Low Rank Sample Reconstruction-based Semi-supervised Feature Subspace Learning

Ruoqi An[1], Ao Li [1], Deyun Chen[1], Xin Liu[1]

{anruoqi77@126.com, dargonboy@126.com, chendeyun@hrbust.edu.cn}

School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China[1]

Abstract. Feature subspaces have been widely applied in image classification. However, majority conventional subspace learning models are supervised，the data containing only partial labeled samples will lead to unsatisfactory classification results in practical applications. To address this problem, in this paper, we design a graph semi-supervised feature subspace model named low rank sample reconstruction-based semi-supervised feature subspace learning, which combines graph-based semi-supervised learning and low lank sample reconstruction into a unified framework. The proposed model implementing low rank constraint on representation coefficient, exploits the graph-based label propagation algorithm to predict the unlabeled labels to obtain the labels of all samples. The experimental results demonstrate that the robustness and the classification efficiency of our model.

Keywords: Image classification; Subspace learning; Graph semi-supervised learning; Low Rank Sample Reconstruction;

1   Introduction

High-dimensional data brings difficulties to image processing, so that how to eliminate redundant data in high-dimensional data becomes a crucial problem to be solved in pattern recognition [1] and computer vision [2]. Dimension reduction can effectively improve performance and efficiency, whose purpose is to eliminate redundant information to enhance the recognition speed. Subspace learning have been extensively used to dimensionality reduction methods and achieved impressive performance. As the unsupervised subspace algorithms, Principal Component Analysis (PCA) aims to seek a subspace whose projection sample has the most massive variance [3]. Meanwhile, Linear Discriminant Analysis (LDA) is an effective method in the supervised subspace learning [4], in which it captures the features with the largest discriminative low dimensional features from high dimensional data, so that the ratio of the dispersion between samples and the dispersion within samples is the largest. Locality Sensitive Discriminant Analysis (LSDA), as a further extension of LDA, aims at preserving the discriminant structure and local geometry of the data [5].

Low-rank representation model (LRR) is introduced for learning valid features from data containing redundant information and mining potential manifold structures. In this decade，LRR has been widely applied to dictionary learning[6], transfer learning[7], [8], domain adaptation [9]. In recent years, particularly, several improved LRR-based subspace models have been proposed. For instance, a subspace clustering approach about low rank structured representation was proposed by Yao et al. [10]. In order to

make the clustering more robust, this method applied the global structure to obtain the low rank representation and generated low rank structure to extract the neighborhood construction. Other LRR-based feature subspace learning models with outstanding performance can be referred to [11], [12], [13].

Most subspace models with the impressive performance are supervised. However, in many practical problems, only a small number of labeled data are available, it caused weak discrimination. Several semi-supervised models for feature subspace have been presented in recent years. The semi-supervised discriminant analysis (SDA) was proposed by Cai et al. [14], in which the method improved the discriminant ability of the model through labeled samples and explored manifold structure through unlabeled samples. Flexible manifold embedding (FME) is a semi-supervised learning framework [15], in which semi-supervised dimension reduction is performed by using the manifold structure of unlabeled sample and information of the labeled sample. Chen et al. proposed the method named rescaled linear square regression (RLSR), which make the feature selection model solve the optimal projection matrix under global and sparse conditions [16].

Inspired by the above semi-supervised model, a semi-supervised model named low rank sample reconstruction-based semi-supervised feature subspace learning is proposed in this paper. The core idea of our model is to combine label propagation and feature subspace learning to obtain better classification results.

The key contributions of the proposed model can be summarized. 1) we proposed a novel semi-supervised feature subspace learning model, which integrates the LRR model, label propagation, and feature subspace learning. The coefficients of LRR simultaneously constrain the prediction of label propagation information and feature subspace learning. 2) Our designed optimization solution applies the Lagrange multiplier [17] to calculate the objective function and guarantee its convergence. 3)We compared different traditional semi-supervised methods to conduct the experiment by using three publicly available datasets and two different classifiers (KNN and SRC). Numerous experiments results show the effectiveness of our model.

