

Received 24 October 2022, accepted 7 November 2022, date of publication 9 November 2022,
date of current version 29 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3221150

RESEARCH ARTICLE

Adaptive Graph Representation for Clustering

MEI CHEN^{ID}, YOSHUAI WANG^{ID}, YONGXU CHEN, HONGYU ZHU, YUE XIE, AND PENGJU GUO

School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Corresponding author: Mei Chen (mei.chen.lzjtu@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62266029, in part by the Gansu Key Research and Development Program under Grant 21YF5GA053, and in part by the Gansu Higher Education Industry Support Program under Grant 2022CYZC-36.

ABSTRACT Many graph construction methods for clustering cannot consider both local and global data structures in the construction of initial graph. Meanwhile, redundant features or even outliers and data with important characteristics are addressed equally in the graph optimization process. These lead to the learned representation graph may not capture the optimal structure when clustering. This paper proposes a novel model for clustering, named adaptive graph construction and low-rank representation of weighted noise (ACLWN), to overcome these problems. ACLWN is composed of an adaptive representation graph construction model named ARG, and an adaptive weighted sparse representation graph learning model named AWSG. In ARG, manifold learning and sparse representation are employed to capture the local structure of data. In AWSG, an adaptive weighted matrix is proposed to strengthen the important features and improve the robustness of the low-dimensional representation graph. Moreover, constraints such as non-negative low-rank, sparsity and distance regularization terms are imposed to capture the local and global structures of data. Comprehensive experimental results show that our method outperforms the compared state-of-the-art methods. The low-dimensional representation graph constructed by ACLWN is more suitable for clustering.

INDEX TERMS Adaptive initial graph construction, adaptive weighted matrix, graph learning, low-rank constraint.

I. INTRODUCTION

Clustering as an unsupervised learning method has long been favored by researchers in machine learning, data mining and pattern recognition. A cluster is a set of data points that are the same as one another within the same cluster and are disparate from the points in other clusters [1], [2], [3]. Spectral clustering [4], [5], [6], the most typical graph learning clustering method, has good performance when dealing with complex high-dimensional data. It first constructs an initial graph to describe the similar relationships among data, then develops a low-dimensional representation matrix based on the initial graph, and eventually obtains the final clusters by k -Means [7]. Spectral clustering can be more accurate and robust only when the initial graph is well constructed. Similarly, the other clustering of graph representation is also done.

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang^{ID}.

The existing graph construction strategies can be roughly categorized into three groups: (1) Capture the similarity between data points by distance metric. Jurusan et al. utilized the straight-line distance between data points to assess similarity [8]; Yin et al. used the cosine function to construct the similarity matrix in the original space [9]; and Ding et al. proposed a random compact Gaussian (RCG) kernel, and used it to measure similarity between data points [10]. But these methods are unable to automatically collect structural information of points suitable for graph learning clustering. (2) Obtain similarity between data points based on global self-representation. Each point is encoded as a weighted combination of all other points, *i.e.*, data point could be represented by its adjacent and reachable indirect neighbors. Yun et al. utilized the weighted tensor nuclear norm to capture the fundamental spatial structure [11]; Shang et al. proposed a self-representation method based on dual-graph regularized feature selection [12]; Weng et al. introduced the Laplace smoothing criteria for graph construction by adopting data

self-representativeness [13]. Nevertheless, these graph construction methods are not ideal for high-dimensional datasets because of their high time complexity. (3) Obtain similarity between data points by describing each point as a linear combination of its neighbors. k NN-Graph [14], [15] methods represent a data point only according to its k -closest points. Roweis et al. developed a locality embedding graph (LLE-graph) [16] by using k NN-Graph to represent a data point. However, because of the various domain parameters, these methods often provide different similarity results. Moreover, the existing initial graph construction methods cannot capture the full local structural features from the original data since they can only reflect structural relationships of data to some extent.

To obtain more information about data structure for graph learning clustering, researchers have achieved a promising effect in extracting the complete structural information from the initial graph. Elhamifar et al. proposed sparse subspace clustering (SSC) [17] by employing a self-expression dictionary and a compressed sensing technology. Since sparse representation can capture the neighborhood structure of data, numerous sparse subspace clustering methods have emerged, such as ILRDFL [18], LRGDDL [19] and SSC+E [20]. Although the sparsity constraint can independently embody the relationships between pairwise points, it fails to reflect the global structure of data. To perform well in capturing the global structure of data, another important researching point, the low-rank representation (LRR) [21], was proposed by Liu et al. LRR can improve the correlation between samples within clusters while weakening the correlation between clusters. Because of the rich complex information in high-dimensional data structures, lots of novel LRR-based methods are proposed to explore the implicit information in data [22], [23], [24], [25]. Among them, Latent low-rank representation [24] is a popular method because it can recover well the hidden effect of inter-point structure and remove damaged points by integrating some latent observations. Xie et al. proposed a low-rank sparse preserving projection method [25], which learned a robust weight matrix by employing LRR-based methods to reduce the influence of outliers and noise. Unfortunately, LRR cannot reflect the adjacency relationship between points.

As a result, exploring how to better capture the both local and global structures has become a hot research topic [26], [27], [28], [29], [30]. Zhu et al. [29] designed a novel subspace clustering method to learn a representation graph by conducting feature selection and subspace learning in self-representation framework. The method addressed the shortcoming that existing methods cannot obtain local and global structures simultaneously. Han et al. [30] combined distance regularization and non-negative regularization to improve the latent LRR model. However, the representation graph learned by these methods brings about a lack of physical interpretation because it contains a large number of negative elements. Moreover, in the initial graph construction, these

models treat redundant features or even outliers as equally significant data. To balance the importance of data, Li et al. adaptively assigned weights to distinguish and filter noise and outliers [31]. Wen et al. proposed an Adaptive Weighted Non-negative Low-Rank Representation (AWNLR) [32] model to assign low weights to redundant features or outliers. They were still unable to obtain the optimal graph representation.

To keep exploring the optimal representation graph for graph learning, we propose a novel model, named adaptive graph construction and low-rank representation of weighted noise (ACLWN), based on manifold learning [33], [34] and sparse representation [35]. The framework of ACLWN model is illustrated in Figure 1. ACLWN includes an adaptive representation graph construction model (ARG) for constructing the initial graph and an adaptive weighted sparse representation graph learning model (AWSG) to obtain the low-dimensional representation matrix of data. First, to generate an initial graph capturing the local structure of data, ARG represents data point by adaptively obtaining its k nearest neighbors. Next, to optimize the initial graph and obtain a low-dimensional representation graph, AWSG jointly employs sparse, non-negative and low-rank constraints to capture the global and local structure of data. Subsequently, this method uses a weight matrix to guide the learning of the low-dimensional representation matrix, which adaptively assigns more weight to important features and less weight to redundant features or outliers. Meanwhile, AWSG uses ℓ_{21} -norm [36] to replace the nuclear norm in AWNLRR to further constrain the representation of noise and outliers. Since ACLWN can obtain both global and local structures of data, and it is robust to noise and outliers, it is suitable for graph learning clustering.

The main contributions of this paper are summarized as follows,

1. A graph learning clustering model is proposed for adaptive graph construction and weighted sparse representation.
2. The model completely utilizes the geometric structural information of data to guide the construction of the initial graph. It overcomes the disadvantages of the initial graph construction based on the complete graph and the k -nearest graph.
3. An adaptive weight matrix is employed to improve robustness to noise and outliers, so that the generated low-dimensional representation matrix can accurately express significant features of data.
4. Extensive experiments on real datasets are conducted to verify that the effectiveness of our framework is superior to those of the other state-of-the-art baseline algorithms.

The remaining sections are organized as follows. The next section briefly surveys the related works. Section III proposes the clustering framework ACLWN. Section IV describes the optimization step of ACLWN. In section V, the framework is analyzed in terms of experimental results on different datasets, parameter sensitivity and convergence. Section VI concludes work.

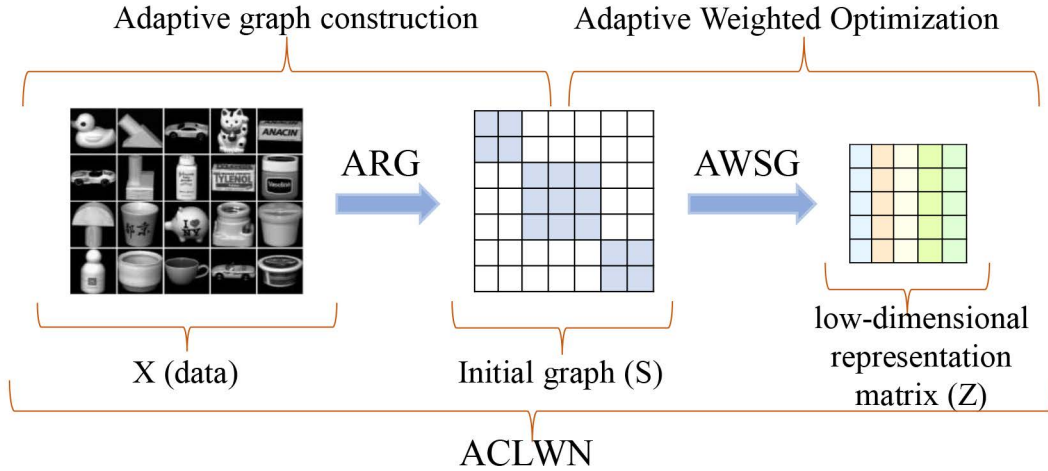


FIGURE 1. The framework of ACLWN model.

