

Research Article

Robust Graph Structure Learning for Multimedia Data Analysis

Wei Zhou ^{1,2}, Zhaoxuan Gong,¹ Wei Guo,¹ Nan Han,³ and Shaojie Qiao ⁴

¹School of Computer, Shenyang Aerospace University, Shenyang 110136, China

²Shenyang Institute of Computing Technology Co. Ltd., CAS, Shenyang 110168, China

³School of Management, Chengdu University of Information Technology, Chengdu 610103, China

⁴School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China

Correspondence should be addressed to Wei Zhou; zhouweineu@outlook.com and Shaojie Qiao; sjqiao@cuit.edu.cn

Received 17 April 2021; Revised 21 May 2021; Accepted 10 June 2021; Published 25 June 2021

Academic Editor: Yuanlong Cao

Copyright © 2021 Wei Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of computer network technology, we can acquire a large amount of multimedia data, and it becomes a very important task to analyze these data. Since graph construction or graph learning is a powerful tool for multimedia data analysis, many graph-based subspace learning and clustering approaches have been proposed. Among the existing graph learning algorithms, the sample reconstruction-based approaches have gone the mainstream. Nevertheless, these approaches not only ignore the local and global structure information but also are sensitive to noise. To address these limitations, this paper proposes a graph learning framework, termed Robust Graph Structure Learning (RGSL). Different from the existing graph learning approaches, our approach adopts the self-expressiveness of samples to capture the global structure, meanwhile utilizing data locality to depict the local structure. Specially, in order to improve the robustness of our approach against noise, we introduce $l_{2,1}$ -norm regularization criterion and nonnegative constraint into the graph construction process. Furthermore, an iterative updating optimization algorithm is designed to solve the objective function. A large number of subspace learning and clustering experiments are carried out to verify the effectiveness of the proposed approach.

1. Introduction

With the rapid growth of information technology and computer network technology, a large number of multimedia data can be collected from many research fields such as computer vision, image processing, and natural language processing. However, most of the multimedia data are represented by the high dimension and complex structures [1, 2]. Therefore, how to accurately analyze these data becomes a vital problem. Inspired by the pattern recognition and machine learning techniques, many multimedia data analysis approaches based on subspace learning and clustering have been put forward recently [3–6]. However, learning or constructing a valuable graph to describe the pairwise similarity or relationship among the samples is a key issue to multimedia data analysis [7].

Nowadays, a series of graph learning approaches have been proposed in which the heat-kernel function is the most widely used graph construction manner, such as k -nearest-

neighborhood graph (k -NN graph) or ϵ -nearest-neighborhood graph (ϵ graph). The edges of vertexes are computed based on the Euclidean distance among samples and then the weights of the edge between two vertexes are estimated by the heat kernel [8]. However, there are two main limitations in these approaches [9]. First, the choice of parameters in these approaches, such as the neighbor number k or radius ϵ , is very challenging, which can impact the final performance of the task. Second, the processes of neighbor selection and weight calculation are independent, which are sensitive to noise and often cannot well reveal the real similarities of samples [10].

To overcome these drawbacks, sparse representation (SR) based graph construction approach has been proposed, which is often called l_1 -graph or sparse graph. In l_1 -graph [11], each sample is regarded as the query sample and the rest of samples are considered as the dictionary to represent the query sample. Therefore, the similarities between the query sample and the remainder samples can

be measured. Since l_1 -graph employs l_1 -norm constraint on the regression model for selecting a few important samples, it has a better discriminability and more robustness to deal with noise. In the past decades, a series of excellent learning approaches based on l_1 -graph have been designed and successfully applied in different areas [12]. Although l_1 -graph can reveal the linear relationship between a single point and other points, there are still some limitations as follows. First, l_1 -graph strictly assumes the dictionary of regression should be overcompleted, which is unsatisfied in many real applications especially for the graph learning. Second, l_1 -graph pays too much attention on the sparsity while it neglects the correlations between the samples, so it cannot offer a smooth data representation. Therefore, SR is not a good choice for graph construction. To overcome the aforementioned problems suffered by SR, Zhang et al. [13] introduced a Collaborative Representation (CR) linear regression approach by employing the l_2 -norm rather than l_1 -norm sparsity regularization. Compared to SR, CR provides more relaxation for regression coefficients and obtains a smoother data representation.

Considering that both SR and CR usually reveal the linear relationship between a single data point and other data points, the global structure of the data is ignored. To address this problem, Liu et al. [14] suggested Low-Rank Representation (LRR) for subspace clustering. The main purpose of LRR is to find a coefficient matrix Z by trying to reconstruct each data point as a linear combination of all the other data points, which is called self-representation. Different from the traditional similarity measurement approaches based on distance, i.e., k -nearest neighborhood or ε -nearest neighborhood, the representation-based approaches, such as SR, CR, or LRR, measure the similarity between data by solving an optimization problem. These approaches improve the image structure to achieve better classification and clustering performance overall. However, the objective function of LRR is not differentiable which has a high computation complexity on solving the rank minimization problem. To efficiently solve the limitation of LRR, Lu et al. [15] proposed Least Squares Regression (LSR) by grouping the highly correlated data together, which is robust to noise. Compared with LRR, LSR is simpler and more efficient.

In recent years, researchers found that the relationships between data in real applications usually show high dimension nonlinear, so the aforementioned linear representation approaches can hardly achieve good performance. Many researchers paid more attention on revealing the nonlinear relationship between data points of interests [16–27]. For example, Wang et al. [28] explored the criterion of Locally Linear Embedding (LLE) and used it to construct the graph by computing the weights between the pairs of samples. Wei and Peng [29] adopted a similar criterion to that LLE to construct a neighborhood-preserving graph for semisupervised dimensionality reduction. Furthermore, Yu et al. [30] have found that the nonzero coefficients of the sparse coding always are assigned to the neighbor samples of the query sample. To encourage the coding to be locality, some local feature-based coding approaches have been proposed, which achieve excellent performance for the classification

and clustering tasks [31]. With the usage of the merits of local constraints, Peng et al. [32] put forward Locality-Constrained Collaborative (LCC) representation, which achieves better classification performance than those nonlocal approaches. Chen and Yi [33] took the local constraint and LSR into consideration and designed Locality-Constrained LSR (LCLSR) for subspace clustering. LCLSR explores both the global structure of data points and the local linear relationship for data points, forcing the representation to prefer the selection of neighborhood points. Although LCLSR considers the locality structure of data, there are still some limitations on the graph structure. On the one hand, the objective function of LCLSR is based on l_2 -norm, which is very sensitive to noise; on the other hand, the process of sample reconstruction ignores the relationships between sample representations. For example, similar original samples should generate similar coding vectors, and this process weakens the effectiveness of graph learning approaches.

To combat these issues, we design a novel graph learning approach, named Robust Graph Structure Learning (RGSL). Specifically, the self-expressiveness of samples and adaptive neighbor selection approach are introduced to preserve both the local and global structures of data. For enhancing the robustness of graph construction, we introduce the $l_{2,1}$ -norm constraint and nonnegative constraint on the adjacency graph weight matrix to reduce the influence of noise points in graph construction. Therefore, the proposed approach can estimate the graph from data alone by self-expressiveness of samples and data locality, which is independent of a priori affinity matrix. We assess the benefits of the proposed approach on the subspace learning and clustering tasks. Extensive experiments verify the effectiveness of the proposed approach over other state-of-the-art approaches. The framework of the proposed approach is shown in Figure 1.

