Neurocomputing 501 (2022) 436–450

Contents lists available at ScienceDirect

Neurocomputing

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

Geometrical structure preservation joint with self-expression maintenance for adaptive graph learning

Yangbo Wang [a], Can Gao [a,b,c,*], Jie Zhou [a,b,c]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, PR China
[b] SZU Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518060, PR China
[c] Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, PR China

ARTICLE INFO

Article history:
Received 23 October 2021
Revised 18 May 2022
Accepted 11 June 2022
Available online 18 June 2022
Communicated by Zidong Wang

Keywords:
Clustering
Geometrical structure preservation
Self-expressive learning
Adaptive graph

ABSTRACT

Locality preserving projection (LPP) is a subspace learning method that uses pairwise distance to measure the similarity between data points. However, when data points from different clusters are adjacent, the pairwise distance may not accurately reflect the similarity between data points. In this study, a novel graph learning model is proposed to alleviate this problem; it is called adaptive graph learning with geometrical structure preservation and self-expression maintenance (GEAGL), in which a discriminative LPP method is used to extract the geometrical structure of data. By integrating self-expressive learning into the LPP method, a similarity graph that preserves the geometrical structure and self-expressive properties of data can be adaptively learned. The learned similarity graph tends to consist of a block diagonal structure when data points are extracted from independent linear subspaces, which alleviates the cluster interweaving problem. In this study, the relationship between the proposed method and k-means clustering was revealed, and the geometrical structure preservation property of the proposed method was theoretically analyzed. Finally, a graph-based clustering method was developed based on the similarity graph produced by GEAGL. Experimental results of benchmark datasets demonstrate the superiority of the proposed method in comparison with state-of-the-art methods.

1. Introduction

In the field of machine learning and pattern recognition, we often face scenarios in which high-dimensional data is intrinsically derived from a low-dimensional structural space [1,2]. It is highly desirable to determine a mapping method that can simultaneously transform high-dimensional data into a low-dimensional space and preserve the inherent structural information of the data to achieve dimensionality reduction and facilitate data visualization. Manifold learning is one of the most commonly used and popular methods to resolve this problem [3–11]. Linear manifold learning methods, including locality preserving projection (LPP) [12] and neighborhood preserving embedding (NPE) [13], attempt to determine a projection matrix to project data linearly from a manifold-embedded high-dimensional space to a low-dimensional space. Nonlinear manifold learning methods, such as locality linear embedding (LLE) [14], Laplacian eigenmaps [1], spectral clustering [15], and nonnegative matrix factorization-based methods [16–18], etc, aim to nonlinearly map data from a manifold-embedded high-dimensional space to a low-dimensional space. In addition, graph-based manifold learning has also received extensive attention. For example, in [19], a multiview graph clustering method was proposed, which also learns a data with clear structure. In [20], graph learning is incorporated into the framework of dimensionality reduction. In [21], the equivalence between graph embedding and matrix factorization is established.

LPP is a manifold learning method that preserves the geometrical structure of data in the projected low-dimensional subspace [12]. He et al. [12] demonstrated that LLP is an optimal linear approximation of the eigenfunctions of the Laplace–Beltrami operator on the manifold. Thus, the geometrical structural information of the data can be sufficiently retained by the LPP. That is, adjacent data points in the manifold-embedded high-dimensional space are also close to one another in the low-dimensional projection space. In the past two decades, LPP and its extensions have been widely used in unsupervised clustering [22–24], feature selection [25,26,6,5,27,28], and multi-view feature selection [29]. In [22], LPP was integrated into clustering with adaptive neighbors (CAN) to adaptively model the geometrical structure of the data. In [23,24], LPP was used to improve the applicability of fuzzy C-

* Corresponding author at: College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, PR China.
E-mail address: 2005gaocan@163.com (C. Gao).

https://doi.org/10.1016/j.neucom.2022.06.045
0925-2312/

means clustering for high-dimensional data. In [25,26], a feature selection framework that unifies LPP and low-rank linear regression was presented to explore the latent relations between neighbors. In [6], a discriminative LPP method was presented to preserve inter-class information. In [27], LPP was integrated into a low-rank representation model to preserve the local and global structure of the data. In [28], a new feature selection method that considers the local manifold structure was proposed by constraining the representation coefficients and projected representations of data points that are close to each other in the original data space to be similar. In [29], a new multi-view unsupervised feature selection method based on the idea of distance relationship preservation was proposed to consider the consistency and diversity information of each view. All these methods aim to construct a projected subspace to obtain the inherent manifold structure of the data. However, the similarity between the data points in the aforementioned LPP and its extensions was only measured based on the pairwise distance. When data points from different clusters are close, the pairwise distance may not accurately reflect the intrinsic similarity between data points.

One potential way to alleviate this problem is to explore the self-representative property of data [30]. The data self-expression method [30] assumes that each data point in a set of subspaces can be properly reconstructed using other points. Elhamifar et al. [31,32] indicated that each data point can always be linearly reconstructed with the points in its subspace and that the similarity matrix formed by the reconstruction coefficient matrix has a block diagonal structure when the subspaces are linear and independent. Thus, many unsupervised clustering methods based on self-representative learning have been proposed to utilize block diagonal structures [31,30,33–37,32,38–40]. Elhamifar et al. [31,30] introduced sparse subspace clustering (SSC) by using a l_1-minimization program to optimize the sparse reconstruction coefficient matrix. Lu et al. [34] obtained a similarity matrix with a strict block diagonal structure, despite the subspace independence assumption not being fully satisfying. Yang et al. [37] improved SSC by optimizing the sparsity and connectivity of the similarity matrix. Zhang et al. [35] proposed a new subspace clustering method by simultaneously considering the self-expressive and local information of the data. Elhamifar et al. [32] further presented the clustering of disjoint subspaces via sparse representation. Kang et al.[38] extended the linear SSC learning method into a nonlinear method to manage data with complex structures. Kang et al.[39] introduced the idea of anchor points to address the out-of-sample problem in traditional SSC methods. In [40], an improved SSC that can explore high-order information was proposed by designing a novel regularizer. All these methods aim to construct a similarity matrix based on the self-expressive property. Although they provide a reference for utilizing the block diagonal structure of the similarity matrix (or expressive coefficient matrix) to address the problem of overlapping data clustering, manifold structure information is not fully exploited.

We propose an adaptive graph learning model, called adaptive graph learning with geometrical structure preservation and self-expression maintenance (GEAGL), to effectively measure the similarity of overlapping cluster data points and fully capture the manifold structure of the data. In this study, a discriminative LPP method was used to model the geometrical structure of the data in the proposed model. By further integrating self-expressive learning into the LPP method, a similarity graph that preserves the geometrical structure and self-expressive properties of the data can be adaptively learned. The learned similarity graph tends to have a block diagonal structure because the self-expressive property is used in the proposed method. The main contributions of this study are as follows:

(1) A new adaptive graph learning model, adaptive graph learning with geometrical structure preservation and self-expression maintenance (GEAGL), was proposed. In the proposed model, the self-expressive property of the data is utilized to recover the block diagonal structure of the similarity matrix. Consequently, the similarity of data points can be measured more accurately, especially when the clusters overlap.
(2) The GEAGL is theoretically analyzed from two aspects: 1) the relationship between the GEAGL model and k-means, and 2) the connection between the generated subspace representation in the proposed method and the eigenfunctions of the Laplace–Beltrami operator on the manifold.
(3) An efficient algorithm was presented to optimize the proposed GEAGL model. A graph-based clustering method was developed based on the similarity graph obtained using GEAGL. Extensive experiments were performed to verify the effectiveness of the proposed model, which presented promising results.

The remainder of this paper is organized as follows. In Section 2, we introduce the related notations and methods. In Section 3, we elaborate on the proposed GEAGL and the corresponding optimization algorithm as well as a clustering method based on GEAGL. The experimental results are presented in Section 4. Finally, Section 5 concludes the study.

2. Related works

2.1. Notations

In this study, all matrices and vectors are indicated by bold uppercase and lowercase letters, respectively, whereas scalars are written in lowercase letters. For example, $\mathbf{A}$ denotes a matrix, whereas $\mathbf{a}$ and a represent a vector and scalar, respectively. $\|\cdot\|_2$ and $\|\cdot\|_F$indicate the L_2 norm of a vector and Frobenius norm of a matrix, respectively, and $Tr(\cdot)$ denotes the trace of a matrix.

2.2. LPP

LPP [12] is a manifold learning method commonly used for dimension reduction. A similarity matrix or nearest neighbor graph was first constructed to capture the geometrical structure of the data [3]. Subsequently, a linear projection matrix that can map the manifold-embedded high-dimensional data to a low-dimensional space was obtained to ensure that the adjacent points in the original space were also close in the projected space. LPP aims to determine an optimal projection matrix that minimizes the following objective function [12]:

$$\min_{\mathbf{W}} \sum_i \sum_j ||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}, \quad \text{s.t.} \quad \mathbf{W}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{W} = \mathbf{I}, \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \Re^{d\times n}$ is the given data matrix, and d and n are the number of features and data points, respectively. $\mathbf{W} \in \Re^{d\times r}$is a projection matrix, r is the dimension of the projection subspace, s_{ij} is the similarity between data points $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathbf{D} \in \Re^{n\times n}$is a diagonal matrix with the ith diagonal element $D_{ii} = \sum_j s_{ij}$, and $\mathbf{I}$ is an identity matrix.

2.3. Sparse subspace clustering

SSC [31,30] is an effective unsupervised clustering method. First, a reconstruction coefficient matrix is constructed by exploiting the self-expressive property of the data. The generated coefficient matrix is then introduced into a spectral clustering method to group each data point. SSC constructs the coefficient matrix by reconstructing each data point as a linear combination of the data

points in its subspace. The reconstruction problem in the SSC can be described as follows:

$$\min_{\mathbf{s}} ||\mathbf{s}||_1, \quad \text{s.t.} \quad \mathbf{x} = \mathbf{X}\mathbf{s}, \tag{2}$$

where $\mathbf{s} \in \Re^{n\times 1}$ is the reconstruction coefficient vector used to reconstruct data point $\mathbf{x}$. When the data points contain noise, (2) can be rewritten as follows:

$$\min_{\mathbf{s}} ||\mathbf{x} - \mathbf{X}\mathbf{s}||_2^2 + \varphi||\mathbf{s}||_1, \tag{3}$$

where φ is a balance parameter.

3. The proposed method

LPP uses pairwise distance to measure the similarity between data points. However, when data points from different clusters are intertwined, the pairwise distance may not accurately reflect the similarity of the data points.

As shown in Fig. 1, there are two clusters indicated by red and blue; data points $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to different clusters, and $\mathbf{x}_1$ is the nearest neighbor of data point $\mathbf{x}_2$. When the reconstruction coefficient is non-negative, the data points that can be used to reconstruct the data point $\mathbf{x}_1$ must be from Cluster1. However, the data point used to reconstruct $\mathbf{x}_2$ are from both Cluster1 and Cluster2. Therefore, the self-expression properties of $\mathbf{x}_1$ and $\mathbf{x}_2$ differ. For certain cases, two similar data points may have different self-expression properties.

Because no true labels can be provided beforehand, we consider utilizing the self-expressive property of the data to alleviate the problem of certain data points from different clusters that are adjacent. On one the hand, we use the LPP method to preserve the geometrical structure; on the other hand, we utilize the self-expressive property of the data to explore the block diagonal structure of the similarity matrix so that the similarity of data points belonging to different clusters is small, despite being close.

3.1. The objective function of GEAGL

To enhance the discriminative information between different clusters, we propose a new discriminative LPP method based on the existing work [6].

Let C_h be a set of data points that belong to the hth cluster and let n_h be the number of data points in the hth cluster.

For the LPP term $\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}$, let

$$s_{ij} = \begin{cases} \frac{1}{n_h}, & \mathbf{x}_i \in C_h \text{ and } \mathbf{x}_j \in C_h \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

minimizing $\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}$ ensures that the distance between the data points in the same cluster is as small as possible.

