

Research Article

Discriminative Label Relaxed Regression with Adaptive Graph Learning

Jingjing Wang ¹, Zhonghua Liu ¹, Wenpeng Lu ², and Kaibing Zhang³

¹Information Engineering College, Henan University of Science and Technology, Luoyang, China

²School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China

³College of Electronics and Information, Xi'an Polytechnic University, Xi'an, China

Correspondence should be addressed to Zhonghua Liu; zhonghualiu@haust.edu.cn

Received 6 August 2020; Revised 3 November 2020; Accepted 27 November 2020; Published 14 December 2020

Academic Editor: Jens Christian Claussen

Copyright © 2020 Jingjing Wang 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.

The traditional label relaxation regression (LRR) algorithm directly fits the original data without considering the local structure information of the data. While the label relaxation regression algorithm of graph regularization takes into account the local geometric information, the performance of the algorithm depends largely on the construction of graph. However, the traditional graph structures have two defects. First of all, it is largely influenced by the parameter values. Second, it relies on the original data when constructing the weight matrix, which usually contains a lot of noise. This makes the constructed graph to be often not optimal, which affects the subsequent work. Therefore, a discriminative label relaxation regression algorithm based on adaptive graph (DLRR_AG) is proposed for feature extraction. DLRR_AG combines manifold learning with label relaxation regression by constructing adaptive weight graph, which can well overcome the problem of label overfitting. Based on a large number of experiments, it can be proved that the proposed method is effective and feasible.

1. Introduction

Information technology is developing rapidly and has become a hot topic in recent years. We can get a lot of information from the data, but the dimension of the data is getting higher and higher [1]. On the one hand, the increase of data dimension makes the description of data samples more comprehensive and provides more bases for further analysis and processing of data samples [2]. On the other hand, the increase in the number of features will bring more redundant features, which not only put forward great requirements for hardware and software equipment to complete data processing but also directly affect the reliability and effectiveness of data analysis and processing results [3]. In order to effectively analyze and process the data, by mapping (or transforming) the original data to the low-dimensional space, the features that best reflect the intrinsic nature of the sample data can be obtained. This process is called feature extraction or data dimensionality reduction [4]. The core task of feature extraction or data

dimensionality reduction is how to find out which features are most effective for the final data analysis and processing and how to retain the useful information in the data transformation process to the maximum extent. Feature extraction is always a key problem in pattern recognition, which will directly affect the design and performance of the classifier [5–7]. Feature extraction can not only reduce the dimension of data but also retain useful information in the data. It is also widely used in the processing and analysis of complex data.

In recent years, the feature extraction method based on manifold learning has made remarkable achievements in nonlinear data analysis and research and has been widely used in nonlinear data processing and analysis. According to the current popular learning methods, the nonlinear characteristics and manifold structures of the sample data are usually distributed in low-dimensional space [8–11], and the traditional nonlinear subspace method is difficult to describe and extract the information effectively [12–14]. At present, the popular learning method is to obtain the embedded

mapping of low-dimensional manifold structure and high-dimensional feature manifold of the sample and then complete the feature extraction of nonlinear information in the sample data [15–18].

Discriminative Fisher embedding dictionary learning (DFEDL) algorithm simultaneously establishes Fisher embedding models on learned atoms and coefficients [19, 20]. In 2000, Tenenbaum from Stanford University published the first manifold learning method based on isometric mapping (ISOMAP) [21] in Science. ISOMAP method is a globally optimized method to preserve the manifold structure between data effectively by maintaining the geometric relationship between sample data (also known as geodesic distance). Roweis proposed a locally linear embedding (LLE) method in literature [22]. By constructing the reconstruction weight between each sample and the neighboring samples, this method can better preserve the manifold features between the neighboring samples when embedding in low-dimensional space [23–26]. Based on Roweis and Tenenbaum’s basic research on manifold learning, researchers proposed some improved feature extraction methods, including Laplacian eigenmap (LE) [27], local learning projection (LLP) [28], linear representation-based classifiers (CRC) [29], and linear regression classification (LRC) [30].

Due to the validity of the least-squares regression method in data analysis and the completeness of statistical theory, it is widely used as a basic tool in many machine learning problems including discriminant analysis, clustering, multiview learning, multilabel classification, and semisupervised learning. Sun et al. [31] proposed a least-squares regression model based on generalized eigenvalue decomposition, and Suzanna et al. [32] proposed a weighted least-squares regression model. When solving the multilabel classification problem with least-squares regression, if the data points belong to different categories, then should be considered as the increase of the distance between the data. For example, in order to increase the distance between different types of data points, Leski [33] proposed a quadratic approximation least-square regression model based on misclassification error. However, this model only considers two categories of classification problems. In multilabel classification, the distances among data points from different classes are also expected to be as large as possible [34, 35], as well as in multilabel feature selection. Therefore, multiple least-squares regression models proposed by Leski [33] can be used simultaneously, but the time cost of algorithm implementation will be relatively high.