The other section in the paper is arranged as follows. In the second section, we briefly introduce graph-based methods and sample reconstruction constraints in related work. And then, in the third section, we introduce the proposed semi-supervised learning method in detail and listed the solution process of the objective function. In the fourth section, we evaluate the proposed method through comparative experiments and parameter analysis with traditional methods. Finally, the conclusion is presented in the fifth section.

2   Related work.

2.1   Graph-based methods

Graph-based methods have excellent potential because of their diverse structures that can represent similarities between samples. In graph-based methods, given a weighted graph $G = (X, E, W)$, where $X = \{x_i\}_{i=1}^n$ represents the vertex set. And, W is the weight of the undirected edge set $E = (x_i, x_j)$. Then, the graph-based algorithms will be defined as:

$$\sum_i \sum_j \|g_i - g_j\|_2^2 W_{ij} \tag{1}$$

where $g_i = g(x_i)$ is the loss function, which is required to be smooth the entire graph and W_{ij} is the coefficient of the weight matrix W. By Gaussian kernel function, W is calculated as follows:

$$W_{ij} = \exp\left(-\mu \|x_i - x_j\|_2^2\right) \tag{2}$$

The labels of unlabeled samples are predicted by the connection between unlabeled samples and labeled samples. The graph-based objective function can be written as follows:

$$\infty \sum_i \|y_i - g_i\|_F^2 + \sum_i \sum_j \|g_i - g_j\|_2^2 W_{ij} \tag{3}$$

Where y_i denotes the predicted labels. The labeled information is transmitted by the similarity between vertices. The more extensive similarity between the sample vertices, the labeled information is more easily propagated. More methods about graph-based can be reference at [18], [19], [20].

2.2 Sample reconstruction constraint

The sample reconstruction constraint based on the core of the Fisher criterion is modeled to eliminate noise interference [21]. Assuming $X = [x_1, x_2, \ldots, x_n]$ is the training set, Z is representation coefficient matrix, the reconstructed constraint item obtained from sample self-representation $X = XZ$ as follows:

$$f(P, Z) = Tr\left(S_w\left(P^T XZ\right)\right) - Tr\left(S_b\left(P^T XZ\right)\right) \tag{4}$$
$$s.t. \ X = XZ + E, P^T P = I_P$$

Where P is feature subspace, which is an orthogonal matrix to reduce redundant information. $S_b\left(P^T XZ\right)$ and $S_w\left(P^T XZ\right)$ denotes the inter-class and within-class scatter divergence matrix, $Tr(\bullet)$ denotes the trace. Since Z is non-convex, adding a relaxation term to the constraint term can be expressed as:

$$f(P, Z) = \left\|P^T XZ(I - D_u)\right\|_F^2 - \left\|P^T XZ(D_u - D_v)\right\|_F^2 + \eta \left\|P^T XZ\right\|_F^2 \tag{5}$$
$$s.t. \ X = XZ + E, P^T P = I_P$$

Where I is an identity matrix, D_u and D_v are two constant-coefficient matrices, and η is regularization parameter. In detail, $B_u(i, j) = 1/n_k$, when x_i and x_j belongs to the same class. Otherwise, $B_u(i, j) = 0$.

3 Our model

The proposed objective function and its detailed description are presented in this section, and we also design the optimal scheme to solve the objective function by using Lagrange multiplier method.

3.1 Proposed method

To identify the noisy information in the original samples and obtain the robust subspace, we implement low rank constraint on representation coefficient. Meanwhile, the low rank constraint is used to predict the labels of unlabeled samples during the graph-based label propagation process. The low rank constraint can effectively eliminate the noise component in the sample. Moreover, the reconstruction of samples can further explore the nearest neighbor relationship, which makes the different classes of samples far away from each other and improve the discriminability of the model. The objective function with low rank representation constraint can be obtained:

$$\min_{P,Z,E} \|Z\|_* + \lambda_2 \|E\|_1 + \lambda_1 \left(\|P^T XZ (I - D_u)\|_F^2 - \|P^T XZ (D_u - D_v)\|_F^2 + \eta \|P^T XZ\|_F^2 \right) \tag{6}$$
$$s.t. \ X = XZ + E, \ P^T P = I_P$$

where $\|Z\|_*$ denotes the nuclear norm of Z. λ_1 and λ_2 denotes two regularization parameters.