Moreover, when $\mathbf{x}_i$ belongs to the $h(i)$th cluster, maximizing

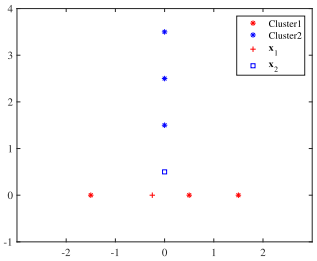

Fig. 1. Two data points that have different self-expression properties are adjacent.

$$\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 - \sum_i(n_{h(i)}\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}) \tag{5}$$

can ensure that the distances between different clusters are as far as possible.

Thus, we attempt to preserve the discriminative information between clusters by solving the following objective function:

$$\max_{\mathbf{W}} \frac{\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 - \sum_i(n_{h(i)}\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij})}{\sum_i(n_{h(i)}\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij})}. \tag{6}$$

And (6) is equal to

$$\max_{\mathbf{W}} \frac{\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2}{\sum_i(n_{h(i)}\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij})}. \tag{7}$$

Problem (7) can further become

$$\min_{\mathbf{W}} \frac{\sum_i(n_{h(i)}\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij})}{\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2}. \tag{8}$$

Because the problem of an imbalanced cluster is beyond the scope of this study, we simplify (8) as follows:

$$\min_{\mathbf{W}} \frac{\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}}{\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2}. \tag{9}$$

For (9), the following equation holds:

$$\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 = 2nTr(\mathbf{W}^T\mathbf{S}_t\mathbf{W}), \tag{10}$$

where $\mathbf{S}_t = \mathbf{X}\mathbf{H}\mathbf{X}^T \in \Re^{d\times d}$ is the total scatter matrix and $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T \in \Re^{n\times n}$ is the centering matrix. Thus, (9) becomes:

$$\min_{\mathbf{W}} \sum_i\sum_j \frac{||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}}{Tr(\mathbf{W}^T\mathbf{S}_t\mathbf{W})}. \tag{11}$$

Furthermore, (11) can be simplified as follows [6]:

$$\min_{\mathbf{W}} \sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij}, \quad \text{s.t.} \quad \mathbf{W}^T\mathbf{S}_t\mathbf{W} = \mathbf{I}. \tag{12}$$

In summary, under the effect of equality constraints $\mathbf{W}^T\mathbf{S}_t\mathbf{W} = \mathbf{I}$, our method enhances the discriminative information between different clusters by maximizing the distance between different clusters and minimizing the distance within the same cluster.

In (12), we measure the similarity between data points using pairwise distance. However, when certain data points from different clusters are close, the pairwise distance cannot fully reflect the similarity between the data points. Because no true labels can be provided beforehand, we consider utilizing the self-expressive property of the data. When each data point is from a set of c independent linear subspaces, a similarity matrix with a block diagonal structure can be constructed by solving (3). To ensure the translation invariance of (3), we obtain the following equation for $\mathbf{s}$ [41]:

$$||(\mathbf{x} + \mathbf{t}) - (\mathbf{X} + \mathbf{t}\mathbf{1}^T)\mathbf{s}||_2^2 = ||\mathbf{x} - \mathbf{X}\mathbf{s}||_2^2, \tag{13}$$

where $\mathbf{t} \in \Re^{d\times 1}$ denotes the translation constant. From (13), $\mathbf{s}^T\mathbf{1} = 1$ [41]. In addition, when the coefficient vector $\mathbf{s}$ is non-negative, it can be used to represent the similarity. Then, (3) becomes:

$$\min_{\mathbf{s}} ||\mathbf{x} - \mathbf{X}\mathbf{s}||_2^2 + \varphi||\mathbf{s}||_1, \quad \text{s.t.} \quad \mathbf{s}^T\mathbf{1} = 1, \mathbf{s} \geqslant \mathbf{0}. \tag{14}$$

Because the constraint terms $\mathbf{s}^T\mathbf{1}=1$ and $\mathbf{s}\geqslant\mathbf{0}$ guarantee the sparsity of the optimal solution $\mathbf{s}$ in (14) [41], (14) can be simplified as follows:

$$\min_{\mathbf{s}}||\mathbf{x}-\mathbf{X}\mathbf{s}||_2^2+\varphi||\mathbf{s}||_2^2,\quad \text{s.t.}\quad \mathbf{s}^T\mathbf{1}=1,\mathbf{s}\geqslant\mathbf{0}. \tag{15}$$

As indicated by [3], high-dimensional data often lie in a very low-dimensional space. We consider recovering the hidden low-dimensional structure with the help of a projection matrix $\mathbf{P}$, such that redundant features can be reduced.

$$\min_{\mathbf{s},\mathbf{P}}||\mathbf{P}^T\mathbf{x}-\mathbf{P}^T\mathbf{X}\mathbf{s}||_2^2+\varphi||\mathbf{s}||_2^2,\quad \text{s.t.}\quad \mathbf{s}^T\mathbf{1}=1,\mathbf{s}\geqslant\mathbf{0},\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I}, \tag{16}$$

where $\mathbf{P}\in\Re^{d\times r}$ is a projection matrix that is different from $\mathbf{W}$, and the constraint term $\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I}$ guarantees the uniqueness of the solution.

In (16), Elhamifar et al. [31] indicated that when n noise-free data points $\{\mathbf{P}^T\mathbf{x}_i\}_{i=1}^n$ lie in c subspaces and the c subspace is linear and independent, the reconstruction coefficient matrix $\mathbf{S}$ has an ideal block diagonal structure, and $s_{ij}=0$ when any two projected points $\mathbf{P}^T\mathbf{x}_i$ and $\mathbf{P}^T\mathbf{x}_j$ belong to different subspaces. Thus, by integrating (12) and (16), we obtain the following:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{P},\mathbf{S}}\sum_i\Big(\sum_j(||\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{x}_j||_2^2s_{ij})+\lambda||\mathbf{P}^T\mathbf{x}_i-\sum_j\mathbf{P}^T\mathbf{x}_js_{ij}||_2^2+\varphi||\mathbf{s}_i||_2^2\Big),\\ \text{s.t. }\mathbf{s}_i^T\mathbf{1}=1,s_{ij}\geqslant0,\mathbf{W}^T\mathbf{S}_t\mathbf{W}=\mathbf{I},\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I},\end{aligned} \tag{17}$$

where λ is a balanced parameter.

In the proposed model, the first term is the LPP model, which preserves the geometrical structure of the data. The second term is the self-expressive term, which is used to recover the block diagonal structure of the similarity graph when data points are from the c independent linear subspaces, so that the problem of cluster interweaving can be alleviated. In particular, the equation $s_{ij}=0$ holds when $\mathbf{P}^T\mathbf{x}_i$ and $\mathbf{P}^T\mathbf{x}_j$ belong to different subspaces; otherwise, the value of s_{ij} increases as the distance between $\mathbf{W}^T\mathbf{x}_i$ and $\mathbf{W}^T\mathbf{x}_j$ decreases. The third term is the regularized term, which ensures the uniqueness of the solution. In contrast to previous graph-based clustering methods, the similarity matrix $\mathbf{S}$ in our method is adaptively learned rather than constructed in advance.

Moreover, for the proposed model (17), two different projection subspaces are used to preserve the geometrical structure and self-expressive property of the data, respectively. The previous methods tend to share the same projection matrix in both terms. However, our paper constructs a projection subspace in each term. This is because the LPP term and the self-expressive term of model (17) linearly approximate the eigenfunction of the Laplacian Beltrami operator from two different aspects. The optimal projection matrices of the two terms are different even for the same similarity matrix. Setting $\mathbf{W}=\mathbf{P}$ directly may impair the ability of the model to preserve the geometrical structure and self-expressive properties of the data. Therefore, we use two different projection matrices to preserve the geometrical structure and self-expressive properties of the data.

3.2. Optimization

Because (17) is convex with respect to $\mathbf{P},\mathbf{W}$, and $\mathbf{S}$; it can be minimized using the following iterative method:

Update P: When fixing all variables except $\mathbf{P}$, (17) is simplified as follows:

$$\min_{\mathbf{P}}\sum_i||\mathbf{P}^T\mathbf{x}_i-\sum_j\mathbf{P}^T\mathbf{x}_js_{ij}||_2^2,\quad \text{s.t.}\quad \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I}.$$

When we denote $\mathbf{y}_i=\mathbf{P}^T\mathbf{x}_i$ and $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n]\in\Re^{r\times n}$, the following equation holds:

$$\begin{aligned}&\sum_i||\mathbf{P}^T\mathbf{x}_i-\sum_j\mathbf{P}^T\mathbf{x}_js_{ij}||_2^2\\ &=\sum_i||\mathbf{y}_i-\sum_j\mathbf{y}_js_{ij}||_2^2\\ &=\sum_i||\mathbf{y}_i-\mathbf{Y}\mathbf{s}_i||_2^2\\ &=Tr(\mathbf{Y}(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})^T\mathbf{Y}^T)\\ &=Tr(\mathbf{P}^T\mathbf{X}(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})^T\mathbf{X}^T\mathbf{P}),\end{aligned} \tag{19}$$

where $\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_n]\in\Re^{n\times n}$ is the similarity matrix.

Then, (18) can be rewritten as the following optimization problem:

$$\min_{\mathbf{P}}Tr(\mathbf{P}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{P}),\quad \text{s.t.}\quad \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I},$$

where $\mathbf{M}=(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})^T$.

The optimal solution $\mathbf{P}$ can be composed of the r eigenvectors of $(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{M}\mathbf{X}^T$ corresponding to the r smallest eigenvalues.

Update W: When all variables except $\mathbf{W}$, (17) are simplified as (12).

And for (12), the following equation holds:

$$\sum_i\sum_j||\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{x}_j||_2^2s_{ij}=2Tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}), \tag{21}$$

where $\mathbf{L}=\mathbf{D}-\mathbf{S}\in\Re^{n\times n}$ denotes the Laplacian matrix of the symmetric similarity matrix $\mathbf{S}$ and matrix $\mathbf{D}\in\Re^{n\times n}$ is a diagonal matrix with the ith diagonal element $D_{ii}=\sum_j s_{ij}$.

Thus, (20) can be rewritten as the following optimization problem:

$$\min_{\mathbf{W}}Tr(\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}),\quad \text{s.t.}\quad \mathbf{W}^T\mathbf{S}_t\mathbf{W}=\mathbf{I}. \tag{22}$$

The optimal solution $\mathbf{W}$ can be composed of the r eigenvectors of $(\mathbf{S}_t)^{-1}\mathbf{X}\mathbf{L}\mathbf{X}^T$ corresponding to the r smallest eigenvalues.

Update S: When fixing $\mathbf{P}$ and $\mathbf{W}$, (17) is simplified as follows:

$$\begin{aligned}\min_{\mathbf{s}}\sum_i\Big(\sum_j(||\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{x}_j||_2^2s_{ij})+\lambda||\mathbf{P}^T\mathbf{x}_i\\ -\sum_j\mathbf{P}^T\mathbf{x}_js_{ij}||_2^2+\varphi||\mathbf{s}_i||_2^2\Big),\quad \text{s.t.}\quad \mathbf{s}_i^T\mathbf{1}=1,s_{ij}\geqslant0.\end{aligned} \tag{23}$$

Due to (23) being independent for each $\mathbf{s}_i$, (23) can be rewritten as:

$$\min_{\mathbf{s}_i}\sum_j(||\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij})+\varphi||\mathbf{s}_i||_2^2+\lambda||\mathbf{P}^T\mathbf{x}_i -\sum_j\mathbf{P}^T\mathbf{x}_j s_{ij}||_2^2,\quad \text{s.t.}\ \ \mathbf{s}_i^T\mathbf{1}=1, s_{ij}\geqslant 0. \tag{24}$$

Because $\mathbf{s}_i^T\mathbf{1}=1$, for the third term in (24), the following equation holds:

$$\begin{aligned}&||\mathbf{P}^T\mathbf{x}_i-\sum_j\mathbf{P}^T\mathbf{x}_j s_{ij}||_2^2\\ &=||\mathbf{y}_i-[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n]\mathbf{s}_i||_2^2\\ &=||\mathbf{y}_i(\sum_j s_{ij})-[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n]\mathbf{s}_i||_2^2\\ &=|||[\mathbf{y}_i,\mathbf{y}_i,\ldots,\mathbf{y}_i]\mathbf{s}_i-[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n]\mathbf{s}_i||_2^2\\ &=\mathbf{s}_i^T\mathbf{A}_i^T\mathbf{A}_i\mathbf{s}_i,\end{aligned} \tag{25}$$

where $\mathbf{A}_i=\mathbf{y}_i\mathbf{1}^T-\mathbf{Y}$.