The traditional label relaxation regression algorithm directly fits the original data without considering the local structure information of the data [36–38]. While the label relaxation regression algorithm for graph regularization considers local geometric information, its performance is largely dependent on graph construction [39, 40]. However, there are two defects in the construction of traditional graphs: first, it is largely influenced by the parameter values; second, it relies on the original data when constructing the weight matrix, which usually contains a lot of noise [41]. This

makes the constructed graph often not to be optimal, which affects the subsequent work. To guarantee the global optima of the latent representation and graphs of all views, we integrate the graph completion and common representation learning into a joint optimization framework. [42–44]. Therefore, it is proposed to combine manifold learning with label relaxation regression and construct weight graph through adaptive method. At the same time, local identification information is added on the basis of original linear identification analysis so that the projection learning can grasp the local identification information to expand the identification of the projection. A label relaxation regression algorithm for image classification and feature extraction based on adaptive graph is proposed. In this paper, the main innovation points are as follows:

- (1) The adaptive graph construction can rightly capture the local structure information of the data
- (2) The problem of overfitting is avoided by introducing adaptive graph into the objective function of label relaxation regression
- (3) In order to take full advantage of discriminant information, the global discriminant information based on the LDA is considered

The rest of this paper is arranged as follows: Section 2 briefly reviews DLSR and structured optimal graph. In Section 3, discriminative label relaxed regression with adaptive graph learning (DLRR_AG) is described in detail, and the convergence of the proposed algorithm is proved. Section 4 mainly provides a lot of experiments to verify the performance of the proposed algorithm. The last Section gives the conclusion.

2. Related Work

2.1. Discriminative Least-Squares Regression (DLSR) Algorithm. The classification training samples are given N numbers, and these samples $\{(x_i, y_i)\}_{i=1}^N$ fall into C ($C \geq 2$) classes, where $y_i \in \{1, 2, \dots, c\}$ is the class label of x_i , and x_i is a data point in R^m . The linear equation can be satisfied as follows:

$$XW + e_N t^T \approx Y, \quad (1)$$

where $X = [x_1, x_2, \dots, x_N]^T \in R^{N \times m}$ and $Y = [f_{y_1}, f_{y_2}, \dots, f_{y_N}]^T \in R^{N \times C}$. For the i th class, $i = 1, 2, \dots, c$, $f_i = [0, \dots, 0, 1, 0, \dots, 0]^T \in R^c$, W is a transformation

matrix in $R^{m \times c}$, t is a translation vector in R^c , and $e_N = [1, 1, \dots, 1] \in R^N$ is a vector with all 1s.

Each column vector in Y is of a binary regression type, with the target of class j th being “+1” and the target of the rest being “0.” We can drag these binary outputs far away along two opposite directions. That is, with a positive slack variable, we hope the output will become for the sample grouped into “1” and for the sample grouped into “0.” This treatment can help enlarge the distance between the classes and mapping data point.

Let $B \in R^{N \times C}$ be a constant matrix, in which the i th row and j th column element B_{ij} is defined as

$$B_{ij} = \begin{cases} +1, & \text{if } y_i = j, \\ -1, & \text{otherwise.} \end{cases} \quad (2)$$

Each element in B corresponds to the direction of the drag. Performing the above ε drag on each element of Y , matrix $M \in R^{N \times C}$ records these $\{\varepsilon\}$. The residual can be obtained as follows:

$$XW + e_N t^T - (Y + B \otimes M) = 0, \quad (3)$$

where \otimes is the Hadamard matrix product operator.

The objective function of DLSR can be obtained as follows:

$$\begin{aligned} \min_{W, t, M} & \|XW + e_N t^T - Y - B \otimes M\|_F^2 + \lambda \|W\|_F^2, \\ \text{s.t.} & M \geq 0, \end{aligned} \quad (4)$$

where λ is a positive regularization parameter. By solving the optimization problem of equation (4), the optimal W and t can be obtained:

$$\begin{aligned} W &= (X^T H X + \lambda I_m)^{-1} X^T H R, \\ t &= \frac{(R^T e_N - W^T X^T e_N)}{N}, \end{aligned} \quad (5)$$

where I_m is a $m \times m$ identity matrix and $H = I_N - (1/N)e_N e_N^T$, in which I_N is a $N \times N$ identity matrix, and $R = Y + B \otimes M \in R^{N \times C}$.

2.2. Structured Optimal Graph. It is well known that the data in high-dimension space is usually embedded in low-dimensional manifold [40]. It is a key success factor to preserve local manifold structure information for graph-based methods. The local manifold structure is captured by the similarity matrix, which determines the ultimate performance of graph-based methods.

Suppose that there are N training samples from c classes which are denoted by $X = [x_1, x_2, \dots, x_N] \in R^{N \times m}$, where N is the number of samples used for classifier training, while all N samples are used for determination of S , where m is the dimension of observed data, and $x_i = (i = 1, 2, \dots, N) \in R^{m \times 1}$ is the i -th sample for sample set X . For any sample x_i , it can be connected by all other samples with probability s_{ij} , where s_{ij} is an element of similarity matrix $S \in R^{N \times N}$. If the distance of two samples is closer, the greater their probability (or similarity) will be, and vice versa. Therefore, the similarity s_{ij} between x_i and x_j is inversely proportional to their distance. The similarity s_{ij} can be obtained by solving the following equation:

$$\begin{aligned} \min_{i, j} & \left(\|x_i - x_j\|_2^2 S_{ij} + \alpha S_{ij}^2 \right), \\ \text{s.t.} & \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1, \end{aligned} \quad (6)$$

where $S_i \in R^{N \times 1}$ is a vector whose j th element is s_{ij} in similarity matrix S and α is a regularization parameter. The second item in equation (6) is mainly to avoid trivial solutions.

