^vnd^v)$. Normalization of $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$ in Formula (27) requires the computation of $O(nd^v+m^vd^v)$. Therefore, the main cost of updating $\boldsymbol{G}^v$ is based on the calculation of $(\boldsymbol{E}^v)^{\mathrm{T}}(\boldsymbol{X}^v)^{\mathrm{T}}\boldsymbol{S}^v$, $(\boldsymbol{E}^v)^{\mathrm{T}}\boldsymbol{G}^v(\boldsymbol{S}^v)^{\mathrm{T}}\boldsymbol{S}^v$, $\boldsymbol{W}^v\boldsymbol{G}^v$, and $\boldsymbol{D}^v\boldsymbol{G}^v$. The cost of $(\boldsymbol{E}^v)^{\mathrm{T}}(\boldsymbol{X}^v)^{\mathrm{T}}\boldsymbol{S}^v$ and $(\boldsymbol{E}^v)^{\mathrm{T}}\boldsymbol{G}^v(\boldsymbol{S}^v)^{\mathrm{T}}\boldsymbol{S}^v$ are $O(d^vn+m^vnd^v)$ and $O(d^vn+(d^v)^2(m^v+n))$, respectively. Here, both $\boldsymbol{W}^v\boldsymbol{G}^v$ and $\boldsymbol{D}^v\boldsymbol{G}^v$ require the computation of $O(Kd^vn)$ and $O(d^vn)$, respectively. The cost for updating $\boldsymbol{G}^v$ is also $O(m^vnd^v)$. When compared to the cost of updating $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$, the computational cost of updating $\boldsymbol{G}_c$ and calculating $\boldsymbol{E}^v$ is negligible; thus, the final cost for GDCMVNMF is $O(tVm^vnd^v)$, where t is the number of iterations.

Algorithm 1 Optimizing scheme for GDCMVNMF

Input: Multi-view data $\boldsymbol{D}_X=\{\boldsymbol{X}^1,\boldsymbol{X}^2,\ldots,\boldsymbol{X}^V\}$ and $\boldsymbol{X}^v\in\mathbf{R}_+^{m^v\times n}$, indicator matrix $\boldsymbol{I}_{\text{block}}\in\mathbf{R}^{n\times d^v}$, Parameters α, β, and γ, and number of k-NNs K

Output: $\boldsymbol{G}^v$ and $\boldsymbol{G}_c$

1: **for** each $v\in V$ **do**
2: Initialize $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$;
3: Construct k-NN graph with heat kernel weight;
4: **end**
5: **Repeat**
6: **for** each $v\in V$ **do**
7: S1: Fix $\boldsymbol{G}^v$, update $\boldsymbol{S}^v$ with Eq. (26);
8: S2: Normalize $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$ with Formula (27);
9: S3: Fix $\boldsymbol{S}^v$, update $\boldsymbol{G}^v$ with Eq. (31);
10: S4: Calculate diagonal matrix $\boldsymbol{E}^v$ as $E^v_{kk}=p/[2\|\boldsymbol{Q}_{\cdot k}\|_2^{2-p}]$;
11: **end**
12: Fix $\boldsymbol{S}^v$ and $\boldsymbol{G}^v$, update $\boldsymbol{G}_c$ with Eq. (33);
13: **Until** convergence

4 ExMVCNMF

As pointed out in the preceding section, the major problem of the previously proposed semi-supervised MVNMFs based on CNMF is that they fail to consider the discriminative information provided by the labeled samples. Although label information has been used to map samples with the same label to one same feature vector (see Fig. 1), the distinction of different classes is not ensured. Additionally, geometric information on data is not considered. To address these issues, ExMVCNMF with a semiMVNMF framework can be developed. Refer to Fig. 2 for the connection between ExMVCNMF and GDCMVNMF.

4.1 Objective function of ExMVCNMF

In each view, the following constraint should be imposed on the auxiliary matrix $\boldsymbol{Z}^v$ to guarantee the discriminative ability of ExMVCNMF:

$$\|\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v\|_{\mathrm{F}}^2 \tag{34}$$

where $\boldsymbol{I}_{\mathrm{disc}} = [\hat{\boldsymbol{I}}; \hat{\boldsymbol{0}}] \in \mathbf{R}^{(n-l+c)\times d^v}$ is a discriminative matrix to ensure the distinction of the features from different classes. Here, $\hat{\boldsymbol{0}} \in \mathbf{R}^{(n-l)\times d^v}$ is a zero matrix corresponding to the unlabeled samples. Specifically, $\hat{\boldsymbol{I}} \in \mathbf{R}^{c\times d^v}$ is defined as follows:

$$\hat{\boldsymbol{I}} = \begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix} \tag{35}$$

The objective function of ExMVCNMF can be expressed as follows:

$$O_{\mathrm{ExMVCNMF}} = \sum_{v=1}^{V} (\|\boldsymbol{X}^v - \boldsymbol{S}^v(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\|_{\mathrm{F}}^2 + \alpha \cdot \|\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v\|_{\mathrm{F}}^2 + \beta \cdot \mathrm{tr}\,((\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v) + \gamma\,\|\boldsymbol{Z}^v - \boldsymbol{Z}_c\|_{\mathrm{F}}^2),$$
$$\text{s.t., } \boldsymbol{S}^v, \boldsymbol{Z}^v, \boldsymbol{Z}_c \geqslant 0 \tag{36}$$

The same strategy, similar to GDCMVNMF, is adopted to align the feature scales of multiple views. Then, Eq. (36) can be further rewritten as follows:

$$O_{\mathrm{ExMVCNMF}} = \sum_{v=1}^{V} (\|\boldsymbol{X}^v - \boldsymbol{S}^v(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\|_{\mathrm{F}}^2 + \alpha \cdot \|\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v\boldsymbol{P}^v\|_{\mathrm{F}}^2 + \beta \cdot \mathrm{tr}\,((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v\boldsymbol{P}^v) + \gamma \cdot \|\boldsymbol{Z}^v\boldsymbol{P}^v - \boldsymbol{Z}_c\|_{\mathrm{F}}^2),$$
$$\text{s.t., } \boldsymbol{S}^v, \boldsymbol{Z}^v, \boldsymbol{Z}_c \geqslant 0 \tag{37}$$

4.2 Optimization algorithm of ExMVCNMF

For the v-th view, the other views are not involved in the optimization of $\boldsymbol{S}^v$ and $\boldsymbol{Z}^v$. Minimizing Eq. (37) for the v-th view is as follows:

$$\min_{\boldsymbol{S}^v, \boldsymbol{Z}^v, \boldsymbol{Z}_c \geqslant 0} \|\boldsymbol{X}^v - \boldsymbol{S}^v(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\|_{\mathrm{F}}^2 + \alpha\,\|\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v\boldsymbol{P}^v\|_{\mathrm{F}}^2 + \beta \cdot \mathrm{tr}\,((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v\boldsymbol{P}^v) + \gamma\,\|\boldsymbol{Z}^v\boldsymbol{P}^v - \boldsymbol{Z}_c\|_{\mathrm{F}}^2 \tag{38}$$

4.2.1 Fixing $\boldsymbol{Z}_c$ and $\boldsymbol{Z}^v$, updating $\boldsymbol{S}^v$

When $\boldsymbol{Z}_c$ and $\boldsymbol{Z}^v$ are fixed, Formula (38) can be expressed as follows:

$$\min_{\boldsymbol{S}^v \geqslant 0} \|\boldsymbol{X}^v - \boldsymbol{S}^v(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\|_{\mathrm{F}}^2 + \alpha \cdot \mathrm{tr}\,((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v)^{\mathrm{T}}\,(\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v)\boldsymbol{P}^v) + \beta \cdot \mathrm{tr}\,((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v\boldsymbol{P}^v) + \gamma \cdot \mathrm{tr}\,((\boldsymbol{P}^v)^{\mathrm{T}}(\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{Z}^v\boldsymbol{P}^v - 2\boldsymbol{Z}_c^{\mathrm{T}}\boldsymbol{Z}^v\boldsymbol{P}^v) \tag{39}$$

Here, $\boldsymbol{U}_1^v$, $\boldsymbol{U}_2^v$, and $\boldsymbol{U}_3^v$ are redefined as follows:

$$\boldsymbol{U}_1^v = \left[(\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v)^{\mathrm{T}}(\boldsymbol{I}_{\mathrm{disc}} \odot \boldsymbol{Z}^v)\right] \odot \boldsymbol{I},$$
$$\boldsymbol{U}_2^v = \left[(\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{A}_{\mathrm{lc}}\boldsymbol{Z}^v\right] \odot \boldsymbol{I} = \boldsymbol{U}_2^{v+} - \boldsymbol{U}_2^{v-},$$
$$\boldsymbol{U}_3^v = \left[(\boldsymbol{Z}^v)^{\mathrm{T}}\boldsymbol{Z}^v\right] \odot \boldsymbol{I} \tag{40}$$