The outline of this paper is as follows. Section 2 reviews some related work briefly. Section 3 gives the proposed approach in details. Section 4 shows extensive experiments to prove the effectiveness of the proposed approach. Section 5 presents some conclusions.

2. Related Work

In this section, first, many classic and widely used graph construction approaches are introduced. Then, two kinds of multimedia data analysis techniques including subspace learning and spectral clustering are presented in detail accordingly.

2.1. Graph Construction Approaches. Recently, many graph construction approaches have been proposed for multimedia data analysis. In this subsection, we will review some graph construction approaches related to our work as below.

Liu et al. [14] proposed a Low-Rank Representation (LRR) graph construction approach, in which each sample can be represented by a linear combination of all samples, and meanwhile, a low-rank constraint of coefficient matrix is imposed. Given a high dimensional database $X = [x_1, x_2, \dots, x_N] \in R^{d \times N}$ in which d is the data dimensionality and N

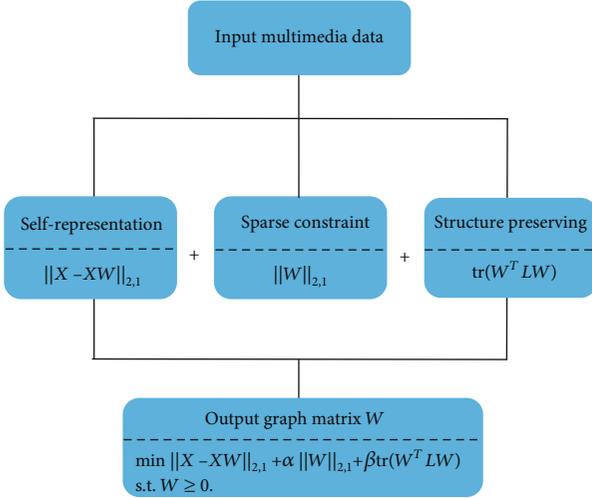


FIGURE 1: The framework of the proposed approach.

is the number of samples. The LRR graph can be obtained by optimizing the following problem:

$$\begin{aligned} \min_W \quad & \|W\|_* \\ \text{s.t.} \quad & X = XW, \end{aligned} \quad (1)$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix, i.e., the sum of the singular values of the matrix. W denotes the coefficient matrix of data X with the lowest rank.

Although LRR graph can obtain the global structure of data, it is very time-consuming to solve the problem of optimal nuclear norm. Hence, Lu et al. [15] utilized the l_2 -graph based on Frobenius norm in place of nuclear norm for fast computing the weight matrix. The LSR graph is defined as

$$\begin{aligned} \min_W \quad & \|X\|_F \\ \text{s.t.} \quad & X = XW, \text{diag}(W) = 0, \end{aligned} \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm. $\text{diag}(\cdot)$ denotes the diagonal operation of a matrix.

In order to make full use of the advantage of locality constraints, Chen and Yi [33] combined LSR and the locality constraints into a unified framework and proposed the LCLSR approach for graph construction. The objective function of LCLSR is

$$\min_W \frac{1}{2} \|X - XW\|_F^2 + \frac{\beta_1}{2} \|W\|_F^2 + \frac{\beta_2}{2} \|W \odot D\|_F^2, \quad (3)$$

where β_1 and β_2 are two balance parameters and the symbol \odot represents the Hadamard product. $D = [d_{ij}]_{N \times N} = [e^{\text{dist}(x_i, x_j)}]_{N \times N}$ denotes the distance matrix between samples where the function $\text{dist}(x_i, x_j)$ is a distance metric, such as the Euclidean distance.

2.2. Subspace Learning. Locality Preserving Projection (LPP) [34] is a well-known subspace learning approach which is used to discover the geometric property of high-dimensional feature space. Suppose that the adjacency graph weight matrix W is given, LPP aims at ensuring that if the original high-dimensional samples x_i and x_j are “close,” then the low-dimensional representations y_i and y_j should be close as well. With the usage of weight matrix W_{ij} as a penalty, LPP is to minimize the following objective function:

$$\frac{1}{2} \sum_{i,j=1}^N \|y_i - y_j\|_2^2 W_{ij} = \text{tr}(W^T X L X^T W), \quad (4)$$

where $L = D - W$ is the Laplacian matrix, in which D is a diagonal matrix with diagonal elements $d_{ii} = \sum_{j=1}^N W_{ij}$. $\text{tr}(S)$ is the trace of matrix S . d_{ii} is used to measure the local density around x_i and the bigger d_{ii} indicates that y_i is more important. Hence, a nature constraint can be imposed as $Y^T D Y = I$. Based on the equation $Y = P^T X$, the LPP model can be rewritten as

$$\begin{aligned} \min_P \quad & \text{tr}(P^T X L X^T P) \\ \text{s.t.} \quad & P^T X D X^T P = I. \end{aligned} \quad (5)$$

The projection matrix P is constructed by the eigenvectors associated with d smallest nonzero eigenvalues, which can be solved by

$$X L X^T p = \lambda X D X^T p. \quad (6)$$

For a new high-dimensional data x , with the usage of the obtained projection matrix P , we can obtain a low-dimensional data representation by $y = P^T x$.

2.3. Spectral Clustering. Spectral clustering is a popular clustering approach that uses eigenvectors of a symmetric matrix derived from the distance between data points [35, 36]. Given a data set consisting of N data points $X \in R^{N \times D}$, spectral clustering approach is aimed at partitioning X into K disjoint clusters by exploiting the top K eigenvectors of the normalized graph Laplacian L . Suppose that the graph matrix W is obtained by graph construction approaches and the new representation $Q \in R^{C \times N}$ can be acquired by optimizing the following objective function:

$$\min_{Q \in R^{C \times N}} \text{Tr}(Q L_W Q^T), \quad (7)$$

where $L_W = D - ((W + W^T)/2)$ is the Laplacian matrix of W , in which $D \in R^{N \times N}$ is the diagonal matrix with $d_{ii} = \sum_j (w_{ij} + w_{ji})/2$. C is the number of selected clusters. Each column in Q represents the new discriminative representation of the corresponding original sample.

Finally, data clustering can be accomplished by performing K -means on the new representation Q .

3. Proposed Method

In this section, some notations are introduced first. Second, we give some detailed descriptions of the proposed RGSL approach. At last, an iterative update algorithm is designed to solve our RGSL approach.

3.1. Notations. Let $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$ be the given high-dimensional original data matrix, where D is the dimensionality of samples and N corresponds to the total number of samples. For a matrix $B \in R^{D \times N}$, the definitions

of Frobenius norm and $l_{2,1}$ -norm are as follows: $\|B\|_F = \sqrt{\sum_{i=1}^D \|b^i\|_2^2} = \sqrt{\sum_{j=1}^N \|b_j\|_2^2}$ and $\|B\|_{2,1} = \sum_{i=1}^D \sqrt{\sum_{j=1}^N b_{ij}^2}$, in which b^i and b_j are the i -th row and the j -th column of B , respectively.