It is the ideal state for each sample to include c -nearest-neighbor numbers. That is to say, each S_{ij} ($i = 1, 2, \dots, N$) in similarity matrix S has exact c connected components. In fact, the obtained similarity matrix S in equation (6) does not meet this requirement in most cases. The problem can be solved as follows. The spectral analysis has an important equation as follows:

$$\sum_{i, j} \|f_i - f_j\|_2^2 S_{ij} = 2 \text{Tr}(F^T L_S F), \quad (7)$$

where $F = [f_1, f_2, \dots, f_N] \in R^{N \times C}$ is a class label matrix corresponding to the observed data X , $L_S = D - (S^T + S)/2$ is the Laplacian matrix, and matrix D is a diagonal matrix whose i th entry is $\sum_j S_{ij} + S_{ij}/2$.

If the rank of Laplacian matrix L_S equals to $N - C$, namely, $\text{rank}(L_S) = N - C$, the obtained similarity matrix S will include exact C connected components [45]. By combining the constraint to equation (6), equation (6) is written as

$$\min \sum_{i, j} \left(\|x_i - x_j\|_2^2 S_{ij} + \alpha S_{ij}^2 \right). \quad (8)$$

In order to solve equation (6), the i th smallest eigenvalue of Laplacian matrix L_S is denoted by $\sigma_i(L_S)$. It is well-known that the solutions of positive semidefinite matrix are more than zero. Laplacian matrix L_S is positive semidefinite, so $\sigma_i(L_S) \geq 0$. Based on $\text{rank}(L_S) = N - C$, $\sum_{i=1}^C \sigma_i(L_S) = 0$ is satisfied [45]. According to KyFan's Theorem [46], we have

$$\sum_{i=1}^C \sigma_i(L_S) = \min_{F^T F = I} \text{Tr}(F^T L_S F). \quad (9)$$

Therefore, based on equation (9), equation (8) can be rewritten as

$$\begin{aligned} \min_{i, j} & \left(\|x_i - x_j\|_2^2 S_{ij} + \alpha S_{ij}^2 \right) + 2\lambda \text{Tr}(F^T L_S F), \\ \text{s.t.} & \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1, F^T F = I. \end{aligned} \quad (10)$$

3. Discriminative Label Relaxation Regression Algorithm Based on Adaptive Graph (DLRR_AG)

In this section, the motivation of our DLRR_AG is firstly introduced. Then, the optimum solution of DLRR_AG is given.

3.1. The Motivation of DLRR_AG. Traditional label relaxation regression algorithms directly fit the original data, which often results in overfitting. In general, to overcome overfitting, a regularization term is added to the target equation, and a regularization factor is used to balance the target equation and

regularization term. In addition to this one, maintaining the information of local manifold structure plays a very important role in improving classification or clustering. There are two kinds of classical nearest-neighbor graphs, one is k -nearest-neighbor graph, and the other is ε -nearest-neighbor graph. A large part of graph-based algorithms rely on these two methods to preserve local manifold structure information. However, two problems often occur when these two methods are used. First, the performance of graph learning is greatly affected by the parameter k or ε , the results of taking different parameter values are sometimes far apart, and the optimal value is not easy to determine, which requires a lot of experiments to obtain, which consumes a lot of time. Second, the traditional

graph construction method requires two complicated steps: first, the corresponding weighted matrix should be constructed in adjacent graphs, and then the relaxation regression should be carried out. However, once the weighted matrix is generated from the most primitive observation data, it will not change any more. There is no flexibility. This kind of weighted matrix generated in advance is not practical in practical application, because the original observation data often contain a lot of errors, which will lead to the destruction of the local manifold structure. To solve this problem, we propose a relaxation regression algorithm for adaptive graph. The objective function of the algorithm is described as follows:

$$\begin{aligned} \min_{W, M} & \|XW - Y - B \otimes M\|_F^2 + \sum_{i,j} \left(\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \alpha s_{ij}^2 \right) + 2\lambda \text{Tr}(F^T L_S F), \\ \text{s.t.} & M \geq 0, \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1, F^T F = I. \end{aligned} \quad (11)$$

In order to take advantage of intraclass discriminative information, the discriminant information based on the

LDA is introduced into the objective function. Equation (11) can be rewritten as

$$\begin{aligned} \min_{W, M} & \|XW - Y - B \otimes M\|_F^2 + \sum_{i,j} \left(\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \alpha s_{ij}^2 \right) + 2\lambda \text{Tr}(F^T L_S F) + \gamma (\text{Tr}(W^T (S_w - S_b) W)), \\ \text{s.t.} & M \geq 0, \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1, F^T F = I, \end{aligned} \quad (12)$$

where S_w and S_b are the within-class scatter matrix and between-class scatter matrix, respectively.

3.2. Optimization of DLRR_AG. Equation (12) of the objective function is convex, so it is difficult to get the global optimal solution. Therefore, we can obtain the local optimal solution through continuous iteration. Because the target function contains four different variables, optimization solution (12) is not directly available; it requires an iterative solution to this problem (12). We propose an iterative algorithm to update the rules to solve these problems.