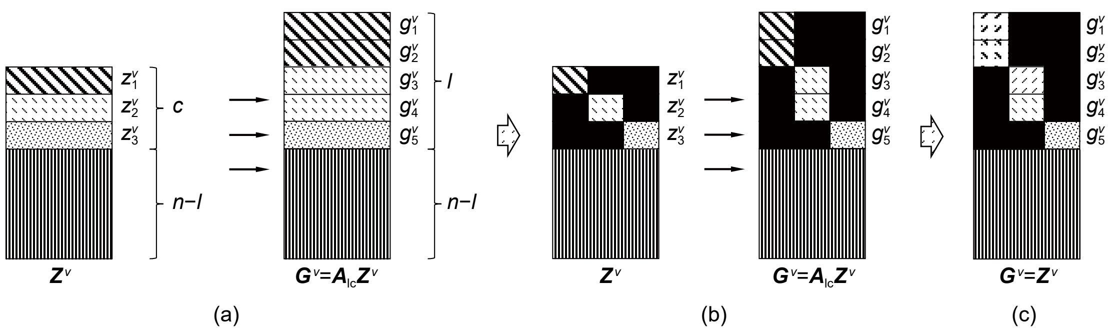


Fig. 2 Z^v and $A_{\mathrm{lc}}Z^v$ of (a) AMVNMF[33], MVCNMF[34], and MVOCNMF[35], (b) ExMVCNMF, and (c) GDCMVNMF. In (a), the distinction of feature vectors from different classes is not guaranteed; in (b), the distinction of different classes is considered in ExMVCNMF, and the samples with the same label are represented by the same feature vectors; In (c), the distinction of different classes is also considered in GDCMVNMF; however, the samples with the same label are represented by different feature vectors. For (c), the label constraint matrix A_{lc} is expanded as an identity matrix, hence $G^v = A_{\mathrm{lc}}Z^v = Z^v$. When p=2, GDCMVNMF can be seen as an extended variant of ExMVCNMF.

where

$$U_2^{v+} = \left[(A_{lc}Z^v)^T D^v A_{lc} Z^v\right] \odot I,$$
$$U_2^{v-} = \left[(A_{lc}Z^v)^T W^v A_{lc} Z^v\right] \odot I.$$

Then, Formula (39) with P^v defined in Eq. (19) can be deformed as follows:

$$\min_{S^v \geqslant 0} \|X^v - S^v(A_{lc}Z^v)^T\|_F^2 + \alpha \cdot \operatorname{tr}(S^v U_1^v (S^v)^T) + \beta \cdot \operatorname{tr}(S^v U_2^v (S^v)^T) + \gamma \cdot \operatorname{tr}(S^v U_3^v (S^v)^T - 2Z_c^T Z^v P^v) \quad (41)$$

For the constraint $S^v = [S_{ih}^v] \geqslant 0$, the Lagrangian multiplier $\Xi = [\xi_{ih}]$ is introduced. Then, the Lagrangian function $\mathcal{L}$ of Formula (41) is obtained as follows:

$$\mathcal{L} = \|X^v - S^v(A_{lc}Z^v)^T\|_F^2 + \alpha \cdot \operatorname{tr}(S^v U_1^v (S^v)^T) + \beta \cdot \operatorname{tr}(S^v U_2^v (S^v)^T) + \gamma \cdot \operatorname{tr}(S^v U_3^v (S^v)^T - 2Z_c^T Z^v P^v) + \operatorname{tr}(\Xi (S^v)^T) \quad (42)$$

Then, the partial derivative of Eq. (42) with respect to S^v can be expressed as follows:

$$\frac{\partial \mathcal{L}}{\partial S^v} = -2X^v A_{lc} Z^v + 2S^v A_{lc} (Z^v)^T A_{lc} Z^v + 2\alpha S^v U_1^v + 2\beta S^v U_2^v + 2\gamma S^v U_3^v - 2\gamma S^v (P^v)^{-1} U_4^v + \Xi \quad (43)$$

where U_4^v is redefined as $U_4^v = [Z_c^T Z^v] \odot I$.

Setting the above expression to zero and using KKT condition of $\xi_{ih} S_{ih}^v = 0$, then the following update rule is obtained as follows:

$$S_{ih}^v = S_{ih}^v \frac{(X^v A_{lc} Z^v + \beta S^v U_2^{v-} + \gamma S^v (P^v)^{-1} U_4^v)_{ih}}{(S^v (A_{lc} Z^v)^T A_{lc} Z^v + \alpha S^v U_1^v + \beta S^v U_2^{v+} + \gamma S^v U_3^v)_{ih}} \quad (44)$$

4.2.2 Fixing Z_c and S^v, updating Z^v

After updating S^v, the columns of S^v are normalized with P^v in Eq. (19), and the norm is compensated to Z^v, that is,

$$S^v \Leftarrow S^v (P^v)^{-1},\ Z^v \Leftarrow Z^v P^v \quad (45)$$

When Z_c and S^v are fixed, Formula (38) is equivalent to the following problem:

$$\min_{Z^v \geqslant 0} \|X^v - S^v(A_{lc}Z^v)^T\|_F^2 + \alpha . \operatorname{tr}((I_{disc} \odot Z^v)^T (I_{disc} \odot Z^v)) + \beta \cdot \operatorname{tr}((A_{lc}Z^v)^T L^v A_{lc} Z^v) + \gamma \cdot \operatorname{tr}(Z^{vT} Z^v - 2Z_c^T Z^v) \quad (46)$$

For the constraint $Z^v = [Z_{jh}^v] \geqslant 0$, the Lagrangian multiplier $\Psi = [\psi_{jh}]$ is introduced. Then, the Lagrangian function $\mathcal{L}$ of Formula (46) is obtained as follows:

$$\mathcal{L} = \|X^v - S^v(A_{lc}Z^v)^T\|_F^2 + \alpha \cdot \operatorname{tr}((I_{disc} \odot Z^v)^T (I_{disc} \odot Z^v)) + \beta \cdot \operatorname{tr}((A_{lc}Z^v)^T L^v A_{lc} Z^v) + \gamma \cdot \operatorname{tr}((Z^v)^T Z^v - 2Z_c^T Z^v) + \operatorname{tr}(\Psi (Z^v)^T) \quad (47)$$

The partial derivative of Eq. (47) with respect to Z^v is expressed as follows:

$$\frac{\partial \mathcal{L}}{\partial Z^v} = -2{A_{lc}}^T (X^v)^T S^v + 2{A_{lc}}^T A_{lc} Z^v (S^v)^T S^v + 2\alpha I_{disc} \odot Z^v + 2\beta {A_{lc}}^T D^v A_{lc} Z^v - 2\beta {A_{lc}}^T W^v A_{lc} Z^v + 2\gamma Z^v - 2\gamma Z_c + \Psi \quad (48)$$

Similarly, letting $\frac{\partial \mathcal{L}}{\partial Z^v} = 0$ and using KKT condition of $\psi_{jh} Z_{jh}^v = 0$, then the following update rule is obtained as follows:

$$Z_{jh}^v = Z_{jh}^v \frac{({A_{lc}}^T (X^v)^T S^v + \beta {A_{lc}}^T W^v A_{lc} Z^v + \gamma Z_c)_{jh}}{({A_{lc}}^T A_{lc} Z^v (S^v)^T S^v + \alpha I_{disc} \odot Z^v + \beta {A_{lc}}^T D^v A_{lc} Z^v + \gamma Z^v)_{jh}} \quad (49)$$

4.2.3 Fixing S^v and Z^v, updating Z_c

Since S^v is normalized in each iteration, the partial derivative of Eq. (37) with respect to Z_c is expressed as follows:

$$\frac{\partial O_{ExMVCNMF}}{\partial Z_c} = \frac{\partial \sum_{v=1}^{V} \gamma \|Z^v - Z_c\|_F^2}{\partial Z_c} = \sum_{v=1}^{V} [-2\gamma Z^v + 2\gamma Z_c] = 0 \quad (50)$$

Then, the exact solution to Z_c is

$$Z_c = \frac{\sum_{v=1}^{V} Z^v}{V} \geqslant 0 \quad (51)$$

Algorithm 2 summarizes the optimizing scheme of ExMVCNMF.