II. RELATED WORKS

In this section, we begin with a brief introduction to some of the symbolic conventions used throughout the paper, and then describe three types of traditional graph construction methods. Finally, two representation-based techniques are introduced, *i.e.*, representation-based subspace clustering (RSC) and representation-based classification methods (RC), which are most closely related to the method proposed in this paper.

A. NOTATIONAL CONVENTIONS

Throughout this paper, given a dataset $X = \{x_1, x_2, \dots, x_n\} \in \mathcal{R}^{d \times n}$ with d features and n instances, its j th column vector and (i, j) th element are denoted by x_j and x_{ij} , respectively. $\|X\|_p$ is the ℓ_p ($p = 1, 2, \{1, 2\}, F$) norm constraints of matrix X , and some typical norm are represented as $\|X\|_1 =$

$$\max_{1 \leq j \leq n} \sum_{i=1}^d |x_{ij}|, \|X\|_F = \sqrt{\sum_{i=1}^d \sum_{j=1}^n x_{ij}^2}, \|X\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^d x_{ij}^2}$$

and $\|X\|_2 = \sqrt{\lambda_{\max}(X^T X)} = \sqrt{\max_{1 \leq i \leq d} |\lambda_i|}$, where X^T is the transposed conjugate matrix of X , λ is the eigenvalue of $X^T X$. $I \in \mathcal{R}^{d \times n}$ is an identity matrix, and $\mathbf{1} \in \mathcal{R}^{d \times 1}$ signifies a column vector with all entries equal to one. $\text{Tr}(X)$ and X^T are the trace and transposition of matrix X , respectively. $\alpha \odot \beta$ denotes the element-wise multiplication of the vector α and β .

B. GRAPH CONSTRUCTION

The traditional graph learning methods [37], [38] convert the dataset X into graph as follows.

- Complete graph. It constructs edges between each point x_i and all the other points in X .
- k -nearest graph [14], [15]. It constructs edges between x_i and its k closest neighbors for each point x_i in X .

Assume S is the initial graph, and each node represents a point. If nodes i and j share an edge connection, three classic definitions of similarity s_{ij} are as follows.

- Binary (0–1) similarity: $s_{ij} = 1$.
- Cosine similarity [9]: $s_{ij} = \frac{(x_i)^T x_j}{\|x_i\| \|x_j\|}$.
- Gaussian kernel similarity [10]: $s_{ij} = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma}\right)$, where σ is the scale parameter.

These measures have some limitations. For example, although the binary similarity is simple, it cannot reflect the similarity between complex data. Cosine similarity cannot take into account the local geometric structure of data. Gaussian kernel similarity is a distance-based measure, which is sensitive to noise, outliers and redundant features.

C. REPRESENTATION-BASED SUBSPACE CLUSTERING

In this paper, representation-based subspace clustering (RSC) is represented by sparse, low-rank constraints and others techniques. Its general framework [17], [21], [39] is defined as

$$\min_{Z, E} \Theta(Z) + \lambda \psi(E) \quad s.t. \quad X = XZ + E, \quad (1)$$

where $\Theta(Z)$ is the regularization term of the variable Z , and $\psi(E)$ represents noise and outliers constrained by various norms. E is the reconstruction error term, while λ is the corresponding regularization parameter of $\psi(E)$. Formula (1) can learn a low-dimensional representation matrix $A \in \mathcal{R}^{n \times c}$, which is useful for clustering. It can also to extract intrinsic geometric structural information from high-dimensional complex data. For various RSC algorithms, their differences lie in the choice of $\Theta(Z)$. Usually, the final clustering result of RSC is obtained by using the following four steps: 1) create the initial graph matrix $M \in \mathcal{R}^{n \times n}$, where m_{ij} denotes the similarity between x_i and x_j ; 2) construct the Laplace matrix $\mathcal{L}_M = D_M - (M^T + M)/2$ of the similarity matrix M , where $D_M = \sum_j (m_{ij} + m_{ji})/2$; 3) compute $\min_{A \in \mathcal{R}^{n \times c}} \text{Tr}(A^T \mathcal{L}_M A)$ to get the low-dimensional representation matrix A ; 4) use k -Means to partition the low-dimensional representation matrix A into c clusters.

D. REPRESENTATION-BASED CLASSIFICATION METHODS

Representation-based classification algorithms are based on the assumption that points in the same class as the test point contribute more than those in other class points in the joint linear representation of the test point. Based on the above assumptions, many famous representation-based supervised classification methods have been proposed, such as sparse representation classification (SRC) [40], cooperative representation classification (CRC) [41], regularization robust coding (RRC) [42], and local constrained linear coding (LLC) [43]. The representation-based classification (RC) [44], [45], [46] method can be uniformly abstracted as

$$\min_{\alpha} \|r \odot (Q - P\alpha)\|_2^2 + \lambda\varphi(d \odot \alpha), \quad (2)$$

where P and Q represent train set and test set respectively. λ denotes the regularization parameter. Under various norm constraints, $\varphi(d \odot \alpha)$ is regarded as the regularization term of α . Prior knowledge is denoted by vectors r and d respectively. The difference between all kinds of RC methods is that they choose the different regularization terms of $\varphi(d \odot \alpha)$ as well as the parameters r and d .

III. ADAPTIVE GRAPH CONSTRUCTION AND LOW-RANK REPRESENTATION OF WEIGHTED NOISE (ACLWN)

The low-dimensional representation matrix can be realized in further optimization learning if the initial graph with high quality is obtained at the early stage, as described above. This section focuses on presenting the graph learning clustering framework for adaptive graph construction and weighted noise low-rank representation (ACLWN).

A. ADAPTIVE REPRESENTATION GRAPH CONSTRUCTION MODEL (ARG)

Traditional initial graph construction methods, as described in Section II-B, suffer from many shortcomings. As a simple explanation for manifold learning, if we choose two very similar data points, they will also represent each other very similarly when constructing the initial graph [33], [34]. Sparse representation [35] is robust to noise and outliers. In this section, we propose a novel graph construction method that combining manifold learning and sparse representation as

$$\min_S \sum_{i,j=1}^n \|x_i - x_j\|_2^2 s_{ij} + \alpha \sum_{i=1}^n \|s_i\|_1$$

$$s.t. \ s_{ii} = 0, \quad 0 \leq s_{ij} \leq 1, \quad (3)$$

where S is the initial graph to be learned, and its i th column vector and (i, j) th element are denoted by s_i and s_{ij} , respectively.

Normally, we measure the similarity of two points by calculating the distance between them.

Here, the first item of Formula (3) calculates the distance between points, and adaptively chooses k -nearest neighbor of current point according to the distance information. In other words, it achieves the sparse representation by representing current point according to the principle of competitive

representation. The ℓ_1 - norm of the second term ensures the sparseness of the representation matrix columns and improves the contribution of high discriminative features to the representation matrix. We also impose constraints $s_{ii} = 0$ and $0 \leq s_{ij} \leq 1$ on the representation matrix to reduce the self-representation contribution of points.

We add the $1^T s_i = 1$ constraint to Formula (3) to ensure that all points receive equal attention in the expression. It leads to the second term being a constant term, which is equal to a sparse constraint on S . Formula (3) can be transformed into

$$\min_S \sum_{i,j=1}^n \|x_i - x_j\|_2^2 s_{ij}$$

$$s.t. \ s_{ii} = 0, \quad 0 \leq s_{ij} \leq 1, \quad 1^T s_i = 1. \quad (4)$$

We denote that Formula (4) has a trivial solution, *i.e.*, x_i has a similarity value of 1 with x_j and 0 with the other points. To address this issue, we add an induction factor to Formula (4) so that the similarity value becomes $\frac{1}{n}$ between x_i and the other points,

$$\min_S \sum_{i,j=1}^n \|x_i - x_j\|_2^2 s_{ij} + \beta \sum_i \|s_i\|_2^2$$

$$s.t. \ s_{ii} = 0, \quad 0 \leq s_{ij} \leq 1, \quad 1^T s_i = 1. \quad (5)$$

B. ADAPTIVE WEIGHTED SPARSE REPRESENTATION GRAPH LEARNING MODEL (AWSG)

Both representation-based clustering and classification algorithms have demonstrated that the data representation matrix includes lots of discriminant information, as explained in Section II-C and Section II-D. To maximize the geometric structure of the data collected in the initial graph and eventually generate the low-dimensional representation matrix comprising discriminant information, we propose the following objective function of AWSG,

$$\min_Z \left\| W^{\frac{1}{2}} \odot (X - XZ) \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2 + \lambda_2 \|Z\|_{2,1}$$

$$s.t. \ W \geq 0, \quad W^T 1 = 1, \quad Z \geq 0, \quad (6)$$

where W is the weight matrix, with $W \geq 0$; Z is the low-dimensional representation matrix that has to be learnt; λ_1 and λ_2 are penalty parameters. Weight matrix W is used to regularize the data reconstruction error. The first two terms of Formula (6) adaptively assign high weights to significant elements and low weights to redundant features or outliers to improve the contribution of high discriminative features of important points in the representation matrix. Hence the method is robust to noise and outliers. Under the constraint ℓ_F - norm, the second term and the first term together constitute the Lasso problem, which is convenient for optimization to find the optimal solution. $W^T 1 = 1$ constraint can ensure all points to be treated equally. Usually, distance information is used to measure the local structure of

data. Therefore, we add the $\sum_{i,j=1}^n \|x_i - x_j\|_2^2$ constraint to Formula (6) in the process of obtaining a low-dimensional representation matrix. Suppose D denotes the distance matrix, $d_{ij} = \sum_{i,j=1}^n \|x_i - x_j\|_2^2$, then $\sum_{i,j=1}^n \|x_i - x_j\|_2^2 z_{ij} = \text{Tr}(D^T Z)$, Formula (6) can be converted to

$$\begin{aligned} \min_Z \left\| W^{\frac{1}{2}} \odot (X - XZ) \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2 \\ + \lambda_2 \|Z\|_{2,1} + \lambda_3 \text{Tr}(D^T Z) \\ \text{s.t. } W \geq 0, \quad W^T \mathbf{1} = 1, \quad Z \geq 0. \end{aligned} \quad (7)$$