3.2. Objective Function. First, in order to enhance the robustness of graph learning algorithm to noise and obtain the discriminative graph structure, the $l_{2,1}$ -norm measure criterion on the traditional LSR is introduced, which is defined as

$$\begin{aligned} \min \sum_{i=1}^N (\|x_i - Xw_i\|_2 + \alpha \|w_i\|_2) &= \|X - XW\|_{2,1} + \alpha \|W\|_{2,1} \\ &= \text{tr} \left((X - XW)^T G (X - XW) \right) \\ &\quad + \alpha \text{tr} (W^T QW), \end{aligned} \quad (8)$$

where $\|\cdot\|_2$ and $\|\cdot\|_{2,1}$ denote l_2 -norm and $l_{2,1}$ -norm, respectively. α is a balance parameter. G and Q are diagonal matrices whose diagonal elements are, respectively, defined as $G_{ij} = 1/(\|x_i - Xw_j\|_2 + \epsilon)$ and $Q_{ii} = 1/(\|w_i\|_2 + \epsilon)$. ϵ is a small non-negative constant for preventing the value of the denominator from being zero.

Second, the relationship between representation coefficients is ignored in the sample reconstruction, i.e., similar original samples should generate similar coding vectors, weakening the effectiveness of graph learning. To solve the abovementioned issue, a modified manifold constraint based on the $l_{2,1}$ -norm is designed, which is defined as

$$\begin{aligned} \min_W \sum_{i,j=1}^N \|w_i - w_j\|_2 s_{ij} &= \sum_{i,j=1}^N \text{tr} \left((w_i - w_j)^T R (w_i - w_j) \right) \\ &= \text{tr} (W^T (D - R) W) = \text{tr} (W^T L W), \end{aligned} \quad (9)$$

where s_{ij} denotes the similarity weight value between sample x_i and sample x_j . The elements in matrix R are defined as $R_{ij} = s_{ij}/(\|w_i - w_j\|_2 + \epsilon)$, and D is a diagonal matrix whose

diagonal elements are $D_{ii} = \sum_{j=1}^N R_{ij}$. L is the Laplacian of the graph matrix W .

At last, the nonnegative constraint is also imposed on the representation coefficients, and the final objective function of the proposed approach is

$$\begin{aligned} \min \quad & \|X - XW\|_{2,1} + \alpha \|W\|_{2,1} + \beta \text{tr} (W^T L W) \\ &= \text{tr} \left((X - XW)^T G (X - XW) \right) + \alpha \text{tr} (W^T QW) \\ &\quad + \beta \text{tr} (W^T L W) \\ \text{s.t.} \quad & W \geq 0, \end{aligned} \quad (10)$$

where α and β are two positive balance parameters.

3.3. Optimization. In this section, we give the optimization procedures for the objective function of the proposed approach in Equation (10). From Equation (10), we can observe that the objective function is related to $l_{2,1}$ -norm. Thus, the variable W in the objective function is nonconvex and a closed form solution to Equation (10) cannot be given. With regard to this limitation, an iterative update algorithm is designed to optimize the objective function.

3.3.1. Fix G , Q , and R , Update W . First, we fix matrices G , Q , and R . After removing the irrelevant terms, the optimization problem with respect to W in Equation (10) can be simplified to

$$\begin{aligned} \min \quad & \text{tr} (X^T G X - 2W^T X^T G X + W^T X^T G X W) \\ &+ \alpha \text{tr} (W^T QW) + \beta \text{tr} (W^T L W) \\ \text{s.t.} \quad & W \geq 0. \end{aligned} \quad (11)$$

The Lagrangian function of Equation (11) is represented as

$$\begin{aligned} \ell(W) &= \text{tr} (X^T G X - 2W^T X^T G X + W^T X^T G X W) \\ &\quad + \alpha \text{tr} (W^T QW) + \beta \text{tr} (W^T L W) + \lambda \text{tr} (\Delta W) \\ \text{s.t.} \quad & W \geq 0. \end{aligned} \quad (12)$$

By computing the derivative of Equation (12) with respect to W and setting it equals to zero

$$\frac{\partial \ell(W)}{\partial W} = -2X^T G X + 2X^T G X W + 2\alpha Q W + 2\beta L W + \lambda \Lambda = 0. \quad (13)$$

According to the KKT condition [37], we update the solution for W as below:

$$W_{ij} = W_{ij} \frac{[X^T G X]_{ij}}{[\alpha Q + \beta L + X^T G X]_{ij}}. \quad (14)$$

1: **Input:** the data matrix $X = [x_1, x_2, \dots, x_N]$, two balance parameters α and β .
2: **Initialize:** set G_t and Q_t to be the identity matrices, random matrix W , and sample similarity matrix S , $t = 1$.
3: **Repeat**
4: Update the matrix $R_{t+1} = \{R_{t+1,ij}\}_{N \times N} = s_{ij}/(\|w_{t,i} - w_{t,j}\|_2 + \varepsilon)$
5: Compute the matrix $D_{t+1,ii} = \sum_{j=1}^N R_{t+1,ij}$ and the Laplacian matrix $L_{t+1} = D_{t+1} - R_{t+1}$
6: Update the matrix $W_{ij} = W_{ij}([X^T GX]_{ij}/[\alpha Q + \beta L + X^T GX]_{ij})$
7: Update the matrix $G_{t+1} = \{G_{t+1,ij}\}_{N \times N} = 1/(\|x_i - Xw_{t+1,j}\|_2 + \varepsilon)$
8: Update the matrix $Q_{t+1} = \{Q_{t+1,ii}\}_{N \times N} = 1/(\|w_{t+1,i}\|_2 + \varepsilon)$
9: $t = t + 1$
10: **Until** convergence
11: **Output:** the graph matrix W

ALGORITHM 1: RGSL.

3.3.2. *Fix W , Update G , Q , and R .* When W is fixed and all the irrelevant terms are removed, the solution can be formulated as

$$G_{ij} = \frac{1}{\|x_i - Xw_j\|_2 + \varepsilon}, \quad (15)$$

$$Q_{ii} = \frac{1}{\|w_i\|_2 + \varepsilon}, \quad (16)$$

$$R_{ij} = \frac{s_{ij}}{\|w_i - w_j\|_2 + \varepsilon}. \quad (17)$$

3.4. *Algorithm.* In conclusion, the proposed optimization algorithm for RGSL can be summarized as below.

In Algorithm 1, the convergence condition is defined as the change of the value of objective function in Equation (10) which is less than a threshold or a predefined maximum iteration number is reached.

4. Experiment and Results

In this section, first, we will introduce the used databases in our experiment. Next, some graph learning comparison approaches are given. At last, subspace learning and clustering tasks are employed for verifying the effectiveness of the proposed approach.

4.1. *Databases.* Four commonly used multimedia databases from the Internet including Yale [38], AR [39], CMU PIE [40], and Extended YaleB [41] are used for verifying the effectiveness of the proposed approach. The detailed statistical information about the four different databases is depicted in Table 1.

Yale database: it contains 165 face images captured from 15 different subjects. Each subject has 11 different images with the varied facial expressions, under different illumination conditions, and wearing glasses or not. Some example images of the Yale database are depicted in Figure 2(a).