Next, by integrating graph learning and low-rank reconstructed subspace learning into the same learning framework，we extend the semi-supervised method to feature subspace learning. In the graph propagation algorithm, the similarity between the two samples is constrained by low rank representation coefficient. The final objective function we can obtain as followed:

$$\min_{F,P,Z,E} \frac{1}{2} \|F - B\|_F^2 + \sum_{ij} Z_{ij} \|F_i - F_j\|_2^2 + \frac{1}{2} \|F - P^T X\|_F^2 + \|Z\|_* + \lambda_2 \|E\|_1$$
$$+ \lambda_1 \left(\|P^T XZ (I - D_u)\|_F^2 - \|P^T XZ (D_u - D_v)\|_F^2 + \eta \|P^T XZ\|_F^2 \right) \tag{7}$$
$$s.t. \ X = XZ + E, \ P^T P = I_P, Z_{ij} \geq 0$$

where $B=[B_1, B_2, ..., B_m]$ denotes the matrix with all the label vectors of unlabeled samples and labeled

samples. $B_i =[-1,...,-1,1,-1,...,-1]^T \in R^C$ is the i-th column of B. The c-th element of B_i is 1 and the

other elements is -1 when B_i belongs to c class. $B_j = [0,...,0,...,0]^T \in R^C$ is the label vector of j

-th unlabeled sample. F represents the learned label matrix through the similarity measure with B. In

equation(7), the first two items learn the complete label information through label propagation, Z_{ij} is

used as similarity weights to constrain the relationship between two labels, we also introduce a non-negative constraint on Z as the non-negative regularization parameter.

We have described the graph-based semi-supervised approach and low rank reconstructed subspace learning jointly. Next, we will introduce the optimal scheme.

3.2   Scheme

Since the minimization for all variables is not a convex problem, we design a solution based on alternating iteration. Aim at relaxing the issue of minimization, we introduce two auxiliary variables R and J. The Eq. (7) can be rewritten as:

$$\min_{F,P,Z,E,R,J} \frac{1}{2}\|F-U\|_F^2 + \sum_{ij} R_{ij}\|F_i-F_j\|_2^2 + \frac{1}{2}\|F-P^T X\|_F^2 + \lambda_2\|E\|_1$$
$$+ \lambda_1\left(\|P^T XZ(I-D_u)\|_F^2 - \|P^T XZ(D_u-D_v)\|_F^2 + \eta\|P^T XZ\|_F^2\right) \tag{8}$$
$$s.t.\ X=XZ+E,\ R_{ij}\geq 0,\ P^T P=I, Z=H, Z=J$$

Next, the Lagrangian function in Eq. (8) can be obtained as the following form:

$$\mathcal{L}(F,P,Z,E,R,J)=\frac{1}{2}\|F-U\|_F^2 + \sum_{ij} R_{ij}\|F_i-F_j\|_2^2 + \frac{1}{2}\|F-P^T X\|_F^2$$
$$+ \lambda_1\left(\|P^T XZ(I-D_u)\|_F^2 - \|P^T XZ(D_u-D_v)\|_F^2 + \eta\|P^T XZ\|_F^2\right) + \eta\|J\|_* + \lambda_2\|E\|_1$$
$$+ \frac{\mu}{2}\left(\|X-XZ-E\|_F^2 + \|Z-J\|_F^2 + \|Z-R\|_F^2\right)$$
$$+ \mathrm{Tr}\left(Y_1^T(X-XZ-E)\right) + \mathrm{Tr}\left(Y_2^T(Z-J)\right) + \mathrm{Tr}\left(Y_3^T(Z-R)\right) \tag{9}$$

Furthermore, Eq. (9) can be converted into the compact form:

$$\mathcal{L}(F,P,Z,E,R,J)=\frac{1}{2}\|F-U\|_F^2 + \sum_{ij} R_{ij}\|F_i-F_j\|_2^2 + \frac{1}{2}\|F-P^T X\|_F^2$$
$$+ \lambda_1\left(\|P^T XZ(I-D_u)\|_F^2 - \|P^T XZ(D_u-D_v)\|_F^2 + \eta\|P^T XZ\|_F^2\right) + \eta\|J\|_* + \lambda_2\|E\|_1$$
$$+ \frac{\mu}{2}\left(\left\|X-XZ-E+\frac{Y_1}{\mu}\right\|_F^2 + \left\|Z-J+\frac{Y_2}{\mu}\right\|_F^2 + \left\|Z-R+\frac{Y_3}{\mu}\right\|_F^2\right) \tag{10}$$

