

Semi-supervised Corrupted Face Classification via Graph Learning

Yisheng Zhong, Ao Li

{951208709@qq.com(Yisheng Zhong), dargonbuy@126.com(Ao Li)}

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

Abstract— Semi-supervised learning aims to training model with both of labeled and unlabeled data by exploring the relationships among them. Graph-based semi-supervised learning is an classical representative method that learning the class indicator matrix by propagating the similarity within the well designed graph constructed by data. However, for face data, they often happen to pixel missing or occlusion, which will degrade the graph learning performance, leading awful semi-supervised classification results. To address this problem, a novel semi-supervised corrupted face classification method via graph learning is proposed, in which the dynamic graph is learned by the completion face data recovered from the low-rank subspace. In our proposed method, the robust data representation and graph learning are implemented alternatively to obtain the overall optimal solutions. Experimental results demonstrate that our proposed method outperforms comparison methods on both of classification accuracy and robustness.

Keywords: Semi-supervised face classification, Graph learning, Self- representation model, Low-rank constraint.

I. Introduction

In the face of massive high-dimensional data, how to conduct effective data analysis and processing has become a major problem in machine learning and other fields [1]. In recent years, studying map automatically is hot, which is one of the important methods for adaptive neighbor method. We construct a matrix by setting each data point as the probability that the current data can be used as the neighborhood of another data, and this probability is used as the similarity between the two data points [2], which does not require similarity measures sensitive to noise and outliers [3], so the result obtained is of high precision.

However, in the process of graph learning, there is noise or interference in the original data, so the graph obtained may be inaccurate or suboptimal, and cannot accurately describe the true relationship between the data. In order to solve this problem, Zhao Kang[4] proposed a new robust graph learning scheme based on the adaptive neighbor method, which decomposes the original data into a low-rank matrix D ("clean data") and a sparse matrix E ("Noise / Error"). They can then use the adaptive neighbor method to build graphs on clean data D . So they can remove image disturbances and learn to map at the same time. However, when there is occlusion or partial absence of data, the results of this method are not reliable or even the learning results are not available. The Laplace Score (LS) method proposed in [5] introduced the analysis of the local structure of the data based on MaxVar. But these two methods only consider the characteristics of the data itself, ignore the correlation between the characteristics, and cannot guarantee the optimal feature subset. Inspired by the self-similarity of images, Zhu [6] believe that images should not only have self-similarity in structure, but also have the ability to express themselves in terms of feature expression. They proposed an unsupervised feature selection method based on regularized self-representation by constructing a regularized self-representation (RSR) model. This method constructs a self-representation model by assuming that each feature in the high-dimensional data can be expressed as a linear combination of other features, and removes insignificant features by adding $\ell_{1,2}$ norm constraints to the feature's weight matrix W . We borrowed the method of regularized self-representation to solve the problem of large-scale image interference or missing.

In the rest of this article, we will introduce graph learning and our multi-source robust graph learning technique in the second section. Details of the algorithm are then given in Section III. The fourth part evaluates the clustering task experimentally. The fifth part discusses semi-supervised applications and compares the data recovery effect of the sixth part.

II. Conventional Graph Learning

Given a sample set $X = \{x_1; x_2; \dots; x_n\} \in R^{n \times m}$, where $x_i \in R^{1 \times m}$ ($i = 1, 2, \dots, n$) represents the i th sample.

As locality preserving projection (LPP) [7] does, we can assign a probability S_{ij} for x_j as the neighborhood of x_i . Thus, S_{ij} characterizes the similarity between x_i and x_j in some sense. Smaller distance $\|x_i - x_j\|_2$ indicates that x_i and x_j are quite similar, thus the bigger value of S_{ij} . To achieve graph S, we can solve the following problem:

$$\min_{s_i} \sum_{j=1}^n \left(\frac{1}{2} \|x_i - x_j\|^2 s_{ij} + \gamma s_{ij}^2 \right) \quad (1)$$

$$s.t. s_i^T \mathbf{1} = 1, 0 \leq s_{ij} \leq 1$$

where γ is a tradeoff parameter. By defining the graph Laplacian matrix $L = D - (S + ST/2)$, where D is a diagonal

matrix with $d_{ii} = \sum_j [(s_{ij} + s_{ji})/2]$, (1) becomes

$$\min_S Tr(XLX^T) + \gamma \|S\|_F^2 \quad (2)$$

$$s.t. S\mathbf{1} = \mathbf{1}, 0 \leq S \leq 1$$

By optimizing the above problem, one can learn S adaptively from the data. However, this method degrades performance when it comes to input data with noise. To solve this problem, Zhao Kang [4] present a principle to robustify graph learning. They assume the data to be decomposed into two parts: 1) the clean data D and 2) the corruptions E. And they proposed robust graph construction (RGC) model can be formulated as

$$\min_{D, E, S} \|D\|_* + \alpha \|E\|_1 + \beta Tr(DLD^T) + \gamma \|S\|_F^2 \quad (3)$$

$$s.t. X = D + E, S\mathbf{1} = \mathbf{1}, 0 \leq S \leq 1$$

This is called manifold RPCA (MRPCA) model in [8].