AR database: it consists of over 4000 facial images obtained from 70 male and 56 female faces. Images of each person were captured with 26 frontal face images with anger, smiling, and screaming, under varied illumination condi-

TABLE 1: Description of databases.

Databases	Image size	Instances	Classes	Instances of per class
Yale	32 × 32	165	15	11
AR	32 × 32	1400	100	14
CMU PIE	32 × 32	1632	68	24
Extended YaleB	32 × 32	2432	38	64

tions, and with sunglass and scarf occlusions. Some examples of the AR database are shown in Figure 2(b).

CMU PIE database: there are 41,368 face images of 68 different subjects. Images of each person are captured under 43 different illumination conditions with 13 different poses and 4 different expressions. Here, we employ a subset of CMU PIE which consists of 24 images per subject. A part of example images is illustrated in Figure 2(c).

Extended YaleB database: there are 38 individuals and each individual has 64 images. For each individual, the face images are taken from different illumination conditions with small changes in head pose and facial expression. Example images from this database are shown in Figure 2(d).

Although we can obtain the graph structure from the proposed approach, it is intractable to assess the graph learning approaches using the estimated graph alone. Hence, we will assess the quality of the learned graph by two kinds of multimedia data analysis tasks including subspace learning and spectral clustering. In our experiments, we first vary the graph construction approaches by fixing the graph learning task and then observe the obtained performance associated with subspace learning and spectral clustering tasks.

4.2. *Comparison among Several Graph Learning Approaches.* To investigate the performance of our approach on subspace learning and clustering, several state-of-the-art graph learning approaches are chosen to compare in our work, which are shown as below:

- (i) KNN graph [8]: the graph edges connected by two vertexes can be generated by the Euclidean distance-based K -nearest neighbor and the heat

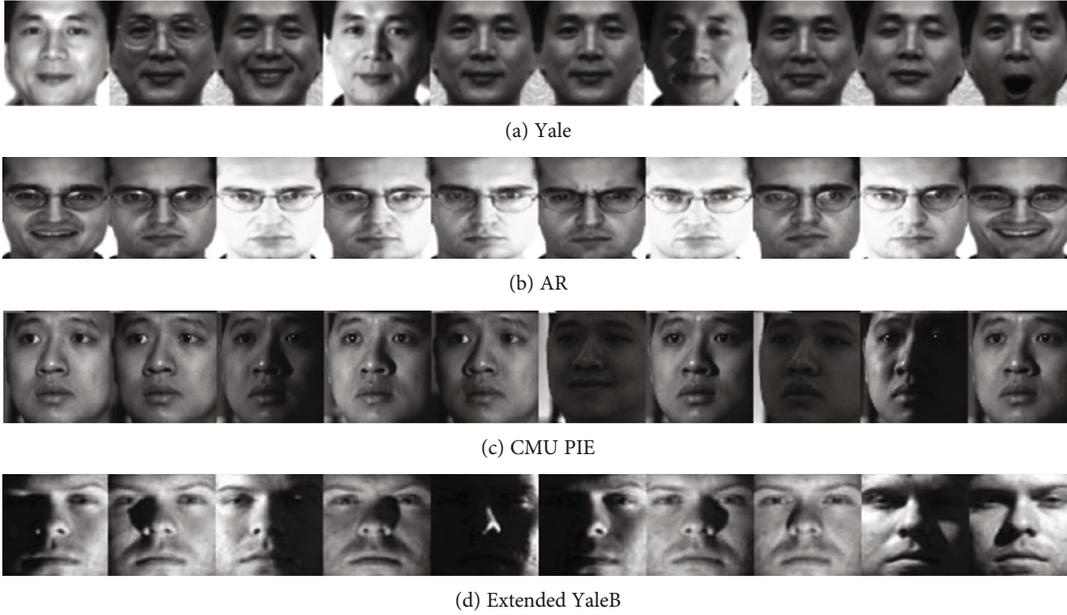


FIGURE 2: Some of the images from the different databases employed in our experiments.

kernel function is used to measure the weight of an edge

- (ii) LLE graph [28]: each sample is linearly reconstructed by its neighbors within a local area to preserve the local manifold structure
- (iii) L1 graph [11]: the locality structure of data by using L1 sparse representation optimization
- (iv) LRR graph [14]: based on self-expressive property, a low-rank graph is obtained
- (v) LSR graph [15]: self-expressive property and Frobenius norm are used for fast computing the weight matrix
- (vi) LCLSR graph [33]: it combines the locality constraint and LSR together to explore both the global structure of data points and the locality linear relationship of data points
- (vii) SGLS graph [42]: it integrates manifold constraints on the unknown sparse codes as a graph regularizer
- (viii) Our proposed RGSL graph: our approach takes the global and local structure information into consideration and also introduces the $l_{2,1}$ -norm regularization criterion and nonnegative constraint into graph construction process to enhance its robustness

4.3. Subspace Learning Experiment and Analysis. In this section, we employ an unsupervised subspace learning approach represented by Locality Preserving Projections (LPP) to verify the effectiveness of the proposed approach. In our experiments, different graphs are employed as W in LPP approach for subspace learning and then the classification accuracy is

used for performance comparison. For each database, we randomly select l images from each class as the training samples. The remaining images are treated as the test samples. The values of l for Yale, AR, CMU PIE, and Extended YaleB databases are set as $\{4, 5, 6\}$, $\{4, 5, 6\}$, $\{6, 8, 10\}$, and $\{10, 15, 20\}$, respectively. In order to more effectively and fairly test the performance of the proposed approach, the random sample selection is repeated 20 times and the average classification accuracy and standard deviation are regarded as the final results for comparison. In this work, we employ the nearest neighbor classifier with Euclidean distance for classification due to its simplicity. To compare the performances of different approaches, the classification accuracy rate is chosen as the evaluation criterion, which is defined as

$$\text{Class_accuracy} = \frac{N_{t_correct}}{N_{total}} \times 100\%, \quad (18)$$

where $N_{t_correct}$ is the number of test samples which are correctly classified using the nearest neighbor classifier. N_{total} is the total number of the test samples.

All the experiments are conducted using MATLAB 2016b on a 3.60 Hz with 8 G RAM. In order to acquire the optimal parameters of different approaches, we employ the grid-search manner in our experiments. Tables 2–5 depict the average classification accuracy rates and standard deviations of the proposed approach on the Yale, AR, CMU PIE, and Extended YaleB databases, respectively. Note that the brackets in Tables 2–5 mean the data dimensionality when achieving the best classification accuracy rates.

From the results depicted in Tables 2–5, we can clearly observe that most of the graph learning approaches perform better than the KNN graph and LLE graph. It indicates that

TABLE 2: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the Yale database.

