Graph Regularized Low-Rank Representation for Semi-Supervised Learning

Cong-Zhe , Xiao-Jun Wu

School of IoT Engineering, Jiangnan University
Wuxi, China

Abstract

Recently, the low-rank representation (LRR) has attracted widely attention as its outstanding performance in exploring the data structure of the low-dimension subspaces. LRR method also widely used in the field of semi-supervised learning. However, the most of the current semi-supervised learning algorithm based on LRR have an obvious disadvantage: in the process of algorithm, the two steps of graph construction and semi-supervised learning are separated. Therefore, the existing label information of data samples in semi-supervised learning cannot be well used to guide the construction of the affinity graph. This will lead to this kind of methods cannot guarantee the obtained results are the global optimal solution. In this paper, we propose a graph regularized low-rank representation for semi-supervised learning, termed as GRLRR. Combing the construction of the affinity graph optimization, the proposed GRLRR method can get the global optimal solution. The experimental results on some benchmark datasets show that the effectiveness of the proposed GRLRR method.

Keywords: Low-Rank and Sparse Representation; Semi-Supervised Learning; Graph Regularization

Email address: youcongzhe@gmail.com ; wu_xiaojun@jiangnan.edu.cn (Cong-Zhe , Xiao-Jun Wu)

1. Introduction

In most pattern recognition computer vision problems, we often face the problem of insufficient samples with label information, while to get the label information is very difficult and high expensive. However, in real practice, the data samples we collected widely are often unlabeled. This brought great difficulties to the machine learning problems. Semi-supervised learning (SSL) is an important method of machine learning, it can make full use of the limited labeled data samples and the large number of unlabeled data samples [1] . Because of this advantage, in recent years, in the area of pattern recognition, machine learning and computer vision, the semi-supervised learning has got a lot of attention. Due to the success in real application and the high computation efficiency, the research of graph based semi-supervised learning (G-SSL) is very active in the field of semi-supervised learning.

For graph based method, we first need to construct a graph $G = (V, E)$, where V represents the vertex set of the graph and E is the edges set of the graph while the edges are associated with the weight matrix W of the dataset. The graph based semi-supervised learning algorithms rely heavily on the construction of the graph as the graph represents the relationships of the data in pairs, and reveals the structure of the dataset. In the learning process, the label information of the labeled data samples can be propagated effectively and efficiently to the rest unlabeled data samples in the dataset through the graph. Thus, how to construct an informative graph which can reveal the true structure of the dataset is quite important for many machine learning tasks including clustering and classification.

For graph based semi-supervised learning algorithm, a premise assumption is the consistency of the label of the data samples. For example, the neighboring data samples tend to have the same labels. As the graph construction is a key point for the graph based semi-supervised learning problem, there are a lot of researches have been done to explore how to construct an informative graph [2, 3, 4, 5] . From the view of machine learning, an informative graph

should have three characteristics [2] : high discriminating power, low sparsity and adaptive neighborhood.

We can discover thousands of various relationship in the dataset, but the validity of the relationship is the most important point. Recently, the studies about sparse representation and low-rank representation suggest that the pair relationship of the data samples is very important. Yan et al. [6, 7] put forward a method to construct the l_1 -graph based on sparse representation (SR) [8] which solves a l_1 optimization problem. However, a disadvantage of the sparse representation based methods is that they cannot describe the global structure of the data. In order to capture the global structure of the whole data set, Liu et al. proposed the low-rank representation (LRR) [9] method, and use the coefficients to construct an undirected affinity graph. Under the global low-rank constraint, the LRR-graph based methods can capture the global structure of the whole data set. The studies have shown that, under the moderate conditions, the LRR method can correctly preserve the membership of the data samples in the same subspace. However, compared with the l_1 -graph, the LRR methods tend to result in a dense graph, and that is not desirable for semi-supervised learning methods based on graph [2]. In addition, as the representation coefficients may be negative, which lack physical meaning for many image processing problems. In fact, the non-negative coefficients are more consistent with the biological modeling of the visual data [10, 11] , and this can lead to a better performance for data representation [11] and graph construction [12]. Zhuang et al. [13] put forward a non-negative low-rank and sparse graph (NNLRS) for semi-supervised learning methods and the edges of the graph is learned by the representation matrix. In this way, the obtained graph can preserve the global mixture structure of the subspaces and the local linear structure of the subspaces as well. In addition, in order to preserve the local structure of the data, a graph regularization term was added to the objective function of the LRR method to propose a graph regularized low-rank representation method [14] , this method can be used to well describe the hyperspectral images.