To eliminate negative effects of low-dimensional representation matrix Z , such as all 0 for some rows of Z and self-representation of points, we introduce the $Z\mathbf{1} = 1$ and $\text{diag}(Z) = 0$ constraints to the Formula (7). Finally, we can derive that the objective function is

$$\begin{aligned} \min_{Z,W} \left\| W^{\frac{1}{2}} \odot (X - XZ) \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2 \\ + \lambda_2 \|Z\|_{2,1} + \lambda_3 \text{Tr}(D^T Z) \\ \text{s.t. } W \geq 0, \quad W^T \mathbf{1} = 1, \quad Z \geq 0, \\ Z\mathbf{1} = 1, \quad \text{diag}(Z) = 0. \end{aligned} \quad (8)$$

C. VERIFY THE FORMULA (6)

In this section, we prove that Formula (6) can generate a sparse non-negative and low-dimension representation graph.

Proposition 1: Suppose $E = X - XZ$ and the elements of each column of E are not all 0, then the optimization sub-problem is simplified to variable W , namely $\min_{Z,W} \left\| W^{\frac{1}{2}} \odot (X - XZ) \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2$, and Formula (6) generates a sparse weighted non-negative matrix W .

Proof: Define $D = E \odot E$, apparently, $D \geq 0$. Then, this problem $\min_{W \geq 0, W^T \mathbf{1} = 1} \left\| W^{\frac{1}{2}} \odot (X - XZ) \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2$ can be converted to $\min_{W \geq 0, W^T \mathbf{1} = 1} \|W\|_F^2 + \frac{2}{\lambda_1} \|D \odot W\|_1$. It is also the same as the n separate sub-problems $\min_{W \geq 0, W^T \mathbf{1} = 1} \|w_i\|_2^2 + \frac{2}{\lambda_1} \|d_i \odot w_i\|_1$ $i = 1, 2, \dots, n$, where d_i and w_i are the i th vectors of D and W respectively. Then, the $\min_{W \geq 0, W^T \mathbf{1} = 1} \|w_i\|_2^2 + \frac{2}{\lambda_1} \|d_i \odot w_i\|_1$ can be regarded as a variant of Lasso's issue [47], which can produce a sparse solution w_i . The penalty parameter $\frac{2}{\lambda_1}$ controls the sparsity. Thus, we can deduce that the sparse weighted matrix W will be produced by solving Formula (6).

Most notably, combining the regularization term $\frac{\lambda_1}{2} \|W\|_F^2$ and boundary constraints $W \geq 0$ and $W^T \mathbf{1} = 1$ can avoid a trivial solution of W [44], [48]. The $Z \geq 0$ constraint can ensure that the graph learned by similarity between points is interpretable. Furthermore, non-negative constraints can improve representation-based graph learning performance [49].

IV. OPTIMIZATION ALGORITHMS

A. OPTIMIZATION ALGORITHM FOR FORMULA (5)

According to the work of Nie et al. [50], we define that the j th column element d_{ij} of distance vector d_i is $d_{ij} = \|x_i - x_j\|_2^2$. Then Formula (5) can be rewritten as

$$\min_{s_i} \left\| s_i + \frac{d_i}{2\beta} \right\|_2^2 \quad \text{s.t. } s_{ii} = 0, \quad 0 \leq s_{ij} \leq 1, \quad \mathbf{1}^T s_i = 1. \quad (9)$$

The Lagrange function in Formula (9) about conditions $0 \leq s_{ij} \leq 1$ and $\mathbf{1}^T s_i = 1$ is written as

$$\mathcal{L}(s_i, \eta, \xi) = \frac{1}{2} \left\| s_i + \frac{d_i}{2\beta} \right\|_2^2 - \eta (\mathbf{1}^T s_i - 1) - \xi^T s_i, \quad (10)$$

where η is the scalar Lagrange coefficient, and ξ is the vector Lagrange coefficient.

We differentiate Formula (10) with respect to s_i and set the partial derivative to 0, then we can obtain the j th element s_{ij} of s_i as follows,

$$s_{ij} + \frac{d_{ij}}{2\beta} - \eta - \xi_i = 0. \quad (11)$$

By multiplying both sides of Eq.(11) by s_{ij} , and according to the Karush-Kuhn-Tucker (KKT) [51] condition that $s_{ij}\xi_j = 0$, we can obtain s_{ij} . Here, we denote it as \bar{s}_{ij} ,

$$\bar{s}_{ij} = \left(\frac{-d_{ij}}{2\beta} + \eta \right)_+. \quad (12)$$

Suppose $d_{i1}, d_{i2}, \dots, d_{in}$ are sorted in ascending order, and s_i needs to satisfy these conditions that $\bar{s}_{ik} > 0, \bar{s}_{ik+1} = 0$, and s_i contains at most k non-zero values, then we have

$$\frac{-d_{ik}}{2\beta} + \eta > 0 \quad \text{and} \quad \frac{-d_{ik+1}}{2\beta} + \eta \leq 0. \quad (13)$$

Impose constraint $\mathbf{1}^T s_i = 1$ on Inequality (13), we can derive

$$\sum_{j=1}^k \left(\frac{-d_{ik}}{2\beta} + \eta \right) = 1 \Rightarrow \eta = \frac{1}{k} + \frac{1}{2k\beta} \sum_{j=1}^k d_{ij}. \quad (14)$$

Learning the self-adaptive \bar{s}_i of k neighbors according to Formula (13) and Formula (14), β is represented by

$$\beta = \frac{k}{2} d_{ik+1} - \frac{1}{2} \sum_{j=1}^k d_{ij}. \quad (15)$$

According to Formula (13), Formula (14) and Formula (15), the j th element s_{ij} of s_i can be defined as

$$\bar{s}_{ij} = \begin{cases} \frac{d_{ik+1} - d_{ij}}{k} & j \leq k \\ kd_{ik+1} - \sum_{h=1}^k d_{ih} & \\ 0 & j > k \end{cases} \quad (16)$$

Finally, we summarize the pseudo code of ARG in Algorithm 1.

Algorithm 1: The Algorithm ARG

Input: Data matrix X , number of adaptive neighbors k
Output: Initial graph matrix S
for $i = 1$ **to** n **do**
 for $j = 1$ **to** k **do**
 Calculate the Lagrange function via Eq.(10);
 Calculate the Lagrange coefficient η via Eq.(14);
 Calculate the maximize parameter β via Eq.(15);
 Calculate the element s_{ij} in s_i via Eq.(16);
 return S ;

B. OPTIMIZATION ALGORITHM FOR FORMULA (8)

Formula (8) is a non-convex optimization problem with two unknown parameters, Z and W . Many methods can address this kind of problem, such as alternating direction method (ADM) [52] and accelerated proximal gradient (APG) [53]. Because of the characteristics of the issue and the efficacy of the ADM approach, we employ the ADM approach to optimally solve the Formula (8). To make the problem simpler, we introduce two auxiliary variables E and U , and then turn Formula (8) into

$$\begin{aligned} \min_Z & \left\| W^{\frac{1}{2}} \odot (X - XZ) \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2 \\ & + \lambda_2 \|Z\|_{2,1} + \lambda_3 \text{Tr}(D^T Z) \\ \text{s.t. } & W \geq 0, W^T 1 = 1, \quad Z \geq 0, Z1 = 1, \\ & \text{diag}(Z) = 0, E = X - XZ, U = Z. \end{aligned} \quad (17)$$

We first rewrite Eq.(17) into Augmented Lagrangian function [54]

$$\begin{aligned} \mathcal{L}(\Upsilon) = & \left\| W^{\frac{1}{2}} \odot E \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2 + \lambda_2 \|U\|_{2,1} \\ & + \lambda_3 \text{Tr}(D^T Z) + \frac{\mu}{2} \left(\left\| X - XZ - E + \frac{C_1}{\mu} \right\|_F^2 \right. \\ & \left. + \left\| Z - U + \frac{C_2}{\mu} \right\|_F^2 \right), \end{aligned} \quad (18)$$

where the penalty parameter is μ ($\mu > 0$), C_1 and C_2 are Lagrange multipliers, Υ is a constrained set, with $\Upsilon = \{Z, W, E, U, C_1, C_2\}$. Then, fixing the remaining variables, the optimal value of the current variable is iteratively obtained.