Method	$l = 4$	$l = 5$	$l = 6$
KNN+LPP	74.86 ± 2.70 (56)	77.56 ± 2.39 (72)	79.07 ± 3.27 (80)
LLE+LPP	77.71 ± 4.22 (56)	79.78 ± 1.55 (56)	80.13 ± 3.23 (80)
L1+LPP	81.52 ± 2.30 (44)	84.79 ± 2.19 (64)	85.07 ± 3.31 (80)
LRR+LPP	79.43 ± 2.70 (56)	82.33 ± 2.59 (72)	82.93 ± 4.06 (76)
LSR+LPP	84.29 ± 4.07 (56)	86.00 ± 2.58 (72)	86.40 ± 5.02 (80)
LCLSR+LPP	82.57 ± 3.84 (52)	85.56 ± 2.92 (72)	85.73 ± 4.71 (80)
SGLS+LPP	83.33 ± 4.15 (56)	85.33 ± 3.18 (72)	85.20 ± 3.74 (76)
RGSL+LPP	84.95 ± 4.12 (56)	86.89 ± 1.72 (64)	86.53 ± 4.14 (64)

TABLE 3: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the AR database.

Method	$l = 4$	$l = 5$	$l = 6$
KNN+LPP	60.76 ± 1.92 (320)	61.40 ± 1.99 (400)	60.11 ± 1.41 (400)
LLE+LPP	66.79 ± 2.28 (320)	67.56 ± 2.15 (380)	66.39 ± 3.17 (400)
L1+LPP	74.12 ± 1.90 (280)	74.96 ± 1.47 (360)	72.46 ± 2.64 (400)
LRR+LPP	77.66 ± 1.16 (380)	78.93 ± 0.95 (400)	82.56 ± 1.53 (400)
LSR+LPP	79.39 ± 1.80 (320)	80.14 ± 0.83 (400)	83.66 ± 1.43 (400)
LCLSR+LPP	80.23 ± 1.54 (380)	81.76 ± 1.01 (400)	85.36 ± 1.06 (400)
SGLS+LPP	79.43 ± 2.06 (380)	84.18 ± 0.97 (400)	85.35 ± 1.01 (400)
RGSL+LPP	83.28 ± 1.25 (380)	85.20 ± 1.06 (400)	90.54 ± 1.23 (380)

TABLE 4: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the CMU PIE database.

Method	$l = 6$	$l = 8$	$l = 10$
KNN+LPP	85.37 ± 0.81 (400)	86.56 ± 0.70 (400)	86.63 ± 0.99 (400)
LLE+LPP	87.55 ± 0.92 (400)	88.05 ± 0.91 (400)	87.95 ± 0.80 (400)
L1+LPP	88.95 ± 0.86 (380)	88.97 ± 0.73 (400)	88.09 ± 1.13 (380)
LRR+LPP	90.08 ± 0.37 (380)	90.18 ± 0.87 (400)	90.11 ± 0.99 (400)
LSR+LPP	89.90 ± 0.55 (400)	90.26 ± 0.81 (400)	90.21 ± 0.89 (320)
LCLSR+LPP	90.32 ± 0.44 (400)	90.68 ± 0.73 (400)	90.68 ± 0.83 (400)
SGLS+LPP	89.72 ± 0.68 (400)	90.08 ± 0.57 (280)	90.09 ± 0.93 (380)
RGSL+LPP	90.46 ± 0.81 (340)	92.72 ± 0.63 (380)	93.60 ± 0.48 (400)

TABLE 5: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the Extended YaleB database.

Method	$l = 10$	$l = 15$	$l = 20$
KNN+LPP	65.38 ± 2.33 (360)	69.62 ± 1.12 (400)	64.49 ± 1.54 (400)
LLE+LPP	73.93 ± 1.71 (360)	75.00 ± 1.18 (400)	72.05 ± 1.17 (400)
L1+LPP	75.77 ± 4.10 (280)	75.97 ± 1.84 (340)	67.70 ± 1.69 (380)
LRR+LPP	82.67 ± 1.02 (360)	85.97 ± 0.71 (400)	84.53 ± 1.19 (380)
LSR+LPP	84.20 ± 0.66 (360)	86.19 ± 0.84 (360)	85.14 ± 0.71 (400)
LCLSR+LPP	85.59 ± 0.75 (360)	88.06 ± 0.75 (400)	87.62 ± 0.62 (380)
SGLS+LPP	84.83 ± 0.73 (360)	86.64 ± 0.73 (380)	87.09 ± 0.94 (340)
RGSL+LPP	86.14 ± 0.65 (360)	90.99 ± 0.48 (400)	92.68 ± 0.54 (400)

graph learning based on Euclidean distance is very sensitive to noise points, weakening the classification performance. Besides, compared to L1 graph learning approach, LRR graph, LSR graph, SGLS graph, and LCLSR graph take the locality structure of data into consideration during the pro-

cess of graph construction to get more excellent performance. At last, the proposed RGSL approach performs best among all of the compared approaches. The main reasons are as follows: first, both the global structure and local structure are essential to the graph learning. Second, $l_{2,1}$ -norm