3.2.1. Fixing W , M , and F to Update S . When other variables are fixed, except S , equation (12) can be transformed into

$$\begin{aligned} \min & \sum_{i,j} \left(\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \alpha s_{ij}^2 \right) + 2\lambda \text{Tr}(F^T L_S F), \\ \text{s.t.} & \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1. \end{aligned} \quad (13)$$

According to equation (7), equation (13) can be rewritten as

$$\begin{aligned} \min & \sum_{i,j} \left(\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \alpha s_{ij}^2 \right) + \lambda \sum_{i,j} \|f_i - f_j\|_2^2 S_{ij}, \\ \text{s.t.} & \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1. \end{aligned} \quad (14)$$

Because it is independent for the similarity vector of each data point, we can solve the optimal problem for each sample as follows:

$$\begin{aligned} \min & \sum_{i,j} \left(\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \alpha s_{ij}^2 \right) + \lambda \sum_{i,j} \|f_i - f_j\|_2^2 S_{ij}, \\ \text{s.t.} & \forall i, s_i^T 1 = 1, 0 \leq s_{ij} \leq 1. \end{aligned} \quad (15)$$

Let $m_{ij} = \|W^T x_i - W^T x_j\|_2^2$, $n_{ij} = \|f_i - f_j\|_2^2$, and $d_{ij} = |m_{ij} + \lambda n_{ij}|$; equation (15) can be rewritten as

$$\begin{aligned} \min & \left\| s_i + \frac{1}{2\alpha} d_i \right\|_2^2, \\ \text{s.t. } & s_i^T 1 = 1, 0 \leq S_{ij} \leq 1. \end{aligned} \quad (16)$$

3.2.2. Fixing S , W , and M to Update F . When other variables are fixed, except F , equation (12) can be transformed into

$$\begin{aligned} \min & \text{Tr}(F^T L_S F), \\ \text{s.t. } & F^T F = I. \end{aligned} \quad (17)$$

The optimal solution of F is formed from the eigenvectors of the c minimum eigenvalues in the L_S .

3.2.3. Fixing S , W , and F to Update M . When other variables are fixed, except M , equation (12) can be transformed into

$$\begin{aligned} \min & \|XW - Y - B \otimes M\|_2^2, \\ \text{s.t. } & M \geq 0. \end{aligned} \quad (18)$$

Let us now consider optimization in terms of $M \in R^{N \times C}$. Given W and T , and let $P = XW - Y$ record the regression error of n data points, then the optimization problem can be solved from the following aspects:

$$\begin{aligned} \min & \|P - B \otimes M\|_F^2, \\ \text{s.t. } & M \geq 0. \end{aligned} \quad (19)$$

According to the square of matrix Frobenius norm, the fact that an element can be decoupled, (10) can be decoupled

equivalently into $n \times c$ subproblems. For the i th row of the matrix and the j th column element M_{ij} , we have

$$\begin{aligned} \min_{M_{ij}} & \|P_{ij} - B_{ij} M_{ij}\|_F^2, \\ \text{s.t. } & M \geq 0, \end{aligned} \quad (20)$$

where P_{ij} and B_{ij} are the i th row and j th elements of P and B , respectively.

Note that $B_{ij}^2 = 1$. Thus, we have $(P_{ij} - B_{ij} M_{ij})^2 = (B_{ij} - P_{ij} M_{ij})^2$. Then, the optimization problem of equation (20) can be rewritten as

$$\begin{aligned} \min_{M_{ij}} & (B_{ij} P_{ij} - M_{ij})^2, \\ \text{s.t. } & M_{ij} \geq 0. \end{aligned} \quad (21)$$

Obviously, the optimal solution of equation (21) is given as follows:

$$M = \max(B \otimes P, 0). \quad (22)$$

3.2.4. Fixing S , M , and F to Update W . Equation (12) is a constrained convex optimization problem. According to the properties of convex optimization, the local optimal solution is also the optimal solution of the whole. Next, this paper uses the iterative method to solve the optimal solution of equation (12).

To solve W , given M , equation (12) is an unconstrained convex optimization problem for W . You just take the derivative of that and you set the derivative to 0 and you get W .

When other variables fix, except W , equation (12) can be transformed into

$$\min_W \|XW - Y - B \otimes M\|_F^2 + \sum_{i,j} \|W^T x_i - W^T x_j\|_2^2 + \gamma (\text{tr}(W^T (S_w - S_b)) W). \quad (23)$$

Let $N = Y + B \otimes M$, then the problem in (23) can be transformed into the following problem:

$$\begin{aligned} \min_W & \|XW - N\|_F^2 + \lambda \text{Tr}(W^T X L X^T W) + \gamma (\text{tr}(W^T (S_w - S_b) W)) \\ & = \text{Tr}(W^T X^T X W - W^T X^T N) + \lambda \text{Tr}(W^T X L X^T W) + \gamma (\text{tr}(W^T (S_w - S_b) W)) \\ & = \text{Tr}(W^T (X^T X + \lambda X L X^T + \gamma (S_w - S_b)) W - W^T X^T N) \\ \Rightarrow & W = (X^T X + \lambda X L X^T + \gamma (S_w - S_b))^{-1} X^T N, \end{aligned} \quad (24)$$

where $L = D - S$ is the graph Laplacian and D is a diagonal matrix whose diagonal elements $D_{ii} = \sum_j S_{ij}$.