4.3 Computational complexity analysis of ExMVCNMF

For the v-th view, updating S^v requires calculating $X^v A_{lc} Z^v$, $S^v (A_{lc} Z^v)^T A_{lc} Z^v$, $S^v U_1^v$, $S^v U_2^v$, $S^v U_3^v$, and $S^v (P^v)^{-1} U_4^v$ in one iteration. The computational cost of $X^v A_{lc} Z^v$ and $S^v (A_{lc} Z^v)^T A_{lc} Z^v$ are $O(d^v n + m^v n (n-l+C))$ and $O((d^v)^2 (m^v + n))$, respectively. The total cost of $S^v U_1^v$, $S^v U_2^v$, $S^v U_3^v$, and $S^v (P^v)^{-1} U_4^v$ is $O(m^v d^v + K d^v n + (d^v)^2 n + (d^v)^2 (n-l+C))$, thus, the cost for updating S^v is $O(m^v n (n-l+C))$. Normalization of S^v

Algorithm 2 Optimizing scheme for ExMVCNMF

Input: Multi-view data $\boldsymbol{D}_X=\{\boldsymbol{X}^1,\boldsymbol{X}^2,\ldots,\boldsymbol{X}^V\}$ and $\boldsymbol{X}^v\in\mathbf{R}_{+}^{m^v\times n}$, indicator matrix $\boldsymbol{I}_{\text{disc}}\in\mathbf{R}^{(n-l+c)\times d^v}$, parameters α, β, and γ, and number of k-NNs K

Output: $\boldsymbol{Z}^v$ and $\boldsymbol{Z}_c$

```
for each v ∈ V do
    Initialize S^v and Z^v;
    Construct k-NN graph with heat kernel weight;
end
Repeat
    for each v ∈ V do
        S1: Fix Z^v, update S^v with Eq. (44);
        S2: Normalize S^v and Z^v with Formula (45);
        S3: Fix S^v, update Z^v with Eq. (49);
    end
    Fix S^v and Z^v, update Z_c with Eq. (51);
Until convergence
```

and $\boldsymbol{Z}^v$ in Formula (45) requires the computation of $O\,((n-l+C)d^v+m^vd^v)$. Therefore, the main cost of updating $\boldsymbol{Z}^v$ is based on the calculation of $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}(\boldsymbol{X}^v)^{\text{T}}\boldsymbol{S}^v$, $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v(\boldsymbol{S}^v)^{\text{T}}\boldsymbol{S}^v$, $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}\boldsymbol{W}^v\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v$, and $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}\boldsymbol{D}^v\boldsymbol{A}\boldsymbol{Z}^v$. The cost of $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}(\boldsymbol{X}^v)^{\text{T}}\boldsymbol{S}^v$ and $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v(\boldsymbol{S}^v)^{\text{T}}\boldsymbol{S}^v$ is $O\,(m^v(n-l+C)n+m^v(n-1+C)d^v)$ and $O\,(d^vn+m^v(n-l+C)n+m^v(n-l+C)d^v)$, respectively. Additionally, both $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}\boldsymbol{W}^v\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v$ and $\boldsymbol{A}_{\text{lc}}{}^{\text{T}}\boldsymbol{D}^v\boldsymbol{A}_{\text{lc}}\boldsymbol{Z}^v$ require the computation of $O\,(Kd^vn)$ and $O\,(d^vn)$, respectively, and the cost for updating $\boldsymbol{Z}^v$ is also $O\,(m^vn(n-l+C))$. The computational cost of updating $\boldsymbol{Z}_c$ is trivial; hence, the final cost for ExMVCNMF is $O\,(tVm^vn(n-l+C))$.

5 Experimental Result

5.1 Datasets

In this section, six real-world multi-view datasets are used to validate the superiority of the proposed methods. Table 2 presents the statistics of these datasets.

Table 2 Database description.

Dataset	Number of samples	Number of views	Feature dimensionality	Number of classes
Yale	165	3	2048/256/1024	15
ORL	400	3	2048/256/1024	40
FEI part 1	700	3	2048/256/1024	50
YaleB	765	3	2048/256/1024	12
ECG	294	2	2560/1281	3
WebKB	1051	2	2949/334	2

(1) **Yale dataset※ :** This dataset includes 165 grayscale images captured from 15 individuals. Each individual has 11 images with different facial expressions or configurations, which are normalized to a 32 pixel × 32 pixel array.

(2) **ORL dataset†:** This dataset has 400 gray scale face images collected from 40 individuals, with 10 images for each individual. These images are captured under different light conditions, with different facial expressions, and with/without glasses.

(3) **FEI part 1 dataset‡:** The FEI part 1 dataset is a subset of the original FEI data base. This dataset has 700 color images collected from 50 individuals. For each individual, 14 images are captured under different views. All images are downsampled from the original resolution of 640 pixel × 480 pixel to 32 pixel × 24 pixel, and the color images are converted into grayscale images.

(4) **YaleB dataset§:** This dataset includes 2414 gray scale face images collected from 38 individuals. Each one has approximately 64 images which are captured under different lighting conditions. The experiments use a subset of this dataset with 12 classes.

(5) **ECG dataset¶:** In this dataset, 162 original ECG records are collected from three classes: arrhythmia (namely ARR), Normal Sinus Rhythm (NSR), and Congestive Heart Failure (CHF), and each of them has 96, 36, and 30 records, respectively. Each record is sampled at 512 s with a rate of 128 Hz. For the ARR record, one segmentation had 20 s, while for the NSR and CHF records, the first 60 s are uniformly segmented to get three 20 s segmentations. Two views are used in this study: the time-doman view and Fourier coefficient view. Finally, a dataset with 294 instances are constructed. Fourier coefficient view and time-domain feature view are adopted as two feature views. In total, 294 records are used for evaluation.

(6) **WebKB dataset☆:** This dataset is a subset of web documents from four universities. It includes 1051 pages with two classes: 230 course pages and 821 non-course pages. Each page has two views: full-text view with 2949 features for the textual content of the web page, and in-link view with 334 features for the anchor text on the hyperlinks pointing to the pages. This

※ http://vision.ucsd.edu/leekc/ExtYaleDatabase/ExtYaleB.html
† http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
‡ http://fei.edu.br/cet/facedatabase.html
§ http://vision.ucsd.edu/content/yale-face-database
¶ https: //github.com/mathworks/physionet_ECG_data/
☆ http://www.cs.cmu.edu/webkb/

dataset is balanced by selecting 241 data points from the second view.

The three views of Yale, ORL, FEI part 1, and YaleB datasets are composed of the 2048D image pixel feature, 256D Gabor feature, and 1024D LBP feature.

5.2 Comparative methods

Various representative NMF-based multi-view clustering methods are compared with the proposed methods to validate their superiority. The compared methods include LP-DiNMF[23], rNNMF[24], MPMNMF_1[20], MPMNMF_2[20], UDNMF[21], AMVNMF[33], MVCNMF[34], MVOCNMF[35], and the following two methods:

VAGNMF: This method conducts GNMF[44] on each view individually, and the average feature of multiple views is treated as the final representation.

VCGNMF: This method conducts GNMF on each view individually and concatenates the features of multiple views as the final representation.

In this study, the clustering performances are evaluated using two metrics: Normalized Mutual Information (NMI)[45] and accuracy (abbreviated as AC)[46]. The optimal parameter settings of the above methods are obtained using grid search. The average results of 10 runs are also reported for all methods to improve the influence of randomness.

5.3 Convergence analysis

Figure 3 shows the convergence curves of ExMVCNMF and GDCMVNMF on six datasets to verify the convergence of these methods. For GDCMVNMF, the curves with $p = 2$ are presented. The proposed methods converge on all datasets in Fig. 3, where the x-coordinate represents the times of iteration and the y-coordinate represents the log value of the objective function.