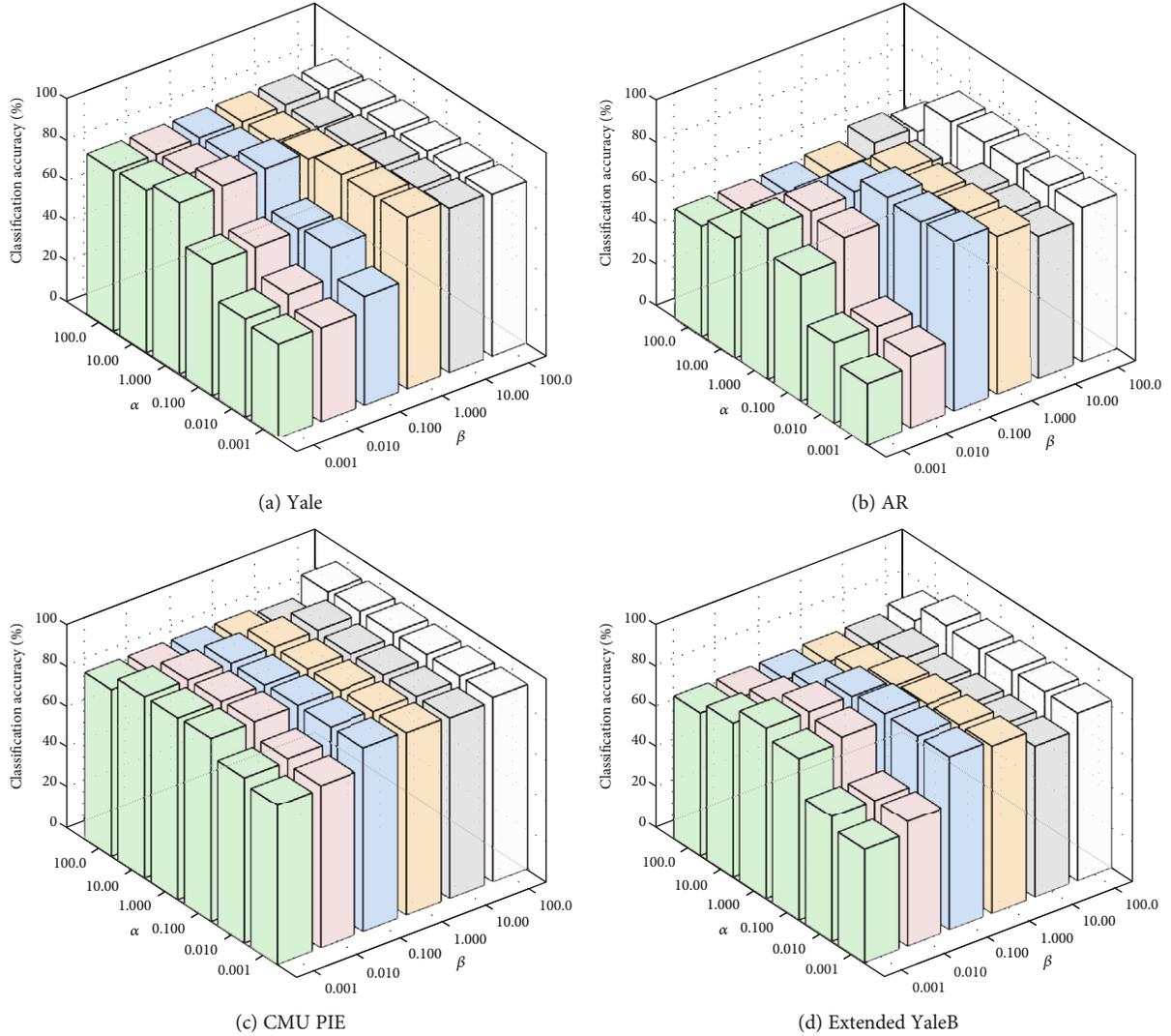


FIGURE 3: The classification accuracy rates of the proposed RGSL approach with different values of parameters α and β on the four different databases.

regularization criterion and nonnegative constraint are introduced into graph construction process to improve the robustness of our approach against noise. Therefore, our approach can improve the classification performance further.

There are two parameters, i.e., α and β in the objective function of our proposed approach. Hence, how to appropriately set their values is very important for our approach. In this study, we tune the values of parameters α and β by searching the grid $\{0.001, 0.01, 0.1, 1, 10, 100\}$ in an alternate manner. The best results of different parameter values on the four databases are shown in Figure 3.

As we can see from Figure 3, when the values of parameters α and β are relatively small, the performance of the proposed approach is relatively small. With the increase of parameters α and β , the performance of the proposed approach will be improved. However, after it achieves its best classification result, the performance of the proposed approach dramatically decreases with the increase of the two parameters. Therefore, the proposed approach can

obtain its best classification results when the values of parameters α and β are set as neither too large nor too small. At last, the convergence curves of our RGSL on the four databases are shown in Figure 4. In this figure, the x-axis and the y-axis are, respectively, denoted as the iteration number and the value of the objective function. As seen from Figure 4, the value of the objective function declines at each iteration and converges very fast on all of the databases.

4.4. Clustering Experiment and Analysis. In spectral clustering, the initialization has a major impact on the performance of the K -means clustering algorithm. Therefore, we carry out the process of clustering 50 times with different random initializations. Then, the average clustering results with standard deviations are used as the final results. In the experiments, three widely employed clustering evaluation indicators including Accuracy (ACC), Normalized Mutual Information (NMI), and Purity are used to evaluate the performance of the proposed approach.

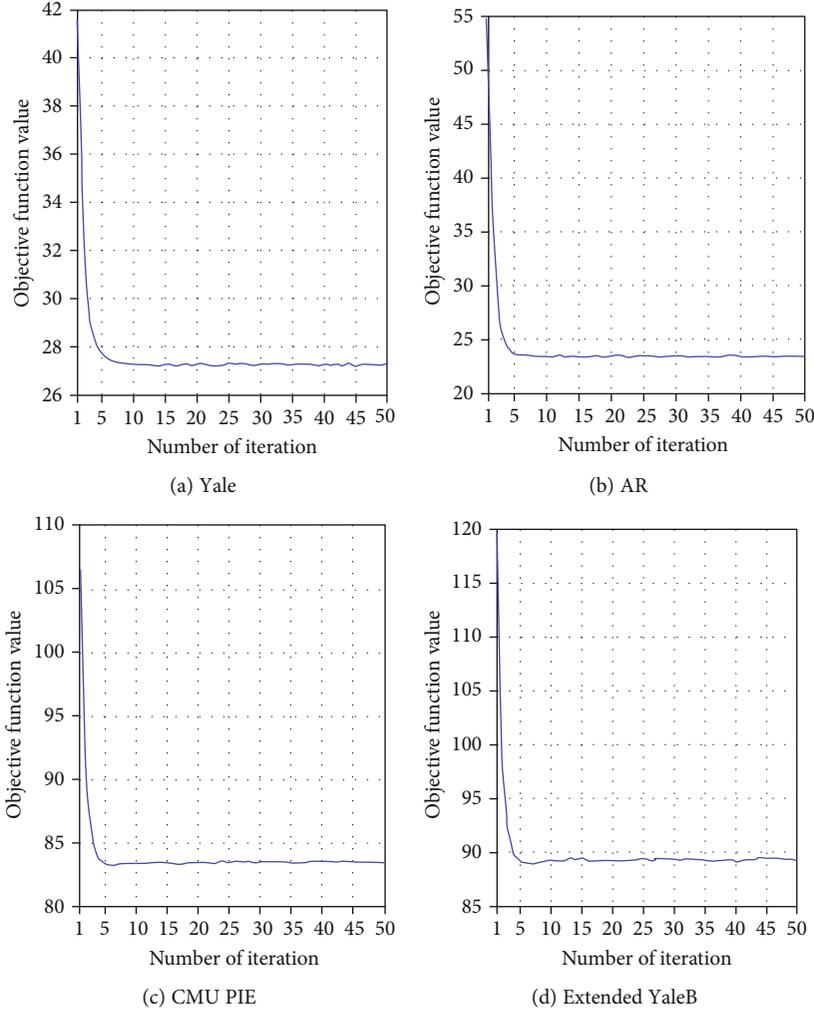


FIGURE 4: The convergence curves of RGSL on the four databases: (a) Yale, (b) AR, (c) CMU PIE, and (d) Extended YaleB.

For a given sample x_i , supposing that the obtained clustering result is p_i and true label is t_i , the clustering accuracy is calculated as

$$\text{ACC} = \frac{\sum_{i=1}^N \delta(t_i, m(p_i))}{N}, \quad (19)$$

where $\delta(x, y) = 1$ if $x = y$, $\delta(x, y) = 0$ otherwise. The function $m(\cdot)$ maps the clustering result to the corresponding ground truth label. N is the number of samples. The Kuhn-Munkres algorithm [37] is employed to find the best mapping result.

Assuming that P and T are, respectively, the clustering result and the true label set obtained by different approaches, the Mutual Information (MI) is defined as

$$\text{MI}(P, T) = \sum_{p_i \in Q, t_i \in T} Q(p_i, t_i) \cdot \log_2 \frac{Q(p_i, t_i)}{Q(p_i) \cdot Q(t_i)}, \quad (20)$$

where $Q(p_i)$ and $Q(t_i)$ represent the probabilities that a sample is randomly selected from the dataset belonging to p_i and t_i , respectively. $Q(p_i, t_i)$ represents the joint probability of a

sample randomly being selected from the dataset belonging to p_i and t_i .