Although the methods based on sparse representation and low-rank repre-

sentation have gained a great success in many applications, this kind of method still have an obvious shortcoming. For graph based semi-supervised learning method, the structure of the graph is often pre-defined. As a result, the construction of the graph and the process of the semi-supervised learning are often two separated steps, while this will lead to the final obtained result of this kind of method cannot guarantee the global optimal solution. As the semi-supervised learning algorithms heavily depend on the construction of the graph, integrating the semi-supervised learning and the graph construction is necessary to solve jointly. Fang et al. [15] put forward a robust semi-supervised subspace clustering algorithm via non-negative low-rank representation (NNLRS) which unifies the two separate steps into a jointly single optimization framework. By using weak label regularized local coordinate coding (LCC) [16] , Wang et al. [17] proposed a face annotation method, in which the graph based weak label regularization and the sparse features were simultaneously used to improve the weal labels of similar facial images.

Motivated by the above analysis, we proposed a novel graph regularized low-rank representation method for semi-supervised learning (GRLRR). The idea of the entire learning process is that, the construction of the graph and the semi-supervised learning should be simultaneously performed to get a global optimal solution. In such a simultaneously learning scheme, the label information of the samples can be propagated accurately in the learning process via the graph structure.

In summary, our main contributions in this paper lie in the following aspects: (1) Unlike previous G-SSL methods, in which the graph structure and the designed algorithm are often independent steps, GRLRR integrates these two tasks into one single optimization step to guarantee an overall optimum. (2) By incorporating graph regularization and sparse constraint into LRR learning, the proposed method takes into account the intrinsic geometrical structure of the recovered data. GRLRR simultaneously captures the intrinsic local and global structure of the high-dimensional data. (3) An efficient optimized strategy is proposed to solve the optimization problem.

The rest of this paper is organized as follows: section 2 briefly reviews the related works. Section 3 introduces the idea of the proposed method GRLRR and its optimization. The extensive experiments performed to show the effectiveness of GRLRR are presented in Section 4. Finally, section 5 gives the conclusion.

2. Related Works

In this section, we briefly introduce the LRR and GLRR [18], and the semi-supervised classification framework used in the paper.

2.1. LRR and GLRR

Let $X = [x_1, x_2, \cdots, x_n] \in R^{d\times n}$ be a set of n data points in d-dimensional space. The goal of the low-rank representation (LRR) is to represent each data sample as a linear combination of the bases in $A = [a_1, a_2, \cdots, a_m] \in R^{d\times m}$ as $X = AZ$, where $Z = [z_1, z_2, \cdots, z_n]$ is the matrix with each z_i being the representation coefficient of sample x_i. Each element in z_i can be regarded as the contribution of the reconstruction of x_i with A as the basis. However, when the dictionary A is over-complete, there will be a lot of feasible solution to this problem. Low-rank representation (LRR) find the lowest rank solution by solving the following optimization problem:

$$\begin{aligned} &min_Z\ rank(Z) \\ &s.t.\ X = AZ \end{aligned} \tag{1}$$

where the low rankness has a good performance in exploiting the global structure of the dataset [9]. The optimal solution of the problem (1) is the lowest rank representation of data matrix X w.r.t. the dictionary A. In this paper, we use the data matrix X itself as the dictionary for simplification. However, due to the discrete nature of rank function, the above problem is NP-hard to solve. Fortunately, in real applications, we often use the nuclear norm to replace the

rank norm, and convert the problem (1) to the following optimization problem:

$$\begin{aligned} &min_Z \ \|Z\|_* \\ &s.t.\ X = AZ \end{aligned} \tag{2}$$

where $\|Z\|_*$ represents the nuclear norm, which is defined as the sum of all the singular values of Z. In real world, the dataset often corrupted by some noise, if we take these situations into consideration, we need add the noise into the objective function of LRR, thus, a more reasonable objective of LRR can be expressed as follows:

$$\begin{aligned} &min_{Z,E} \ \|Z\|_* + \lambda \|E\|_{2,1} \\ &s.t.\ X = AZ + E \end{aligned} \tag{3}$$

where the $l_{2,1}$-norm is defined as $\|E\|_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{d} e_{ij}^2}$ and parameter λ is used to balance the effect of low-rank term and error term.

in order to preserve the intrinsic manifold structure of the dataset, a graph regularization term is introduced into the objective function of LRR, and proposed the GLRR [18] method, the objective is defined as follows:

$$\begin{aligned} &min_{Z,E} \ \|Z\|_* + \lambda \|E\|_{2,1} + \beta tr(ZLZ^T) \\ &s.t.\ X = AZ + E \end{aligned} \tag{4}$$

where L is the graph Laplacian constructed by "HeatKernel" function in the Euclidean space. This model emphasizes the importance of the local consistency of the data while ignore the repulsion information of the dataset.