Finally, the minimized problem of Eq. (7) can be rewritten as:

$$\min_{F,P,Z,E,R,J} \mathcal{L}(F,P,Z,E,R,J)$$
$$s.t.\ R_{ij}\geq 0,\ P^T P=I \tag{11}$$

We use an iterative method to solve each variable. At the k-th iteration, variables other than F are fixed, and the objective function formula for F is:

$$\min_F \frac{1}{2}\|F-P^{T(k)}X\|_F^2 + \frac{1}{2}\|F-U\|_F^2 + \sum_{ij} R_{ij}^k\|F_i-F_j\|_2^2 \tag{12}$$

To simplify the solution, we convert Eq. (12) to the following form:

$$\min_F \frac{1}{2}\|F-P^{T(k)}X\|_F^2 + \frac{1}{2}\|F-U\|_F^2 + \mathrm{Tr}\left(FLF^T\right) \tag{13}$$

where $L=D-R$ presents the matrix of graph Laplacian, D denotes the diagonal matrix ($D_{ii}=\frac{\sum R_{*i}+\sum R_{i*}}{2}$).Eq. (13) can easily solved by setting the derivative to 0:

$$F=\left(U+P^{T(k)}X\right)(2I+L)^{-1} \tag{14}$$

Then, we update the objective function about eliminating the terms irrelevant to P:

$$\min_P \frac{1}{2}\|F-P^T X\|_F^2 + \lambda_1\left(\|P^T XZ(I-D_u)\|_F^2 - \|P^T XZ(D_u-D_v)\|_F^2 + \eta\|P^T XZ\|_F^2\right) \tag{15}$$

Due to the orthogonal constraint is contained in Eq. (15), the derivative can be obtained firstly:

$$\frac{\partial \mathcal{L}_P}{\partial P} = X^T X P - X F + 2\lambda_1 X Z \left(I - D_H\right)\left(I - D_H\right)^T Z^T X^T$$
$$+ 2\lambda_2 X Z \left(D_u - D_v\right)\left(D_u - D_v\right)^T Z^T X^T P + 2\lambda_1 \eta X Z Z^T X^T P \qquad (16)$$

Up to now, Eq. (16) can be solved regarding the methods in [22].

After fixing the variables independent of J in Eq. (10), we can obtain:

$$\min_{J} \left\| J - \left(Z^k + \frac{Y_2^k}{\mu}\right)\right\|_F^2 + \eta \|J\|_* \qquad (17)$$

Eq. (17) is a classical rank minimization problem by using existing technology to solve [23]. Removing the terms irrelevant to R, we obtain:

$$\min_{R} \left\| Z^k - R + \frac{Y_3^k}{\mu}\right\|_F^2 + \sum_{ij} R_{ij} \left\| F_i^{k+1} - F_j^{k+1}\right\|_2^2 \qquad (18)$$

The Eq. (18) can be rewritten as:

$$\min_{R} \left\| R - \left(Z^k + \frac{Y_3^k}{\mu}\right)\right\|_F^2 + S^{k+1} \otimes R \qquad (19)$$
$$s.t.\ R_{ij} \geq 0$$

where S^{k+1} is a matrix consists of $S_{ij}^{k+1} = \left\| F_i^{k+1} - F_j^{k+1}\right\|_2^2$. Eq. (19) can be transformed as follows:

$$\min_{R} \left\| R - \left(Z^k + \frac{Y_3^k}{\mu}\right)\right\|_F^2 + \left\| S^{k+1} \otimes R\right\|_1 \qquad (20)$$
$$s.t.\ R_{ij} \geq 0$$