According to (25), (24) can be transformed into

$$\min_{\mathbf{s}_i}\mathbf{s}_i^T(\lambda\mathbf{A}_i^T\mathbf{A}_i+\varphi\mathbf{I})\mathbf{s}_i+\mathbf{s}_i^T\mathbf{d}_i,\quad \text{s.t.}\ \ \mathbf{s}_i^T\mathbf{1}=1, s_{ij}\geqslant 0, \tag{26}$$

where $\mathbf{d}_i\in\Re^{n\times 1}$ is the distance vector between data point $\mathbf{x}_i$ and other data points with $d_{ij}=||\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{x}_j||_2^2$.

Problem (26) is a quadratic programming problem, that can be solved by using the primal dual interior point method [42].

The aforementioned procedure for solving the proposed model can be described by the following algorithm.

Algorithm 1: Geometrical Structure Preservation Joint with Self-Expression Maintenance for Adaptive Graph Learning Algorithm (GEAGL)

Input: Data matrix $\mathbf{X}=[x_1,x_2,\ldots,x_n]\in\Re^{d\times n}$, the number of clusters c, the projection dimension r, the parameter λ, the number of neighbors k, the number of iterations T;
Output: The similarity matrix $\mathbf{S}\in\Re^{n\times n}$;
1: Initialize $\mathbf{S}$ and φ;
2: Initialize $\mathbf{P}$ by (20);
3: Initialize $\mathbf{W}$ by (22);
4: **while**not converge and number of iterations $t\leqslant T$**do**
5: For each i, update the ith row of $\mathbf{S}$ by solving (26);
6: $\mathbf{S}=(\mathbf{S}+\mathbf{S}^T)/2$;
7: Update $\mathbf{P}$ by solving (20);
8: Update $\mathbf{W}$ by solving (22);
9: **end while**
10: **return**: $\mathbf{S}$.

3.3. Convergence and complexity analysis

In Algorithm 1, the optimal solutions for $\mathbf{P},\mathbf{W}$, and $\mathbf{S}$ can be obtained by solving problems (20), (22), and (26). Then,

$$\begin{aligned}&J_{GEAGL}(\mathbf{P}^{(t)},\mathbf{W}^{(t)},\mathbf{S}^{(t)})\\ &\leqslant J_{GEAGL}(\mathbf{P}^{(t-1)},\mathbf{W}^{(t)},\mathbf{S}^{(t)})\\ &\leqslant J_{GEAGL}(\mathbf{P}^{(t-1)},\mathbf{W}^{(t-1)},\mathbf{S}^{(t)})\\ &\leqslant J_{GEAGL}(\mathbf{P}^{(t-1)},\mathbf{W}^{(t-1)},\mathbf{S}^{(t-1)}),\end{aligned} \tag{27}$$

where $t-1$ and t represent the $(t-1)$th and tth iteration, respectively. Inequality (27) indicates that the proposed GEAGL model gradually converges during the iterative process.

Let T be the maximum number of iterations for Algorithm 1. The time complexity of updating $\mathbf{P}$ and $\mathbf{W}$ is at most $O\left(T\left(dn^2+d^3+d^2n\right)\right)$ and $O\left(T\left(dn^2+d^3+d^2n\right)\right)$, respectively. By contrast, the complexity of updating $\mathbf{S}$ is $O(nTT_1)$ if $\mathbf{S}$ is sparse and $||\mathbf{s}||_1\leqslant 0$, where T_1 is a constant required to solve quadratic programming with k variables. Therefore, the overall computational complexity of Algorithm 1 is approximately $O\left(T\left(dn^2+d^3+d^2n+nT_1\right)\right)$.

3.4. A clustering based on GEAGL

In the GEAGL model, we can obtain similarity matrix $\mathbf{S}$. The obtained graph structure can be used to complete the clustering. There are generally two types of clustering frameworks based on similarity graphs: one constrains the number of connected components in the similarity graph (each connected component represents a cluster) [22,43], and the other exploits the spectral clustering framework [15]. Considering that the latter often requires k-means as post-processing and k-means are sensitive to initialization, we selected the former. By integrating the technique of constrained Laplacian rank (CLR) [43], we proposed constrained Laplacian rank clustering based on GLEAL (GEALGL-CLR). The objective function is expressed as follows [43]:

$$\min_{\mathbf{A}}||\mathbf{A}-\mathbf{S}||_2^2,\quad \text{s.t.}\ \ \mathbf{a}_i^T\mathbf{1}=1, a_{ij}\geqslant 0, rank(\mathbf{L}_A)=n-c, \tag{28}$$

where a_{ij} is the redefined similarity between data points $\mathbf{x}_i$ and $\mathbf{x}_j$ that satisfies the constrained optimization.

In (28), the similarity matrix $\mathbf{A}$ is symmetric and non-negative, which satisfies the following Theorem 1 [44]:

Theorem 1. *If the similarity matrix* $\mathbf{A}$ *is symmetric and non-negative, the multiplicity* c *of the eigenvalue 0 of the Laplacian matrix* $\mathbf{L}_A$ *is equal to the number of connected components in the graph with the similarity matrix* $\mathbf{A}$.

According to Theorem 1, the constraint term $rank(\mathbf{L}_A)=n-c$ ensures that the number of connected components in $\mathbf{A}$ is equal to the number of clusters [22,43]. By utilizing Ky Fan's Theorem [45] and KKT conditions, (28) can be solved [41].

3.5. The properties of the proposed model

3.5.1. Connection to k-means

When the projection dimension r is fixed to c and replaces the constraint $\mathbf{W}^T\mathbf{S}_t\mathbf{W}=\mathbf{I}$ in (17) with constraint $\mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W}=\mathbf{I}$, (17) can be transformed into the following:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{P},\mathbf{S}}\sum_i(\sum_j(||\mathbf{W}^T\mathbf{x}_i-\mathbf{W}^T\mathbf{x}_j||_2^2 s_{ij})+\lambda||\mathbf{P}^T\mathbf{x}_i-\sum_j\mathbf{P}^T\mathbf{x}_j s_{ij}||_2^2)+\varphi||\mathbf{S}||_F^2,\\ \text{s.t.}\ \ \mathbf{s}_i^T\mathbf{1}=1, s_{ij}\geqslant 0, \mathbf{W}\in\Re^{d\times c}, \mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W}=\mathbf{I}, \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I}.\end{aligned} \tag{29}$$

We denote $\mathbf{f}_i=\mathbf{W}^T\mathbf{x}_i$ and $\mathbf{F}=[\mathbf{f}_1,\mathbf{f}_2,\ldots,\mathbf{f}_n]^T\in\Re^{n\times c}$, and $\mathbf{F}=\mathbf{X}^T\mathbf{W}$. Then, (29) can be rewritten as follows:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{P},\mathbf{S}}\sum_i(\sum_j(||\mathbf{f}_i-\mathbf{f}_j||_2^2 s_{ij})+\lambda||\mathbf{P}^T\mathbf{x}_i-\sum_j\mathbf{P}^T\mathbf{x}_j s_{ij}||_2^2)+\varphi||\mathbf{S}||_F^2,\\ \text{s.t.}\ \ \mathbf{s}_i^T\mathbf{1}=1, s_{ij}\geqslant 0, \mathbf{F}^T\mathbf{F}=\mathbf{I}, \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}=\mathbf{I}.\end{aligned} \tag{30}$$

For the first term in (30), the following equation holds:

$$\sum_i \sum_j ||\mathbf{f}_i - \mathbf{f}_j||_2^2 s_{ij} = 2Tr(\mathbf{F}^T\mathbf{L}\mathbf{F}),$$

where $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix of the similarity matrix $\mathbf{S}$, and $\mathbf{D}$ is a diagonal matrix with the ith diagonal element $D_{ii} = \sum_j s_{ij}$.

According to Ky Fan's Theorem [45],

$$\min_{\mathbf{F}\in\Re^{n\times c}, \mathbf{F}^T\mathbf{F}=\mathbf{I}} Tr(\mathbf{F}^T\mathbf{L}\mathbf{F}) = \sum_i^c \sigma_i(\mathbf{L}), \tag{32}$$

where $\sigma_1(\mathbf{L}), \sigma_2(\mathbf{L}), \ldots, \sigma_c(\mathbf{L})$ are the c minimum eigenvalues of $\mathbf{L}$.

Because $\mathbf{L}$ is positive semidefinite, $\sigma_i(\mathbf{L}) \geqslant 0$ [22,43]. Then, (30) can be transformed into the following:

$$\min_{\mathbf{W},\mathbf{P},\mathbf{S}} 2\sum_i^c \sigma_i(\mathbf{L}) + \lambda \sum_i (||\mathbf{P}^T\mathbf{x}_i - \sum_j \mathbf{P}^T\mathbf{x}_j s_{ij}||_2^2) + \varphi||\mathbf{S}||_F^2. \tag{33}$$

When $\varphi \ll 1$ and $\lambda \ll 1$,

$$\sum_i^c \sigma_i(\mathbf{L}) = 0, \tag{34}$$

which is equal to $rank(\mathbf{L}) = n - c$. According to Theorem 1, the number of connected components in a similarity graph $\mathbf{S}$ can be found to be exactly equal to c [22].

Let $\varphi \gg \lambda$, then

$$s_{ij} = \begin{cases} \frac{1}{n_h}, & \mathbf{x}_i \in C_h \text{ and } \mathbf{x}_j \in C_h \\ 0, & \text{otherwise} \end{cases}, \tag{35}$$

where C_h is a set of data points that belong to the hth cluster (each connected component in the similarity graph represents a cluster), and n_h is the number of data points in the hth cluster.

When $\mathbf{x}_i$ belongs to the $h(i)$th cluster, (33) can be transformed into the following:

$$\min_{\mathbf{P},C_h} \sum_i ||\mathbf{P}^T(\mathbf{x}_i - \frac{1}{n_{h(i)}} \sum_{\mathbf{x}_j \in C_{h(i)}} \mathbf{x}_j)||_2^2, \quad \text{s.t.} \quad \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P} = \mathbf{I}. \tag{36}$$

By denoting $\mathbf{v}_{h(i)} = \frac{1}{n_{h(i)}}(\sum_{\mathbf{x}_j \in C_{h(i)}} \mathbf{x}_j)$ as the clustering center of the $h(i)$th cluster, (36) can be rewritten as follows:

$$\min_{\mathbf{P},C_h} \sum_i ||\mathbf{P}^T(\mathbf{x}_i - \mathbf{v}_{h(i)})||_2^2, \quad \text{s.t.} \quad \mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P} = \mathbf{I}. \tag{37}$$

Problem (37) can be regarded as an objective function of the projection k-means clustering method. Therefore, the GEAGL model can be regarded as an extension of the k-means method. Moreover, the proposed method sufficiently addresses the problem of k-means being unable to learn the manifold structure.