2.2. Semi-Supervised Learning

In this section, we present a kind of very popular semi-supervised learning method, Gaussian Fields and Harmonic Functions (GFHF) [19]. Suppose $Y \in R^{n \times c}$ is the label matrix, where $Y_{ij} = 1$ if sample x_i is associated with label j for $j \in \{1, 2, \cdots, c\}$ and $Y_{ij} = 0$ otherwise. $F \in R^{n \times c}$ is the predicted label

matrix, and it is estimated on the graph which takes the abel fitness and the manifold smoothness into consideration. Let us denote F_i and Y_i as the ith rows of F and Y, respectively. GFHF minimizes the following objective function

$$min_F \ \frac{1}{2}\sum_{i,j=1}^{n} \|F_i - F_j\|^2 S_{ij} + \lambda_\infty \sum_{i=1}^{u} \|F_i - Y_i\|^2 \tag{5}$$

where λ_∞ is a very large number such that $\sum_{i=1}^{u} \|F_i - Y_i\|^2 = 0$ is approximately satisfied and F is the predicted labels for all the samples. $S^{n\times n}$ is the graph weight matrix which represents the similarity of a pair of training samples. The above problem can be also reformulated as

$$min_F \ \frac{1}{2} tr(F^T L F) + tr((F-Y)^T U (F-Y)) \tag{6}$$

where $L \in R^{n\times n}$ is the graph Laplacian matrix and calculated as $L = D - S$, where $D_{ii} = \sum_j S_{ij}$ is a diagonal matrix. $U \in R^{n\times n}$ is also a diagonal matrix with the first u and the rest $n-u$ diagonal elements as λ_∞ and 0, respectively.

3. Graph Regularized Low-Rank Representation for Semi-Supervised Learning

In this section, the graph regularized low-rank representation for semi-supervised learning (GLRSC) is introduced. The goal of the proposed GLRSC method is, under a unified optimization framework, to perform the construction of the graph and the semi-supervised learning at the same time, thus, we can get an overall optimal solution.

3.1. Objective function of GLRSC

In GLRSC, the graph learning and semi-supervised learning are simultaneously completed within one step. Based on low-rank representation theory and

GFHF, we propose the following objective function of GLRSC:

$$\begin{aligned} min_{F,Z,E} \sum_{i=1}^{n}\sum_{j=1}^{n} \|F_i - F_j\|^2 Z_{ij} + tr((F-Y)^T U(F-Y)) + \\ \|Z\|_* + \alpha \|Z\|_1 + \beta tr(ZLZ^T) + \gamma \|E\|_{2,1} \\ s.t.\ X = AZ + E\ ,\ Z \geqslant 0 \end{aligned} \tag{7}$$

where α, β , γ are the parameters, which are used to balance the importance of the corresponding in the objective function. The first two items are a semi-supervised learning framework. The third term uses the low-rank constraint to guarantee the affinity matrix Z to capture the global mixture structure of the subspaces. The fourth term uses the l_1 -norm to enable the sparsity of the coefficients. The fifth term is the Laplacian regularizer, it takes into account the intrinsic geometrical structures within the data. As for the effect of noise, we use $l_{2,1}$-norm, the $l_{2,1}$-norm encourages the columns of E to be zero, which assumes that the corruptions are "sample-specific", i.e., some data vectors are corrupted and the others are clean. The non-negative constraint on Z aims to guarantee that the coefficients are meaningful and better embody the dependency among the data points.