Step 1. Update W and fix the other variables:

$$\min_{W \geq 0, W^T 1 = 1} \left\| W^{\frac{1}{2}} \odot E \right\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2. \quad (19)$$

Since E is a fixed variable, Formula (19) can be converted into

$$\begin{aligned} \min_{W \geq 0, W^T 1 = 1} & \sum_{i=1}^m \sum_{j=1}^n \left(w_{ij} e_{ij}^2 + \frac{\lambda_1}{2} w_{ij}^2 \right) \\ \Leftrightarrow & \min_{W \geq 0, W^T 1 = 1} \sum_{i=1}^m \sum_{j=1}^n \left(w_{ij} + \frac{e_{ij}^2}{\lambda_1} \right)^2. \end{aligned} \quad (20)$$

Formula (20) and Formula (9) are optimized in a similar way, for further details, please see Formula (9). Finally, variable W can be calculated as follows,

$$w_i = \max \left(\eta_i 1 - \frac{1}{\lambda_1} f_i, 0 \right). \quad (21)$$

Because of $w_i^T 1 = 1$ constraint, we have

$$\sum_{j=1}^m \left(\eta_i 1 - \frac{1}{\lambda_1} f_{ij} \right) = 1 \Rightarrow \eta_i = \frac{1}{m} + \frac{1}{m\lambda_1} \sum_{j=1}^m f_{ij}. \quad (22)$$

Step 2. Update E and fix the other variables:

$$\min_E \left\| W^{\frac{1}{2}} \odot E \right\|_F^2 + \frac{\mu}{2} \left\| X - XZ - E + \frac{C_1}{\mu} \right\|_F^2. \quad (23)$$

Define $G = X - XZ + \frac{C_1}{\mu}$, Formula (23) can turn into

$$\begin{aligned} \min_E & \left\| W^{\frac{1}{2}} \odot E \right\|_F^2 + \frac{\mu}{2} \|E - G\|_F^2 \\ \Leftrightarrow & \min_{e_{ij}} \sum_{i=1}^m \sum_{j=1}^n \left(w_{ij} e_{ij}^2 + \frac{\mu}{2} (e_{ij} - g_{ij})^2 \right) \\ \Leftrightarrow & \sum_{i=1}^m \sum_{j=1}^n \min_{e_{ij}} \left(e_{ij} - \frac{\mu g_{ij}}{\mu + 2w_{ij}} \right)^2. \end{aligned} \quad (24)$$

We can obtain each optimization value of e_{ij} in e using Eq.(24),

$$e_{ij} = \frac{\mu g_{ij}}{\mu + 2w_{ij}}. \quad (25)$$

Step 3. Update U and fix the other variables:

$$\min_U \lambda_2 \|U\|_{2,1} + \frac{\mu}{2} \left\| Z - U + \frac{C_2}{\mu} \right\|_F^2. \quad (26)$$

Define $M = Z + \frac{C_2}{\mu}$, Formula (26) can be turned into

$$\min_U \lambda_2 \|U\|_{2,1} + \frac{\mu}{2} \|M - U\|_F^2. \quad (27)$$

We can obtain a closed solution by simplifying Formula (27),

$$U_{:,i}^* = \begin{cases} \frac{\|M_{:,i}\|_2 - \frac{\lambda_2}{\mu}}{\|M_{:,i}\|_2} M_{:,i} & \text{if } \|M_{:,i}\|_2 > \frac{\lambda_2}{\mu} \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

Step 4. Update Z and fix the other variables:

$$\begin{aligned} \min_Z & \lambda_3 \text{Tr}(D^T Z) + \frac{\mu}{2} \left(\left\| X - XZ - E + \frac{C_1}{\mu} \right\|_F^2 \right. \\ & \left. + \left\| Z - U + \frac{C_2}{\mu} \right\|_F^2 \right) \\ \text{s.t. } & \text{diag}(Z) = 0, \quad Z \geq 0, \quad Z1 = 1. \end{aligned} \quad (29)$$

Find a latent solution \hat{Z} by minimizing the following Formula (30)

$$\hat{Z} = \left(X^T X + I \right)^{-1} \left(X^T * \left(X - E + \frac{C_1}{\mu} \right) \right. \\ \left. + U - \frac{C_2 + \lambda_3 * D}{\mu} \right), \quad (30)$$

then we can obtain a closed solution as

$$\min_{\text{diag}(Z)=0, Z \geq 0, Z1=1} \|Z - \hat{Z}\|_F^2. \quad (31)$$

Referring to Formula (20), the optimization process of Formula (31) can be solved,

$$z_i = \max\left(\zeta_i \bar{1}^T + \bar{z}_i, 0\right), \quad (32)$$

where the Lagrange multiplier is ζ_i , with $\zeta_i = (n-1)^{-1}(1 - \bar{z}_i)$; \bar{z}_i is the i th row of \hat{Z} . Eventually, we can obtain an ideal Z by plugging the value of ζ_i into Formula (32).

Step 5. Update C_1 , C_2 and μ , and fix the other variables:

$$\begin{aligned} C_1 &= C_1 + \mu(X - XZ - E), \\ C_2 &= C_2 + \mu(Z - U), \\ \mu &= \begin{cases} \min(\mu_{\max}, \rho\mu), & \text{if } \pi < 0.01 \\ \mu, & \text{otherwise,} \end{cases} \end{aligned} \quad (33)$$

where $\pi = \max\left(\frac{\|Z_k - Z_{k-1}\|_F, \|U_k - U_{k-1}\|_F, \|E_k - E_{k-1}\|_F}{\|X\|_F}, Z_k, U_k, E_k\right)$ and $Z_{k-1}, U_{k-1}, E_{k-1}$ are the value of Z, U, E at the k th iteration (the current step) and $(k-1)$ th iteration (the previous step), respectively. μ_{\max} and ρ are positive parameters.

Finally, we summarize our AWSG in Algorithm 2.

Algorithm 2: The Algorithm AWSG

Input: Initial graph matrix S and penalty parameters λ_1, λ_2 , and λ_3 ;

Output: W and Z ;

Initialize: $W, Z, E = X - XZ, U,$

$C_1 = C_2 = 0, \rho, \mu, \mu_{\max}$;

while not converged do

 Update W via Eq.(21);

 Update E via Eq.(25);

 Update U via Eq.(28);

 Update Z via Eq.(32);

 Update C_1, C_2 and μ via Eq.(33);

return W and Z ;

C. TIME COMPLEXITY

When constructing the initial graph in Algorithm 1, it only calculates the distance vectors between the current data and its k -nearest neighbors, where $k \ll n$ and n is the number of points. Hence its time complexity can be ignored compared with the other steps. Obviously, there are five steps in Algorithm 2. Since steps 1, 2, and 5 are considered the element-wise operations, they are quickly solved and have low time complexity. The matrix inversion process in step 4 has high time complexity. However, because the inversion process is performed outside the iterative loop in our method, the time complexity is minimal. In Step 3, the procedure of iteratively updating U has a high time complexity, $O(t \cdot n^2)$, where t is the number of iterations. Therefore, the overall time complexity of ACLWN is $O(t \cdot n^2)$.

TABLE 1. Description of datasets.

	Datasets	Instances	Dimensions	Clusters	
Object	COIL20	1440	1024	20	
	Face	YaleB	2414	1024	38
		ORL	400	1024	40
		MSRA25	1799	256	12
UCI	Cars	392	8	3	
	Vehicle	846	18	4	
	Yeast	1484	8	10	
	UCSC	Cora	2708	1433	7
Handwritten Digit	Mnist	1000	784	10	
	Handwritten	2000	216	10	

V. EXPERIMENTS AND ANALYSIS

A. EXPERIMENTS ENVIRONMENT

In this section, we compare the performance of ACLWN to those of its baseline methods on face, object and non-image datasets, and quantify it with two common evaluation measures. The baseline methods include k -Means [7], Ncut [55], SSC [17], LRR [21], NSHLRR [56], AWNLRR [32], FTRR [57] and SGL [58]. Among them, k -Means directly runs on the original features, while the other methods cluster by learning various representation graphs from datasets. In the experiments, Ncut constructs the adjacency graph through gaussian kernel function. SSC, LRR, AWNLRR and NSHLRR perform the spectral clustering on the learned graphs. AWNLRR and NSHLRR utilize the k -nearest graph method to construct the initial graph. FTRR uses a low-pass filter to obtain similarity matrix and then uses k -Means to cluster. SGL is one of the latest graph learning methods. Moreover, since k -Means is unstable, we run it 30 times and then calculated its mean values of the measures as its final results. These experiments are run on Windows 10 system and MATLAB 2019a, hardware platform of Intel(R) Core(TM) i7-10510U CPU and 20GB RAM.

B. DATASETS

1) Object datasets: We chose COIL20 from the Columbia University Object dataset.¹ As shown in Figure 2(a), the dataset includes 1440 32×32 gray-scale images of 20 distinct points. To verify the proposed clustering performance of model on the dataset with different numbers of object classes, we use the top k subsets of the dataset for experiments, with $k = 4, 6, \dots, 20$.

2) Face datasets: Face image dataset is the most typical image dataset for performance testing of graph learning clustering. We chose three common face datasets in this paper, which are Extended YaleB,² ORL³ and MSRA25 [50]. As shown in Figure 2(c), YaleB contains 2414 images of 38 faces in various lighting conditions; ORL includes 400 faces of 40 people in various times, illuminations and expressions; MSRA25 contains 1799 images of faces in

¹<http://www.cs.columbia.edu/CAVE/software/softlib>

²<http://vision.ucsd.edu/content/yale-face-database>

³<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

12 categories. For the convenience of the experiment, the image is uniformly processed as a 32×32 gray-scale image.

3) UCI Datasets:⁴ To test the performance of our model, we chose three UCI datasets, which are Cars, Vehicle and Yeast.

4) Handwritten Digit Datasets: As shown in Figure 2(b), we chose the Mnist [57] dataset and the Profile view of the Handwritten [59] datasets, with 10 classes ranging from “0” to “9” and 216 pixel gray levels.