Let $H(P)$ and $H(T)$ be the entropies of P and T , respectively. The Normalized Mutual Information (NMI) is calculated as

$$\text{NMI} = \frac{\text{MI}(P, T)}{\max(H(P), H(T))}. \quad (21)$$

Purity is defined as follows:

$$\text{Purity} = \frac{1}{k} \sum_{i=1}^k \frac{|C_i^d|}{|C_i|}, \quad (22)$$

where k represents the number of clusters, $|C_i^d|$ is the number of elements in the most numerous category in cluster C_i , and $|C_i|$ is the number of elements in cluster C_i .

Tables 6–9 show the best values of ACC, NMI, and Purity of eight approaches, respectively, on the Yale, AR, CMU PIE, and Extended YaleB databases. According to the results as shown in Tables 6–9, the following conclusions can be

TABLE 6: The clustering results obtained by different approaches on the Yale database.

Method	ACC	NMI	Purity
KNN	0.6887 ± 0.0173	0.7098 ± 0.0069	0.6753 ± 0.0155
LLE	0.7028 ± 0.0032	0.7147 ± 0.0048	0.7085 ± 0.0028
L1	0.6998 ± 0.0191	0.7396 ± 0.0145	0.7145 ± 0.0160
LRR	0.7320 ± 0.0119	0.7580 ± 0.0102	0.7358 ± 0.0111
LSR	0.7509 ± 0.0123	0.7790 ± 0.0092	0.7520 ± 0.0122
LCLSR	0.7782 ± 0.0220	0.7848 ± 0.0183	0.7865 ± 0.0202
SGLS	0.7353 ± 0.0284	0.7811 ± 0.0201	0.7550 ± 0.0219
RGSL	0.8532 ± 0.0243	0.8665 ± 0.0321	0.8563 ± 0.0304

TABLE 7: The clustering results obtained by different approaches on the AR database.

Method	ACC	NMI	Purity
KNN	0.3907 ± 0.0060	0.6556 ± 0.0032	0.4023 ± 0.0062
LLE	0.3908 ± 0.0072	0.6645 ± 0.0043	0.4090 ± 0.0064
L1	0.6451 ± 0.0123	0.8057 ± 0.0068	0.6816 ± 0.0096
LRR	0.6763 ± 0.0121	0.8182 ± 0.0082	0.7130 ± 0.0096
LSR	0.7236 ± 0.0165	0.8703 ± 0.0078	0.7547 ± 0.0144
LCLSR	0.8108 ± 0.0147	0.9050 ± 0.0052	0.8326 ± 0.0103
SGLS	0.8130 ± 0.0151	0.9186 ± 0.0056	0.8361 ± 0.0121
RGSL	0.8301 ± 0.0124	0.9243 ± 0.0036	0.8485 ± 0.0132

TABLE 8: The clustering results obtained by different approaches on the CMU PIE database.

Method	ACC	NMI	Purity
KNN	0.4825 ± 0.0118	0.7186 ± 0.0059	0.5244 ± 0.0082
LLE	0.5884 ± 0.0124	0.7520 ± 0.0049	0.6108 ± 0.0090
L1	0.6284 ± 0.0168	0.7484 ± 0.0091	0.6644 ± 0.0130
LRR	0.7805 ± 0.0193	0.8694 ± 0.0065	0.7992 ± 0.0151
LSR	0.8077 ± 0.0146	0.8774 ± 0.0049	0.8165 ± 0.0096
LCLSR	0.8159 ± 0.0114	0.8733 ± 0.0059	0.8262 ± 0.0084
SGLS	0.8276 ± 0.0147	0.8717 ± 0.0069	0.8380 ± 0.0109
RGSL	0.8399 ± 0.0095	0.8886 ± 0.0070	0.8490 ± 0.0081

obtained. First, since KNN graph and LLE graph are based on Euclidean distance, they are very sensitive to the noise points, outliers, and parameter values. So the clustering performance based on KNN graph and LLE graph is lower than that based on other compared approaches. Second, the performance of LRR graph, LSR graph, SGLS graph, and LCLSR graph is superior to that of L1 graph because of taking the locality structure of data into consideration during the process of graph construction. However, these objective functions are all based on l_2 -norm, so it is very sensitive to the noise data.

TABLE 9: The clustering results obtained by different approaches on the Extended YaleB database.

Method	ACC	NMI	Purity
KNN	0.4195 ± 0.0074	0.5191 ± 0.0069	0.4311 ± 0.0061
LLE	0.4611 ± 0.0090	0.5460 ± 0.0070	0.4757 ± 0.0070
L1	0.4793 ± 0.0168	0.5981 ± 0.0087	0.5082 ± 0.0123
LRR	0.7146 ± 0.0147	0.7603 ± 0.0085	0.7181 ± 0.0139
LSR	0.7043 ± 0.0181	0.7496 ± 0.0102	0.7086 ± 0.0174
LCLSR	0.7157 ± 0.0226	0.7656 ± 0.0140	0.7186 ± 0.0219
SGLS	0.7220 ± 0.0235	0.7768 ± 0.0122	0.7256 ± 0.0226
RGSL	0.7402 ± 0.0276	0.7897 ± 0.0163	0.7487 ± 0.0194

Besides, the relationship between the representation coefficients is ignored in the sample reconstruction, i.e., similar original samples should generate similar coding vectors, weakening the effectiveness of graph learning. To overcome these disadvantages, our RGSL approach combines $l_{2,1}$ -norm with manifold constraints on the coding coefficients to learn a locality and smoothness representation. Therefore, the performance of the proposed approach is superior to that of all of the comparison approaches.

Similar to the subspace learning experiment, we also tune the values of parameters α and β by searing the grid $\{0.001, 0.01, 0.1, 1, 10, 100\}$ in an alternate manner. From the objective function, we can learn that there are three terms. When the values of parameters α and β are set as small, the effectiveness of the second and third terms in the objective function will be weakened, and the role of the first term will be overemphasized. On the contrary, the second and third terms in the objective function will play a major role, reducing the effect of the first term. Therefore, the proposed RGSL approach can achieve the best performance when parameters α and β are set as moderate values, which is similar to the discussions of subspace learning.

5. Conclusions

This paper proposes a novel graph learning framework, named Robust Graph Structure Learning (RGSL) for effective multimedia data analysis. In order to preserve both local and global structures of data, we employ the data self-representativeness to capture the global structure and adaptive neighbor approach to describe the local structure. Furthermore, we also introduce the $l_{2,1}$ -norm regularization criterion and nonnegative constraint into graph learning to improve the robustness of the model against noise. Extensive experimental results associated with subspace learning and clustering tasks show that the proposed approach performs better performance than the state-of-the-art graph learning approaches. Since our proposed approach will be affected by the graph construction when the dimensionality of data is high, in the future, we will take the dimensionality reduction, subspace learning, and graph learning into a united framework to address this issue.