3.5.2. Local manifold structure preservation

Mathematically, the manifold is a topological space that is similar to Euclidean space near each point. When the low-dimensional manifold is embedded in high-dimensional space, the data points in the high-dimensional space near each point will continue to have the properties of Euclidean space. Thus, we can establish a local dimensionality reduction-mapping relationship.

Let $\mathscr{M}$ be a smooth, compact, r dimensional Riemannian manifold embedded in $\mathbb{R}^d$ and let $f : \mathscr{M} \to \Re^{1\times 1}$ be a quadratic differentiable function on the manifold. Belkin and Niyogi [1,2] demonstrated that the best mapping with local smoothness can be determined by optimizing the following problem [1]:

$$\arg\min_{||f||_{\mathscr{L}^2(\mathscr{M})}=1} \int_{\mathscr{M}} ||\nabla f(\mathbf{x})||^2, \tag{38}$$

where gradient $\nabla f(\mathbf{x})$ is a vector field on the manifold [2]. And the distances between $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ are small if $\mathbf{x}_i$ and $\mathbf{x}_j$ are adjacent on the high-dimensional space [2].

For a vector field $\mathbf{X}$ and a function f on the manifold, the following equation holds [1]:

$$\int_{\mathscr{M}} \langle \mathbf{X}, \nabla f(\mathbf{x}) \rangle = -\int_{\mathscr{M}} div(\mathbf{X}) f, \tag{39}$$

where $div(\mathbf{X})$ is the divergence of $\mathbf{X}$. Thus, the following equation holds [1]:

$$\int_{\mathscr{M}} ||\nabla f(\mathbf{x})||^2 = -\int_{\mathscr{M}} \mathscr{L}(f) f, \tag{40}$$

where $\mathscr{L} \overset{def}{=} div\ \nabla f$ is the Laplace–Beltrami operator on the manifold $\mathscr{M}$. Belkin and Niyogi [1,2] pointed out that the optimal solution of (38) must be an eigenfunction of $\mathscr{L}$ since $\mathscr{L}$ is positive semidefinite.

According to the series expansion of the mean value of a function on a sphere or ball [46], we can consider the Laplacian operator of a function f as the difference between the function value at a point $\mathbf{x}_i$ and the average function value over a small sphere around $\mathbf{x}_i$. Let $\mathbf{e}_l \in \Re^d$ $(l = 1, 2, \ldots, d)$ be the standard coordinate vectors and h be the radius of the sphere. The discrete Laplace operator can be defined as follows [46]:

$$\Delta f(\mathbf{x}_i) = \frac{1}{h^2} \sum_{l=1}^d [f(\mathbf{x}_i - h\mathbf{e}_l) - 2f(\mathbf{x}_i) + f(\mathbf{x}_i + h\mathbf{e}_l)]. \tag{41}$$

Denoting $\mathscr{N}(\mathbf{x}_i) = \{\mathbf{x}_i - h\mathbf{e}_1, \mathbf{x}_i + h\mathbf{e}_1, \ldots, \mathbf{x}_i - h\mathbf{e}_d, \mathbf{x}_i + h\mathbf{e}_d\}$, where d is the dimension of the ambient space, (41) can be rewritten as follows:

$$\Delta f(\mathbf{x}_i) = \frac{1}{h^2} \left[\sum_{\mathbf{x}_j \in \mathscr{N}(\mathbf{x}_i)} f(\mathbf{x}_j) - 2d f(\mathbf{x}_i) \right]. \tag{42}$$

On this basis, if we use s_{ij} to represent the similarity between data points $\mathbf{x}_i$ and $\mathbf{x}_j$, and assume the similarity between $\mathbf{x}_i$ and its neighbors to be equal, i.e., $s_{ij} = \frac{1}{2d}$, the discrete Laplacian operator of function f can be redefined as follows:

$$\Delta f(\mathbf{x}_i) = \frac{2d}{h^2} \left[\sum_{\mathbf{x}_j \in \mathscr{N}(\mathbf{x}_i)} (f(\mathbf{x}_j) s_{ij}) - (\sum_{\mathbf{x}_j \in \mathscr{N}(\mathbf{x}_i)} s_{ij}) f(\mathbf{x}_i) \right]. \tag{43}$$

The vector form of (43) is follows:

$$\Delta f(\mathbf{x}_i) = -\frac{2d}{h^2} \mathbf{L}_{i,} \boldsymbol{f}, \tag{44}$$

where $\boldsymbol{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n)]^T \in \Re^{n\times 1}$, $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix of the similarity matrix $\mathbf{S}$, $\mathbf{L}_{i,}$ express the ith row of $\mathbf{L}$.

According to (44), one has:

$$\Delta f = -\frac{2d}{h^2} \mathbf{L}\boldsymbol{f}, \tag{45}$$

where $\Delta f = [\Delta f(\mathbf{x}_1), \Delta f(\mathbf{x}_2), \ldots, \Delta f(\mathbf{x}_n)]^T$.

Then, the discrete form of $-\int_{\mathscr{M}} \mathscr{L}(f) f$ is as follows:

$$-\sum_{i=1}^n \Delta f(\mathbf{x}_i) f(\mathbf{x}_i) = -(\Delta f)^T \boldsymbol{f} = \frac{2d}{h^2} \boldsymbol{f}^T \mathbf{L}\boldsymbol{f}. \tag{46}$$

When the mapping is linear, that is, $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ and $\boldsymbol{f} = \mathbf{X}^T\mathbf{a}$, according to (46), the $\mathbf{X}^T\mathbf{a}$ that minimizes $\mathbf{a}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}$ must be the optimal linear approximation of the eigenfunctions of $\mathscr{L}$ [1], and

$\mathbf{W} = \mathbf{a}$ is the optimal solution of (22) when $r = 1$. Therefore, the local manifold structure can be preserved in the generated projection subspace of the proposed GEAGL model.

Moreover, because $\mathbf{L}$ and $\mathbf{L}^2$ have the same eigenvectors, the optimal $\boldsymbol{f}$ for minimizing $\boldsymbol{f}^T\mathbf{L}^2\boldsymbol{f}$ is also the best solution for minimizing $\boldsymbol{f}^T\mathbf{L}\boldsymbol{f}$. For the proposed GEAGL model, the formulas $\mathbf{L} = \mathbf{I} - \mathbf{S}$ and $\mathbf{M} = \mathbf{L}^2$ apply; thus, the $\mathbf{X}^T\mathbf{a}$ that minimizes $\mathbf{a}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{a}$ is also a linear approximation of the eigenfunctions of $\mathscr{L}$. However, $\mathbf{P} = \mathbf{a}$ is also the optimal solution of (20) when $r = 1$. Namely, the LPP and self-expressive terms in the proposed model achieve a linear approximation of the eigenfunctions of the Laplace–Beltrami operator in two different manners [13]. Therefore, in the proposed model, the geometrical structure was explored in two different ways, and self-expressive learning was seamlessly integrated into the LPP.

4. Experiments

In this section, we first evaluate the effectiveness of the proposed GEAGL-CLR clustering on some benchmark data sets and then visualize the generated projection subspace to show the discriminability. All experiments are conducted on a personal computer with i5-9500 CPU @ 3.00 GHz and 8 GB RAM. The codes are implemented in MATLAB R2018b 64 bit.

4.1. Data sets

Five data sets from UCI repository [47], i.e., Movement, Teaching, Thyroid, Userknowledge and Wine, six image data sets, i.e., AR, COIL20 [48], COIL100 [49], ORL2116, Palm and PIE, and two gene data sets [50], i.e., Lung Cancer and SRBCT were selected for experiments. The details of these data sets are shown in Table 1.

4.2. Comparison methods and parameter settings

Some clustering and dimensionality reduction methods are selected for comparison, including k-means [51], RCut [52], Locality Preserving Projections (LPP) [12], Neighborhood Preserving Embedding (NPE) [13], Clustering with Adaptive Neighbors (CAN) [22], Projected Clustering with Adaptive Neighbors (PCAN) [22], Graph-Based Constrained Laplacian Rank Algorithm (CLR_L2) [43], Structured Graph Optimization for Unsupervised Feature Selection (SOGFS) [5], Robust Graph Learning from Noisy Data (RGC) [53] and Low-Rank Adaptive Graph Embedding for Unsupervised Feature Extraction (LRAGE) [26]. All selected methods, except k-means, can learn the manifold structure of the given data. Among these, RCut, LPP, PCAN, SOGFS, RGC, and LRAGE can preserve the geometrical structure of the data in the generated low-dimensional subspace, whereas NPE can exploit the self-expressive property of the data.

For the selected high-dimensional datasets, including the movement, image and gene datasets, the PCA technique [54] was utilized to reduce the dimensionality to 60 to speed up the learning process. In addition, the dimensions of the projection subspace vary within the range of $[9,\ 12,\ 15, \ldots,\ 60]$. For the UCI datasets called Teaching, Thyroid, Userknowledge, and Wine, the projection dimension is ranged from 2 to $min(n, d-1)$, where n and d are the number of data points and features, respectively. The similarity matrix plays a critical role in these graph-based clustering methods. For RCut, a similarity matrix was constructed using the self-turning Gaussian method [55]. For the NPE, a similarity matrix was constructed using the LLE method [14]. For CAN, PCAN, CLR, SOGFS, LRAGE, and GEAGL-CLR, the similarity matrix was constructed using the adaptive neighbors method (the number of neighbors was set to $k = 10$) [22]. In GEAGL-CLR, the parameter-free strategy in [22] was utilized to determine the regularization parameter φ. The balanced parameter λ was set to the range of $[0.01,\ 0.1,\ 1,\ 10,\ 100,\ 1000]$. For SOGFS, RGC and LRAGE, the balanced parameter was set to the range of $[0.001,\ 0.01,\ 0.1,\ 1,\ 10,\ 100,\ 1000]$. Considering that the k-means method is sensitive to initialization, in the related methods, that is, k-means, RCut, LPP, NPE, and LRAGE, k-means was repeated 20 times and the result with the smallest loss function value was selected to be the final clustering result.

4.3. Experimental results

4.3.1. Clustering performance

Tables 2–4 present the ACC, NMI, and Purity values obtained by the different methods on the selected data sets, respectively. The best results are shown in bold, and the second-best results are identified in brackets; the symbol "–" indicates timeout. The last row of Tables 2–4 presents the average ACC, NMI and Purity of each method on all selected data sets.

As shown in Tables 2–4, owing to the negative impact of redundant features when computing the distances between data points and cluster centers, k-means does not perform sufficiently on most datasets. For RCut, LPP and NPE, the construction of the similarity matrix and the optimization of the low-dimensional representation are performed in two separate stages. The pre-constructed similarity matrix is often not optimal, such that its performance is generally worse than that of PCAN and SOGFS. The performance of CAN and CLR is generally worse than that of PCAN when managing high-dimensional data, especially for the ORL2116 and Palm image datasets. This is because the projected features are more discriminable when compared with the original features. Therefore, PCAN and SOGFS appear to perform sufficiently on all datasets. However, they are inferior to GEAGL-CLR when certain data points from different clusters are close to one another because their similarity measure adopt pairwise distances.

As shown in Table 3, the performance of the proposed method is inferior to that of the SOGFS model, which may be because the $L_{2,1}$-norm penalty term can guarantee the structured sparseness of the projection and reduce the effect of noise in the SOGFS model. Moreover, the performance of the proposed method is also inferior to that of the PCAN model, which may be owing to the fact that the PCAN can learn the similarity

Table 1
The benchmark data set.

Data sets	Samples	Dimensions	Clusters	Data sets	Samples	Dimensions	Clusters
Movement	360	90	15	COIL100	7200	1024	100
Teaching	151	5	3	ORL2116	400	2116	40
Thyroid	215	5	3	Palm	2000	256	100
Userknowledge	403	5	4	PIE	1632	1024	68
Wine	178	13	3	LungCancer	181	12533	2
AR	2400	2000	120	SRBCT	83	2308	4
COIL20	1440	1024	20	–	–	–	–

Table 2
ACC of different methods on the selected data sets.