5) UCSC Datasets: We chose the Cora dataset from the UCSC⁵ website to evaluate the clustering performance of ACLWN and its baseline methods, and then turned it into a form suitable for graph learning clustering. The dataset includes 2708 scientific papers split into seven categories, which are case-based, genetic algorithm, neural network, probabilistic technique, reinforcement learning, rule learning, and theory. Each paper has a vector of 1433 words, representing the 1433 features.

C. EVALUATION METRICS

Accuracy and normalized mutual information (*NMI*) are two standard measures that are used to quantitatively evaluate clustering performance. The ground truth and predicted labels are denoted by Y and P , respectively.

Accuracy is defined as

$$Accuracy(Y, P) = \frac{\sum_i \tau(\text{map}(P_i) = Y_i)}{n}, \quad (34)$$

where $\tau(\cdot)$ denotes indicator function, P is mapped to its best group label using Kuhn-Munkres method.

Normalized mutual information is defined as

$$NMI(Y, P) = \frac{I(Y, P)}{\sqrt{H(Y) \cdot H(P)}}, \quad (35)$$

where $I(\cdot)$ stands for mutual information, while $H(\cdot)$ stands for information entropy. We can realize standardization by calculating the information gain.

D. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results of different clustering methods on the datasets introduced in Section V-B are quantitatively evaluated by *Accuracy* and *NMI*. The clustering *Accuracys* are listed in Tables 2-5, where the bold black numbers indicate the best results. Tables 2-4 shows the results of our model ACLWN and its baseline methods on the top k subsets of COIL20, YaleB, and ORL datasets, respectively. Table 5 summarizes the clustering *Accuracys* of different methods on the rest of the datasets. Figure 3 illustrates the *NMI* results of our model and its baseline methods on all the datasets. It can be found that our model ACLWN has achieved the best results on these datasets. Specific analysis as follows:

(1) It is clear from Tables 2-5 and Figure 3 that the clustering *Accuracys* of k -Means are generally lower than those

of the representation-based clustering models. Therefore, we can draw a conclusion that representation-based methods are superior to the methods clustering directly on the original datasets. Since the representation-based methods eliminate redundant features to some extent, the representation matrixes obtained by these methods are more discriminating.

(2) Table 2, Table 4 and Figure 3 exhibit the comparison of *Accuracy* and *NMI* of ACLWN and its baseline methods on COIL20 and ORL datasets. Ncut outperforms the representation-based methods, SSC, LRR, NSHLRR, SGL and AWNLRR in most cases. It is because Ncut is a distance-based method which captures more local structure of data than those of methods based on sparse and low-rank representation.

(3) As can be found from Figure 3 and Table 4 that SSC and LRR have lower clustering *Accuracy* and *NMI* in most cases compared to NSHLRR, SGL, AWNLRR and ACLWN. Since SSC employs sparse representation, it only captures the local structure of data. LRR only obtains the global structure of data by using low-rank representation. As a result, SSC and LRR cannot capture both global and local data structures, and their final representation graphs are weaker than those of other methods.

(4) Although NSHLRR, SGL, AWNLRR and ACLWN all use the non-negative sparse low-rank representation, AWNLRR and ACLWN employ a weight matrix to efficiently reduce the representation of noise and outliers while improving the weight value of important features. Besides, the k -nearest neighbor method based on Euclidean distance is still used in the initial graph construction of AWNLRR, hence its experimental results are not perfect. To produce a more robust initial graph, ACLWN uses an adaptive neighbor graph construction method based on the local structures of data. Therefore, as shown in Tables 2-5 and Figure 3, our method has the best clustering *Accuracy* and *NMI* results on the various datasets.

(5) We also select the most recent method, FTRR, to further prove that the effectiveness of our adaptive graph construction method is superior to most other graph construction methods. In FTRR, a low-pass filter construction method is used to create the representation graph. As demonstrated in Tables 2-5 and Figure 3, clustering *Accuracy* and *NMI* of our method are generally greater than FTRR, indicating that our method is still better than the method FTRR.

Therefore, ACLWN has the following advantages in graph representation: 1) ACLWN jointly employs sparse, low-rank, distance regular terms and non-negative constraints to obtain the geometric structure of data. It is the efficient integration of these constraints that guides the generation of low-dimensional representation graphs suitable for clustering. 2) ACLWN uses the weight matrix to eliminate noise and outliers while increasing the weight of important features, resulting in a more accurate low-dimensional representation graph. 3) ACLWN uses manifold learning and sparse representation graph construction methods to produce a high-quality initial graph.

⁴<http://www.escience.cn/people/fpnie/papers.html>

⁵<https://linqs.so.e.ucsc.edu/data>



(a) Typical images of the object dataset.

(b) Typical images of the handwritten digit dataset.



(c) Typical images of the faces dataset.

FIGURE 2. Typical images of the datasets.

TABLE 2. Clustering Accuracy(%) of different methods on the subsets of COIL20 containing different numbers of clusters.

k	k -Means	Ncut	SSC	LRR	NSHLRR	SGL	AWNLR	FTRR	Ours
4	67.01	61.81	67.36	64.93	84.72	81.25	100	99.65	100
6	47.92	47.22	56.25	54.17	66.44	67.13	87.27	82.87	95.6
8	43.23	60.59	58.68	55.21	50.69	62.85	85.07	77.95	94.97
10	45.42	66.53	59.03	53.89	45.00	65.56	82.50	83.61	90.14
12	55.79	61.11	64.00	60.30	52.78	69.91	81.71	84.72	92.82
14	57.94	67.06	63.39	60.12	47.02	69.44	87.20	86.31	92.56
16	61.55	67.88	60.94	63.28	52.52	73.44	90.20	88.89	92.45
18	66.05	71.45	65.05	62.27	55.17	75.54	91.90	88.66	92.75
20	64.58	74.93	62.43	59.51	52.57	72.57	83.47	82.85	84.10

TABLE 3. Clustering Accuracy(%) of different methods on the subsets of YaleB containing different numbers of clusters.

k	k -Means	Ncut	SSC	LRR	NSHLRR	SGL	AWNLR	FTRR	Ours
2	50.78	93.75	99.22	54.69	100	95.31	99.22	100	99.22
8	17.19	46.88	90.63	36.52	58.59	34.96	87.50	90.23	92.58
14	14.29	49.77	86.85	41.16	26.64	24.49	96.03	88.44	97.62
20	10.22	48.34	83.20	33.04	13.71	23.61	90.10	86.69	90.49
26	9.48	49.27	77.76	38.40	23.57	16.34	91.56	86.39	92.59
32	10.00	46.95	73.69	46.35	26.06	17.34	91.63	80.34	93.20
38	8.33	44.53	64.79	46.64	10.11	16.61	87.95	78.13	90.02

E. PARAMETER SENSITIVITY ANALYSIS

There are four parameters in ACLWN model, including the adaptive number of neighbors k in ARG method and the three penalty parameters, λ_1 , λ_2 and λ_3 in AWSG method. This section analyzes the influence of the four parameters on the clustering accuracy of YaleB, COIL20 and Cora datasets.

Figure 4 shows the relationship between the number of nearest neighbors k and the clustering accuracy on YaleB, COIL20 and Cora datasets when the penalty parameters λ_1 , λ_2 and λ_3 are fixed. In Figure 4(a), when k takes different values, the minimum and the maximum accuracies are 87.49% and 88.77% respectively. It is only 1.28% difference

TABLE 4. Clustering Accuracy(%) of different methods on the subsets of ORL containing different numbers of clusters.

k	k -Means	Ncut	SSC	LRR	NSHLRR	SGL	AWNLR	FTRR	Ours
5	54.00	72.00	66.00	70.00	80.00	80.00	78.00	80.00	80.00
10	55.00	65.00	72.00	63.00	69.00	76.00	73.00	76.00	78.00
15	54.67	71.33	63.33	64.00	60.00	78.00	76.00	77.33	78.00
20	51.50	62.50	64.00	66.00	71.00	69.50	69.00	72.00	76.00
25	51.20	64.00	58.00	64.80	47.20	65.60	70.40	70.00	72.80
30	53.67	66.67	61.33	65.33	52.33	67.67	71.67	71.33	75.33
35	53.43	69.14	60.00	64.00	31.71	60.57	71.14	72.00	73.43
40	55.50	68.50	61.50	63.75	31.00	62.00	70.50	70.25	75.75

TABLE 5. Clustering Accuracy(%) of different methods on the other datasets.