Data Availability

The data are derived from public domain resources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported in part by grants from the National Natural Science Foundation of China (Nos. 61903262, 62062040, 61772091, 61802035, and 61962006), the China Postdoctoral Science Foundation (No. 2019M661117), the Scientific Research Fund Project of Liaoning Provincial Department of Education (Nos. JYT19040 and JYT19053), the Scientific Research Funds of Shenyang Aerospace University under Grant (18YB01 and 19YB01), the Natural Science Foundation of Liaoning Province Science and Technology Department (No. 2019-ZD-0234), the Sichuan Science and Technology Program under Grant Nos. 2021JDJQ0021, 2020YFG0153, 2020YJ0481, 2020YFS0466, 2020YJ0430, 2020JDR0164, and 2019YFS0067, the CCF-Huawei Database System Innovation Research Plan under Grant No. CCF-HuaweiDBIR2020004A, the Natural Science Foundation of Guangxi under Grant No. 2018GXNSFDA138005, and the Major Project of Digital Key Laboratory of Sichuan Province in Sichuan Conservatory of Music under Grant No. 21DMAKL02.

References

- [1] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Non-negative matrix factorization with locality constrained adaptive graph," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 427–441, 2020.
- [2] Y. Yi, J. Wang, W. Zhou, Y. Fang, J. Kong, and Y. Lu, "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognition*, vol. 92, pp. 258–273, 2019.
- [3] W. Zhou, C. Wu, J. Wang, X. Yu, and Y. Yi, "Double regularized matrix factorization for image classification and clustering," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, Article ID 49, 2018.
- [4] W. Wei, H. Dai, and W. Liang, "Regularized least squares locality preserving projections with applications to image recognition," *Neural Networks*, vol. 128, pp. 322–330, 2020.
- [5] S. Affeldt, L. Labiod, and M. Nadif, "Spectral clustering via ensemble deep autoencoder learning (SC-EDAE)," *Pattern Recognition*, vol. 108, p. 107522, 2020.
- [6] Y. Yi, Y. Chen, J. Wang, G. Lei, J. Dai, and H. Zhang, "Joint feature representation and classification via adaptive graph semi-supervised nonnegative matrix factorization," *Signal Processing: Image Communication*, vol. 89, p. 115984, 2020.
- [7] Z. Kang, C. Peng, Q. Cheng et al., "Structured graph learning for clustering and semi-supervised classification," *Pattern Recognition*, vol. 110, p. 107627, 2021.
- [8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [9] F. Dornaika, L. Weng, and Z. Jin, "Structured sparse graphs using manifold constraints for visual data analysis," *Neuro-computing*, vol. 315, pp. 107–114, 2018.
- [10] C. Tang, M. Bian, X. Liu et al., "Unsupervised feature selection via latent representation learning and manifold regularization," *Neural Networks*, vol. 117, pp. 163–178, 2019.
- [11] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with ℓ^1 -graph for image analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 858–866, 2010.
- [12] P. Chen, L. Jiao, F. Liu, Z. Zhao, and J. Zhao, "Adaptive sparse graph learning based dimensionality reduction for classification," *Applied Soft Computing*, vol. 82, p. 105459, 2019.
- [13] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *2011 International Conference on Computer Vision*, pp. 471–478, Barcelona, Spain, November 2011.
- [14] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27-Th International Conference on Machine Learning*, pp. 663–670, Haifa, Israel, 2010.
- [15] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *European Conference on Computer Vision*, pp. 347–360, Springer, 2012.
- [16] X. Peng, J. Feng, S. Xiao, W. Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Transaction on Image Processing*, vol. 27, no. 10, pp. 5076–5086, 2018.
- [17] Y. Chen, L. Zhang, and Z. Yi, "Subspace clustering using a low-rank constrained autoencoder," *Information Sciences*, vol. 424, pp. 27–38, 2018.
- [18] K. G. Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang, "Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 5747–5756, Venice, Italy, October 2017.
- [19] W. Zhou, Y. Yi, J. Bao, and W. Wang, "Adaptive weighted locality-constrained sparse coding for glaucoma diagnosis," *Medical & Biological Engineering & Computing*, vol. 57, no. 9, pp. 2055–2067, 2019.
- [20] L. Zhang, W. Lu, X. Liu, W. Pedrycz, and C. Zhong, "Fuzzy c-means clustering of incomplete data based on probabilistic information granules of missing values," *Knowledge-Based Systems*, vol. 99, pp. 51–70, 2016.
- [21] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, and J. Lu, "Fuzzy regression transfer learning in Takagi-Sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1795–1807, 2017.
- [22] W. Zhou, S. Qiao, Y. Yi, N. Han, Y. Chen, and G. Lei, "Automatic optic disc detection using low-rank representation based semi-supervised extreme learning machine," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 1, pp. 55–69, 2020.
- [23] Z. Shu, X. Wu, C. Hu, C. Z. You, and H. H. Fan, "Deep semi-nonnegative matrix factorization with elastic preserving for data representation," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 1707–1724, 2021.
- [24] W. Zhou, C. Wu, D. Chen, Y. Yi, and W. du, "Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method," *IEEE Access*, vol. 5, pp. 2563–2572, 2017.

- [25] Z. Kang, H. Pan, S. C. H. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1833–1843, 2019.
- [26] S. Huang, Z. Kang, I. W. Tsang, and Z. Xu, "Auto-weighted multi-view clustering via kernelized graph learning," *Pattern Recognition*, vol. 88, pp. 174–184, 2019.
- [27] P. Jing, Y. Su, Z. Li, J. Liu, and L. Nie, "Low-rank regularized tensor discriminant representation for image set classification," *Signal Processing*, vol. 156, pp. 62–70, 2019.
- [28] J. Wang, F. Wang, C. Zhang, H. C. Shen, and Long Quan, "Linear neighborhood propagation and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1600–1615, 2009.
- [29] J. Wei and H. Peng, "Neighbourhood preserving based semi-supervised dimensionality reduction," *Electronics Letters*, vol. 44, no. 20, pp. 1190–1192, 2008.
- [30] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *Advance in Neural Information Processing Systems, NIPS*, pp. 2223–2231, 2009.
- [31] D. Arpit, G. Srivastava, and Y. Fu, "Locality-constrained low rank coding for face recognition," in *International Conference on Pattern Recognition*, pp. 1687–1690, Tsukuba, Japan, 2012.
- [32] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognition*, vol. 47, no. 9, pp. 2794–2806, 2014.
- [33] Y. Chen and Z. Yi, "Locality-constrained least squares regression for subspace clustering," *Knowledge-Based Systems*, vol. 163, pp. 51–56, 2019.
- [34] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.
- [35] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern Recognition*, vol. 41, no. 1, pp. 176–190, 2008.
- [36] K. Song, X. Yao, F. Nie, X. Li, and M. Xu, "Weighted bilateral K-means algorithm for fast co-clustering and fast spectral clustering," *Pattern Recognition*, vol. 109, p. 107560, 2021.
- [37] H. Zhu and M. C. Zhou, "Efficient role transfer based on Kuhn-Munkres algorithm," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 2, pp. 491–496, 2012.
- [38] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.
- [39] A. Martinez, *The AR Face Database*, CVC, Luxembourg City, Luxembourg, 1998, Technical Report 24.
- [40] Terence Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.
- [41] Y. Yi, C. Bi, X. Li, J. Wang, and J. Kong, "Semi-supervised local ridge regression for local matching based face recognition," *Neurocomputing*, vol. 167, pp. 132–146, 2015.
- [42] F. Dornaika and L. Weng, "Sparse graphs with smoothness constraints: application to dimensionality reduction and semi-supervised classification," *Pattern Recognition*, vol. 95, pp. 285–295, 2019.