Data Sets	*k*-means	RCut	NPE	LPP	CAN	PCAN	CLR	SOGFS	RGC	LRAGE	Ours
Movement	45.00	52.22	49.72	51.11	48.33	50.83	48.33	(57.22)	50.00	53.06	**61.11**
Teaching	39.74	37.09	42.38	43.05	38.41	39.07	38.41	(49.01)	43.05	46.36	**52.32**
Thyroid	86.05	71.16	82.79	84.19	74.88	83.26	87.44	86.05	70.70	(90.70)	**95.81**
Userknowledge	55.58	48.14	43.67	59.55	59.06	(76.18)	57.32	59.55	53.60	59.31	**80.40**
Wine	70.22	72.47	64.61	70.79	72.47	72.47	72.47	(82.58)	73.03	70.79	**100.00**
AR	31.42	45.88	67.33	55.67	41.21	47.46	40.17	68.50	48.33	(69.17)	**73.46**
COIL20	67.78	79.79	58.68	77.36	(88.06)	84.51	84.86	75.63	85.90	76.39	**100.00**
COIL100	47.78	59.06	42.83	55.13	74.00	73.79	(75.78)	–	73.43	54.28	**83.29**
ORL2116	69.25	78.75	70.00	86.00	78.00	(82.50)	76.00	80.50	75.50	81.25	**89.00**
Palm	71.05	83.80	81.10	84.15	76.70	82.85	78.40	78.15	(85.90)	81.95	**98.85**
PIE	16.42	26.04	(56.99)	22.98	15.63	28.98	21.57	53.31	31.99	55.21	**87.32**
LungCancer	54.68	69.95	63.05	74.88	79.80	88.18	(89.66)	80.30	72.41	76.35	**90.64**
SRBCT	46.99	68.67	59.04	(74.70)	54.22	69.88	60.24	60.24	71.08	62.65	**77.11**
Avg.	54.00	61.00	60.17	64.58	61.60	67.69	63.90	(69.25)	64.22	67.50	**83.79**

Table 3
NMI of different methods on the selected data sets.

Data Sets	*k*-means	RCut	NPE	LPP	CAN	PCAN	CLR	SOGFS	RGC	LRAGE	Ours
Movement	59.07	64.13	64.27	62.62	68.47	66.03	68.50	(68.95)	62.41	64.25	**70.10**
Teaching	2.63	2.27	4.94	5.40	6.36	5.53	6.36	**9.28**	2.88	6.65	(9.05)
Thyroid	49.45	30.02	41.62	39.91	29.76	42.85	52.65	49.45	33.71	(64.19)	**79.91**
Userknowledge	35.71	30.74	20.40	43.96	49.32	(64.63)	32.63	49.57	31.80	49.34	**66.90**
Wine	42.88	43.72	39.95	41.93	39.48	39.48	39.48	(51.69)	42.14	41.93	**100.00**
AR	63.55	70.76	87.11	81.57	65.56	76.46	65.11	87.91	72.70	(88.21)	**92.47**
COIL20	78.27	88.21	74.85	84.94	(96.10)	92.66	94.61	83.24	92.44	83.75	**100.00**
COIL100	77.08	83.85	75.69	79.59	92.30	93.03	(94.07)	–	89.56	79.69	**96.88**
ORL2116	86.39	88.89	85.36	(92.82)	92.06	92.81	90.85	90.18	89.80	91.46	**94.82**
Palm	90.65	94.88	94.47	(96.24)	94.80	96.08	95.23	92.48	94.89	94.17	**99.40**
PIE	43.16	55.13	(74.30)	49.83	33.58	54.93	39.18	73.58	60.36	73.02	**88.43**
LungCancer	48.82	61.26	57.96	64.99	62.10	**73.55**	70.55	66.00	54.10	63.12	(73.13)
SRBCT	21.01	47.02	27.42	(57.91)	45.52	56.21	52.49	39.48	51.43	47.57	**64.58**
Avg.	53.74	58.53	57.56	61.67	59.65	(65.71)	61.67	63.48	59.86	65.18	**79.67**

Table 4
Purity of different methods on the selected data sets.

Data Sets	*k*-means	RCut	NPE	LPP	CAN	PCAN	CLR	SOGFS	RGC	LRAGE	Ours
Movement	54.17	58.33	63.89	57.78	**68.33**	61.94	**68.33**	(64.72)	55.28	60.83	**68.33**
Teaching	50.33	66.23	47.02	43.71	(78.15)	**83.44**	(78.15)	66.89	48.34	47.68	56.29
Thyroid	86.05	71.63	86.98	84.19	(92.56)	86.51	87.44	86.05	70.70	90.70	**95.81**
Userknowledge	55.58	51.61	46.65	62.03	**88.34**	(84.12)	79.16	60.30	64.27	60.55	83.87
Wine	70.22	72.47	62.92	70.79	72.47	72.47	72.47	(82.58)	73.03	70.79	**100.00**
AR	42.46	49.96	73.75	66.79	55.71	72.75	58.25	74.04	55.58	(75.63)	**82.13**
COIL20	72.85	82.43	74.51	80.21	(98.06)	93.68	94.86	79.86	91.67	81.18	**100.00**
COIL100	63.72	72.39	63.60	66.79	92.81	91.15	**95.51**	–	81.51	66.18	(95.29)
ORL2116	75.25	80.25	78.25	83.75	86.00	(86.75)	(86.75)	84.00	84.75	85.50	**90.25**
Palm	78.35	87.90	92.30	(92.80)	88.25	91.35	89.40	83.05	86.90	87.85	**98.85**
PIE	18.57	31.19	65.07	28.00	48.71	(74.39)	50.74	65.20	37.68	63.42	**87.32**
LungCancer	57.14	71.92	68.47	72.91	(93.10)	91.13	92.61	80.30	74.38	76.35	**93.60**
SRBCT	60.24	68.67	59.04	73.49	**87.95**	79.52	81.93	66.27	71.08	68.67	(86.75)
Avg.	60.38	66.54	67.88	67.94	80.80	(82.25)	79.66	74.44	68.86	71.95	**87.58**

matrix and cluster structure in one step without any post-processing by constraining the number of connected components in the similarity graph. Although the performance of the proposed method is inferior to that of the SOGFS model for the Teaching dataset and inferior to that of the PCAN model for the LungCancer dataset, our method achieves better results on most datasets and has the best average NMI value.

Based on Tables 2–4, we can conclude that our proposed GEAGL-CLR method is superior to the other methods on most datasets. In particular, our proposed methods has the highest ACC for all the datasets. Regarding the datasets of Wine and COIL20, GEAGL-CLR achieved impressive ACC results of 1. The cause of these excellent results was considered. Certain adaptive graph learning methods, including SOGFS, LRAGE, CLR, and PCAN, fail to achieve satisfactory results using pairwise distances to determine the similarity between data points. Therefore, the manifold structure of the data may not be effectively recovered by using only the distance relationship in certain scenarios. Our method uses the LPP and self-expression terms to learn the distance and reconstruction relationships between the data points, and further uses the learned structural information to determine the similarity between data points. Because the two terms use different projection matrices, the manifold structural information of the data can be learned accurately, which may be the main reason for the higher discriminability of the proposed model on the Wine and COIL20 datasets.

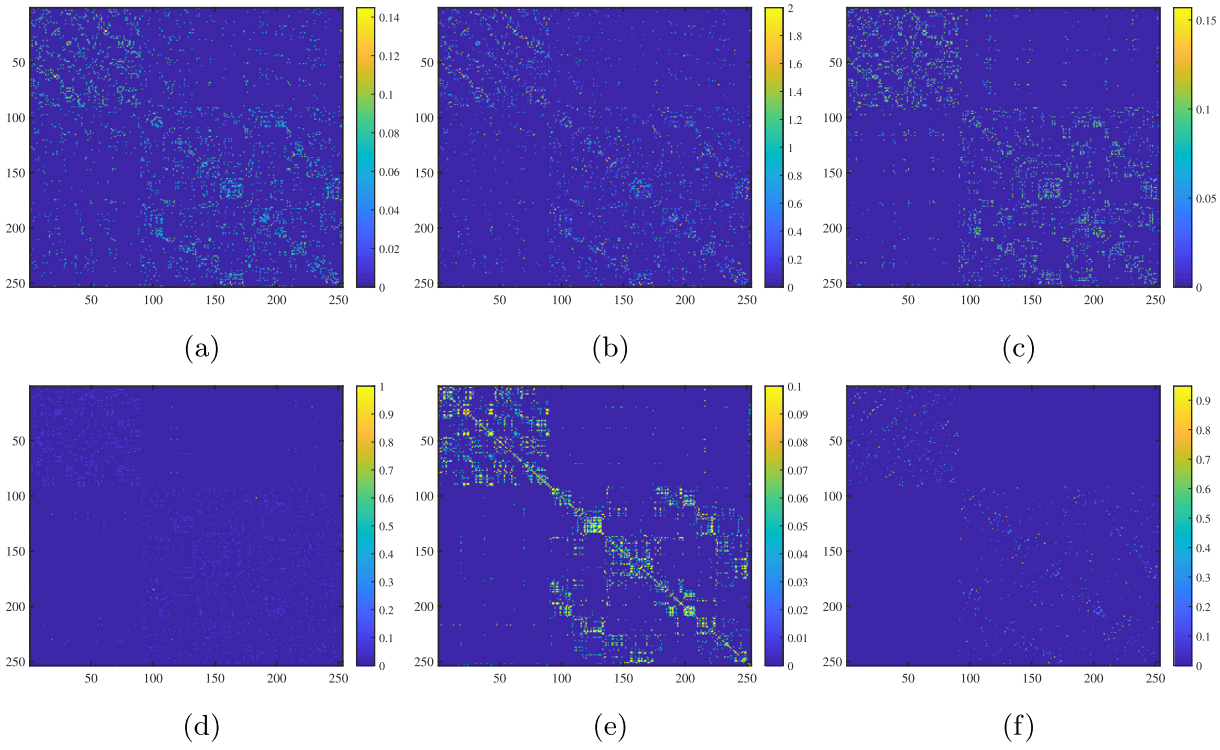

Fig. 2. The similarity matrix of different methods on Ovarian. (a) Self-Turing; (b) LLE; (c) PCAN; (d) SOGFS; (e) LRAGE; (f) Ours (GEALG).

4.3.2. Visualization of the similarity matrix

For NCut and NPE, a similarity matrix was constructed in advance. Conversely, the similarity matrices in PCAN, SOGFS, LRAGE and GEAGL were adaptively learned. To demonstrate the difference in constructing the similarity matrices using the two mechanisms, we visualized the constructed similarity matrices in Fig. 2.

By observing the results in Fig. 2 (a)-(b), we found that certain data points belonging to different clusters have high similarities, whereas Fig. 2 (c)-(f) do not demonstrate this trend. Therefore, compared to the pre-construction mechanism, the adaptive learning of the similarity matrix is generally better and improves the discriminability of the model. Furthermore, the result in Fig. 2 (f) present a clearer block-diagonal structure than that in Fig. 2 (c)-(e). This can be attributed to the fact that the proposed GEAGL can preserve both the self-expressive property and the geometrical structure of the data simultaneously, whereas other methods retain only one of them.

4.3.3. Visualization of projection subspace

To examine the discriminability of the projection subspace generated by the various methods, we used the t-SNE [56] method to visualize the dataset in the original feature space and generated the projection subspaces. The results are shown in Fig. 3.

In Fig. 3, GEAGL (W) and GEAGL (P) represent the projection subspace obtained by projection matrices **W** and **P** in our method, respectively. In the original feature space, certain data points from different clusters are observed to be intertwined. Compared to Fig. 3 (b)-(f), the results in Fig. 3 (g) and 3 (h) appear to be more discriminative. This is because by utilizing the self-expressive property of the data, the reconstruction coefficient matrix **S** in the proposed method tends to have an ideal block diagonal structure such that the similarity between the data points that belong to different clusters is small, despite being close.