Datasets	k -Means	Ncut	SSC	LRR	NSHLRR	SGL	AWNLR	FTRR	Ours
Cars	20.99	63.18	62.76	62.76	63.01	61.48	66.07	60.71	66.58
Vehicle	30.60	46.10	48.23	41.73	40.54	41.02	48.70	43.74	50.71
Yeast	25.52	34.84	37.60	40.16	40.03	40.57	46.56	42.79	47.10
MSRA25	10.82	54.14	51.53	57.09	53.25	58.48	56.86	61.76	79.43
Handwrite	67.20	75.80	73.20	58.75	57.40	69.30	73.35	85.00	90.67
Cora	14.00	27.95	30.06	38.37	29.84	43.02	30.54	—	53.36
Mnist	55.10	10.44	43.00	54.00	56.89	64.30	59.11	—	71.00

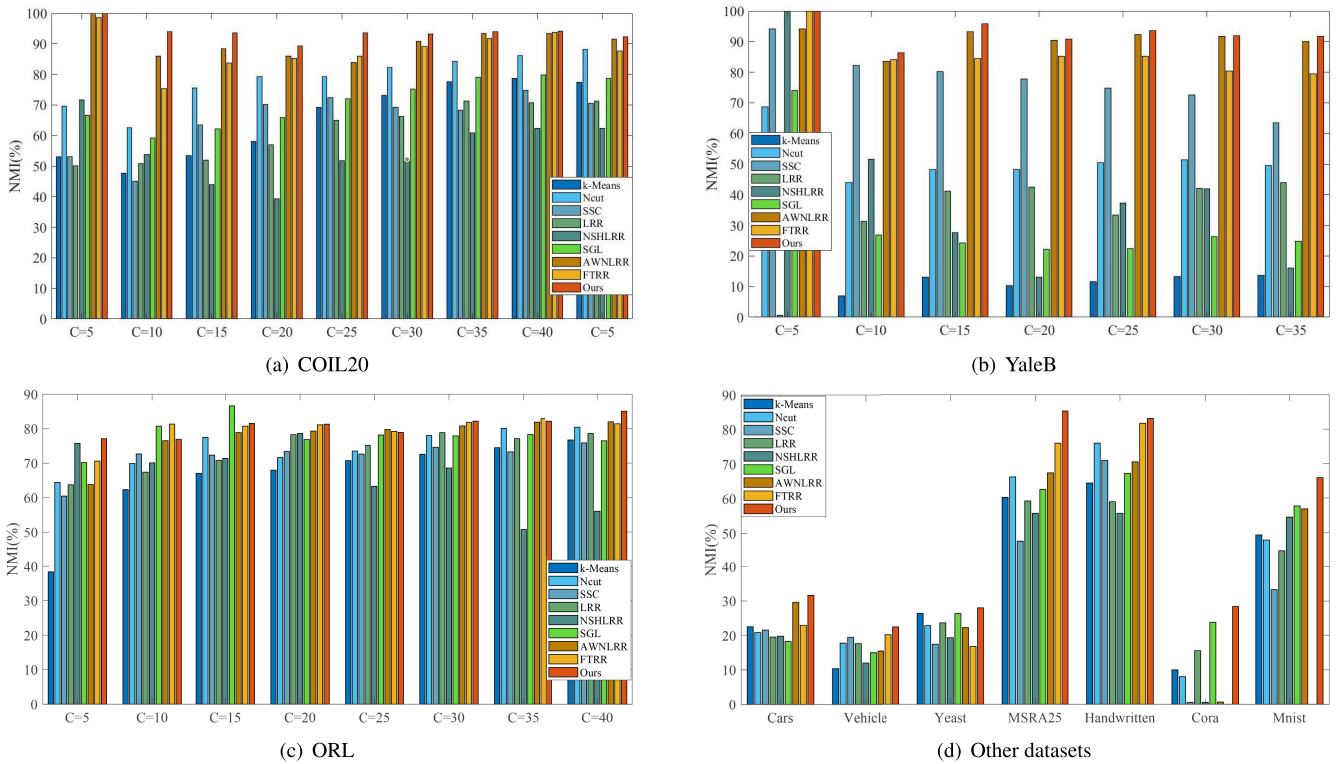


FIGURE 3. Clustering NMI(%) of our method and baseline methods on the (a) COIL20, (b) YaleB, (c) ORL, and (d) other datasets.

between the minimum accuracy and the maximum accuracy on YaleB dataset. Besides, the difference between the minimum accuracy and the maximum accuracy is 3.96% and 7.34% on COIL20 and Cora datasets respectively. As a result, the adaptive nearest neighbor number k has little influence on the representation of the current data, and the clustering accuracy generally tends to be stable when k has different values. That is, ARG is highly insensitive to the nearest neighbors number k . We recommend setting k within the

range of [8, 14] to ensure high clustering accuracy and lower time complexity.

Next, we analyze the sensitivity of the parameters, λ_1 , λ_2 and λ_3 , in our method. For the sake of illustration, we define the parameter coarse set Λ as $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4\}$. λ_1 acts as the harmonic parameter of the weight matrix W , which avoids meaningless trivial solutions. Figures 5, 6 and 7 show the relationship between the three parameters in our method

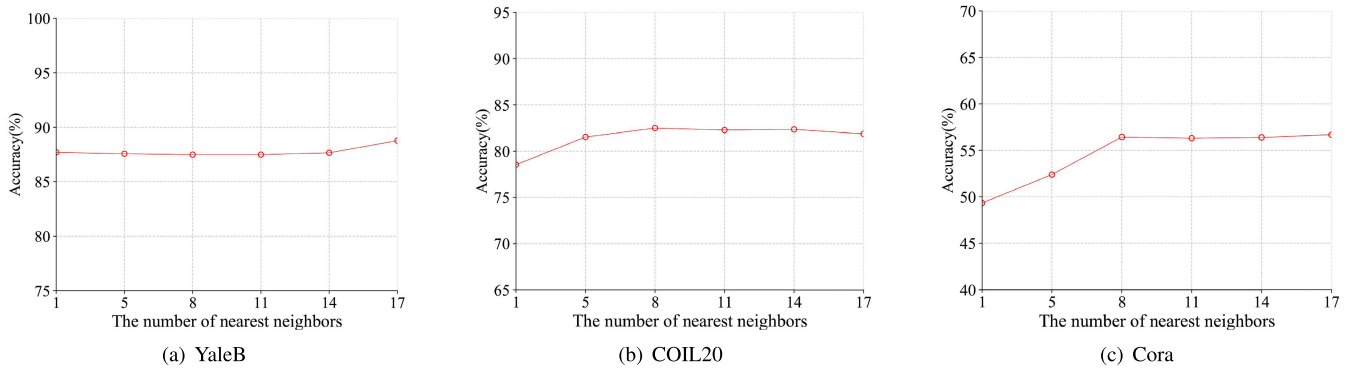


FIGURE 4. The relationship between the number of nearest neighbours k and clustering Accuracy on the YaleB, COIL20 and Cora datasets.

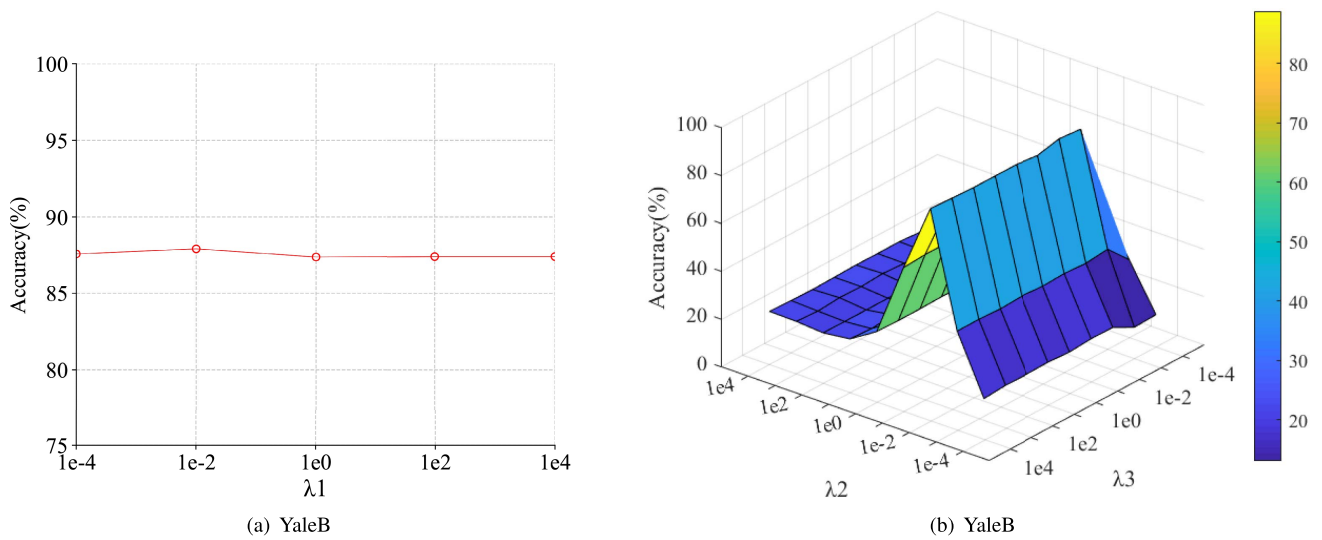


FIGURE 5. The relationship between different values of parameters (λ_1 , λ_2 and λ_3) and clustering Accuracy on the YaleB dataset.