3.2. LADMAP for solving GLRSC

In order to put forward an effective method to solve the problem (7), we use the linearized alternating direction method with adaptive penalty (LADMAP) [20]. In order to make the objective function separable, we introduce two auxiliary variables W and J.Thus the optimization problem can be rewritten as follows:

$$\begin{aligned} min_{F,Z,E} \sum_{i=1}^{n}\sum_{j=1}^{n} \|F_i - F_j\|^2 W_{ij} + tr((F-Y)^T U(F-Y)) + \\ \|Z\|_* + \alpha \|J\|_1 + \beta tr(ZL_1Z^T) + \gamma \|E\|_{2,1} \\ s.t.\ X = AZ + E\ ,\ Z = J\ ,\ Z = W\ ,\ Z \geqslant 0 \end{aligned} \tag{8}$$

To remove three linear constraints in (8), we can introduce three Lagrange multiplier Y_1, Y_2 and Y_3, therefore, the optimization problem can be rewritten as the following unconstrained minimization problem:

$$\begin{aligned} min_{F,Z,E} \sum_{i=1}^{n}\sum_{j=1}^{n} \|F_i - F_j\|^2 W_{ij} + tr((F-Y)^T U(F-Y)) + \\ \|Z\|_* + \alpha \|J\|_1 + \beta tr(ZL_1Z^T) + \gamma \|E\|_{2,1} + \\ \langle Y_1, X - AZ - E\rangle + \langle Y_2, Z - W\rangle + \langle Y_3, Z - J\rangle + \\ \frac{\mu}{2}\left(\|X - AZ - E\|_F^2 + \|Z - W\|_F^2 + \|Z - J\|_F^2\right) \end{aligned} \tag{9}$$

where $\psi(Z, W, J, E, Y_1, Y_2, Y_3) = \beta tr(ZL_1Z^T) + \frac{\mu}{2}\left\|X - AZ - E + \frac{1}{\mu}Y_1\right\|_F^2 + \frac{\mu}{2}\left\|Z - W + \frac{1}{\mu}Y_2\right\|_F^2 + \frac{\mu}{2}\left\|Z - J + \frac{1}{\mu}Y_3\right\|_F^2$ and $\langle A, B\rangle = tr(A^TB)$. $\mu \geq 0$ is a penalty parameter. This problem can be easily solved by alternately updating one variable while others fixed. Then, the multipliers are subsequently updated and the whole optimizing procedure is done in an iterative way till the convergence conditions are met.

A. Computation of Z

Solving (9) w.r.t. Z is equivalent to optimizing the following objective:

$$Z_{k+1} = argmin_Z \|Z\|_* + \langle \nabla_Z \psi(Z_k)\rangle + \frac{\theta\mu_k}{2}\|Z - Z_k\|_F^2 \tag{10}$$

where $\nabla_Z \psi$ is the partial differential of ψ with respect to Z. (10) has a closed-form solution given by:

$$Z_{k+1}^* = \Theta_{\frac{1}{\theta\mu_k}}(Z_k - \nabla_Z \psi(Z_k)/\theta) \tag{11}$$

where $\theta = \|A\|_F^2$, $\Theta(\cdot)$ denotes the singular value thresholding operator (SVT) [21].

B. Computation of J

Similarly, solving (9) w.r.t. J is equivalent to optimizing the following objective, while other variables are fixed to their current value

$$J_{k+1} = argmin_J \ \alpha \left\| J_k \right\|_1 + \frac{\mu}{2} \left\| J_k - \frac{1}{\mu_k} Y_{3,k} \right\|_F^2 \tag{12}$$

The sub-problem (12) has the following objective function:

$$J_{k+1} = max \left\{ S_{\frac{\alpha}{\mu}} (Z_{k+1} + \frac{1}{\mu_k} Y_{3,k}), 0 \right\} \tag{13}$$

where $S(\cdot)$ denotes the shrinkage operator [22].

C. Computation of E

The sub-problem for updating E can be recast as:

$$E_{k+1} = argmin_E \ \gamma \left\| E \right\|_{2,1} + \frac{\mu_k}{2} \left\| X - AZ_{k+1} + \frac{Y_{1,k}}{\mu_k} - E \right\|_F^2 \tag{14}$$

The solution is defined by

$$E_{k+1} = \Gamma_{\frac{\gamma}{\mu}} (X - AZ_{k+1} + \frac{1}{\mu_k} Y_{1,k}) \tag{15}$$

where $\Gamma(\cdot)$ denotes the $l_{2,1}$ minimization operator [9].