5.4 Experimental results and analysis

In this section, the performances of ExMVCNMF and GDCMVNMF are evaluated and compared with several representative recently proposed unsupervised and semi-supervised multi-view NMF approaches on six datasets. In Tables 3 and 4, GDCMVNMF with 0% and 10% labeled data points are compared with several representative unsupervised multi-view NMF methods, and results of GDCMVNMF with p = 0.5, 1, and 2 are reported. From Tables 3 and 4, we can see that GDCMVNMF with 10% labeled data points significantly outperforms the other unsupervised methods, regardless of whether p = 0.5, 1, or 2. This indicates that the discriminative information learned from the label information helps learn a better compact representation for multi-view datasets. However, it cannot be directly inferred whether supervised methods always behave better than the unsupervised methods. In Table 5 and 6, ExMVCNMF and GDCMVNMF are compared with several recently proposed semi-supervised multi-view methods. Similarly, the results of GDCMVNMF with p = 0.5, 1, and 2 are reported. Comparing the results in Tables 3–6, it is observed that, on the Yale dataset, the performances of semi-supervised methods AMVNMF, MVCNMF, and MVOCNMF are comparable with the results of several unsupervised methods, such as LP-DiNMF, rNNMF, MPMNMF_1, and MPMNMF_2. However, VCGNMF outperforms these supervised methods. When labeling ratios are less than 20% on the ORL and ECG datasets, AMVNMF, MVCNMF, and MVOCNMF have no advantage over several unsupervised methods in many

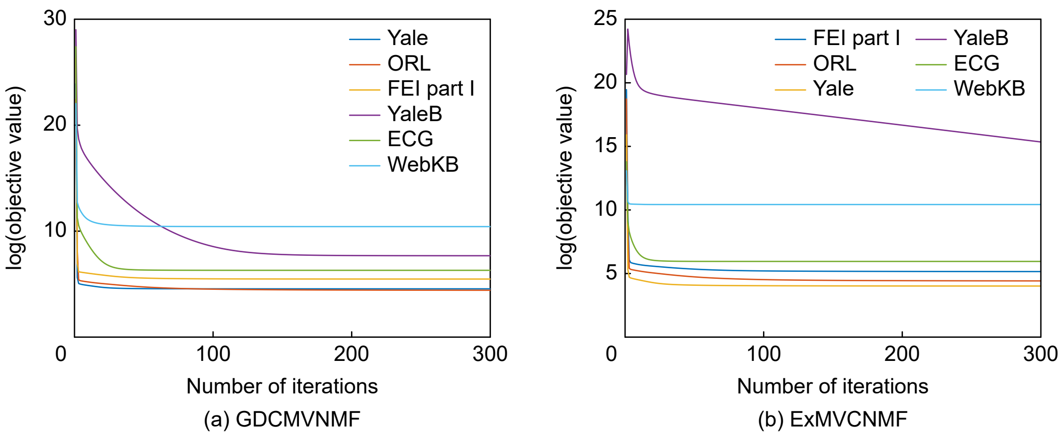


Fig. 3 Convergence curves of GDCMVNMF (p=2) and ExMVCNMF on six datasets.

Table 3 AC on four datasets compared with state-of-the-art unsupervised methods. The best results are in bold, while the second best results are in bold italic. GDCMVNMF$_{l=10\%}$ and GDCMVNMF$_{l=10\%}$ denote GDCMVNMF with 0% and 10% labeled samples, respectively, when p is set to 0.5, 1, and 2.

(%)

Method	Yale	ORL	FEI part 1	YaleB	ECG	WebKB
VAGNMF	48.55±2.95	68.78±3.17	69.41±2.67	44.29±4.00	55.41±4.58	77.02±1.24
VCGNMF	53.52±3.14	70.28±2.26	71.44±2.25	43.91±4.59	56.43±2.93	79.57±0.70
LP-DiNMF	51.82±3.03	68.48±3.33	68.24±3.25	44.77±4.16	54.90±5.69	76.49±0.79
rNNMF	51.15±5.49	64.35±2.74	55.26±2.70	21.07±1.00	57.38±0.45	78.30±1.17
MPMNMF_1	50.00±3.05	70.25±2.87	69.19±3.17	45.48±4.70	57.86±3.60	81.11±1.77
MPMNMF_2	49.52±3.67	69.48±2.89	70.39±2.80	47.87±3.89	58.20±3.71	78.98±2.10
UDNMF	47.76±2.83	61.73±1.56	70.24±2.10	33.36±9.59	56.84±3.11	79.34±4.80
GDCMVNMF$_{l=0\%,\,p=0.5}$	51.27±3.04	66.70±1.69	72.49±2.10	45.46±2.39	57.04±6.52	75.51±0.70
GDCMVNMF$_{l=10\%,\,p=0.5}$	***61.89±2.46***	74.24±1.20	80.00±0.68	***85.98±2.78***	66.88±6.92	***93.22±0.32***
GDCMVNMF$_{l=0\%,\,p=1}$	50.61±2.64	67.20±1.94	72.60±1.31	45.35±3.27	59.18±6.88	82.57±1.90
GDCMVNMF$_{l=10\%,\,p=1}$	61.28±1.26	**76.12±1.29**	***80.15±0.87***	**85.99±2.81**	***67.00±6.90***	**93.33±1.33**
GDCMVNMF$_{l=0\%,\,p=2}$	51.88±2.87	69.85±2.12	72.40±1.57	45.78±3.46	55.44±4.46	84.13±6.11
GDCMVNMF$_{l=10\%,\,p=2}$	**62.27±2.35**	***75.67±1.90***	**81.92±0.66**	85.98±2.64	**67.16±6.89**	91.94±0.64

Table 4 NMI on four datasets compared with state-of-the-art unsupervised methods. The best results are in bold, while the second best results are in bold italic. GDCMVNM$_{l=0\%}$ and GDCMVNMF$_{l=10\%}$ denote GDCMVNMF with 0% and 10% labeled samples, respectively, when p is set to 0.5, 1, and 2.

(%)

Method	Yale	ORL	FEI part 1	YaleB	ECG	WebKB
VAGNMF	52.38±2.70	82.88±1.68	84.95±1.07	46.60±3.25	22.82±4.34	26.84±2.28
VCGNMF	56.79±2.27	83.67±1.01	85.99±0.81	48.61±4.62	26.36±4.43	29.42±1.42
LP-DiNMF	53.73±2.67	82.78±1.80	84.76±1.17	46.93±3.16	22.52±7.80	26.14±1.56
rNNMF	53.00±4.40	79.79±1.31	74.13±1.23	17.22±1.74	23.80±1.24	24.63±5.85
MPMNMF_1	52.89±2.78	83.65±0.99	85.06±1.64	47.10±3.66	19.72±4.60	34.22±2.46
MPMNMF_2	54.00±2.57	83.64±1.31	85.78±1.02	49.28±3.95	20.41±4.42	30.99±5.15
UDNMF	49.88±2.73	78.54±1.03	86.18±0.74	37.21±7.87	22.28±3.25	29.98±9.00
GDCMVNMF$_{l=0\%,\,p=0.5}$	55.03±2.00	83.34±0.88	87.42±0.92	44.71±3.23	24.31±7.42	21.06±1.14
GDCMVNMF$_{l=10\%,\,p=0.5}$	61.94±2.81	86.20±0.37	89.17±0.27	78.36±1.93	35.15±7.66	***64.64±1.24***
GDCMVNMF$_{l=0\%,\,p=1}$	55.13±1.76	82.93±0.73	87.53±0.64	44.32±3.33	22.85±6.40	33.93±4.60
GDCMVNMF$_{l=10\%,\,p=1}$	***63.04±1.48***	**86.70±0.46**	***89.24±0.25***	**78.49±1.88**	***35.33±7.57***	**65.60±4.41**
GDCMVNMF$_{l=0\%,\,p=2}$	55.61±1.59	84.15±1.00	87.25±0.83	45.55±3.38	19.42±5.53	44.20±14.89
GDCMVNMF$_{l=10\%,\,p=2}$	**63.32±2.14**	***86.69±0.64***	**89.62±0.22**	***78.44±1.68***	**35.57±7.67**	62.34±1.93

cases. These cases are even worse on the FEI part 1 and YaleB datasets. Note that all unsupervised methods have used geometric information on the data, whereas AMVNMF, MVCNMF, and MVOCNMF do not. This indicates that geometric information is crucial when the labeling ratio is low. Additionally, from Table 3 and 4, it can be observed that the performance of GDCMVNMF$_{l=0\%}$ (with 0% labeled samples) is comparable to the most of the recently proposed unsupervised methods. This proves the effectiveness of the adopted feature normalizing strategy. In Tables 5 and 6, it can be seen that, when compared with other semi-supervised methods, ExMVCNMF and GDCMVNMF have obvious advantages under different labeling ratios. These cases indicate that the proposed methods effectively use the label information to obtain more discriminative representations. The proposed methods perform better than AMVNMF, MVCNMF, and MVOCNMF. This is because the former methods use both geometric and discriminative information. Additionally, ExMVCNMF outperforms GDCMVNMF (especially when p = 2) in most cases,

Table 5 AC on six datasets compared with state-of-the-art semi-supervised methods. The best results are in bold, while the second best results are in bold italic.