4. Experiments

In this section, we perform a large number of experiments, and the experimental results can prove that our proposed DLRR_AG algorithm can achieve high classification accuracy. In order to prove the effectiveness of our DLRR_AG algorithm, six public image databases are used to validate our method. For the sake of contrast, the proposed DLRR_AG algorithm is compared with other typical feature extraction methods as follows:

- (i) Collaborative representation-based classification (CRC) [29]: CRC is a combined result of machine learning and compressed sensing, which shows its good classification performance on face image data. Considering the training samples from a specific class and the query set as two linear subspaces, the classwise prototypes most correlated with the query set are learned, resulting in a condensed gallery set.
- (ii) Linear regression classification (LRC) [30]: LRC has attracted a great amount of attention owing to its promising performance in face recognition. However, its performance will dramatically decline in the scenario of limited training samples per class, particularly when only single training sample is available for a specific person.
- (iii) Flexible manifold embedding (FME) [45]: FME is a semisupervised manifold learning framework with good applicability. It can effectively utilize label information from labeled data as well as a manifold structure from both labeled and unlabeled data.
- (iv) Joint global and local structure discriminant analysis (JGLDA) [46]: for linear dimension reduction, it preserves the local intrinsic structure, which characterizes the geometric properties of similarity and diversity of data by two quadratic functions.
- (v) Flexible linear regression classification (FLRC) [47]: the inferences are based on the least-squares estimators of the model which have been shown to be coherent with the interval arithmetic defining the model and to verify good statistical properties.
- (vi) Discriminative least-squares regression (DLSR) [34]: DLSR is to embed class label information into the LSR formulation such that the distances between classes can be enlarged. In order to implement this idea, a technique called ϵ -dragging is introduced to force the regression targets of different classes moving along with opposite directions.

4.1. Experiments on YALE Database. Yale faces a database of 15 people containing 165 photos. They were taken by 15 people under different lighting conditions and with different facial expressions. Each person took 11 pictures. In our experiment, each image was manually cropped to a size of 50

by 40 pixels. Figure 1 shows the sample images of one of them.

In this experiment, the first 2, 3, to 6 images from each object are used for training set, and the rest is utilized for test set. In order to evaluate the algorithm more objectively, we will eliminate random effects on the algorithm in the process of implementation, and all the methods are repeated 10 times. The recognition rates of all algorithms are shown as Table 1.

It can be seen from Table 1 that we can draw two points. Firstly, the recognition performance of the proposed DLRR_AG method is better than DLSR method irrespective of the number of training samples. Secondly, our method is superior to all other methods, except when the training sample is 4.

4.2. Experiments on ORL Database. The ORL dataset includes 400 face images from 40 different objects, and each object has 10 face images. For some people, their images were taken at different time and different light; image content includes different facial expression and facial details. Figure 2 shows the sample images of one of them.

In this experiment, the training set was the first 3, 4, 5, and 6 images of each person, and the test set was the remaining images. All algorithms are repeated 10 times. The recognition rates of each algorithm are shown in Table 2.

We can clearly see from Table 2 that the proposed DLRR_AG is superior to CRC, LRC, FLRC, FME, LRR JGLDA, and DLSR.

4.3. Experiments on Georgia Tech Database. The Georgia Tech face database contains photos of 50 people taken during two or three sessions and produced at Georgia Tech. In the database, each individual took 15 color JPEG images with a cluttered background and a resolution of 640 by 480 pixels. The faces in these pictures may be front and tilted, or they may be front or tilted. These images include the different expressions, illuminations, and proportions. Each image is manually cropped to 60 by 50 pixels. All images are converted to grayscale images in the experiment. Figure 3 shows the sample images of one of them.

In this experiment, the training set was the first 4, 5, to 8 images of each person, and the test set was the remaining images. Repeat the algorithm 10 times. The recognition rates of each algorithm are shown in Table 3.

As can be seen from Table 3, DLRR_AG performs well compared with all other algorithms on the Georgia Tech database. In particular, the performance of DLRR_AG is much higher than that of CRC.

4.4. Experiments on CMU PIE Database. The CMU PIE database contains 41,368 facial images. The images were taken by 68 people with different expressions, decorations, and postures. The acquisition of multiple images of each object is based on the premise of fixed expression and attitude, and the illumination is changed to obtain 14 face

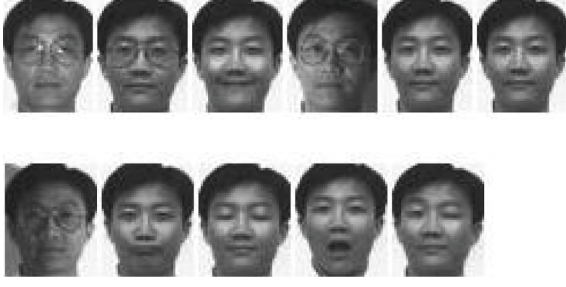


FIGURE 1: Sample face images from Yale database.

TABLE 1: Average recognition rate of different methods on the Yale database (%).

