

Discriminative Graph Learning via Low-Rank Constraint for Semi-Supervised Learning

Ying Wang¹, Ao Li¹, Gunaglu Sun¹, Xin Liu¹

{vicwangying@163.com, dargonboy@126.com, sunguangl@hrbust.edu.cn}

School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China¹

Abstract. Graph learning is more and more widely used in discovering associations and mining relationships between data and data. Our paper proposed a discriminative graph learning for semi-supervised learning method discriminative graph learning with low rank constraints. Traditional semi-supervised learning methods exist the problem of similarity between similar species is not high. How to improve the similarity between similar species has become a problem we need to solve in this field. In order to solve this problem, this paper proposed a new semi-supervised learning means, which mainly includes two parts: 1) Representing the local structure of the research sample by low-rank representation, compressing similar content, and easily finding the focus of image classification; 2) In order to improve the similarity between classes, a model that uses similarity constraints between the data to be classified. The proposed means is used for comparative experiments of classification on multiple data sets. Through experiments show that the means proposed in this paper performs better than traditional semi-supervised classification methods in terms of accuracy of classification.

Keywords: Image classification, graph learning, low-rank constraints, semi-supervised.

1 Introduction

With the increasing application of face recognition and image classification technology in our daily lives, how to extract features more conducive to classification from high-dimensional and complex data has become the focus of today's research [1]. Recently, because supervised learning and unsupervised learning have many problems in real life and cannot practical applications, researchers have combined supervised learning with semi-supervised learning and proposed semi-supervised learning [2]. Semi-supervised learning is to learn the features of known label images to extract the features that have a higher proportion of image classification, and classify the image data of unknown labels according to the learning results[3][4][5].

Inspired by the methods proposed by the above researchers, this paper proposes a discriminative graph learning with low rank constraints for semi-supervised learning. This method puts low rank constraints and semi-supervised learning into the same framework and expresses low rank Coefficients and discriminant graph learning errors refer to semi-supervised learning as two different constraints, which improves the similarity between classes, makes semi-supervised learning more accurate, and reduces the time to mark unknown label data.

The main contributions of this paper are:

1) Put low-rank constraints and semi-supervised learning into the same method, and use low-rank representations to transform high-dimensional data into low-rank representations, so that semi-

supervised learning becomes less computational and takes less time.

2) In order to improve the accuracy of semi-supervised learning for the classification of unknown label data, a discriminative graph learning error is constructed, which improves the accuracy of semi-supervised learning.

2 Method description

2.1 Low rank constraints for semi-supervised learning

2.1.1 Low rank constraints

Since the low rank model was proposed, the low rank model has been applied to various fields such as subspace learning[6], subspace clustering[7], and image processing[8][9]. The low-rank model can represent the sample data with a low-dimensional structure. This representation can play a role in eliminating the noise of the sample data to a certain extent. Compared with the method of directly processing the dry data by using the low-rank model to eliminate the noise of the sample data, the performance of this method is more prominent and increases the robustness of semi-supervised learning. The function of this model is:

$$\|Z\|_* + \|E\|_1 \quad (1)$$

The above formula $\|Z\|_*$ represents the kernel norm of the matrix and $\|E\|_1$ represents the l_1 norm of the matrix, which is often used for data containing outliers. The model $\|Z\|_* + \|E\|_1$ is a low-rank representation model. By applying a low rank constraint to the coefficient representation matrix Z , it is easier to find the local similarity of the training sample itself [10], and it is beneficial to eliminate noise interference in the sample data.

2.1.2 Discriminative graph learning error

In order to further enhance the accuracy of semi-supervised learning so that the method can maintain high performance even under noisy conditions, we propose a discriminant graph learning error constraint term. The constraint term is obtained by training the sample data and finding the matrix

$D = [d_1 \cdots d_n] \in R^{K \times N}$ according to the training result. The matrix D is composed of 0 and 1.

When only one value in a column is 1 and the other values are 0, the image belongs to this kind. For

example, suppose $Z = [z_1 \cdots z_4]$ exists, where z_1 belong to class 1, z_2, z_3 belong to class two,

z_4 belong to class three, then D should be defined as

$$D \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

According to D defined above, we construct the discriminant graph learning error as

$$\|WZ - D\|_F^2 \quad (2)$$

The purpose of constructing this constraint term is to reduce the error when classifying unknown label data.

The idea of constructing this term is that we want to reduce the error in data classification of unknown labels as much as possible, so ideally it is $\|Z - D\|_F^2$, but because of the error, we construct the coefficient matrix W , and the term construction is complete.