Eq. (20) can be regarded as the non-negative weight l_1-norm minimization that can be solved efficiently in [24]. Then, after solving the auxiliary variables J and R, the variables other than Z are fixed, and the objective function for Z is:

$$\min_{Z} \lambda_1 \left(\left\| P^T X Z\left(I - D_u\right)\right\|_F^2 - \left\| P^T X Z\left(D_u - D_v\right)\right\|_F^2 + \eta\left\| P^T X Z\right\|_F^2\right)$$
$$+ \frac{\mu}{2}\left\| X - X Z - E^k + \frac{Y_1^k}{\mu}\right\|_F^2 + \left\| Z - J^{k+1} + \frac{Y_2^k}{\mu}\right\| + \left\| Z - R^{k+1} + \frac{Y_3^k}{\mu}\right\|_F^2 \qquad (21)$$

The solution can be obtained simply by derivation.

Finally, we express the objective function about the error matrix E by fixing other variables:

$$\min_{E} \lambda_2 \|E\|_1 + \frac{\mu}{2}\left\| E - \left(X - X Z^{k+1} + \frac{Y_1^k}{\mu}\right)\right\|_F^2 \qquad (22)$$

Through defining $\Omega = X - X Z^{k+1} + \frac{Y_1^k}{\mu}$, the i-th column of E^{k+1} is presented as:

$$E_i^{k+1} = \begin{cases} \dfrac{\|\Phi_i\|_2 - \lambda}{\|\Phi_i\|_2}, & if\ \lambda < \|\Phi_i\|_2 \\ 0, & otherwise \end{cases} \qquad (23)$$

We describe the detailed information of our scheme in **Algorithm 1**

Algorithm 1

Input: training set X, label U, $Z = R = J = 0$, $E = 0$, $Y_1 = Y_2 = Y_3 = 0$,

$\mu = 0.6$, $\mu_{max} = 10^{10}$, $\rho = 1.1$

Output: F, P

While not convergence do

    1. Update F^{k+1} using (12)

    2. Update P^{k+1} using (15);

    3. Update J^{k+1} using (17);

    4. Update R^{k+1} using (18);

    5. Update Z^{k+1} using (21);

    6. Update the Y_1^{k+1}, Y_2^{k+1}, Y_3^{k+1} and μ;

$$Y_1^{k+1} = Y_1^k + \mu\left(X - XZ^{k+1} - E^{k+1}\right)$$

$$Y_2^{k+1} = Y_2^k + \mu\left(Z^{k+1} - J^{k+1}\right)$$

$$Y_3^{k+1} = Y_3^k + \mu\left(Z^{k+1} - H^{k+1}\right)$$

$$Y_4^{k+1} = Y_4^k + \mu\left(Z^{k+1} - R^{k+1}\right)$$

$$\mu = \min\left(\mu_{max}, \rho\mu\right);$$

    end while

---

4   Experimental results and discussion

4.1   Experimental results

We choose 3 datasets to verify the method in the experiment, respectively named Extended YaleB, COIL20, and USPS in this section. According to the experimental criteria of the semi-supervised model,

half of the whole samples were labeled and the rest of samples were not labeled. The detailed description of the dataset is shown below:

Extended YaleB   Extended YaleB was photographed under a variety of controlled lighting conditions, including totally over 2, 000 frontal faces images. Part of its frontal face image is shown in Fig 1(a). Each image is adjusted to 32×32 pixels. For each class, we choose 32 face images for each class and half of the facial images as training samples randomly, and the others were chosen as test samples in our experiment.

COIL20   The COIL20 contains 20 objects, each of which is rotated 360 degrees horizontally to take an image every 5 degrees and 72 images of each object. We cut the size of the image data to 32×32 before the experiment as shown in Fig 1(b). 10 face images are chosen as training samples for each class.