(%)

Dataset	Ratio	AMVNMF	MVCNMF	MVOCNMF	ExMVCNMF	GDCMVNMF$_{p=0.5}$	GDCMVNMF$_{p=1}$	GDCMVNMF$_{p=2}$
Yale	10	50.40±0.89	50.99±1.53	51.33±1.62	***62.05±3.67***	61.89±2.46	61.28±1.26	**62.27±2.35**
	20	55.38±1.97	57.68±1.12	57.84±1.44	**72.07±2.49**	70.80±2.31	70.86±2.57	***71.03±1.91***
	30	61.56±1.70	62.63±1.56	63.70±2.63	**77.56±1.74**	***77.53±1.51***	76.18±1.71	75.75±1.63
ORL	10	63.15±0.65	66.48±0.80	66.77±1.29	**76.86±1.46**	74.24±1.20	***76.12±1.29***	*75.67±1.94*
	20	68.41±0.54	70.18±0.85	70.27±0.55	**85.69±3.02**	83.37±1.60	***85.18±1.96***	85.01±2.94
	30	71.99±1.20	74.02±1.41	75.27±0.54	**89.13±1.47**	88.45±1.41	86.87±2.38	***89.00±1.17***
FEI part 1	10	56.55±0.79	59.36±0.89	55.61±1.15	**82.45±1.02**	80.00±0.68	80.15±0.87	***81.92±0.66***
	20	60.76±0.74	64.07±0.89	57.86±1.24	**86.98±1.29**	84.13±0.57	84.18±0.68	***86.11±0.61***
	30	65.91±0.61	69.59±1.36	64.97±0.94	**91.76±0.33**	88.76±0.58	90.15±0.28	***90.27±0.33***
YaleB	10	25.34±0.85	24.12±1.09	23.71±1.02	80.00±2.39	***85.98±2.78***	**85.99±2.81**	***85.98±2.64***
	20	31.16±0.92	29.30±1.49	29.53±0.79	90.30±1.01	91.20±1.07	***91.54±0.89***	**91.61±0.89**
	30	39.24±1.29	36.64±0.90	34.26±0.51	88.26±4.09	93.23±0.68	***93.29±0.41***	**93.39±0.57**
ECG	10	50.90±0.96	53.78±0.56	58.69±1.15	**70.63±6.19**	66.88±6.92	67.00±6.90	***67.16±6.89***
	20	54.81±2.34	57.50±1.57	63.74±1.84	**82.26±3.25**	***79.39±4.58***	78.22±4.21	77.31±4.50
	30	55.67±3.63	66.22±1.49	70.29±2.65	**84.97±1.24**	83.86±2.24	***84.00±2.43***	83.81±2.40
WebKB	10	82.44±1.54	82.73±2.39	83.29±1.96	92.09±1.46	***93.22±0.32***	**93.33±1.33**	91.94±0.64
	20	85.90±1.94	84.44±2.05	81.87±2.61	**92.35±2.58**	91.46±4.11	***92.21±4.24***	90.11±4.49
	30	90.17±1.00	91.06±0.73	90.45±0.68	95.03±1.68	***95.87±0.71***	**96.21±1.17**	94.55±1.18

Table 6 NMI on six datasets compared with state-of-the-art semi-supervised methods. The best results are in bold, while the second best results are in bold italic.

(%)

Dataset	Ratio	AMVNMF	MVCNMF	MVOCNMF	ExMVCNMF	GDCMVNMF$_{p=0.5}$	GDCMVNMF$_{p=1}$	GDCMVNMF$_{p=2}$
Yale	10	53.51±0.75	54.54±1.59	54.61±1.10	**63.74 2.72**	61.94±2.81	63.04±1.48	***63.32±2.14***
	20	59.08±2.04	61.28±1.24	61.34±0.90	**70.14 2.33**	68.06±2.43	***68.82±1.92***	68.24±2.08
	30	65.45±1.68	67.00±1.02	67.57±1.40	**74.95 2.08**	***74.47±2.12***	72.88±2.47	72.40±2.32
ORL	10	79.16±0.51	81.41±0.40	81.22±0.56	**87.09±0.58**	86.20±0.37	***86.70±0.46***	86.69±0.64
	20	82.37±0.29	83.76±0.28	83.81±0.40	**91.58±1.38**	89.96±0.79	90.91±0.63	***91.09±1.15***
	30	84.51±0.42	85.98±0.50	86.55±0.13	**92.68±1.03**	92.08±0.64	91.23±1.38	***92.30±0.82***
FEI part 1	10	75.96±0.40	77.38±0.34	73.93±0.33	**89.97±0.68**	89.17±0.27	89.24±0.25	***89.62±0.22***
	20	78.39±0.44	80.37±0.43	75.50±0.96	**92.09±0.59**	89.93±0.30	90.05±0.55	***91.47±0.31***
	30	82.29±0.24	84.37±0.58	80.59±0.77	**94.66±0.26**	92.68±0.52	93.49±0.25	***93.65±0.21***
YaleB	10	20.03±1.06	16.73±1.20	16.44±1.26	74.47±2.00	78.36±1.93	**78.49±1.88**	***78.44±1.68***
	20	26.51±2.25	22.05±1.19	22.28±1.04	83.56±1.71	84.40±1.57	***84.93±1.42***	**85.05±1.42**
	30	36.10±1.25	29.93±1.34	28.40±0.64	84.90±2.86	87.97±1.00	***87.98±0.86***	**88.17±1.05**
ECG	10	13.90±1.19	26.23±0.87	26.31±1.39	**40.52±6.79**	35.15±7.66	35.33±7.57	***35.57±7.67***
	20	18.93±1.51	28.86±0.89	25.51±1.58	**56.08±6.23**	***51.17±6.51***	50.23±7.24	48.91±7.22
	30	19.20±3.54	31.46±1.24	36.43±2.16	***58.45±2.62***	58.20±3.46	**58.60±3.68**	57.94±4.10
WebKB	10	33.26±3.55	38.54±4.39	39.89±3.84	63.64±5.18	***64.64±1.24***	**65.60±4.41**	62.34±1.93
	20	41.58±5.12	43.13±3.01	39.35±3.54	***62.12±8.44***	61.57±8.66	**64.75±9.18**	58.75±8.87
	30	53.84±3.19	59.09±1.66	57.14±1.58	71.92±7.23	***75.48±3.23***	**77.02±5.47**	70.90±4.99

indicating that the intra-class compactness is necessary for discovering discriminant information. Although, both ExMVCNMF and GDCMVNMF have attempted to use the discriminative information of data, the property of CNMF adopted in ExMVCNMF requires tighter intra-class compactness than that in

GDCMVNMF. From Tables 3–6, it can be observed that in most cases, GDCMVNMF with p=1 and 2 outperform that with $p = 0.5$.

5.5 Parameter sensitivity analysis

There are several parameters in the proposed GDCMVNMF and ExMVCNMF. Here, α, β, and γ are the parameters to balance the influence of the discriminative regularizing term $\Omega_l(\cdot)$, the geometric information regularizing term $\Omega_g(\cdot)$, and the multiple views aligning term $\Omega_a(\cdot)$, respectively (see Eq. (11)). Furthermore, K is the number of k-NNs in the Laplacian graph of each view. For GDCMVNMF, there is an additional parameter, that is p in $\ell_{2,p}$-norm. In this section, the performances of ExMVCNMF and GDCMVNMF are evaluated against these parameters on some selected datasets to test the influence of the abovementioned parameters in the proposed methods.

The performances of GDCMVNMF and ExMVCNMF on ORL, FEI part 1, ECG, and WebKB datasets, with the labeling ratios of 10%, 20%, and 30%, versus parameters α, β, and γ are illustrated in Figs. 4–7, respectively: α controls the discriminative abilities of the proposed methods, thus the AC reduces with the decreasing α; β balances the influence of the geometric information regularizing term. When α is small, β should be neatly selected to retain a relatively good performance. However, when α is large enough, the influence of β is relatively minimized. Therefore, this results in a "ridge" when β is small and a "plateau" when β is large. This implies that the geometric information becomes important with rarer discriminative information; moreover, when the discriminative information is adequately explored, the influence of the geometric information becomes insignificant.