D. Computation of W

Solving (9) w.r.t. W is equivalent to optimizing the following objective:

$$W_{k+1} = argmin_{W \geqslant 0} \ tr(\Xi(R \odot W)) + \frac{\mu_k}{2} \left\| W - (Z_{k+1} + \frac{1}{\mu_k} Y_{2,k}) \right\|_F^2 \tag{16}$$

where $R_{ij} = \frac{1}{2} \left\| F_i - F_j \right\|^2$, $\odot$ is a Hadamard product operator of matrix and Ξ is a matrix with all elements are ones. We decompose problem (16) into n independent sub-problems each of which can be formulated as a weighted

non-negative sparse coding problem, namely

$$min_{W_i} \sum_{g=1}^{n} (W_k)_g^i \odot R_g^i + \frac{\mu_k}{2} \left\| (W_k)^i - (Z_{k+1} + \frac{Y_{2,k}}{\mu_k})^i \right\|_2^2 \quad s.t.\ W \geq 0 \tag{17}$$

where $(W_k)_g^i$ and $(R)_g^i$ are the g-th elements of i-th columns of matrix W_k and R respectively.

E. Computation of F

The sub-problem for updating F can be recast as:

$$\begin{aligned} F_{k+1} &= arg\ min_F \sum_{i=1}^{n} \sum_{i=1}^{n} \|F_{i,k} - F_{j,k}\|^2 W_{ij,k} + \\ &\qquad tr((F_k - Y)^T U (F_k - Y)) \\ &= arg\ min_F tr((F_k)^T L_2 F_k) + tr((F_k - Y)^T U (F_k - Y)) \end{aligned} \tag{18}$$

where $L_2 \in R^{n \times n}$ is the graph Laplacian matrix and calculated as $L_2 = D - W$, $D_{ii} = \sum_j W_{ij}$ is the diagonal matrix. It is straightforward to set the derivative of (18) with respect to F to zero, namely

$$\partial(min_F tr((F_k)^T F_k) + tr((F_k - Y)^T U (F_k - Y)))/\partial F_k = 0 \tag{19}$$

Then, we have

$$F_{k+1} = (L + U)^{-1} U Y \tag{20}$$

The detailed algorithm procedure of solving the proposed Laplacian regularized problem can be described in Algorithm 1.

Algorithm 1: LADMAP for solving GLRSC

Input: Data matrix X ; Label indicator matrix Y ; parameters α , β and γ

Initialization: $Z_0 = W_0 = J_0 = 0$; $Y_1 = Y_2 = Y_3 = 0$; $\mu_0 = 0.1$; $\mu_{max} = 10^7$; $\rho_0 = 1.01$; $\epsilon_1 = 10^{-7}$; $\epsilon_2 = 10^{-6}$; $\theta = \|A\|_F^2$; $k = 0$

while not converged do

1. Fixed the others and update Z by solving (10)

2. Fixed the others and update J by solving (12)

3. Fixed the others and update E by solving (14)

4. Fixed the others and update W by solving (16)

5. Fixed the others and update F by solving (18)

6. Update the multipliers as follows

$Y_{1,k+1} = Y_{1,k} + \mu_k(X - AZ_k - E_k)$

$Y_{2,k+1} = Y_{2,k} + \mu_k(Z_k - W_k)$

$Y_{3,k+1} = Y_{3,k} + \mu_k(Z_k - J_k)$

7. Update the parameter μ follows

$\mu_{k+1} = min(\mu_{max}, \rho\mu_k)$

where $\rho = \begin{cases} \rho_0 & \text{if } \mu_k\Omega / \|X\|_F \leq \epsilon_2 \\ 1 & otherwise \end{cases}$

8. Check the convergence conditions

$\|X - AZ_k - E_k\|_F \leq \epsilon_1$ or $\mu_k\Omega / \|X\|_F \leq \epsilon_2$

where $\Omega = max(\sqrt{\theta}\|Z_k - Z_{k+1}\|_F, \|J_k - J_{k+1}\|_F, \|W_k - W_{k+1}\|_F, \|E_k - E_{k+1}\|_F, \|F_k - F_{k+1}\|_F)$

9. Update k

$k \leftarrow k + 1$

end while

Output: F , Z , E

---

4. Experiments

4.1. Experiment setup

Datasets: we test our proposed method on three public datasets for evaluation: Extended Yale B, CMU PIE and USPS.