USPS   USPS database is a database for digital handwriting recognition. There are 9298 images from zero to nine in the data set. For our experiment, the whole images are 16×16 gray pixel values, which have been normalized. As shown in Fig 2(c), 10 images of per digit are selected for training and the others are test data.
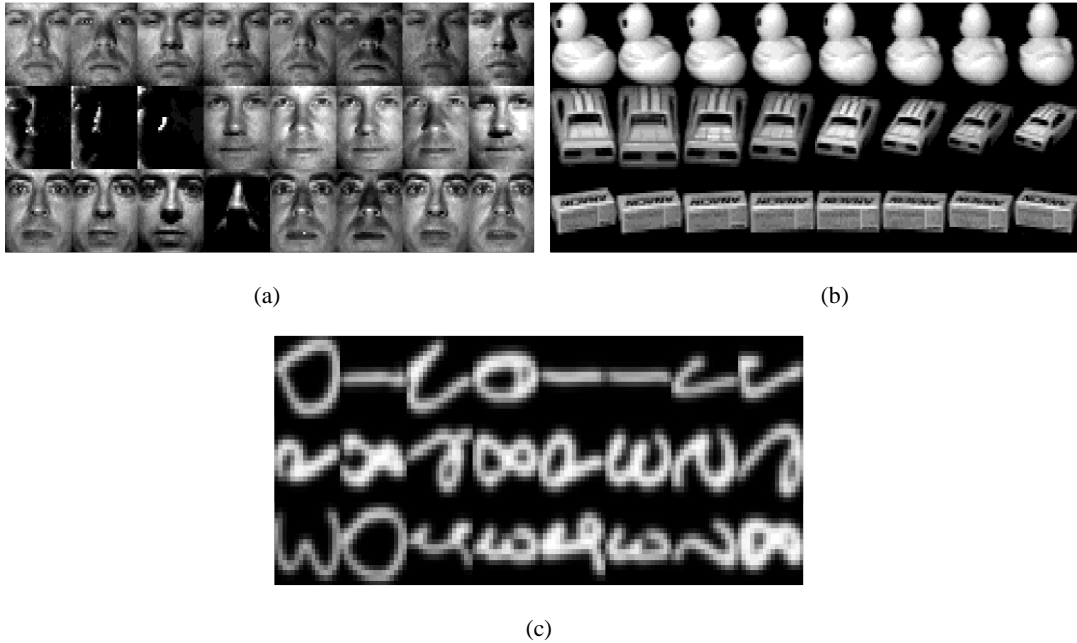

(a)   (b)

(c)

Fig 1. Sample images. (a) Extended YaleB, (b) COIL20, (c) USPS

Two representative methods of feature subspace learning and several novel semi-supervised approaches are chosen for comparison in experiments, including PCA, LDA, FME, RLSR, and SDA. To ensure that the experiment does not lose its generality, we use sparse representation classifier (SRC) and K nearest neighbors (KNN) to verify all methods on the test set. All experiments were performed 5 times to calculate standard deviation and mean as shown in Table 1 and 2. It is noted that, in Table 1 and 2,"Unlabeled" and "Test" represent the test sequence of the data with no labeled part of the training set and the testing set in our experiment.

Table1.   The classification results of experimental datasets with KNN (%)

Methods	Extended YaleB		COIL20		USPS	
	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test

Methods	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
PCA	---	70.38 ± 0.69	---	84.19 ± 0.45	---	70.20 ± 1.36
LDA	---	85.75 ± 0.63	---	84.23 ± 1.82	---	78.12 ± 2.30
FME	79.84 ± 1.42	84.12 ± 1.32	77.53 ± 1.87	73.26 ± 1.41	75.78 ± 2.83	64.52 ± 2.13
RLSR	75.78 ± 1.84	76.42 ± 1.32	82.84 ± 2.24	81.86 ± 1.85	67.79 ± 3.52	70.52 ± 3.28
SDA	80.95 ± 1.65	77.15 ± 1.23	76.85 ± 2.97	73.51 ± 2.28	63.54 ± 3.23	63.98 ± 2.13
Ours	80.95 ± 1.65	86.39 ± 1.91	84.28 ± 1.86	85.33 ± 1.24	76.99 ± 2.62	71.26 ± 1.77

Table2. The classification results of experimental datasets with SRC (%)

| Methods	Extended YaleB		COIL20		USPS	
	Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
PCA	---	80.68 ± 1.42	---	70.58 ± 1.86	---	77.10 ± 1.51
LDA	---	85.5 ± 1.05	---	66.0 ± 2.74	---	78.9 ± 2.56
FME	78.35 ± 0.35	81.54 ± 1.88	73.25 ± 3.29	68.86 ± 1.68	75.43 ± 2.35	75.20 ± 1.52
RLSR	83.34 ± 1.26	84.55 ± 1.35	76.08 ± 3.57	65.91 ± 2.45	75.86 ± 3.56	75.84 ± 2.41
SDA	84.98 ± 0.54	81.23 ± 1.29	72.17 ± 2.35	74.65 ± 3.14	72.89 ± 2.34	71.52 ± 1.61
Ours	86.67 ± 1.53	84.99 ± 0.89	78.27 ± 1.46	75.52 ± 3.16	79.23 ± 2.82	79.98 ± 2.94