Methods	The number (or proportion) of the training samples per class				
	2 (18.18%)	3 (27.27%)	4 (36.36%)	5 (45.45%)	6 (54.54%)
CRC	67.14	66.79	65.83	70.06	74.52
LRC	65.73	62.94	61.89	69.76	73.79
FLRC	66.17	65.98	67.85	70.85	74.85
FME	60.73	62.94	70.89	71.86	76.34
JGLDA	68.55	69.45	69.64	75.48	75.82
DLSR	67.38	67.62	66.07	70.60	75.48
DLRR_AG	69.17	70.71	70.60	73.10	76.55

Bold values indicate the highest recognition rate of all methods.



FIGURE 2: Sample face images from ORL face database.

TABLE 2: Average recognition rate of different methods on the ORL database (%).

Methods	The number (or proportion) of the training samples per class			
	3 (30%)	4 (40%)	5 (50%)	6 (60%)
CRC	87.86	91.25	92.00	93.75
LRC	82.14	85.42	89.50	95.60
FLRC	86.56	88.50	93.20	95.61
FME	83.00	86.42	90.12	95.33
JGLDA	88.55	89.63	92.93	93.61
DLSR	88.45	90.23	92.66	93.25
DLRR_AG	88.93	91.67	94.50	95.63

Bold values indicate the highest recognition rate of all methods.

images, and then these images are cut to 32×32 pixels. Figure 4 shows multiple images from an object.

In this experiment, the first 4, 5, 6, 7, and 8 face images of each object are selected to act as training set, and the rest are

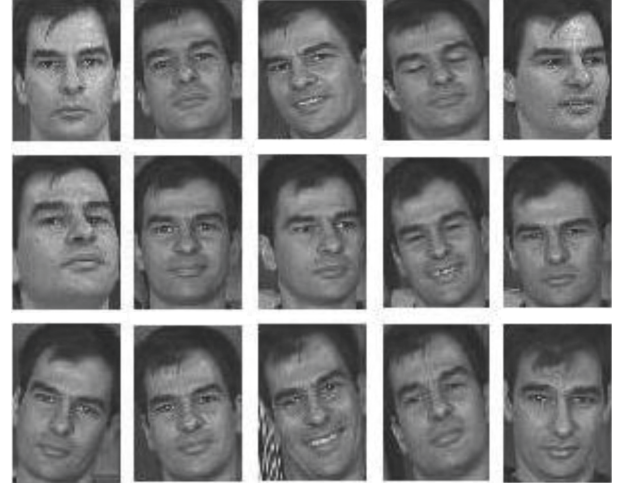


FIGURE 3: Sample face images from Georgia Tech face database.

TABLE 3: Average recognition rate of different methods on the Georgia Tech database (%).

Methods	The number (or proportion) of the training samples per class				
	4 (26.67%)	5 (33.33%)	6 (40.00%)	7 (46.67%)	8 (53.33%)
CRC	57.64	61.43	68.11	71.25	75.40
LRC	56.36	59.60	66.67	70.00	74.57
FLRC	58.22	60.18	67.10	71.38	75.20
FME	53.28	58.56	67.55	70.22	73.55
JGLDA	57.68	59.32	66.79	68.95	72.84
DLSR	50.55	54.40	65.33	69.88	72.10
DLRR_AG	59.27	63.60	68.22	73.00	75.42



FIGURE 4: Some face images of one object from CMU PIE database.

taken as test set. The proposed algorithms are repeated for 10 times. The recognition rates of all algorithms are shown in Table 4.

From Table 4, we can know that the proposed DLRR_AG can respectively obtain the best recognition performance in the corresponding comparison algorithms irrespective of the variations of training sample size.

4.5. Experiments on AR Database. The AR face database [44] contains color face images of 120 people, and the total number of face images exceeds 4,000. Among them, 120 subjects were photographed twice with different facial expressions, light conditions, and shade, with a 14-day interval,

TABLE 4: Average recognition rate of different methods on the CMU PIE database (%).

Methods	The number (or proportion) of the training samples per class				
	4	5	6	7	8
	(19.05%)	(23.81%)	(28.57%)	(33.33%)	(38.10%)
CRC	79.43	84.27	87.68	88.99	92.20
LRC	70.72	79.61	82.86	87.57	89.62
FLRC	78.22	80.23	84.36	88.45	91.33
FME	70.93	77.52	81.33	85.07	86.96
JGLDA	78.55	79.78	79.95	88.63	88.69
DLSR	74.25	80.29	85.98	88.74	91.95
DLRR_AG	81.50	84.71	88.15	89.75	92.59

Bold values indicate the highest recognition rate of all methods.



FIGURE 5: Sample face images from AR database.

and each person produced 26 images. In our experiment, out of the 26 facial images of 120 people, seven were selected from each stage, or 14 face images per person. Each image is manually cropped to 50 by 40 pixels. All images are converted to grayscale images. Figure 5 shows the sample images from one object.

In this experiment, we obtained 14 images of unmasked faces from the first and second experiments. The images in the first stage (ranging from 3 to 7) were used as training images, and the face images in the second stage were used as test images. The proposed algorithms are repeated for 10 times. The experiment results are shown in Table 5.