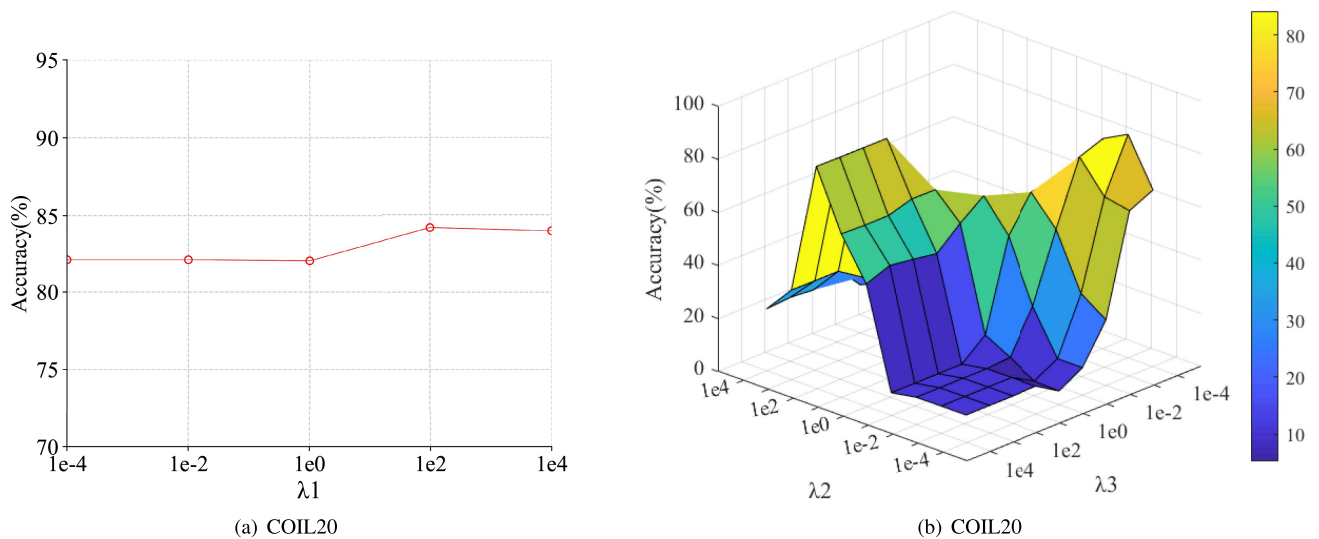


FIGURE 6. The relationship between different values of parameters (λ_1 , λ_2 and λ_3) and clustering Accuracy on the COIL20 dataset.

and the clustering accuracy on YaleB, COIL20 and Cora datasets. From Figures 5(a), 6(a) and 7(a), we can see that

when the parameters λ_2 and λ_3 are fixed, the clustering accuracy is less affected by the selection of parameter λ_1 in

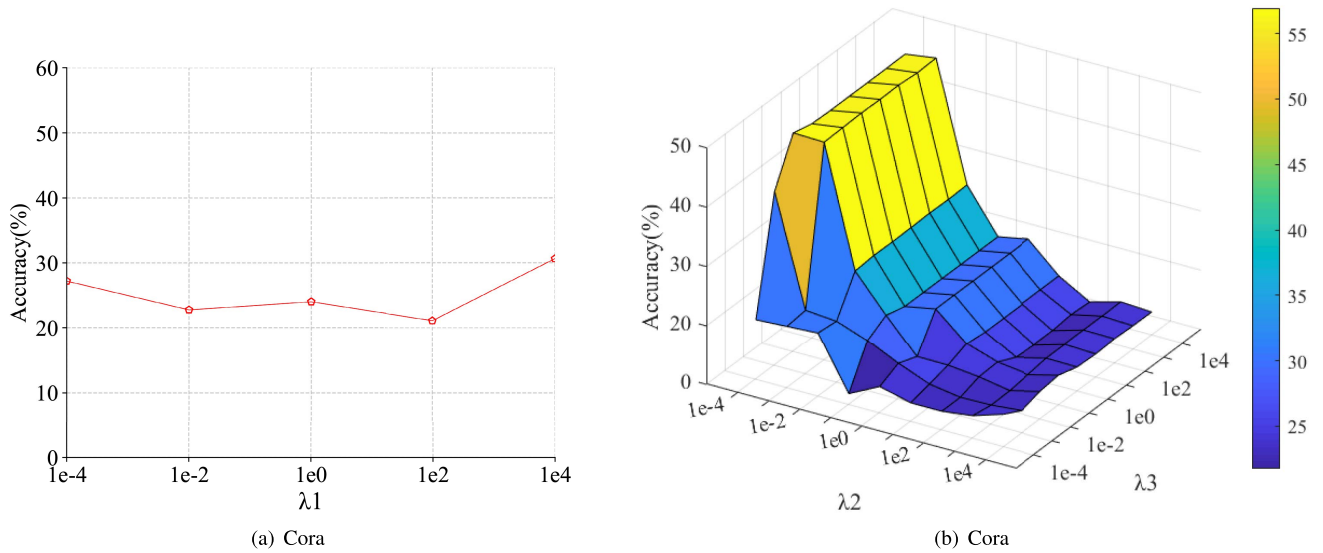


FIGURE 7. The relationship between different values of parameters (λ_1 , λ_2 and λ_3) and clustering Accuracy on the Cora dataset.

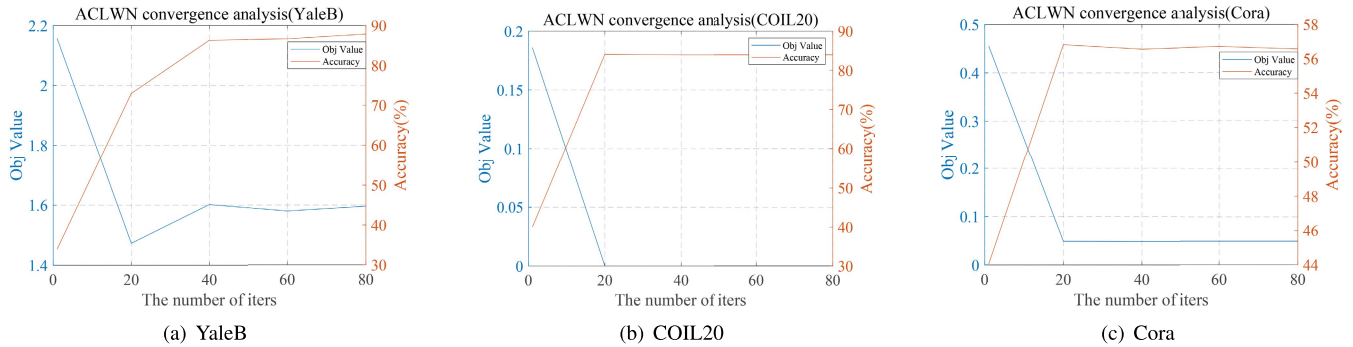


FIGURE 8. The changing trend of objective function value and clustering Accuracy of the ACLWN on the YaleB, COIL20 and Cora datasets with the increase of iteration times.

the coarse set Λ . Hence, λ_1 is very insensitive to our method. In comparison with parameter λ_1 , parameters λ_2 and λ_3 have a huge influence on clustering performance. As shown in Figures 5(b), 6(b) and 7(b), when fixing λ_1 and choosing various combination parameters of λ_2 and λ_3 , clustering accuracy are not always maximal, and the gap between the maximum clustering accuracy and the minimum clustering accuracy is irregular. This occurs because the parameter directly affects the quality of the corresponding representation matrix, and λ_2 and λ_3 directly determine the function of the corresponding term in the process of learning graphs. Thus, these three parameters need to be optimized to achieve the best clustering performance. Finding common optimal values for all three parameters is challenging due to the diversity of datasets. A simple and effective method for determining the optimal solution is presented in this paper. Based on the previous analysis, first we can fix λ_1 to a value such as 1, then find the ideal combination of λ_2 and λ_3 from the coarse set Λ . According to the best combination of the two parameters, we can further define an ideal candidate

set which may contain the optimal values for these two parameters. The method is then rerun by selecting a variety of combinations of the two parameters from the optimal candidate set. Only in this manner, we can determine the ideal penalty parameter to ensure the optimal clustering performance.

F. CONVERGENCE ANALYSIS

In this section, we demonstrate the convergence of the ACLWN using the YaleB, COIL20 and Cora datasets by iterating it. The objective function value of ACLWN is $obj = \|W^{\frac{1}{2}} \odot E\|_F^2 + \frac{\lambda_1}{2} \|W\|_F^2 + \lambda_2 \|U\|_{2,1} + \lambda_3 Tr(D^T Z) + \|X - XZ - E\|_F^2 + \|Z - U\|_F^2$. In Figure 8(a), obj decreases monotonically and rapidly in the first few iterations, then briefly increases, and finally trends towards stability. In Figures 8(b) and 8(c), obj decreases before reaching equilibrium and then shows a convergence tendency. It can also be shown in Figures 8(a), 8(b) and 8(c) that clustering accuracy increases significantly at first and after to be stable trends.

Therefore, ACLWN has good convergence, which enables it to obtain the local optimal solution.

VI. CONCLUSION

The present paper proposes a new graph learning clustering framework ACLWN, which consists of two parts, *i.e.*, ARG and AWSM. ARG is used to generate the initial graph adaptively with the local structure of data. AWSG introduces an adaptive weight matrix into the graph optimization process to effectively eliminate representations of redundant features, noise and outliers. ACLWN jointly employs sparse representation, low-rank representation, distance regularization term and non-negative constraints to obtain the geometric structure of data, so as to form a low-dimensional graph with more discriminative characteristics. Experiments on objects, faces, UCI non-image and the other datasets show that the performance of our ACLWN is better than those of the baseline clustering methods.