The above facts can also be determined from a different viewpoint. From Fig. 6, it can be observed that as the area of the "plateau" enlarges with the increasing labeling ratioes in GDCMVNMF and

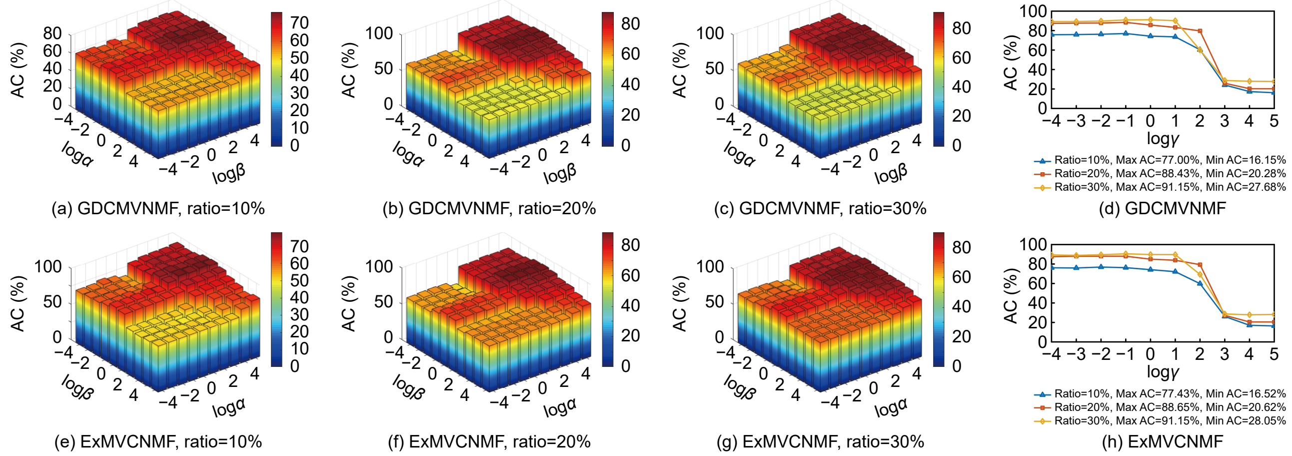


Fig. 4 GDCMVNMF and ExMVCNMF vs. parameters α, β, and γ on ORL dataset (for GDCMVNMF, p = 2).

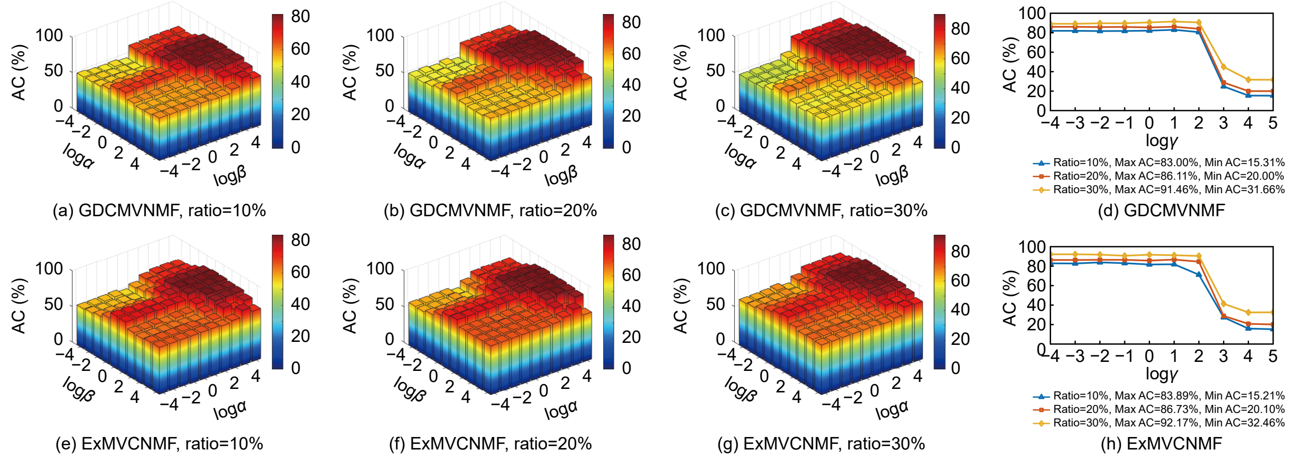


Fig. 5 GDCMVNMF and ExMVCNMF vs. parameters α, β, and γ on FEI part 1 dataset (for GDCMVNMF, p = 2).

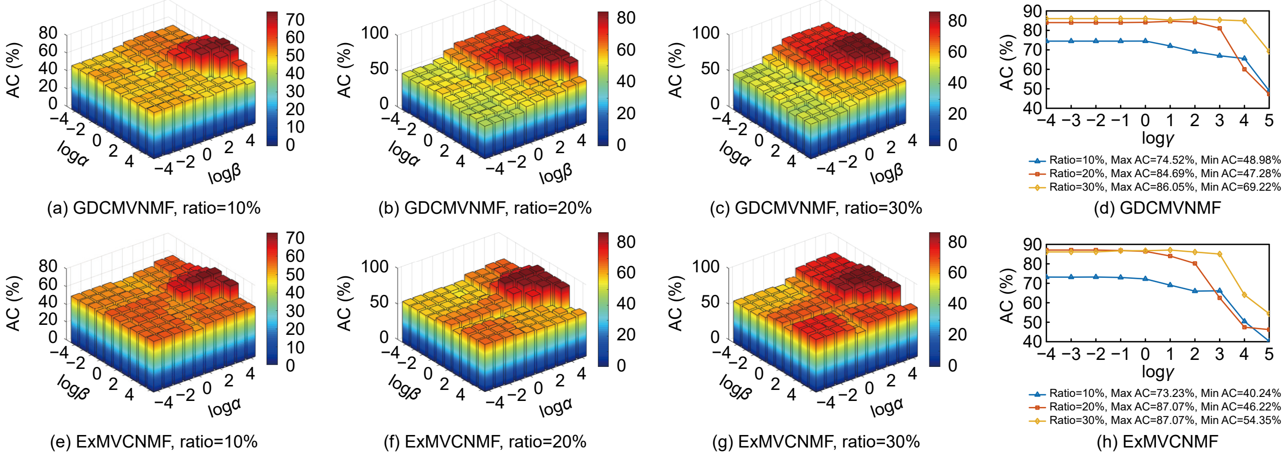


Fig. 6　GDCMVNMF and ExMVCNMF vs. parameters α, β, and γ on ECG dataset (for GDCMVNMF, $p = 2$).

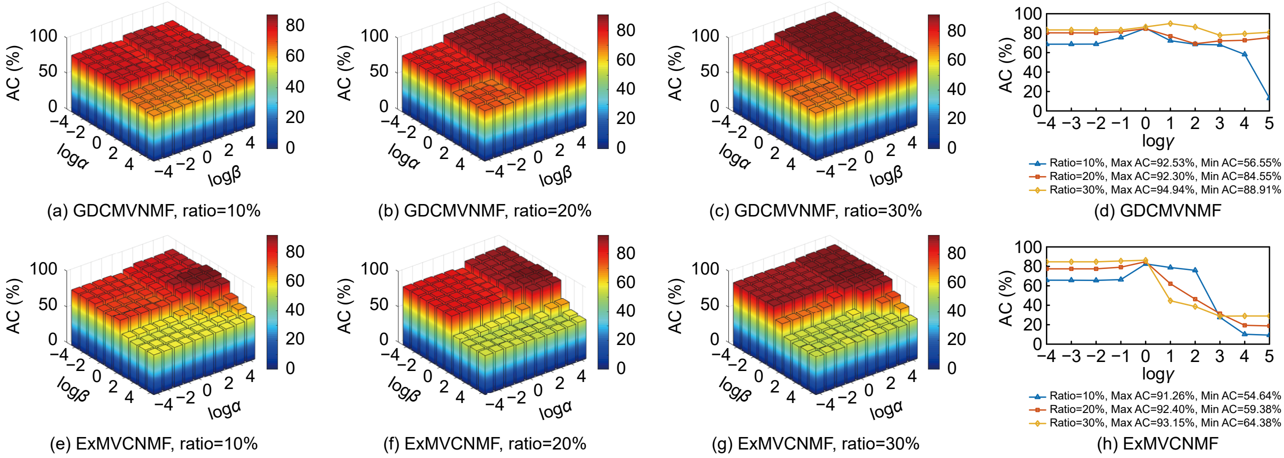


Fig. 7　GDCMVNMF and ExMVCNMF vs. parameters α, β, and γ on WebKB dataset (for GDCMVNMF, $p = 2$).

ExMVCNMF. This indicates that the influence of the geometric information (β) minimizes with the increasing “power” of the discriminative information (α). Similarly, from Figure 8, it can be observed that when the labeling ratio is low, the number of k-NNs should be carefully selected to obtain a better performance. Figure 8 presents a special case of GDCMVNMF with no label information, i.e., ratio = 0%. When the labeling ratio increases, the performance of the proposed methods become relatively stable; that is, the fluctuation of the curves becomes smaller. From the last columns of Figs. 4–7, it can be observed that the performances of these methods degrade when γ is set very large. This is because if the value of γ is set too high, the effect of this term will overwhelm the other terms, which may cause the proposed methods not to effectively use the label information (block-diagonal structure term) and the geometric information (graph regularization term) of the data.

The influence of the p-norm on GDCMVNMF is tested on ORL and ECG datasets. Here, p varies in the range of $\{0.5, 1, 1.5, 2, 2.1, 2.2\}$. Although, the authors of RSNMF[42] claimed that when $p = 0.5$, RSNMF is more robust to noises. From Fig. 9, it can be observed that the performance of GDCMVNMF is relatively stable in the entire range of p. This maybe because, in multi-view settings, the different views can provide complementary information for each other, which enhances the robustness of the method to some extent.