As can be seen from Table 5, compared with the proposed DLRR_AG algorithm, the classification results of CRC, LRC, FLRC, FME LRR, and JGLDA are poor. In other words, the proposed method can achieve the best recognition performance.

4.6. Experiments on UMIST Database. The UMIST face database contains a total of 575 facial images of 20 people. The 575 photos had all the poses of 20 people in the database, a mix of race, gender, and physical appearance. The views are different for each topic, between 19 and 48 views for each topic. The size of the face image in the view is 56 by 48 pixels. Figure 6 shows example images of a person.

In this experiment, the first 1, 2 to 5 face images of each person are generally selected as the training set, while other images are used as the test images. The algorithm is repeated 10 times for each test. The experimental results are shown in Table 6.

TABLE 5: Average recognition rate of different methods on the AR database (%).

Methods	The number (or proportion) of the training samples per class				
	3	4	5	6	7
	(21.43%)	(28.57%)	(35.71%)	(42.86%)	(50.00%)
CRC	70.71	70.60	73.10	76.55	77.38
LRC	66.79	65.83	70.06	74.52	75.12
FLRC	67.62	66.07	70.60	75.48	75.71
FME	70.88	70.96	73.30	76.85	77.84
JGLDA	68.55	69.45	69.64	75.48	75.82
DLSR	67.62	66.07	70.60	75.48	75.71
DLRR_AG	71.50	72.19	74.67	77.29	79.31

Bold values indicate the highest recognition rate of all methods.



FIGURE 6: Sample face images from the UMIST database.

TABLE 6: Average recognition rate of different methods on the UMIST database (%).

Methods	The number (or proportion) of the training samples per class				
	1	2	3	4	5
	(21.05%)	(26.31%)	(31.58%)	(36.84%)	(42.10%)
CRC	46.77	60.00	71.55	74.68	84.33
LRC	48.53	60.56	71.72	75.63	84.60
FLRC	49.93	60.88	72.87	76.24	83.97
FME	56.25	58.28	69.79	75.31	83.29
JGLDA	53.13	59.67	69.66	78.59	87.80
DLSR	55.45	60.00	74.25	80.23	85.98
DLRR_AG	56.78	61.69	73.79	81.61	88.28

Bold values indicate the highest recognition rate of all methods.

As can be seen from Table 6, when the training sample size is 3, the recognition rate of DLRR_AG algorithm is slightly lower than that of DLSR algorithm. However, the recognition rate of the proposed DLRR_AG algorithm is higher than other algorithms when the training sample size is not 3.

5. Conclusion

This paper presents a discriminative label relaxation regression algorithm based on adaptive graph (DLRR_AG) algorithm, which can effectively alleviate the overfitting problem caused by label relaxation by correctly capturing the local structure information of the original observed data. The main innovation of this paper has the following points. (1) The adaptive neighborhood graph can well capture the essential local structural information of the original data. (2) Label relaxation, manifold learning, and discriminant analysis are integrated into a unified framework. A large number of experiments in six public image databases show that the proposed method is superior to other related

methods. Therefore, the proposed method is effective and feasible.

Data Availability

The image databases used in this paper are publicly available for scientific research.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Natural Science Foundation of China (U1504610, 61971339, and 61471161), the Key Project of the Natural Science Foundation of Shaanxi Province (2018JZ6002), Scientific and Technological Innovation Team of Colleges and Universities in Henan Province (20IRTSTHN018), and the Doctoral Startup Foundation of Xi'an Polytechnic University (BS1616).