Among them, $\|E\|_F^2$ is the Frobenius specification, which is commonly used to process data containing Gaussian noise. The matrix W is the coefficient matrix.

According to the definition of the Frobenius-norm, equation (2) is equivalent to the following equation:

$$Tr\left\{(WZ - D)^T \wedge (WZ - D)\right\} \quad (3)$$

According to the two constraints described above, the function model can be defined as

$$\begin{aligned} \min_{Z,E,W} \|Z\|_* + \lambda_1 \|E\|_1 + Tr\left\{(WZ - D)^T \wedge (WZ - D)\right\} \\ s.t. X = XZ + E \end{aligned} \quad (4)$$

2.2 Optimization

To solve the convergence of the objective function (4), we use the inexact ALM algorithm. In order to make our solution more convenient, the auxiliary variable J is introduced, so the objective function can be written as:

$$\begin{aligned} \min_{Z,E,J,W,D} \|J\|_* + \|E\|_1 + Tr\left\{(WZ - D)^T \wedge (WZ - D)\right\} \\ s.t. X = XZ + E, Z = J \end{aligned} \quad (5)$$

According to the Lagrangian multiplier method, the Lagrangian function of equation (5) is:

$$\begin{aligned} \min_{Z,E,J,W} \|J\|_* + \|E\|_1 + Tr\left\{(WZ - D)^T \wedge (WZ - D)\right\} + Tr(R^T (X - XZ - E)) \\ + \frac{\mu}{2} \|X - XZ - E\|_F^2 + Tr(S^T (Z - J)) + \frac{\mu}{2} \|Z - J\|_F^2 \\ s.t. X = XZ + E, Z = J \end{aligned} \quad (6)$$

In order to solve the above formula, the ALM algorithm that has been proposed is used to solve the unknown variable in iterations and find the minimum value. The approach we take is to solve only one unknown variable during each solution and treat the other unknown variables as constants.

We first solve the low rank representation J . When solving J , we consider other unknown variables as fixed values. After K iterations, equation (6) can be written as:

$$\min_J \|J\|_* + Tr(S^T (Z - J)) + \frac{\mu}{2} \|Z - J\|_F^2 \quad (7)$$

The above formula can be rewritten as:

$$\min_J \frac{1}{\mu} \|J\|_* + \frac{1}{2} \|J - (Z + R/\mu)\|_F^2 \quad (8)$$

We use the (SVT) method in the literature [11] to solve problem (8).

To initialize W , we use a multiple ridge regression model [12]. The model consists of the quadratic loss and L_2 norm regularization. So W can be updated as:

$$W = \arg \min_W \|WX - Q\|^2 + \lambda_2 \|W\|_2^2 \quad (9)$$

Ignore the variable (5) that has nothing to do with Z , the original formula can be rewritten as:

$$\begin{aligned} \min_{Z,E,J,W} Tr \left\{ (WZ - D)^T \wedge (WZ - D) \right\} + Tr(R^T (X - XZ - E)) \\ + \frac{\mu}{2} \|X - XZ - E\|_F^2 + Tr(S^T (Z - J)) + \frac{\mu}{2} \|Z - J\|_F^2 \end{aligned} \quad (10)$$

Find the partial derivative of the above formula about Z , and set the partial derivative equal to 0 to solve Z .

Similarly, in order to solve E , we omit variables not related to E , then we can rewrite equation (5) as:

$$E = \min_E \frac{\lambda_1}{\mu} \|E\|_1 + \frac{1}{2} \|E - (X - AZ + Y/\mu)\|_F^2 \quad (11)$$

The solution of the above formula is proposed in [12]. When we let $\Psi = X - AZ + Y/\mu$, then then i -th first column in E^{k+1} can be expressed as:

$$E^{k+1}(:, i) = \begin{cases} \frac{\|\Psi_i\| - \lambda_1}{\|\Psi_i\|} \Psi_i, & \text{if } \lambda_1 < \|\Psi_i\| \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

According to the previous analysis, we summarize the previous solution method as the following algorithm:

Algorithm 1:

Input: data set X , parameter λ_1 , $Z = J = 0$, $E = 0$, $W = 0$, $\mu = 0.1$, $\mu_{\max} = 10^{10}$

$$S = 0, R = 0$$

Output: Z_k, W_k, E_k

1. While not convergence do
2. Updated J_{k+1} using (8).
3. Updated W_{k+1} using (9).
4. Updated Z_{k+1} using (10).
5. Updated E_{k+1} using (12).
6. Updated R_{k+1} and S_{k+1}

$$R_{k+1} = R_k + \mu_k (X - XZ_{k+1} - E_{k+1})$$

$$S_{k+1} = S_k + \mu_k (Z_{k+1} + J_{k+1})$$

7. $k = k + 1$

8. End while

3. Experiments

In the experiment, we compare the method proposed in this paper with several existing semi-supervised learning methods such as FME, RLSR, RRPC, and SDA. So as to have a more accurate understanding of the performance of ours.