III. Proposed Model

This paper proposes a new robust learning graph scheme. We use the self-representation property of data [9], that is, every data point in the space can be effectively reconstructed through linear combination of other points in the data set to construct the similarity matrix. Specifically expressed as

$$X = XZ \quad (4)$$

Where $Z = \{z_1, z_2, \dots, z_3\} \in R^{n \times n}$ is a similarity matrix and each z_i is a linear representation of the original data point x_i .

We propose a new robust learning graph scheme, and we introduced theta (4)., and then (3) becomes

$$\min_{Z, X, E, S} \|Z\|_* + \alpha \|E\|_1 + \beta Tr(XZZ^T X^T) + \gamma \|S\|_F^2 \quad (5)$$

$$s.t. X = XZ + E, S\mathbf{1} = \mathbf{1}, 0 \leq S \leq 1$$

We then used the adaptive neighbor method to build the graph on clean XZ. We used a joint learning method to optimize both S and XZ, which enabled us to construct the clustering graph while recovering low-rank clean data. The finally learned graph can be transformed into a block diagonal matrix through matrix transformation:

$$\begin{bmatrix} S_1^{n \times n} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & S_k^{m \times m} \end{bmatrix}$$

To summary, compared with the existing work in the literature, the main contributions of this paper are as follows.

- 1) Through the establishment of self-presentation model, the vacancy of graph learning method in solving data interference or lack is filled, making the adaptive map learning method more robust.
- 2) Through joint solution, denoising and graph learning can be carried out simultaneously.
- 3) A large number of experiments have been performed on face clustering, document clustering, face / target recognition, and face image shadow removal, etc., which verifies the effectiveness of the method.

IV. Numerical Scheme

To solve (5), we first introduce auxiliary variable R to facilitate the solution of Z . Then, (5) can be written as:

$$\begin{aligned} \min_{Z, X, E, S} & \|R\|_* + \alpha \|E\|_1 + \|R - Z\|_F^2 + \beta \text{Tr}(XZZ^T X^T) + \gamma \|S\|_F^2 \\ \text{s.t.} & X = XZ + E, S1 = 1, 0 \leq S \leq 1, R = Z \end{aligned} \quad (6)$$

It can be solved via alternating direction method of multipliers. Removing the equality constraints on X and R , we obtain the augmented Lagrange function as follows:

$$L(XZ, E, S, Z, Y_1, Y_2) = \|Z\|_* + \alpha \|E\|_1 + \beta \text{Tr}(RLR^T) + \gamma \|S\|_F^2 + \frac{\mu}{2} (\|XZ + E - X + \frac{Y_1}{\mu}\|_F^2 + \|XZ - R + \frac{Y_2}{\mu}\|_F^2) \quad (7)$$

According to the above derivation, we can get the following algorithm flow:

Algorithm

Input: Data matrix X , parameters $\alpha > 0, \beta > 0, \gamma, \mu > 0$.

Initialize: $R = E = 0, Y_1 = Y_2 = 0$.

While not converge do

1: Calculate R by $\min_R \|R\|_* + \mu \|Z - R\|_F^2$.

2. Calculate Z by $\left\| XZ + E - X + \frac{Y_1}{\mu} \right\|_F^2 + \left\| XZ - R + \frac{Y_2}{\mu} \right\|_F^2 + \|R - Z\|_F^2$;

3: Update E according to $\min_E \alpha \|E\|_1 + \frac{\mu}{2} \left(\left\| E - \left(X - XZ - \frac{Y_1}{\mu} \right) \right\|_F^2 \right)$.

4: Update S using the scenario in [4];

5: Update Lagrange multipliers Y_1 and Y_2 as

$$Y_1 = Y_1 + \mu (XZ + E - X)$$

$$Y_2 = Y_2 + \mu (Z - R)$$

End while.

V. Experimental results

In this part, we will evaluate the semi-supervised classification algorithm proposed in this paper on two datasets. We first obtain the graph matrix S from our proposed method and then use the graph-based propagation method to perform the semi-supervised classification task [10].