In addition, compared to NPE and LPP (Fig. 3 (b) and Fig. 3 (c)), the results of our proposed method (Fig. 3 (g) and Fig. 3 (h)) have higher discriminability, which demonstrates that using a unified model to connect the two generated subspaces is improves their discriminability. Therefore, the LPP and self-expressive terms are mutually beneficial in GEAGL.

4.3.4. Convergence analysis

As proved in Section 3.3, the proposed algorithm can converge inevitably in execution. Fig. 4 shows the convergence curves of GEAGL on AR, COIL20, PIE, and SRBCT. In Fig. 4 (a)-(d), GEAGL converges approximately after 10 iterations, which also indicates that the proposed GEAGL is effective.

4.3.5. Parameter sensitivity

The subspace projection technique is used in LPP, NPE, PCAN, SOGFS, and the proposed method. To examine the sensitivity of these methods to the projection dimension, we performed experiments on the AR, COIL20, ORL2016, and Palm datasets under different dimensions r; the clustering results are shown in Fig. 5.

As shown in Fig. 5 (a)-(d), the NPE is sensitive to the projection dimensions, and its clustering performance improves as the projection dimension increases. This is because the NPE utilizes the self-expressive property to capture the subspace structure of the data. When the dimension of the projection subspace is small, the assumption of subspace independence is difficult to satisfy. Considering $r = 1$ as an example, the independent subspace is unique. However, the other methods do not exhibit this trend because the LPP term can be used to preserve the geometrical structure of the data in the low-dimensional space such that their performance can be ensured even in a low-dimensional space.

For the proposed GEAGL-CLR, the factor λ and projection dimension r must be provided in advance, where the parameter λ balances the importance of the pairwise distance and the self-expressive error in reflecting the similarity between the data

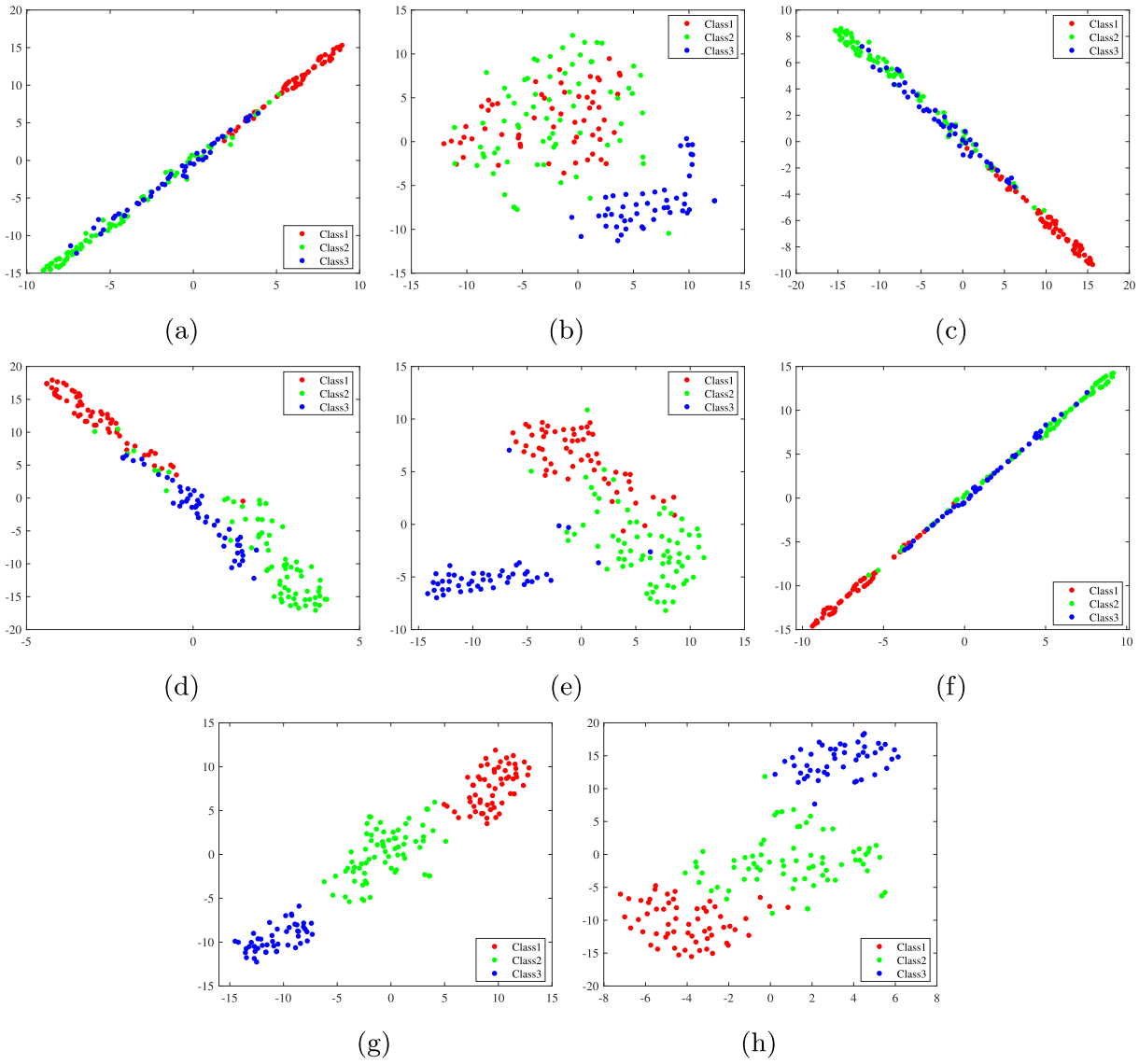

Fig. 3. The visualizations of data set in the original feature space and in the projected subspace obtained by different methods on Wine using t-SNE technique. (a) Original data set; (b) NPE; (c) LPP; (d) NPE; (e) SOGFS; (f) LRAGE; (g) GEALG (W); (h) GEAGL (P).

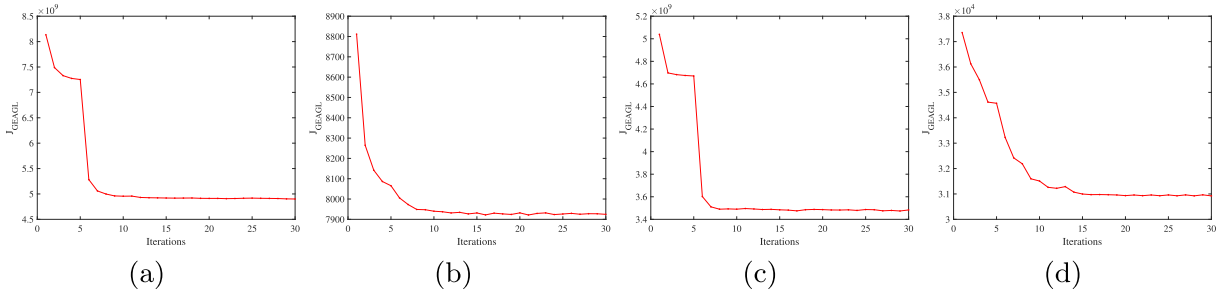

Fig. 4. The convergence curves of GEAGL on different data sets. (a) AR; (b) COIL20; (c) PIE; (d) SRBCT.

points. As shown in Fig. 6 (a) and (c), GEAGL-CLR is sensitive to the parameters λ and r. From Fig. 6 (a), when $r = 12$, GEAGL-CLR usually performs better when λ is sufficiently large. By observing all results, is seems that the value of ACC is generally higher when the value of (λ, r) is (10, 36).

To further observe the variation trend of the performance of GEAGL-CLR when varying the parameter λ, we performed the experiments on the datasets AR, Palm, ORL2016 under different λ; the clustering results are shown in Fig. 7.

As shown in Fig. 7, when $\lambda = 10$, the performance of GEAGL-CLR is the best; however, when the value of λ is excessively large or small, the performance of GEAGL-CLR is not optimal. This is because, when the value of λ is too large, the proposed GEAGL-CLR model can only preserve the self-expressive properties of the

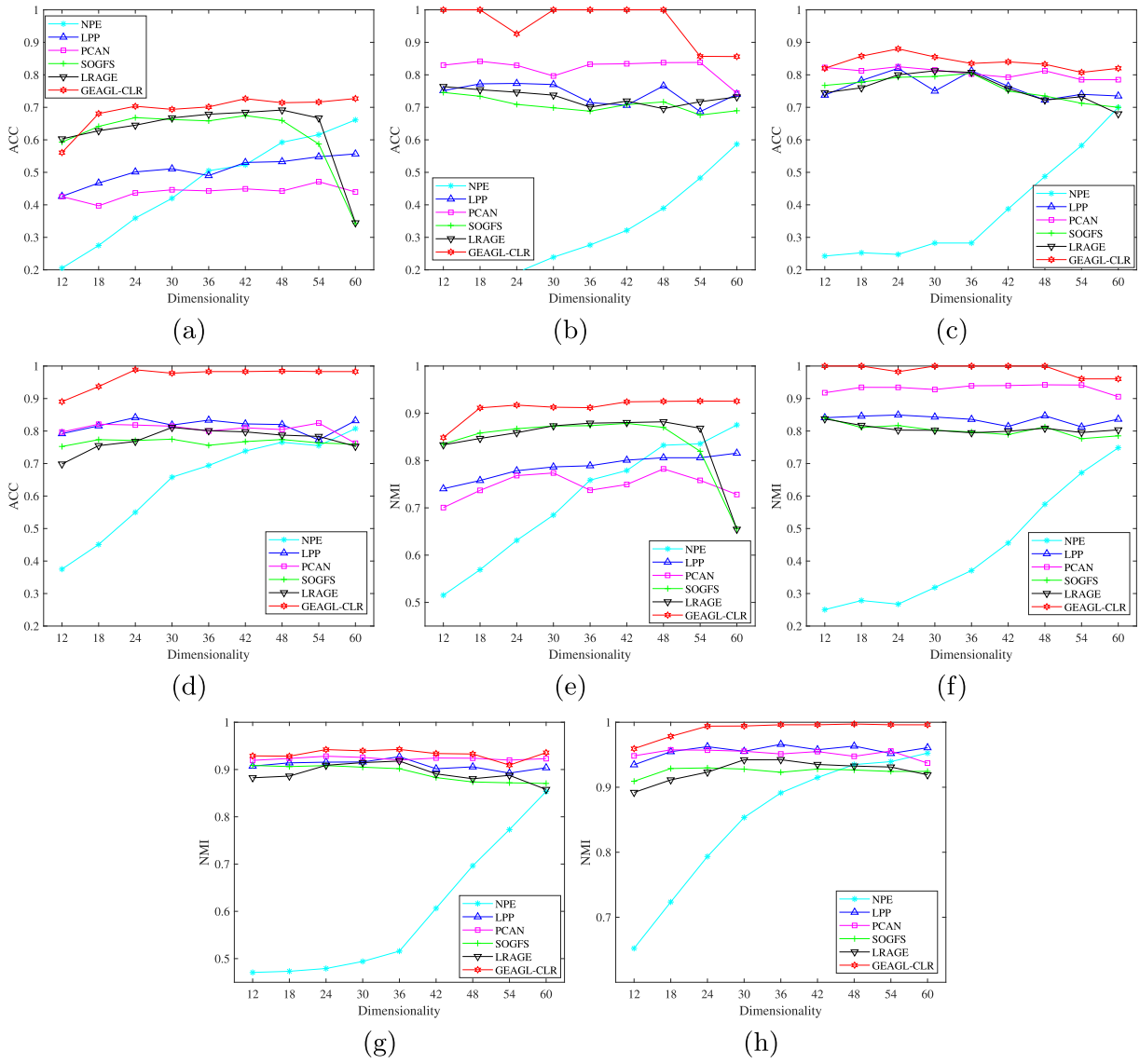

Fig. 5. The performance of the selected subspace projection methods on different data sets when varying the projection dimension. (a) AR (ACC); (b) COIL20 (ACC); (c) ORL2116 (ACC); (d) Palm (ACC); (e) AR (NMI); (f) COIL20 (NMI); (g) ORL2116 (NMI); (h) Palm (NMI).

data; when the value of λ is too small, GEAGL-CLR can only preserve the geometric structure of the data. When $\lambda = 10$, the GEAGL-CLR model can simultaneously preserve the geometric structure and self-expressive properties of the data, which improves the clustering performance of the GEAGL-CLR model.