6　Conclusion

This study proposes a general discriminatively constrained S^{2}MVNMF with a novel feature alignment strategy. In this algorithm, the discriminative information on the multi-view data is effectively explored. Two specific implementations of this algorithm are presented along with their detailed optimizing procedure, i.e., GDCMVNMF and ExMVCNMF. GDCMVNMF with $p = 2$ reduces to

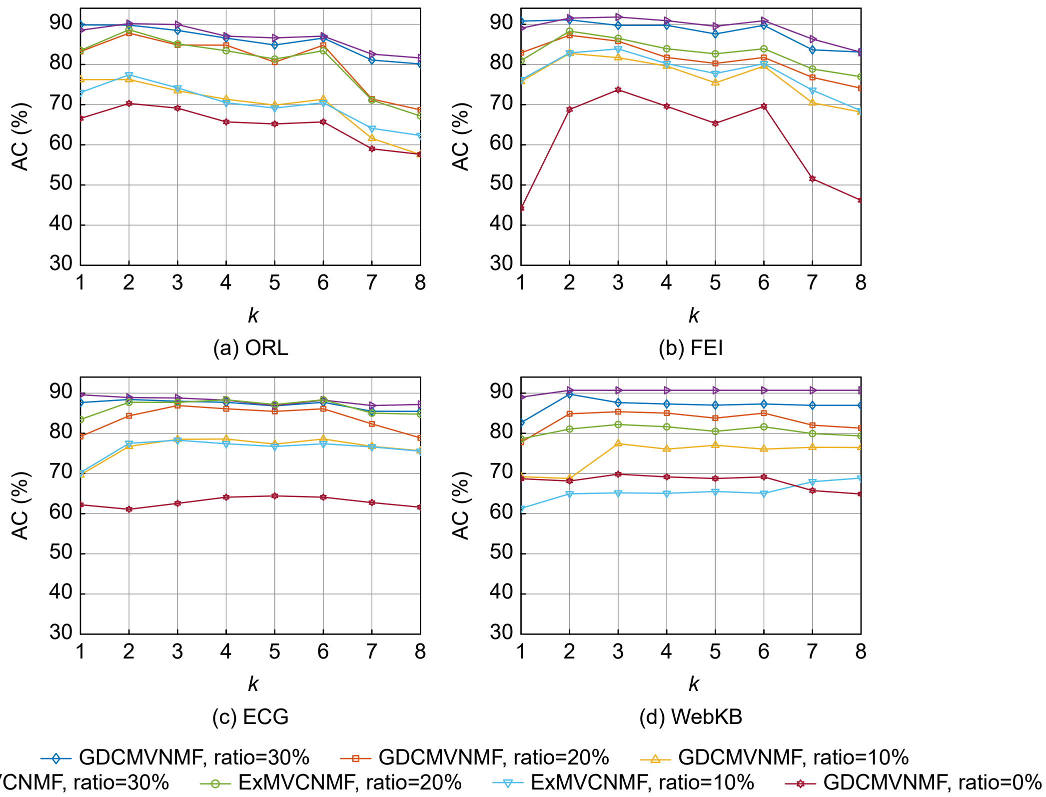


Fig. 8 GDCMVNMF and ExMVCNMF vs. parameter k on ORL, FEI part 1, ECG, and WebKB datasets.

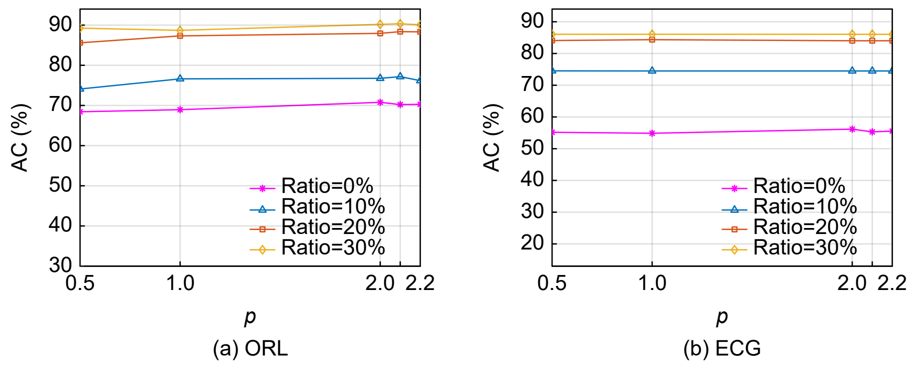


Fig. 9 GDCMVNMF vs. parameter p on ORL and ECG datasets with different labeling ratios, i.e., 0%, 10%, 20%, and 30%.

ExMVCNMF by extending the label constraint matrix in GDCMVNMF as an identity matrix. The experimental results show that in most cases, ExMVCNMF outperforms GDCMVNMF (especially when $p = 2$). This indicates that intra-class compactness plays a crucial role in discovering the discriminative information of data. ExMVCNMF has retained the CNMF property; hence it imposes tighter intra-class compactness constraints than GDCMVNMF. The influence of $\ell_{2,p}$-norm to the model in the experiments is also explored under the same configuration. In most cases, the performance of GDCMVNMF is better when p is set to 1 or 2. The superiority of the presented methods has been validated by comparing them with several representative works on six real-world multi-view datasets.

Acknowledgment

This work was supported by the National Key Research and Development Project of China (No. 2019YFB2102500), the Strategic Priority CAS Project (No. XDB38040200), the National Natural Science Foundation of China (Nos. 62206269 and U1913210), the Guangdong Provincial Science and Technology Projects

(Nos. 2022A1515011217 and 2022A1515011557), and the Shenzhen Science and Technology Projects (No. JSGG20211029095546003).

References

[1] H. Liang, C. Jiang, D. Feng, X. Chen, H. Xu, X. Liang, W. Zhang, Z. Li, and L. Van Gool, Exploring geometry-aware contrast and clustering harmonization for self-supervised 3D object detection, in *Proc. IEEE/CVF Int. Conf. Computer Vision*, Montreal, Canada, 2021, pp. 3293–3302.

[2] J. Yin and S. Sun, Incomplete multi-view clustering with cosine similarity, *Patt. Recogn.*, vol. 123, p. 108371, 2022.

[3] D. Wu, X. Dong, F. Nie, R. Wang, and X. Li, An attention-based framework for multi-view clustering on Grassmann manifold, *Patt. Recogn.*, vol. 128, p. 108610, 2022.

[4] Y. Chen, X. Xiao, C. Peng, G. Lu, and Y. Zhou, Low-rank tensor graph learning for multi-view subspace clustering, *IEEE Trans. Circ. Syst. Video Technol.*, vol. 32, no. 1, pp. 92–104, 2022.

[5] S. Zhou, X. Liu, J. Liu, X. Guo, Y. Zhao, E. Zhu, Y. Zhai, J. Yin, and W. Gao, Multi-view spectral clustering with optimal neighborhood Laplacian matrix, in *Proc. AAAI Conf. Artificial Intelligence*, Palo Alto, CA, USA, pp. 6965–6972, 2020.

[6] Y. Li, M. Yang, and Z. Zhang, A survey of multi-view representation learning, *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, 2019.

[7] X. Yan, Y. Ye, X. Qiu, and H. Yu, Synergetic information bottleneck for joint multi-view and ensemble clustering, *Inform. Fusion*, vol. 56, pp 15–27, 2020.

[8] B. Cui, H. Yu, T. Zhang, and S. Li, Self-weighted multi-view clustering with deep matrix factorization, in *Proc. Eleventh Asian Conf. Machine Learning*, Nagoya, Japan, 2019, pp. 567–582.

[9] J. Xu, H. Tang, Y. Ren, L. Peng, X. Zhu, and L. He, Multi-level feature learning for contrastive multi-view clustering, in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 16051–16060.

[10] J. Han, J. Xu, F. Nie, and X. Li, Multi-view k-means clustering with adaptive sparse memberships and weight allocation, *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 816–827, 2022.

[11] N. Zhao and J. Bu, Robust multi-view subspace clustering based on consensus representation and orthogonal diversity, *Neural Netw.*, vol. 150, pp. 102–111, 2022.

[12] S. Wei, J. Wang, G. Yu, C. Domeniconi, and X. Zhang, Multi-view multiple clusterings using deep matrix factorization, in *Proc. AAAI Conf. Artificial Intelligence*, Palo Alto, CA, USA, vol. 34, pp. 6348–6355, 2020.

[13] J. Liu, F. Cao, and J. Liang, Centroids-guided deep multi-view *K*-means clustering, *Inform. Sci.*, vol. 609, pp. 876–896, 2022.