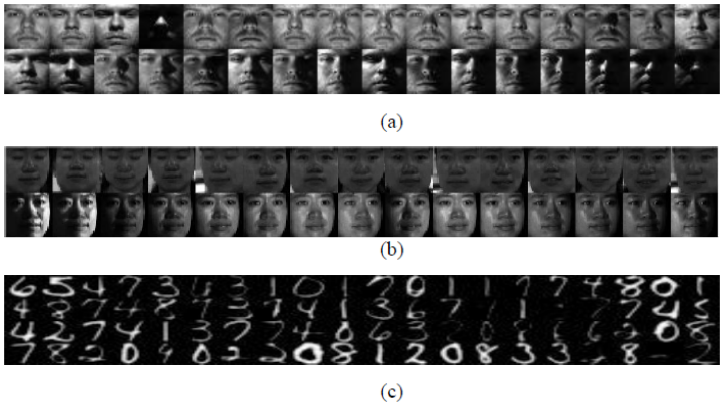

Figure 1: Sample images from ORL, Extended Yale B, CMU PIE and USPS datasets. (a) Extended Yale B (b) CMU PIE (c) USPS

-The extended Yale B dataset contains 38 subjects and there are about 64 frontal images for each individual. The facial images were taken in different light conditions. The face images of each subject correspond to a low-dimensional subspace. In this paper, because of the computational cost involved, we select the first 15 classes of the extended Yale B dataset as our test dataset. Thus, the test dataset contains in total 640 images and each image is resized to 32×32 pixels.

-CMU PIE: This face dataset contains 41368 images of 68 subjects with different poses, illumination and expressions. We select the first 15 subjects and only use their images in five near frontal poses (C05, C07, C09, C27, C29) and under different illuminations and expressions. Each image is manually cropped and normalized to a size of 32×32 pixels.

-USPS: The handwritten digit dataset contains 9298 handwritten digit images in total, each having 16×16 pixels. We only use the images of digits 1, 2, 3 and 4 as four classes, each having 1269, 926, 824 and 852 samples, respectively. So there are 3874 images in total.

Comparison Methods: We compare our proposed graph construction meth-

ods with the following state-of-the-art baseline methods:

-KNN-graph: We use the Euclidean distance as the similarity measure, and use the Gaussian kernel function to weight the edges of the graph. The Gaussian kernel parameter σ is set to 1. In this paper, we adopt two kinds of construction of graphs, termed as KNN1 and KNN2, where the number of nearest neighbors are set to 5 and 8, respectively.

-LLE-graph: Following the lines of [23], we construct two LLE-graphs, denoted as LLE1 and LLE2, where the numbers of nearest neighbors are 8 and 10, respectively. Since the weights W of LLE-graph may be negative and symmetric, similar to [7], we symmetrize them by $W = (|W| + \left|W^T\right|)/2$.

-l_1-graph: Following the lines of [7], we construct the l_1-graph. Since the graph weights W of l_1-graph is asymmetric, we also symmetrize it as suggested in [7].

-SPG: In essence, the SPG problem is a lasso problem with the non-negativity constraint, without considering corruption errors. Here we construct the SPG graph following the lines of [12].

-LRR-graph: Following [9], we construct the LRR-graph, and symmetrize it as we do for l_1-graph. The parameters of LRR are the same as those in [9].

-GLRR-graph: Following the lines of [18], we construct the GLRR-graph.

-NNLRS-graph: For NNLRS-graph, the two regularization parameters are set to $\beta = 0.2$ and $\lambda = 10$ according to [13].

4.2. Experimental studies

The purpose of semi-supervised learning task is to reveal more unlabeled information with limited known labeled data. Therefore, we select the percentage of labeled samples range from 10% to 60% and the rest as unlabeled samples. The parameters of the GLRSC method are set as $\alpha = 0.5$, $\beta = 0.01$, $\gamma = 10$. For fair comparison, we record the indices of the randomly selected labeled samples under each level and use these indices for all above mentioned methods. For each configuration, we conduct 50 independent runs for each algorithm. Table 1 to 3 report the experimental results.

Table 1 Classification accuracy rates (%) on Extended Yale B dataset

Labeled samples	10 %	20 %	30 %	40 %	50 %	60 %
KNN1	67.11	68.91	71.44	73.65	75.22	77.02
KNN2	63.16	64.41	66.46	69.03	70.27	71.42
LLE1	71.00	74.16	77.76	80.18	82.39	84.25
LLE2	70.24	73.35	77.17	80.10	82.35	84.06
l_1-graph	53.18	49.67	47.67	42.84	34.21	22.44
SPG	83.63	87.61	90.43	92.93	94.37	95.58
LRR	71.78	75.54	77.67	80.58	81.96	83.91
GLRR	73.83	75.62	77.91	80.77	82.34	84.33
NNLRS	94.44	94.69	95.71	96.25	96.00	96.77
GLRSC	94.98	95.36	96.16	96.84	96.61	97.49