Compared with the comparison method, the method we proposed shows better performance on all datasets. Furthermore, the experimental results under both classifiers show that our proposed approach has more robustness and effectiveness than other methods.

4.2　Discussion on parameters and convergence

In our model, there are three regularization parameters λ_1, λ_2 and η. The COIL20 dataset was selected to discuss the effects of changes in these three parameters on the model performance by KNN classifier and the classification rate curve with the changes of parameters was drawn in Fig 2. It can be indicated from the figure that as the λ_1, λ_2 and η change within a certain range, the classification rate does not change significantly. The fluctuation of the classification rate shows the performance stability in our model.
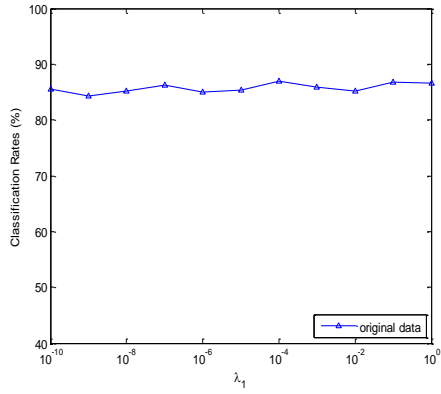
(a)
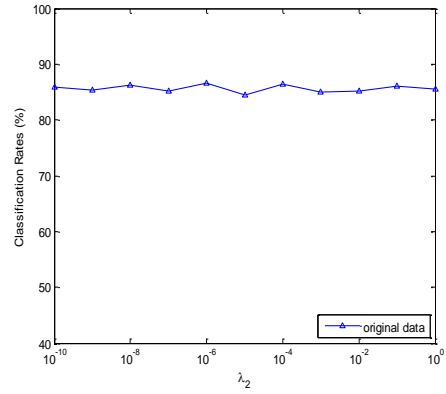
(b)
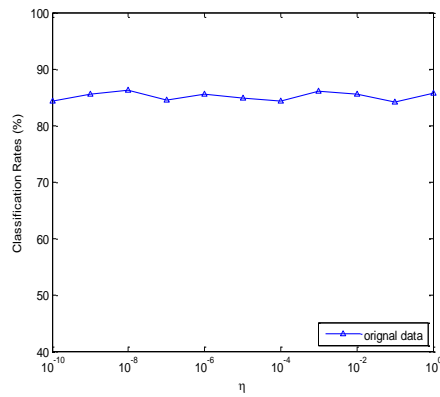
(c)

Fig 2. Classification results versus variational (a) λ_1 (b) λ_2 (c) η

In order to prove the convergence of the proposed method, we select Extended YaleB and used KNN classifier to draw the convergence curve with the increase of iterations. The convergence curve in our model can rapidly converge and the convergence curve will be stable within 50 iterations as shown in Fig 3.
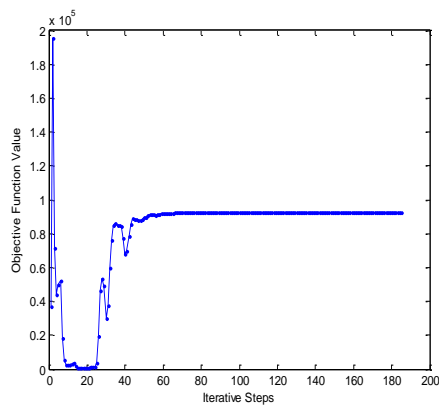

Fig 3. Objective function values versus iterative steps.