4.3.6. Ablation experiment

To analyze the effect of each component in the objective function of the proposed methods, we defined the method after eliminating the second and third items of GEAGL as GEAGL-01, and the method after eliminating the second item as GEAGL-02. Their performances with different values of r on AR, COIL20, ORL2016, and Palm are shown in Fig. 8.

As shown in Fig. 8 (a)-(h), compared to GEAGL-01, GEAGL-02 achieves better results regardless of the value of r, which demonstrates that adopting adaptive graph technology is helpful for learning the optimal similarity matrix. As shown in Fig. 8 (c) and (g), when $r = 54$, the performance of GEAGL-02 drops sharply. The possible reason is that when $r = 54$, the pairwise distance between data points cannot fully reflect the similarity of data points. However, in the proposed GEAGL-CLR model, the second term is used to preserve the self-expressive properties of the data. On the one hand, when the data points are from the independent linear subspaces, based on the self-expressive properties of the data, the similarity between data points can be accurately measured, which enhances the performance of the proposed model when data points from different clusters are close. On the other hand, data points coming from the same cluster are easier to represent each other and the adjacent data points may have different self-expression properties. GEAGL-02 ignores the reconstruction relationship of the data based on the self-expression properties, whereas GEAGL-CLR utilizes both the distances and reconstruction relationships between samples to more accurately measure the similarity between data points.

For nearly all selected datasets, the performance of the GEAGL-CLR method was significantly better than that of GEAGL-01 and GEAGL-02 under different dimensions, which demonstrates that the introduction of the self-expressive item is helpful in alleviating the problem of cluster interweaving and prompting clustering performance.

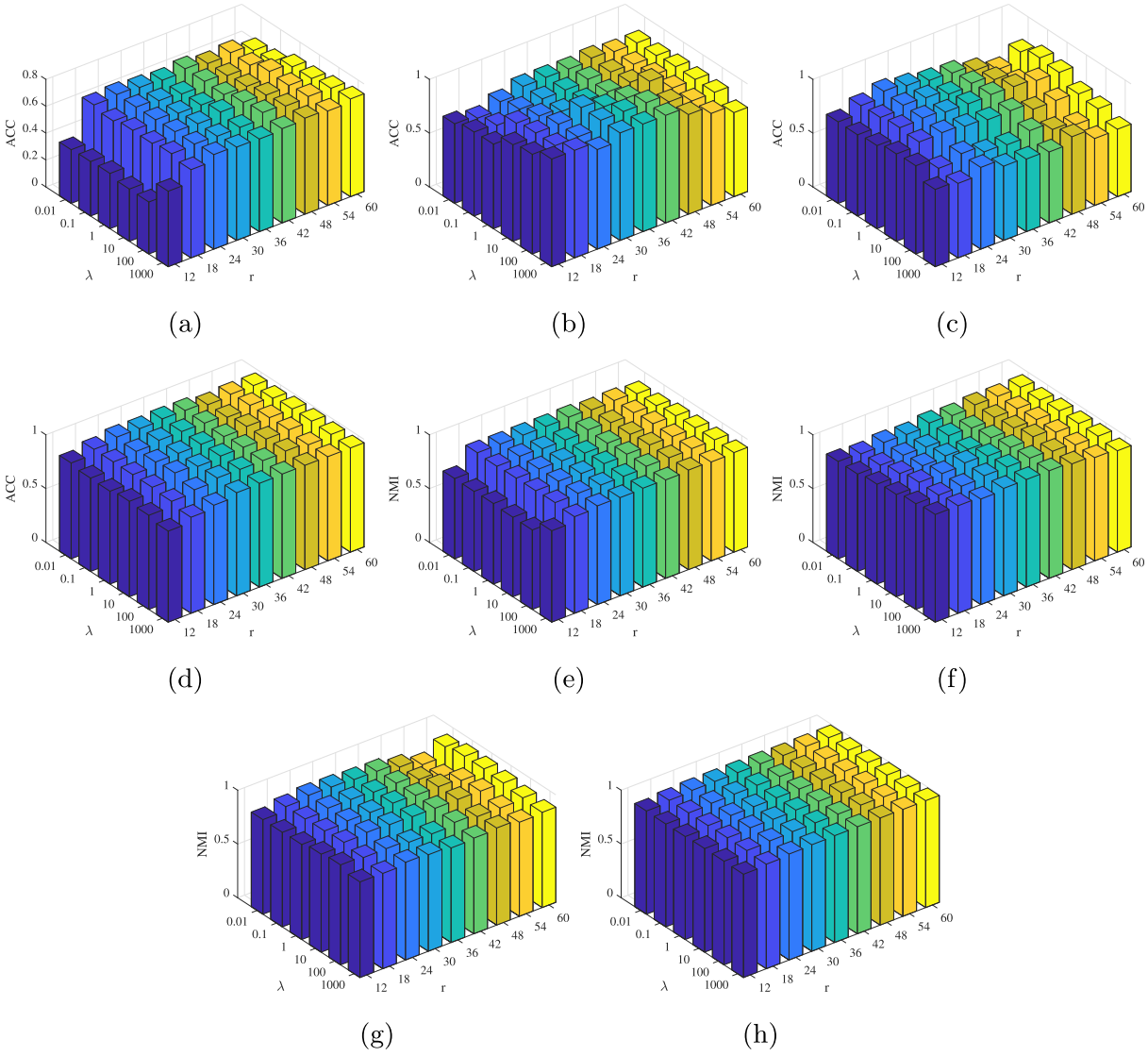

Fig. 6. The performance of the proposed method on the selected data sets when varying the parameters λ and r. (a) AR (ACC); (b) COIL20 (ACC); (c) ORL2116 (ACC); (d) Palm (ACC); (e) AR (NMI); (f) COIL20 (NMI); (g) ORL2116 (NMI); (h) Palm (NMI).

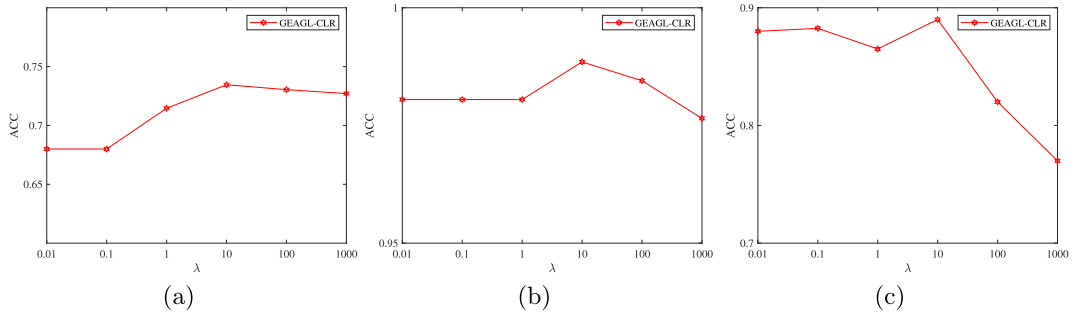

Fig. 7. The performance of the proposed method on the selected data sets when varying the parameter λ. (a) AR (ACC); (b) Palm (ACC); (c) ORL2116 (ACC).

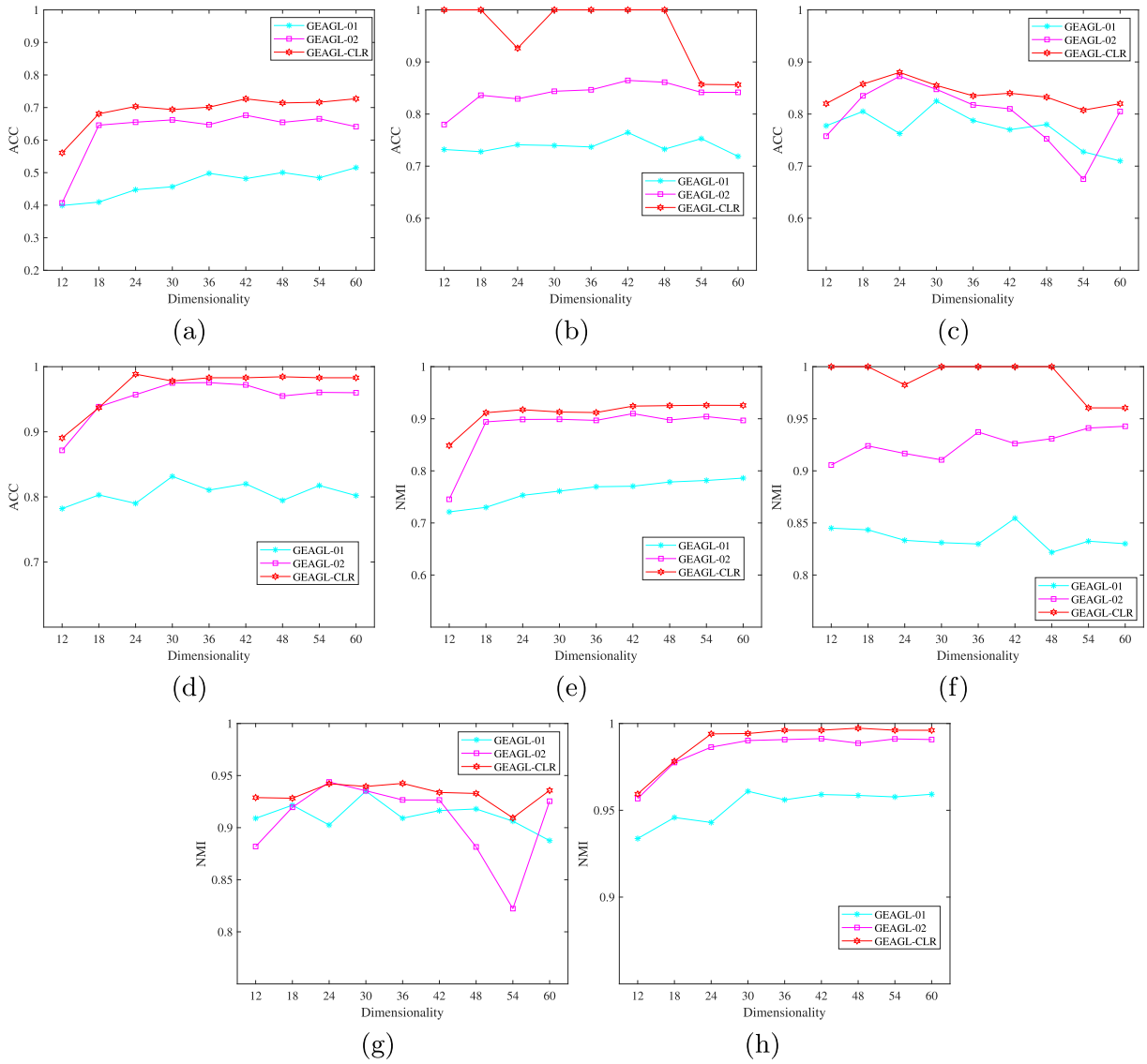

Fig. 8. The performance of GEAGL-01, GEAGL-02 and GEAGL-CLR on the selected data sets when varying the parameter *r*. (a) AR (ACC); (b) COIL20 (ACC); (c) ORL2116 (ACC); (d) Palm (ACC); (e) AR (NMI); (f) COIL20 (NMI); (g) ORL2116 (NMI); (h) Palm (NMI).