A. Datasets description

We evaluated the effectiveness of our method in face recognition using our proposed graph learning in the frequently used YALE and JEFFE datasets. Specifically, the YALE face database has 15 people, each of whom has 11 near-frontal images taken under different lighting conditions; The JEFFE face dataset contains 10 individuals, each with seven different facial expressions, including six basic facial expressions and one neutral expression. Figure 3(a) shows some sample photos of the two cubes. In order to demonstrate the robustness of our method, we randomly added large areas of missing or interference to these two data sets, as shown in figure 3(b).



Fig.1. Sample images of (a)YALE, (b)YALE with defects.

B. Comparison Algorithms

We compare the proposed method to two existing excellent methods.

- 1)LGC [10]: LGC is a widely used semisupervised classification method.
- 2) Semisupervised Classification With Adaptive Neighbors (SCAN) [11]: This recently developed method uses adaptive neighbors approach to construct the similarity graph. Moreover, the graph construction and clustering are formulated into a unified framework to improve the performance.

C. Experimental Results

After using different percentages of samples as labels many times, we found that the accuracy of all methods improved as the number of label samples increased. The experimental results are shown in Table 1. It can be seen that our proposed method is superior to other existing technologies when the data contains large area missing or interference. This proves the robustness of the data self-representation model, because when there are missing data, similar data can help reconstruct the data while eliminating noise, so as to construct the map more accurately, which illustrates the importance of removing noise and data reconstruction.

TABLE I
CLASSIFICATION ACCURACY(%) ON FACE DATASETS (MEAN±STANDARD DEVIATION).

| Data | Labeled Percentage(%) | LGC[10] | SCAN[11] | Proposed method |
|---------------------|-----------------------|-------------|-------------|-----------------|
| YALE | 10 | 47.33±13.96 | 45.07±1.30 | 47.38±2.57 |
| | 30 | 63.08±2.20 | 60.92±4.03 | 65.82±1.93 |
| | 50 | 69.56±5.42 | 68.94±4.57 | 70.74±3.48 |
| YALE with defects. | 10 | 39.54±23.33 | 37.58±12.40 | 46.23±17.45 |
| | 30 | 49.12±19.12 | 46.17±2.67 | 56.46±4.83 |
| | 50 | 63.31±9.63 | 58.85±6.32 | 62.26±9.62 |
| JAFFE | 10 | 96.65±7.76 | 96.92±1.68 | 95.84±4.74 |
| | 30 | 98.86±1.14 | 98.20±1.22 | 98.74±0.97 |
| | 50 | 99.29±0.94 | 99.25±5.79 | 99.36±3.83 |
| JAFFE with defects. | 10 | 93.1±3.79 | 95.93±2.43 | 95.14±5.93 |
| | 30 | 97.46±2.14 | 97.83±4.10 | 98.56±2.50 |
| | 50 | 98.36±2.11 | 99.03±1.35 | 99.33±1.96 |

VI. Conclusion

This paper presents a model for learning robust graphs from data to serve the semi-supervised classification. It solves the problem of robustness of adaptive neighbor graph learning. We improve the conventional model by introducing the self-representation model and the idea of joint solution. Therefore, our proposed framework can not only enhance performance of semi-supervised classification, but also improve low-rank recovery from damaged data. Experiments on several public face benchmark datasets show that the robustness of our method is superior to various advanced existing graphing learning techniques.

Acknowledgments. 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] Rodriguez G V F, Luque E J A, Chica O M, et al. Feature selection approaches for predictive modelling of groundwater nitrate pollution: an evaluation of filters, embedded and wrapper methods [J]. Science of The Total Environment, 2018, 624 (15): 661-672.
- [2] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in Proc. 20th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining, 2014, pp. 977-986.
- [3] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in Proc. IJCAI, 2015, pp. 3569-3575.
- [4] Kang Z , Pan H , Hoi S C H , et al. Robust Graph Learning from Noisy Data[J]. IEEE Transactions on Cybernetics, PP(99):1-11.

- [5] He Xiaofei, Cai Deng, Niyogi Partha. Laplacian score for feature selection [C]// Proceedings of Advances in Neural Information Processing Systems. Cambridge: MIT, 2005: 507-514.
- [7] X. He and P. Niyogi, "Locality preserving projections," in Proc. Adv. Neural Inf. Process. Syst., 2004, pp. 153–160.
- [8] N. Shahid, V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Robust principal component analysis on graphs," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 2812–2820.
- [9] Elhamifar E, Vidal R. Sparse subspace clustering: algorithm, theory, and applications [J]. IEEE Transactions on Software Engineering, 2013, 35 (11): 2765-2781.
- [10] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 321–328.
- [11] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi- supervised classification with adaptive neighbours," in *Proc. AAAI*, 2017, pp. 2408–2414.