REFERENCES

- [1] S. Huang, Z. Kang, and Z. Xu, "Auto-weighted multi-view clustering via deep matrix decomposition," *Pattern Recognit.*, vol. 97, Jan. 2020, Art. no. 107015.
- [2] X. Shen, W. Liu, I. Tsang, F. Shen, and Q.-S. Sun, "Compressed k -means for large-scale clustering," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2528–2530.
- [3] Z. Lin, Z. Kang, L. Zhang, and L. Tian, "Multi-view attributed graph clustering," *IEEE Trans. Knowl. Data Eng.*, early access, Aug. 6, 2021, doi: 10.1109/TKDE.2021.3101227.
- [4] K. K. Sharma and A. Seal, "Multi-view spectral clustering for uncertain objects," *Inf. Sci.*, vol. 547, pp. 723–745, Feb. 2021.
- [5] K. Xia, X. Gu, and Y. Zhang, "Oriented grouping-constrained spectral clustering for medical imaging segmentation," *Multimedia Syst.*, vol. 26, no. 1, pp. 27–36, Feb. 2020.
- [6] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 101–114, Jan. 2021.
- [7] J. Cai, D. Liao, J. Chen, X. Chen, T. Liu, and J. Xi, "Research on data security protection method based on improved K -means clustering algorithm," in *Proc. 4th Int. Conf. Big Data Res. (ICBDR)*, Nov. 2020, pp. 7–11.
- [8] M. Nishom, "Perbandingan akurasi Euclidean distance, Minkowski distance, dan Manhattan distance pada algoritma K -means clustering berbasis chi-square," *Jurnal Informatika, Jurnal Pengembangan IT*, vol. 4, no. 1, pp. 20–24, Jan. 2019.
- [9] J. Yin and S. Sun, "Incomplete multi-view clustering with cosine similarity," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108371.
- [10] X. Ding, J. Liu, F. Yang, and J. Cao, "Random compact Gaussian kernel: Application to ELM classification and regression," *Knowl.-Based Syst.*, vol. 217, Apr. 2021, Art. no. 106848.
- [11] Y. Yun, W. Xia, Y. Zhang, Q. Gao, and X. Gao, "Self-representation and class-specificity distribution based multi-view clustering," *Neurocomputing*, vol. 437, pp. 9–20, May 2021.
- [12] R. Shang, Z. Zhang, L. Jiao, C. Liu, and Y. Li, "Self-representation based dual-graph regularized feature selection clustering," *Neurocomputing*, vol. 171, pp. 1242–1253, Jan. 2016.
- [13] L. Weng, F. Dornaika, and Z. Jin, "Graph construction based on data self-representativeness and Laplacian smoothness," *Neurocomputing*, vol. 207, pp. 476–487, Sep. 2016.
- [14] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [15] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [16] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [17] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
- [18] P. Zhou, Z. Lin, and C. Zhang, "Integrated low-rank-based discriminative feature learning for recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 1080–1093, May 2015.
- [19] H. Du, L. Ma, G. Li, and S. Wang, "Low-rank graph preserving discriminative dictionary learning for image recognition," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104823.
- [20] L. Bai and J. Liang, "Sparse subspace clustering with entropy-norm," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 561–568.
- [21] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [22] L. Wang, Z. Zhang, S. Li, G. Liu, C. Hou, and J. Qin, "Similarity-adaptive latent low-rank representation for robust data representation," in *Proc. Pacific Rim Int. Conf. Artif. Intell.* Cham, Switzerland: Springer, 2018, pp. 71–84.
- [23] Z. Liu, W. Ou, W. Lu, and L. Wang, "Discriminative feature extraction based on sparse and low-rank representation," *Neurocomputing*, vol. 362, pp. 129–138, Oct. 2019.
- [24] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1615–1622.
- [25] L. Xie, M. Yin, X. Yin, Y. Liu, and G. Yin, "Low-rank sparse preserving projections for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5261–5274, Nov. 2018.
- [26] X. Yang, X. Jiang, C. Tian, P. Wang, F. Zhou, and H. Fujita, "Inverse projection group sparse representation for tumor classification: A low rank variation dictionary approach," *Knowl.-Based Syst.*, vol. 196, May 2020, Art. no. 105768.
- [27] J. Lu, H. Wang, J. Zhou, Y. Chen, Z. Lai, and Q. Hu, "Low-rank adaptive graph embedding for unsupervised feature extraction," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107758.
- [28] W. Zhu and B. Peng, "Sparse and low-rank regularized deep subspace clustering," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106199.
- [29] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang, and Y. Fang, "Low-rank sparse subspace for spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1532–1543, Aug. 2019.
- [30] Y.-N. Han, J.-W. Liu, and X.-L. Luo, "An improved latent low rank representation for automatic subspace clustering," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 5188–5189.
- [31] M. Li, H. Zhou, Q. Liu, and G. Wang, "SW: A weighted space division framework for imbalanced problems with label noise," *Knowl.-Based Syst.*, vol. 251, Sep. 2022, Art. no. 109233.
- [32] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted non-negative low-rank representation," *Pattern Recognit.*, vol. 81, pp. 326–340, Sep. 2018.
- [33] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
- [34] H. Wang, Y. Yang, and T. Li, "Multi-view clustering via concept factorization with local manifold regularization," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 1245–1250.
- [35] J. Xu, W. An, L. Zhang, and D. Zhang, "Sparse, collaborative, or nonnegative representation: Which helps pattern classification?" *Pattern Recognit.*, vol. 88, pp. 679–688, Apr. 2019.
- [36] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 1–9.
- [37] S. Huang, Z. Kang, I. W. Tsang, and Z. Xu, "Auto-weighted multi-view clustering via kernelized graph learning," *Pattern Recognit.*, vol. 88, pp. 174–184, Apr. 2019.
- [38] H. Wang, Y. Yang, B. Liu, and H. Fujita, "A study of graph-based system for multi-view clustering," *Knowl.-Based Syst.*, vol. 163, pp. 1009–1019, Jan. 2019.
- [39] X. Peng, S. Xiao, J. Feng, W.-Y. Yau, and Z. Yi, "Deep subspace clustering with sparsity prior," in *Proc. IJCAI*, 2016, pp. 1925–1931.
- [40] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [41] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.

- [42] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.
- [43] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.
- [44] J. Zheng, P. Yang, S. Chen, G. Shen, and W. Wang, "Iterative re-constrained group sparse face recognition with adaptive weights learning," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2408–2423, May 2017.
- [45] Z. Kang, C. Peng, Q. Cheng, X. Liu, X. Peng, Z. Xu, and L. Tian, "Structured graph learning for clustering and semi-supervised classification," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107627.
- [46] Z. Zhang, M. Zhao, and T. W. S. Chow, "Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 9, pp. 2362–2376, Sep. 2015.
- [47] G. Perez, S. Ament, C. Gomes, and M. Barlaud, "Efficient projection algorithms onto the weighted l_1 ball," *Artif. Intell.*, vol. 306, May 2022, Art. no. 103683.
- [48] F. Nie, X. Wang, M. Jordan, and H. Huang, "The constrained Laplacian rank algorithm for graph-based clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2016, vol. 30, no. 1, pp. 1–8.
- [49] R. He, W.-S. Zheng, B.-G. Hu, and X.-W. Kong, "Nonnegative sparse coding for discriminative semi-supervised learning," in *Proc. CVPR*, Jun. 2011, pp. 2849–2856.
- [50] J. Wen, N. Han, X. Fang, L. Fei, K. Yan, and S. Zhan, "Low-rank preserving projection via graph regularized reconstruction," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1279–1291, Apr. 2019.
- [51] A. A. Shaikh, A. K. Mukhopadhyay, S. Poddar, and S. Samui, "Towards power-efficient design of myoelectric controller based on evolutionary computation," 2022, *arXiv:2204.02179*.
- [52] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 1–9.
- [53] G. Bareilles and F. Lutzeler, "On the interplay between acceleration and identification for the proximal gradient algorithm," *Comput. Optim. Appl.*, vol. 77, no. 2, pp. 351–378, Nov. 2020.
- [54] N. T. V. Hang, B. S. Mordukhovich, and M. E. Sarabi, "Augmented Lagrangian method for second-order cone programs under second-order sufficiency," *J. Global Optim.*, vol. 82, no. 1, pp. 51–81, Jan. 2022.
- [55] Z. A. E. Mouden and A. Jakimi, "K-eNSC: K-estimation for normalized spectral clustering," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, Jun. 2020, pp. 1–5.
- [56] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 504–517, Mar. 2016.
- [57] Z. Ma, Z. Kang, G. Luo, L. Tian, and W. Chen, "Towards clustering-friendly representations: Subspace clustering via graph filtering," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 3081–3089.
- [58] Z. Kang, Z. Lin, X. Zhu, and W. Xu, "Structured graph learning for scalable subspace clustering: From single view to multiview," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8976–8986, Sep. 2022.
- [59] S. Pashine, R. Dixit, and R. Kushwah, "Handwritten digit recognition using machine and deep learning algorithms," 2021, *arXiv:2106.12614*.



YOSHUAI WANG received the bachelor's degree from the School of Electronics and Information Engineering, Lanzhou Jiaotong University, in June 2020, where he is currently pursuing the Graduate degree. His current research interests include data mining and graph learning clustering.



YONGXU CHEN is currently pursuing the master's degree in computer technology with the School of Electronic and Information Engineering, Lanzhou Jiaotong University. His research interests include community detection and clustering.



HONGYU ZHU was born in Chongqing, China, in 1996. She is currently pursuing the master's degree with Lanzhou Jiaotong University. Her research interest includes clustering in data mining.



YUE XIE is currently pursuing the master's degree in computer technology with the School of Electronic and Information Engineering, Lanzhou Jiaotong University. His research interests include community detection and clustering.



MEI CHEN received the Ph.D. degree in computer science from Lanzhou University, in 2016. She is currently a Professor with the School of Electronic and Information Engineering, Lanzhou Jiaotong University. She has published over 20 research papers in many conferences and journals, such as IEEE ACCESS, *Pattern Recognition*, *Atmospheric Environment*, *Frontiers of Computer Science*, and *WAIM*. Her research interests include artificial intelligence and data mining. She is a member of CCF.



PENGJU GUO was born in Changzhi, Shanxi, in 1995. He received the master's degree from Lanzhou Jiaotong University. His research interests include clustering and community discovery in complex networks.

...