5 Conclusion

In this paper, we propose a semi-supervised subspace learning model based on low rank sample reconstruction. To obtain complete label information in incomplete labels and improve the robustness of the model, we jointly learn low rank sample reconstruction and graph-based semi-supervised learning. To calculate the objective function and ensure its convergence, we designed an optimization solution. Model optimization can be achieved during iterative processing. We conduct comparative experiments with some classical feature subspace methods and other novel semi-supervised approaches on Extended YaleB, COIL20, and USPS and perform analysis experiments on the convergence and parameters of the model. Experimental results indicate that the effectiveness and robustness performance of this model is superior to other methods.

References

[1] Fagan J F.: The origins of facial pattern recognition. Psychological development from infancy. Routledge. 83-113 (2017)

[2] Szegedy C, Vanhoucke V, Ioffe S, et al.: Rethinking the inception architecture for computer vision.Proceedings of the IEEE conference on computer vision and pattern recognition. 2818-2826 (2016)

[3] Turk M, Pentland A: Eigenfaces for recognition. Journal of Cognitive Neuroscience.3(1): 71-86 (1991)

[4] Belhumeur P N, Hespanha J P, Kriegman D J: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis & Machine Intelligence. (7): 711-720 (1997)

[5] Cai D, He X, Zhou K, Han J, Bao H: Locality sensitive discriminant analysis. International Joint Conference on Artificial Intelligence. 1713-1726 (2007)

[6] Zhou T, Liu F, Bhaskar H, et al: Robust visual tracking via online discriminative and low-rank dictionary learning. IEEE transactions on cybernetics. 48(9): 2643-2655 (2017)

[7] Wang L, Ding Z, Fu Y: Low-rank transfer human motion segmentation. IEEE Transactions on Image Processing. 28(2): 1023-1034 (2018)

[8] Wang S, Zhang L, Zuo W: Class-specific reconstruction transfer learning via sparse low-rank constraint.Proceedings of the IEEE International Conference on Computer Vision Workshops. 949-957 (2017)

[9] Wang M, Zhang D, Huang J, et al: Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation. IEEE Transactions on Medical Imaging. PP(99). 1-1 (2019)

[10] Sui Y, Wang G, Zhang L: Sparse subspace clustering via low-rank structure propagation. Pattern Recognition. 95: 261-271 (2019)

[11] Brbić M, Kopriva I: Multi-view low-rank sparse subspace clustering. Pattern Recognition.73: 247-258 (2018)

[12] Li A, Liu X, Wang Y, et al: Subspace structural constraint-based discriminative feature learning via nonnegative low rank representation. PloS one.14(5) (2019)

[13] Fei L, Xu Y, Fang X, et al: Low rank representation with adaptive distance penalty for semi-supervised subspace classification. Pattern Recognition.67: 252-262 (2017)

[14] Cai D, He X, Han J: Semi-supervised Discriminant Analysis. IEEE International Conference on Computer Vision. 4408856 (2007)

[15] Nie F, Xu D, Tsang I W H, et al: Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction. IEEE Transactions on Image Processing 19(7): 1921-1932 (2010)

[16] Chen X, Yuan G, Nie F, et al: Semi-supervised feature selection via rescaled linear regression. International Joint Conference on Artificial Intelligence. 1525-1531 (2017)

[17] Lin Z, Chen M, Ma Y: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv.1009.5055 (2010)

[18] Azran A: The rendezvous algorithm: Multiclass semi-supervised learning with markov random walks. International Conference on Machine Learning. 49-56 (2007)

[19] Szummer M, Jaakkola T: Partially labeled classification with Markov random walks. Advances in Neural Information Processing Systems. 945-952 (2002)

[20] Belkin M, Niyogi P, Sindhwani V: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. Journal of Machine Learning Research. 2399-2434 (2006)

[21] Li S, Fu Y: Learning robust and discriminative subspace with low-rank constraints. IEEE transactions on neural networks and learning systems. 27(11): 2160-2173 (2015)

[22] Wen Z, Yin W: A feasible method for optimization with orthogonality constraints. Mathematical Programming. 142(1-2): 397-434 (2013)

[23] Candès EJ, Li X, Ma Y, Wright J: Robust principal component analysis? Journal of the ACM. 58(3): 11 (2011)

[24] Yang J, Zhang Y: Alternating direction algorithms for \ell_1-problems in compressive sensing. SIAM journal on scientific computing. 33(1): 250-278 (2011)