Table 2 Classification accuracy rates (%) on CMU PIE dataset

Labeled samples	10 %	20 %	30 %	40 %	50 %	60 %
KNN1	65.72	66.94	69.89	71.54	73.04	74.91
KNN2	63.58	63.89	66.49	67.85	69.55	70.91
LLE1	67.75	69.58	73.48	76.38	78.35	80.44
LLE2	67.47	69.17	72.99	75.99	77.78	79.98
l_1-graph	78.29	82.82	87.94	90.59	93.39	94.87
SPG	80.25	84.55	89.29	91.75	93.71	95.05
LRR	68.74	70.18	74.39	76.14	78.76	79.95
GLRR	69.73	71.25	75.15	76.98	79.91	81.20
NNLRS	87.78	89.37	90.18	92.92	96.00	95.00
GLRSC	88.57	90.49	91.03	94.16	96.92	96.31

Table 3 Classification accuracy rates (%) on USPS dataset

Labeled samples	10 %	20 %	30 %	40 %	50 %	60 %
KNN1	96.87	97.78	98.45	98.80	99.18	99.35
KNN2	96.79	97.90	98.47	98.82	99.14	99.28
LLE1	72.31	77.57	80.82	83.38	85.72	87.39
LLE2	64.34	71.04	74.70	77.47	79.99	82.31
l_1-graph	66.48	73.58	81.08	83.36	88.33	91.11
SPG	93.08	95.96	97.31	98.12	98.86	99.17
LRR	96.51	98.17	98.78	99.08	99.39	99.51
GLRR	96.57	98.17	98.81	98.99	99.38	99.51
NNLRS	97.20	98.38	98.87	99.12	99.41	99.52
GLRSC	97.75	98.83	99.08	99.39	99.60	99.64

From the experimental results, we can observe that:

1. In most cases, compared to other graph based semi-supervised learning algorithms, the proposed GLRSC method can consistently get the highest classification accuracy, even with low labeled samples rate.

2. Compared with NNLRS method which also use the sparse and low-rank constraints to construct affinity graph, the proposed GLRSC method is able to use the label information to construct affinity matrix effectively. In most cases, the improvement of the classification accuracy is obvious.

3. Among the compared methods, l_1-graph uses the sparse constraint, SPG-graph imposes non-negative sparse constraint on the affinity matrix, such constraint only captures locally linear structure of the data. LRR-graph imposes the low-rank constraint which can capture the global mixture of subspaces structure, however, it often results in a dense graph which is undesirable for G-SSL. The proposed GLRSC method integrate the advantages of low-rank and sparse representation. The experimental results have also proven the effectiveness.

There are three parameters affecting the performance of our proposed GLRSC method. α and β are parameters to control the impact of sparse constraint and local affinity constraint respectively. γ is to deal with the gross corruption errors in the data. Similar to the previous experimental settings, we run GLRSC on each combination of parameters 50 independent times on Extended Yale B dataset. We select 50% samples as labeled and the remaining as the unlabeled samples. Fig. 2 shows the experimental results.

From Fig. 1, we can see that, the performance of GLRSC is much stable when α , β and γ vary in relative large ranges. α is used to balance the sparsity, when the value of α is small, the performance also decreases. This means that both low rankness and sparsity property are important for graph construction. As for β , when we set a big value, the accuracy will decrease. γ is to deal with the gross corruption errors in the data. And the experimental suggest a wide range is appropriate for the selection of γ .

