2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science

Graph Regularized Low-Rank Representation for Semi-Supervised learning

Cong-Zhe You, Xiao-Jun Wu
School of IoT Engineering
Jiangnan University
Wuxi, China
congzhe_you@foxmail.com,
wu_xiaojun@jiangnan.edu.cn

Vasile Palade
School of Computing, Electronics, and Mathematics
Coventry University
Coventry, United Kingdom
vasile.palade@coventry.ac.uk

***Abstract*—Due to the excellent performance in exploring the structure of low-dimensional subspaces, the low-rank representation (LRR) has recently attracted wide attention of the researchers. However, in most current semi-supervised learning problems based on LRR method, the two steps of graph construction and semi-supervised learning are separated. Therefore, the existing label information cannot be well used to guide the construction of the affinity graph. Thus, these methods cannot guarantee the results are the global optimal solutions. In this paper, we propose a graph regularized low-rank representation for semi-supervised learning, termed as GLRSC. Combing the construction of the affinity graph and the semi-supervised learning, and solving the joint optimization, the proposed GLRSC method can get the global optimal solution. The experimental results on some benchmark datasets show that the effectiveness of the proposed GLRSC method.**

Keywords- low-rank and sparse representation; semi-supervised learning; graph construction

I. INTRODUCTION

In most of the computer vision and pattern recognition problems, we often face the problem of insufficient labeled data, while the acquisition of the label information is very difficult and expensive. However, the data we can widely obtain is often unlabeled. This brought great difficulties to the machine learning problem. Semi-supervised learning (SSL) can make full use of the limited labeled data and the large number of the unlabeled data. In the current semi-supervised learning methods, the graph-based SSL (G-SSL) is particularly attractive, mainly because of its success in the practical application and the computational efficiency.

The graph based semi-supervised learning methods heavily rely on the construction of a graph $G = (V, E)$ which can represent the data structures, where V is the vertex set of the dataset and E is the edges set of the graph associated with the weight matrix W. Through the graph, information of the labeled samples in the dataset can be efficient and effective propagated to the rest of the unlabeled samples in the dataset. Therefore, for many machine learning tasks such as clustering and classification, it is very important to construct a good graph which can represent the structure of the dataset.

Although from the data, we can discover millions of such a kind of relationship, recent studies on low-rank and sparse representation show that the selection of pairwise relationship are very important. Yan et al. [1,2] put forward a method to construct the l_1-graph based on sparse representation (SR) [3] 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) [4] 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. 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. In addition, as the representation coefficients may be negative, which lack physical meaning for many image processing problems. 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 [5], this method can be used to well describe the hyperspectral images.

Although the methods based on low-rank representation and sparse representation have gained a great success, these methods still have several obvious disadvantages. In most graph-based semi-supervised learning methods, the structure of the graph is often pre-defined. Therefore, the graph construction and the semi-supervised learning are often two independent steps, in this way, these algorithms cannot to obtain an overall 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.

Motivated by the above analysis, we proposed a novel graph regularized low-rank representation method for semi-supervised learning (GLRSC). 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, GLRSC 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. GLRSC simultaneously captures the intrinsic local and global structure of the high-dimensional data.

978-1-5386-7445-1/18/$31.00 ©2018 IEEE
DOI 10.1109/DCABES.2018.00033

II. RELATED WORKS

In this section, we briefly introduce the LRR and GLRR [6], and the semi-supervised classification framework used in the paper.

A. 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:

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

where $\|Z\|_*$ represents the nuclear norm, which is defined as the sum of all the singular values of Z, 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 [6] method, the objective is defined as follows:

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

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.

B. Semi-supervised classification

In this section, we present a kind of very popular semi-supervised learning method, Gaussian Fields and Harmonic Functions (GFHF) [7]. 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 label 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

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

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.

III. 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.

A. 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

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

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.

B. 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). 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:

$$\begin{aligned} &min_{F,Z,E} \ \sum_{i=1}^{n} \sum_{j=1}^{n} \|F_i - F_j\|^2 W_{ij} + \\ &tr((F - Y)^T U(F - Y)) + \\ &\|Z\|_* + \alpha \|J\|_1 + \beta tr(ZL_1 Z^T) + \gamma \|E\|_{2,1} \\ &\text{s.t. } X = AZ + E, \ Z = W, \ Z = J, \ J \geqslant 0 \end{aligned} \tag{8}$$

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:

$$\begin{aligned} &min_{F,Z,E} \ \sum_{i=1}^{n} \sum_{j=1}^{n} \|F_i - F_j\|^2 W_{ij} + \\ &tr((F - Y)^T U(F - Y)) + \\ &\|Z\|_* + \alpha \|J\|_1 + \beta tr(ZL_1 Z^T) + \gamma \|E\|_{2,1} + \\ &\langle Y_1, X - AZ - E\rangle + \langle Y_2, Z - W\rangle + \langle Y_3, Z - J\rangle + \\ &\frac{\mu}{2} \left(\|X - AZ - E\|_F^2 + \|Z - W\|_F^2 + \|Z - J\|_F^2 \right) \end{aligned} \tag{9}$$

$$\psi(Z,W,J,E,Y_1,Y_2,Y_3) = \beta tr(ZL_1Z^T) + \frac{\mu}{2}\left\|X - AZ - E + \frac{1}{\mu}Y_1\right\|_F^2 + \frac{\mu}{2}\left\|Z - W + \frac{1}{\mu}Y_2\right\|_F^2 +$$

where $\frac{\mu}{2}\left\|Z - J + \frac{1}{\mu}Y_3\right\|_F^2$

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:

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

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

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

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

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

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

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

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

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

C. Computation of E

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

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

The solution is defined by

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

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:

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

where $R_{ij} = \frac{1}{2}\|F_i - F_j\|^2$, $\odot$ 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

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

$$\text{s.t. } W \geq 0$$

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:

$$F_{k+1} = arg\ min_F \sum_{i=1}^n\sum_{i=1}^n \|F_{i,k} - F_{j,k}\|^2 W_{ij,k} + tr((F_k - Y)^TU(F_k - Y)) = arg\ min_F tr((F_k)^TL_2F_k) + tr((F_k - Y)^TU(F_k - Y)) \quad (18)$$

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

$$\partial(min_F tr((F_k)^TF_k) + tr((F_k - Y)^TU(F_k - Y)))/\partial F_k = 0 \quad (19)$$

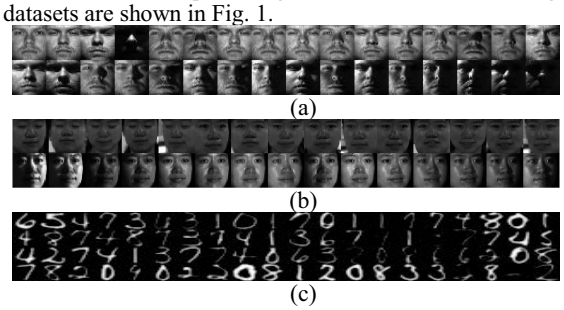
Then, we have

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

IV. EXPERIMENTS

A. Experiment setup

Datasets: we test our proposed method on three public datasets for evaluation: Extended Yale B, CMU PIE and USPS. Some sample images from these four image datasets are shown in Fig. 1.



(a) (b) (c)

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

Comparison Methods: We compare our proposed graph construction methods with the following state-of-the-art baseline methods: KNN-graph, LLE-graph[8], l_1-graph[2], SPG[9], LRR-graph[4], GLRR-graph[6], NNLRS-graph[10].

B. 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 I. CLASSIFICATION ACCURACY (%) ON EXTENDED YALE B

Labeled samples	10%	20%	30%	40%	50%	60%
KNN(5)	67.11	68.91	71.44	73.65	75.22	77.02
KNN(8)	63.16	64.41	66.46	69.03	70.27	71.42
LLE(8)	71	74.16	77.76	80.18	82.39	84.25
LLE(10)	70.24	73.35	77.17	80.1	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	96.77
GLRSC	94.98	95.36	96.16	96.84	96.61	97.49

TABLE II. CLASSIFICATION ACCURACY (%) ON CMU PIE

Labeled samples	10%	20%	30%	40%	50%	60%
KNN(5)	65.72	66.94	69.89	71.54	73.04	74.91
KNN(8)	63.58	63.89	66.49	67.85	69.55	70.91
LLE(8)	67.75	69.58	73.48	76.38	78.35	80.44
LLE(10)	67.47	69.17	72.99	75.99	77.78	79.98
-graph	78.29	82.82	87.94	90.99	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	81.2
NNLRS	87.78	89.37	90.18	92.92	96	95
GLRSC	88.57	90.49	91.03	94.16	96.92	96.31

TABLE III. CLASSIFICATION ACCURACY (%) ON USPS

Labeled samples	10%	20%	30%	40%	50%	60%
KNN(5)	96.87	97.78	98.45	98.8	99.18	99.35
KNN(8)	96.79	97.9	98.47	98.82	99.14	99.28
LLE(8)	72.31	77.57	80.82	83.38	85.72	87.39
LLE(10)	64.34	71.04	74.7	77.47	79.99	82.31
-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	97.2	98.38	98.87	99.12	99.41	99.52
GLRSC	97.75	98.83	99.08	99.39	99.6	99.64

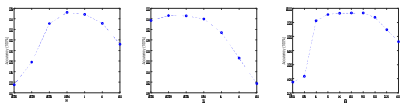
From the experimental results, we can observe that:

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.



(a) (b) (c)

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

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.

V. 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.

REFERENCES

[1] Yan, Shuicheng, and Huan Wang. "Semi-supervised Learning by Sparse Representation." SDM. 2009.

[2] Cheng, Bin, et al. "Learning with-graph for image analysis." Image Processing, IEEE Transactions on 19.4 (2010): 858-866.

[3] Wright, John, et al. "Robust face recognition via sparse representation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.2 (2009): 210-227.

[4] Liu, Guangcan, Zhouchen Lin, and Yong Yu. "Robust Subspace Segmentation by Low-Rank Representation." ICML. 2010.

[5] Yin, Ming, Junbin Gao, and Zhouchen Lin. "Laplacian regularized low-rank representation and its applications." IEEE transactions on pattern analysis and machine intelligence 38.3 (2016): 504-517.

[6] Lu, Xiaoqiang, Yulong Wang, and Yuan Yuan. "Graph-regularized low-rank representation for destriping of hyperspectral images." Geoscience and Remote Sensing, IEEE Transactions on 51.7 (2013): 4009-4018.

[7] Zhu, Xiaojin, Zoubin Ghahramani, and John Lafferty. "Semi-supervised learning using gaussian fields and harmonic functions." ICML. Vol. 3. 2003.

[8] Wang, Jingdong, et al. "Linear neighborhood propagation and its applications." Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.9 (2009): 1600-1615.

[9] He, Ran, et al. "Nonnegative sparse coding for discriminative semi-supervised learning." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[10] Zhuang, Liansheng, et al. "Non-negative low rank and sparse graph for semi-supervised learning." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.