5. Conclusions

In this study, a novel similarity graph learning model called adaptive graph learning with geometrical structure preservation and self-expression maintenance (GEAGL) is proposed. In the proposed model, a graph structure that retains the manifold structure and maintains the self-expression characteristics of the data is constructed, and two projection subspaces are formed. Because the projection representation of the two subspaces linearly approximates the eigenfunctions of the Laplace–Beltrami operator on the manifold in two different ways, the LPP and self-expressive terms are mutually beneficial. A clustering method was developed based on the graph structure produced by the proposed GEAGL model. Extensive experimental results on benchmark datasets demonstrate that the proposed clustering method achieves a better performance than other representative methods. The proposed GEAGL model constructs two projection subspaces. We will further analyze the relationship between the two projection subspaces and try to develop a new dimensionality reduction method by minimizing the difference between them.

CRediT authorship contribution statement

Yangbo Wang: Conceptualization, Methodology, Writing – original draft. **Can Gao:** Methodology, Software, Data curation, Validation. **Jie Zhou:** Methodology, Software, Visualization, Writing – review & editing.

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.

Acknowledgements

The author would like to thank the Editor-in-Chief, editors and anonymous reviewers for their kind help and valuable comments. This work was supported in part by the National Natural Science Foundation of China (Nos.62076164, 61806127), Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515011861),

and Shenzhen Science and Technology Program (No. JCYJ20210324094601005).

References

[1] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Advances in Neural Information Processing Systems 14, Vancouver, British Columbia, Canada, 2001, pp. 585–591..

[2] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neurocomputing 15 (6) (2003) 1373–1396.

[3] D. Cai, X. He, X. Wu, J. Han, Non-negative matrix factorization on manifold, in: Proceedings of the Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 63–72.

[4] M. Chen, Q. Wang, X. Li, Adaptive projected matrix factorization method for data clustering, Neurocomputing 306 (2018) 182–188.

[5] F. Nie, W. Zhu, X. Li, Structured graph optimization for unsupervised feature selection, IEEE Trans. Knowl. Data Eng. 33 (3) (2019) 1210–1222.

[6] H. Zhao, Q. Li, Z. Wang, F. Nie, Joint adaptive graph learning and discriminative analysis for unsupervised feature selection, Cogn. Comput. 14 (3) (2021) 1211–1221.

[7] Q. Wang, Z. Qin, F. Nie, X. Li, Spectral embedded adaptive neighbors clustering, IEEE Trans. Neural Networks Learn. Syst. 30 (4) (2019) 1265–1271.

[8] X. Li, M. Chen, Q. Wang, Discrimination-aware projected matrix factorization, IEEE Trans. Knowl. Data Eng. 32 (4) (2020) 809–814.

[9] J. Wang, F. Xie, F. Nie, X. Li, Unsupervised adaptive embedding for dimensionality reduction, IEEE Trans. Neural Networks Learn. Syst. (2021) 1–12, https://doi.org/10.1109/TNNLS.2021.3083695.

[10] Z. Li, F. Nie, D. Wu, Z. Hu, X. Li, Unsupervised feature selection with weighted and projected adaptive neighbors, IEEE Trans. Cybern. (2021) 1–12, https://doi.org/10.1109/TCYB.2021.3087632.

[11] Z. Li, F. Nie, X. Chang, L. Nie, H. Zhang, Y. Yang, Rank-constrained spectral clustering with flexible embedding, IEEE Trans. Neural Networks Learn. Syst. 29 (12) (2018) 6073–6082.

[12] X. He, P. Niyogi, Locality preserving projections, in: Proceedings of the 16th International Conference on Neural Information Processing Systems, 2003, pp. 153–160.

[13] X. He, D. Cai, S. Yan, H.-J. Zhang, Neighborhood preserving embedding, in: Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 2005, pp. 1208–1213.

[14] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[15] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, in: Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2003, pp. 849–856..

[16] S. Huang, Z. Xu, F. Wang, Nonnegative matrix factorization with adaptive neighbors, in: 2017 International Joint Conference on Neural Networks, Anchorage, AK, USA, 2017, pp. 486–493.

[17] S. Huang, Z. Xu, J. Lv, Adaptive local structure learning for document co-clustering, Knowl.-Based Syst. 148 (2018) 74–84.

[18] S. Huang, Z. Xu, Z. Kang, Y. Ren, Regularized nonnegative matrix factorization with adaptive local structure learning, Neurocomputing 382 (2020) 196–209.

[19] X. Li, M. Chen, F. Nie, Q. Wang, A multiview-based parameter free framework for group detection, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017, pp. 4147–4153..

[20] X. Li, M. Chen, F. Nie, Q. Wang, Locality adaptive discriminant analysis, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017, pp. 2201–2207..

[21] M. Chen, X. Li, Robust matrix factorization with spectral embedding, IEEE Trans. Neural Networks Learn. Syst. 32 (12) (2021) 5698–5707.

[22] F. Nie, X. Wang, H. Huang, Clustering and projected clustering with adaptive neighbors, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2014, pp. 977–986.

[23] J. Zhou, W. Pedrycz, X. Yue, C. Gao, Z. Lai, J. Wan, Projected fuzzy c-means clustering with locality preservation, Pattern Recogn. 113 (2021) 107748.

[24] J. Zhou, W. Pedrycz, C. Gao, Z. Lai, J. Wan, Z. Ming, Robust jointly sparse fuzzy clustering with neighborhood structure preservation, IEEE Trans. Fuzzy Syst. 30 (4) (2022) 1073–1087.

[25] Y. Chen, Z. Lai, W.K. Wong, L. Shen, Q. Hu, Low-rank linear embedding for image recognition, IEEE Trans. Multimedia 20 (12) (2018) 3212–3222.

[26] J. Lu, H. Wang, J. Zhou, Y. Chen, Z. Lai, Q. Hu, Low-rank adaptive graph embedding for unsupervised feature extraction, Pattern Recogn. 113 (2021) 107758.

[27] J. Wen, N. Han, X. Fang, L. Fei, K. Yan, S. Zhan, Low-rank preserving projection via graph regularized reconstruction, IEEE Trans. Cybern. 49 (4) (2019) 1279–1291.

[28] C. Tang, X. Liu, X. Zhu, J. Xiong, M. Li, J. Xia, X. Wang, L. Wang, Feature selective projection with low-rank embedding and dual laplacian regularization, IEEE Trans. Knowl. Data Eng. 32 (9) (2020) 1747–1760.

[29] C. Tang, X. Zheng, X. Liu, W. Zhang, J. Zhang, J. Xiong, L. Wang, Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection, IEEE Trans. Knowl. Data Eng. (2021) 1–12, https://doi.org/10.1109/TKDE.2020.3048678.

[30] E. Elhamifar, R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2765–2781.

[31] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, 2009, pp. 2790–2797.

[32] E. Elhamifar, R. Vidal, Clustering disjoint subspaces via sparse representation, in: Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas, 2010, pp. 1926–1929.

[33] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.

[34] C. Lu, J. Feng, Z. Lin, T. Mei, S. Yan, Subspace clustering by block diagonal representation, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2) (2019) 487–501.

[35] Z. Zhang, J. Ren, S. Li, R. Hong, Z. Zha, M. Wang, Robust subspace discovery by block-diagonal adaptive locality-constrained representation, in: Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 2019, pp. 1569–1577.

[36] G. Liu, Z. Zhang, Q. Liu, H. Xiong, Robust subspace clustering with compressed data, IEEE Trans. Image Process. 28 (10) (2019) 5161–5170.

[37] J. Yang, J. Liang, K. Wang, P.L. Rosin, M.-H. Yang, Subspace clustering via good neighbors, IEEE Trans. Pattern Anal. Mach. Intell. 42 (6) (2020) 1537–1544.

[38] Z. Kang, C. Peng, Q. Cheng, X. Liu, X. Peng, Z. Xu, L. Tian, Structured graph learning for clustering and semi-supervised classification, Pattern Recogn. 110 (2021) 107627.

[39] Z. Kang, Z. Lin, X. Zhu, W. Xu, Structured graph learning for scalable subspace clustering: From single view to multiview, IEEE Trans. Cybern. (2021) 1–11, https://doi.org/10.1109/TCYB.2021.3061660.

[40] Z. Lin, Z. Kang, L. Zhang, L. Tian, Multi-view attributed graph clustering, IEEE Trans. Knowl. Data Eng. (2021) 1–9, https://doi.org/10.1109/TKDE.2021.3101227.

[41] J. Huang, F. Nie, H. Huang, A new simplex sparse learning model to measure data similarity for clustering, in: Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015, pp. 3569–3575.

[42] S. Wright, Primal-Dual Interior-Point Methods, Soc. Ind. Appl. Math. (1997) 21–47.

[43] F. Nie, X. Wang, M. Jordan, H. Huang, The constrained laplacian rank algorithm for graph-based clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA, 2016, pp. 1969–1976.

[44] B. Mohar, Y. Alavi, G. Chartrand, O. Oellermann, The laplacian spectrum of graphs, Graph Theory, Combinatorics, and Applications, Wiley (1991) 871–898.

[45] K. Fan, On a theorem of weyl concerning eigenvalues of linear transformations, Proc. Nat. Acad. Sci. 36 (1) (1950) 31–35.

[46] J.S. Ovall, The laplacian and mean and extreme values, Am. Math. Monthly 123 (3) (2016) 287–291.

[47] P. Franti, Clustering datasets, (2015). [Online]. Available: http://cs.uef.fi/sipu/datasets/..

[48] S.A. Nene, S.K. Nayar, H. Murase, Columbia object image library (coil-20), Dept. Comput Sci, Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-005-96, 1996..

[49] S.A. Nene, S.K. Nayar, H. Murase, Columbia object image library (coil-100), Dept. Comput Sci., Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-006-96, 1996..

[50] Z. Zhu, Y.-S. Ong, M. Dash, Markov blanket-embedded genetic algorithm for gene selection, Pattern Recogn. 40 (11) (2007) 3236–3248.

[51] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

[52] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.

[53] Z. Kang, H. Pan, S.C.H. Hoi, Z. Xu, Robust graph learning from noisy data, IEEE Trans. Cybern. 50 (5) (2020) 1833–1843.

[54] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics Intell. Lab. Syst. 2 (1–3) (1987) 37–52.

[55] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, in: Proceedings of the 17th International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2004, pp. 1601–1608.

[56] L.V.D. Maaten, G. Hinton, Visualizing data using t-sne, J. Mach. Learn. Res. 9 (11) (2008) 2579–2965.

Yangbo Wang received the B.E. degree in network engineering from Liaoning Technical University, China, in 2020. He is currently pursuing the M.E. degree with the School of College of Computer Science and Software Engineering, Shenzhen University, China. His current research interests include computer vision and machine learning.

Can Gao received the Ph.D. degree in pattern recognition and intelligence system from Tongji University, China, in 2013. From 2010 to 2011, he was a visiting scholar, supported by China Scholarship Council, with the Department of Electrical and Computer Engineering (ECE), University of Alberta, Canada. He was a research associate, post-doctoral fellow, and research fellow since 2015 at The Hong Kong Polytechnic University. Now, he is an assistant professor with the College of Computer Science and Software Engineering, Shenzhen University, China. His research interests include machine learning, computer vision, and granular computing.

Jie Zhou received the Ph.D. degree from the Tongji University, in 2011. From 2010 to 2011, he was a visiting scholar in the Department of Electrical and Computer Engineering (ECE) at University of Alberta, Edmonton, Canada. From 2017 to 2018, he was an associate researcher in The Hong Kong Polytechnic University, Kowloon, Hong Kong. He is now an associate researcher in the College of Computer Science and Software Engineering, Shenzhen University, China. His current major research interests include uncertainty analysis, pattern recognition, data mining, and intelligent systems.