First, we compared our method with several means in Extended YaleB[13]. When we set the training sample to 10, the comparison data is as follows:

Table3.1 Comparison of accuracy in Extended YaleB

Method	Accuracy (%)
FME	77.56
RLSR	83.47
RRPC	82.81
SDA	74.67
Ours	88.49

We compared our method with several other methods in AR Face[14].

Table3.2 Comparison of accuracy in Extended AR Face

Method	Accuracy (%)
FME	85.69
RLSR	84.43
RRPC	85.41
SDA	74.67
Ours	91.74

We compared our method with several other methods in COIL20.

Table3.3 Comparison of accuracy in Extended COIL20

Method	Accuracy (%)
FME	69.74
RLSR	75.26
RRPC	65.84
SDA	72.15
Ours	76.12

From the above experimental results, we can see that our proposed method has higher accuracy than other semi-supervised learning methods.

4. Conclusion

This paper proposes a discriminant graph learning with low rank constraint for semi-supervised learning to classify data of unknown labels. This method designs two constraint terms based on low rank representation, discriminant graph learning error and iterative solution method. ALM effectively solves the convergence problem of the objective function, and compares it with the existing semi-supervised learning methods on three different data sets. Through experiments we learned that the method proposed in this paper is better than some current semi-supervised learning methods.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grant 61501147, in part by the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province under Grant UNPYSCT-2018203, in part by the Natural

Science Foundation of Heilongjiang Province under Grant YQ2019F011, in part by the Fundamental Research Foundation for University of Heilongjiang Province under Grant LGYC2018JQ013, and in part by the Postdoctoral Foundation of Heilongjiang Province under Grant LBH-Q19112.

References

- [1]. Huan Liu and Lei Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):491–502, 2005.
- [2]. Samuel H Huang. Supervised feature selection: A tutorial. *Artificial Intelligence Research*, 4(2):22, 2015.
- [3]. Zheng Zhao and Huan Liu. Semi-supervised feature selection via spectral analysis. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 641–646, 2007.
- [4]. Jidong Zhao, Ke Lu, and Xiaofei He. Locality sensitive semi-supervised feature selection. *Neurocomputing*, 71:1842 – 1849, 2008.
- [5]. J. Xu, B. Tang, H. He, and H. Man. Semi supervised feature selection based on relevance and redundancy criteria. *IEEE Transactions on Neural Networks and Learning Systems*, PP (99):1–11, 2016.
- [6]. Candès E J, LI Xiaodong, MA Yi, et al. Robust principal component analysis? [J]. *Journal of the ACM (JACM)*, 2011, 58(3): 11-49. doi: 10.1145/1970392.1970395.
- [7]. YANG Junfeng and ZHANG Yin. Alternating direction algorithms for ℓ_1 -problems in compressive sensing [J]. *SIAM Journal on Scientific Computing*, 2011, 33(1): 250-278. doi: 10.1137/090777761.
- [8]. YANG Junfeng, YIN Wotao, ZHANG Yin, et al. A fast algorithm for edge-preserving variational multichannel image restoration [J]. *SIAM Journal on Imaging Sciences*, 2009, 2(2): 569-592. doi: 10.1137/080730421.
- [9]. WEN, Zaiwen and YIN Wotao. A feasible method for optimization with orthogonality constraints [J]. *Mathematical Programming*, 2013, 142(1-2): 397-434. doi: 10.1007/s10107-012-0584-1.
- [10]. LIN Zhouchen, CHEN Minming, and MA Yi. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices [J]. *arXiv preprint arXiv:1009.5055*, 2010. doi: 10.1016/j.jab.2012.10.010.
- [11]. J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [12]. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, “Robust recovery of subspace structures by low-rank representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [13]. A. Georghiades, P. Belhumeur, and D. Kriegman, “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, June 2001.
- [14]. A. Martinez and R. Benavente, “The AR Face Database,” *CVC Technical Report 24*, 1998.