[14] X. Si, Q. Yin, X. Zhao, and L. Yao, Consistent and diverse multi-view subspace clustering with structure constraint, *Patt. Recogn.*, vol. 121, p. 108196, 2022.

[15] L. Hu, N. Wu, and X. Li, Feature nonlinear transformation non-negative matrix factorization with Kullback-Leibler divergence, *Patt. Recogn.*, vol. 132, p. 108906, 2022.

[16] J. Liu, C. Wang, J. Gao, and J. Han, Multi-view clustering via joint nonnegative matrix factorization, in *Proc. 2013 SIAM Int. Conf. Data Mining*, Philadelphia, PA, USA, 2013, pp. 252–260.

[17] G. Cui, R. Wang, D. Wu, and Y. Li, Incomplete multiview clustering using normalizing alignment strategy with graph regularization, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 8126–8142, 2023.

[18] G. Chao, S. Sun, and J. Bi, A survey on multiview clustering, *IEEE Trans. Artif. Intell.*, vol. 2, no. 2, pp. 146–168, 2021.

[19] G. A. Khan, J. Hu, T. Li, B. Diallo, and H. Wang, Multi-view data clustering via non-negative matrix factorization with manifold regularization, *Int. J. Mach. Learn. Cybernet.*, vol. 13, no. 3, pp. 677–689, 2022.

[20] X. Wang, T. Zhang, and X. Gao, Multiview clustering based on non-negative matrix factorization and pairwise measurements, *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3333–3346, 2019.

[21] Z. Yang, N. Liang, W. Yan, Z. Li, and S. Xie, Uniform distribution non-negative matrix factorization for multiview clustering, *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3249–3262, 2021.

[22] Z. Yang, Y. Xiang, K. Xie, and Y. Lai, Adaptive method for nonsmooth nonnegative matrix factorization, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 948–960, 2017.

[23] J. Wang, F. Tian, H. Yu, C. H. Liu, K. Zhan, and X. Wang, Diverse non-negative matrix factorization for multiview data representation, *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2620–2632, 2018.

[24] F. Chen, G. Li, S. Wang, and Z. Pan, Multiview clustering via robust neighboring constraint nonnegative matrix factorization, *Math. Probl Eng.*, vol. 2019, p. 6084382, 2019.

[25] G. Du, L. Zhou, K. Lü, and H. Ding, Deep multiple non-negative matrix factorization for multi-view clustering, *Intell. Data Anal.*, vol. 25, no. 2, pp. 339–357, 2021.

[26] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, A deep matrix factorization method for learning attribute representations, *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, 2017.

[27] H. Zhao, Z. Ding, and Y. Fu, Multi-view clustering via deep matrix factorization, in *Proc. Thirty-First AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2017, pp. 2921–2927.

[28] S. Huang, Z. Kang, and Z. Xu, Auto-weighted multi-view clustering via deep matrix decomposition, *Patt. Recogn.*, vol. 97, p. 107015, 2020.

[29] K. Luong, R. Nayak, T. Balasubramaniam, and M. A. Bashar, Multi-layer manifold learning for deep non-negative matrix factorization-based multi-view clustering, *Patt. Recogn.*, vol. 131, p.108815, 2022.

[30] Y. Jiang, J. Liu, Z. Li, and H. Lu, Semi-supervised unified latent factor learning with multi-view data, *Mach. Vision*

Appl., vol. 25, no. 7, pp. 1635–1645, 2014.

[31] J. Liu, Y. Jiang, Z. Li, Z. H. Zhou, and H. Lu, Partially shared latent factor learning with multiview data, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1233–1246, 2015.

[32] N. Liang, Z. Yang, Z. Li, S. Xie, and C. Y. Su, Semi-supervised multi-view clustering with graph-regularized partially shared non-negative matrix factorization, *Knowl. Based Syst.*, vol. 190, p. 105185, 2020.

[33] J. Wang, X. Wang, F. Tian, C. H. Liu, H. Yu, and Y. Liu, Adaptive multi-view semi-supervised nonnegative matrix factorization, in *Proc. 23^{rd} Int. Conf. Neural Information Processing*, Kyoto, Japan, 2016, pp. 435–444.

[34] H. Cai, B. Liu, Y. Xiao, and L. Lin, Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization, *Knowl. Based Syst.*, vol. 182, p. 104798, 2019.

[35] H. Cai, B. Liu, Y. Xiao, and L. Lin, Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization, *Inform. Sci.*, vol. 536, pp. 171–184, 2020.

[36] S. Wang, J. Cao, F. Lei, Q. Dai, S. Liang, and B. W. K. Ling, Semi-supervised multi-view clustering with weighted anchor graph embedding, *Comput. Intell. Neurosci.*, vol. 2021, p. 4296247, 2021.

[37] F. Nie, G. Cai, J. Li, and X. Li, Auto-weighted multi-view learning for image clustering and semi-supervised classification, *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, 2018.

[38] N. Liang, Z. Yang, Z. Li, S. Xie, and W. Sun, Semi-supervised multi-view learning by using label propagation based non-negative matrix factorization, *Knowl. Based Syst.*, vol. 228, p. 107244, 2021.

[39] W. Zhao, C. Xu, Z. Guan, and Y. Liu, Multiview concept learning via deep matrix factorization, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 814–825, 2021.

[40] R. Chen, Y. Tang, W. Zhang, and W. Feng, Deep multi-view semi-supervised clustering with sample pairwise constraints, *Neurocomputing*, vol. 500, pp. 832–845, 2022.

[41] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, Constrained nonnegative matrix factorization for image representation, *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1299–1311, 2012.

[42] Z. Li, J. Tang, and X. He, Robust structured nonnegative matrix factorization for image representation, *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1947–1960, 2018.

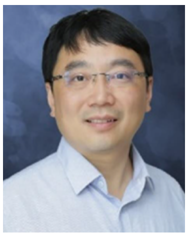
[43] H. Huang, Y. Luo, G. Zhou, and Q. Zhao, Multi-view data representation via deep autoencoder-like nonnegative matrix factorization, in *IEEE Int. Conf. Acoustics, Speech and Signal Processing* (*ICASSP*), Singapore, 2022, pp. 3338–3342.

[44] D. Cai, X. He, X. Wu, and J. Han, Non-negative matrix factorization on manifold, in *Proc. 8th IEEE Int. Conf. Data Mining*, Pisa, Italy, 2008, pp. 63–72.

[45] G. Cui and Y. Li, Nonredundancy regularization based nonnegative matrix factorization with manifold learning for multiview data representation, *Informa. Fusion*, vol. 82, pp. 86–98, 2022.

[46] M. You, A. Yuan, M. Zou, D. Jian He, and X. Li, Robust unsupervised feature selection via multi-group adaptive graph representation, *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 3030–3044, 2023.

Guosheng Cui received the PhD degree from University of Chinese Academy of Sciences, Beijing, China in 2018. He is currently an assistant research fellow at SIAT, CAS, Shenzhen, China. His current research interests include artificial intelligence and machine learning.



Ye Li is currently a full professor at Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), China. He received the PhD degree in electrical engineering from Arizona State University, AZ, USA in 2006. He has been working at CAS since 2008. His research interests include medical big data, artificial intelligence, and health informatics. He has published more than 150 papers in prestigious journals and conferences, like *Information Fusion*, *AI in Medicine, IEEE Journal on Selected Areas in Communications*, *IEEE Internet of Things Journal*, *IEEE Journal of Biomedical and Health Informatics*, etc. He has served as an editorial board member of *Information Fusion* since 2018, the program chair of IEEE UIC 2021, etc.

Jianzhong Li is a chair professor at SIAT, CAS, China, and a professor at Harbin Institute of Technology, China. His current research interests include big data computation and wireless sensor networks. He has published more than 400 papers in refereed journals and conference proceedings, such as *VLDB Journal*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Parallel and Distributed Systems*, SIGMOD, VLDB, ICDE, and INFOCOM. His papers have been cited more than 20 000 times and his H-index is 65. He has been involved in the program committees of major computer science and technology conferences, including SIGMOD, VLDB, ICDE, and INFOCOM. He has also served on the editorial boards for distinguished journals, such as *IEEE Transactions on Knowledge and Data Engineering*.

Jianping Fan is an academician of the International Eurasian Academy of Sciences, the president of SIAT, CAS, China. He received the PhD degree from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China in 1990. He was elected as a fellow of China Computer Federation (CCF) in 2013. He is mainly engaged in the research of high-performance computers, cloud computing, and parallel and distributed computing. He is one of the founders of Dawning high-performance computer, and presided over the research of "Dawning No. 1", "Dawning 1000", and a series of Dawning scalable parallel computer systems. He has published 4 monographs and 159 papers, and possesses 81 patents.