References

- [1] C. Lohrmann, P. Luukka, M. Jablonska-Sabuka, and T. Kauranne, "A combination of fuzzy similarity measures and fuzzy entropy measures for supervised feature selection," *Expert Systems with Applications*, vol. 110, pp. 216–236, 2018.
- [2] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2018.
- [3] X. Peng, J. Lu, and R. Yan, "Automatic subspace learning via principal coefficients embedding," *IEEE Transactions on Cybernetics*, vol. 47, no. 11, pp. 3583–3596, 2017.
- [4] P. Chen, L. Jiao, F. Liu, J. Zhao, Z. Zhao, and S. Liu, "Semi-supervised double sparse graphs based discriminant analysis for dimensionality reduction," *Pattern Recognition*, vol. 61, pp. 361–378, 2017.
- [5] R. Ji, H. Liu, L. Cao, D. Liu, Y. Wu, and F. Huang, "Toward optimal manifold hashing via discrete locally linear embedding," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5411–5420, 2017.
- [6] A. K. Jain, "Statistical pattern recognition," *IEEE Transaction of Fattern Analysis and Machine Intllilnce*, vol. 22, no. 1, pp. 4–37, 2018.
- [7] W. K. Wong, Z. Lai, J. Wen, X. Fang, and Y. Lu, "Low-rank embedding for robust image feature extraction," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2905–2917, 2017.
- [8] H. Du, X. Zhang, Q. Hu, and Y. Hou, "Sparse representation-based robust face recognition by graph regularized low-rank sparse representation recovery," *Neurocomputing*, vol. 164, pp. 220–229, 2015.
- [9] Z. Zhang, Y. Zhang, S. Li et al., "Robust unsupervised flexible auto-weighted local-coordinate concept factorization for image clustering," in *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2092–2096, Brighton, UK, May 2019.
- [10] W. Chen and Q. Wang, "Local and non-local feature-based kernel nonnegative matrix factorization method for face recognition," in *Proceedings of the International Conference on Computational Intelligence and Security*, pp. 107–110, IEEE, Wuxi, China, December 2017.
- [11] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognition*, vol. 43, no. 12, pp. 4165–4176, 2010.
- [12] L. Wang, H. Wu, and C. Pan, "Manifold regularized local sparse representation for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 4, pp. 651–659, 2015.
- [13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2019.
- [14] R. He, W. Zheng, and B. Hu, "Maximum correntropy criterion for bobust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561–1576, 2011.
- [15] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [16] Y. Xu, X. Fang, Q. Zhu, Y. Chen, J. You, and H. Liu, "Modified minimum squared error algorithm for robust classification and face recognition experiments," *Neurocomputing*, vol. 135, no. 5, pp. 253–261, 2014.
- [17] X. Cao and W. Shen, "Illumination invariant extraction for face recognition using neighboring wavelet coefficients," *Pattern Recognition*, vol. 45, pp. 1299–1305, 2018.
- [18] Y. Chen, J. Tao, L. Liu et al., "Research of improving semantic image segmentation based on a feature fusion model," *Journal of Ambient Intelligence and Humanized Computing*, 2020.
- [19] Z. Li, Z. Zhang, J. Qin, Z. Zhang, and L. Shao, "Discriminative Fisher embedding dictionary learning algorithm for object recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 786–800, 2020.
- [20] W. Lu, Y. Zhang, S. Wang et al., "Concept representation by learning explicit and implicit concept couplings," *IEEE Intelligent Systems*, 2021.
- [21] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2020.
- [22] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [23] L. Zhou, L. Wang, and C. Shen, "Feature selection with redundancy constrained class separability," *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 853–858, 2014.
- [24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2017.
- [25] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Learning a nonnegative sparse graph for linear regression," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 24, no. 9, pp. 2760–2771, 2015.
- [26] L. Wang, J. Zhu, and H. Zou, "Hybrid huberized support vector machines for microarray classification," in *Proceedings of the 24th International Conference on Machine learning—ICML'07*, pp. 983–990, Corvallis, OR, USA, 2017.
- [27] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral tech-niques for embedding and clustering," in *Advances in Neural Information Processing Systems*, pp. 585–591, MIT Press, Cambridge, MA, USA, 2001.

- [28] M. Wu, K. Yu, and S. Yu, "Local learning projections," in *Proceedings of the International Conference on Machine Learning*, pp. 1039–1046, Corvallis, OR, USA, June 2007.
- [29] M. Wu, J. Zhou, and J. Sun, "Query-expanded collaborative representation based classification with class-specific prototypes for object recognition," *Pattern Recognition*, vol. 47, no. 11, pp. 3585–3596, 2014.
- [30] Y. Peng, J. Ke, S. Liu, J. Li, and T. Lei, "An improvement to linear regression classification for face recognition," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 9, pp. 2229–2243, 2019.
- [31] J. Sun, B. Lu, and S. Wang, "Enhanced least squares regression adaption for semi-supervised learning," *Neural Networks*, vol. 65, pp. 1–17, 2015.
- [32] B. Suzanna and G. Zoubin, "Weighted least-squares regression via joint L2, 1-norms minimization," in *Advances in Neural Information Processing Systems*, pp. 1813–1821, MIT Press, Cambridge, MA, USA, 2010.
- [33] J. Leski, "A quadratic approximation least-square regression model," *Pattern Recognition*, vol. 24, no. 14, pp. 2281–2290, 2003.
- [34] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1738–1754, 2012.
- [35] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2018.
- [36] X. Li and Y. Pang, "Deterministic column-based matrix decomposition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 1, pp. 145–149, 2010.
- [37] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2108–2118, 2015.
- [38] M. Kalakech, P. Biela, L. Macaire, and D. Hamad, "Constraint scores for semi-supervised feature selection: a comparative study," *Pattern Recognition Letters*, vol. 32, no. 5, pp. 656–665, 2015.
- [39] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2408–2414, San Francisco, CA, USA, 2017.
- [40] M. Kalakech, P. Biela, L. Macaire, and D. Hamad, "Constraint scores for semi-supervised feature selection: a comparative study," *Pattern Recognition Letters*, vol. 32, no. 5, pp. 656–665, 2011.
- [41] F. Oveis, S. Oveis, A. Erfanian, and I. Patras, "Tree-structured feature extraction using mutual information," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 127–137, 2012.
- [42] J. Wen and Z. Zhang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Transactions on Cybernetics*, 2020, In press.
- [43] J. Wen, K. Yan, Z. Zhang et al., "Adaptive graph completion based incomplete multi-view clustering," *IEEE Transactions on Multimedia*, 2020, In press.
- [44] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multi-view clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1774–1782, 2019.
- [45] F. Nie and D. Xu, "Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1738–1754, 2012.
- [46] Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint global and local structure discriminant analysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 626–635, 2013.
- [47] A. Blanco-Fernández and G. González-Rodríguez, "Inferential studies for a flexible linear regression model for interval-valued variables," *International Journal of Computer Mathematics*, vol. 93, no. 4, pp. 658–675, 2014.