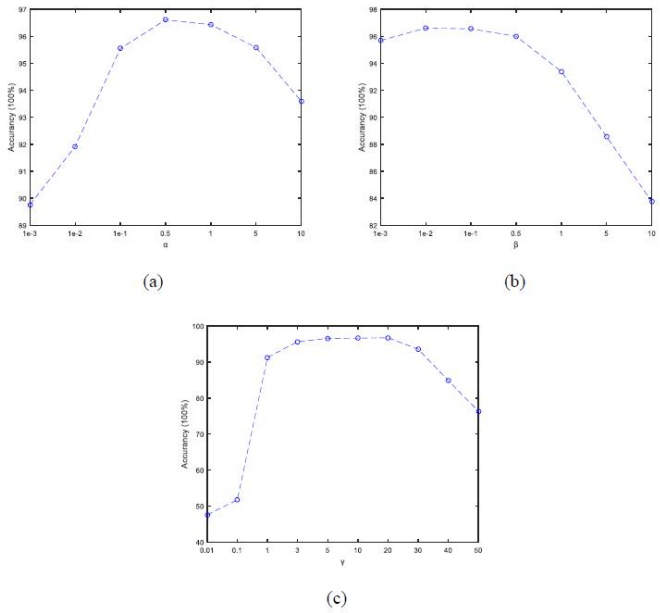

Figure 2: Classification accuracy with varied parameters (a) α (b) β (c) γ

5. Conclusion

In this paper, we propose a novel semi-supervised subspace clustering method named GLRSC, in which the label information is used to guide the affinity construction. Moreover, GLRSC integrates the affinity construction and semi-supervised subspace clustering into one step to guarantee an overall optimum. An associated efficient iteratively linearized ADM with adaptive penalty (LADMAP) is introduced to solve the optimization problem, which uses less auxiliary variables and less matrix inversion. The experimental results on three datasets show that our novel method compared with the state-of-the-art approaches is more effective through a set of evaluations both classification and recognition.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61373055) and the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20130093110009).

References

[1] X. Zhu, Semi-supervised learning literature survey, world 10 (2005) 10.

[2] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, Proceedings of the IEEE 98 (6) (2010) 1031–1044.

[3] S. I. Daitch, J. A. Kelner, D. A. Spielman, Fitting a graph to vector data, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 201–208.

[4] T. Jebara, J. Wang, S.-F. Chang, Graph construction and b-matching for semi-supervised learning, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 441–448.

[5] P. P. Talukdar, K. Crammer, New regularized algorithms for transductive learning, in: Machine Learning and Knowledge Discovery in Databases, Springer, 2009, pp. 442–457.

[6] S. Yan, H. Wang, Semi-supervised learning by sparse representation., in: SDM, SIAM, 2009, pp. 792–801.

[7] B. Cheng, J. Yang, S. Yan, Y. Fu, T. S. Huang, Learning with-graph for image analysis, Image Processing, IEEE Transactions on 19 (4) (2010) 858–866.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, Pattern Analysis and Machine Intelligence, IEEE Transactions on 31 (2) (2009) 210–227.

[9] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation., in: ICML, 2010, pp. 663–670.

[10] P. O. Hoyer, Modeling receptive fields with non-negative sparse coding, Neurocomputing 52 (2003) 547–552.

[11] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.

[12] R. He, W.-S. Zheng, B.-G. Hu, X.-W. Kong, Nonnegative sparse coding for discriminative semi-supervised learning, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 2849–2856.

[13] L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang, N. Yu, Non-negative low rank and sparse graph for semi-supervised learning, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2328–2335.

[14] M. Yin, J. Gao, Z. Lin, Laplacian regularized low-rank representation and its applications, IEEE Transactions on Pattern Analysis & Machine Intelligence 1 (1) (2013) 1–1.

[15] X. Fang, Y. Xu, X. Li, Z. Lai, W. Wong, Robust semi-supervised subspace clustering via non-negative low-rank representation., IEEE transactions on cybernetics.

[16] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: Advances in neural information processing systems, 2009, pp. 2223–2231.

[17] D. Wang, S. C. Hoi, Y. He, J. Zhu, T. Mei, J. Luo, Retrieval-based face annotation by weak label regularized local coordinate coding, Pattern Analysis and Machine Intelligence, IEEE Transactions on 36 (3) (2014) 550–563.

[18] X. Lu, Y. Wang, Y. Yuan, Graph-regularized low-rank representation for destriping of hyperspectral images, Geoscience and Remote Sensing, IEEE Transactions on 51 (7) (2013) 4009–4018.

[19] X. Zhu, Z. Ghahramani, J. Lafferty, et al., Semi-supervised learning using gaussian fields and harmonic functions, in: ICML, Vol. 3, 2003, pp. 912–919.

[20] Z. Lin, R. Liu, Z. Su, Linearized alternating direction method with adaptive penalty for low-rank representation, in: Advances in neural information processing systems, 2011, pp. 612–620.

[21] J.-F. Cai, E. J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM Journal on Optimization 20 (4) (2010) 1956–1982.

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

[23] J. Wang, F. Wang, C. Zhang, H. C. Shen, L. Quan, Linear neighborhood propagation and its applications, Pattern Analysis and Machine Intelligence, IEEE Transactions on 31 (9) (2009) 1600–1615.