Abstract

Sparse Representation (SR) shows powerful discriminating power when the training samples are sufficient to construct an over-complete dictionary. However, in the lack of training samples case, the dictionary is too small to sparsely represent the test sample which restricts the classification performance of sparse representation. In order to address this problem, we propose a new sparse representation-based classifier named Sparse Graph-based Classifier (SGC) via introducing graph-based transductive learning to the sparse representation. In this algorithms, we apply sparse representation to measure the correlation between each two samples and then construct a graph Laplacian which encodes the sparse representation relationships of all samples. Finally, the obtained graph Laplacian is plugged into the conventional graph-based transduction to infer the labels of test samples. Four popular image databases are adapted to evaluate our works. The results demonstrate that SGC achieve a promising performance in comparison with the-state-of-art classifiers particular in the small training sample size case.

Keywords: Image Classification, Sparse Representation, Graph Learning, Transductive Learning, Semi-supervised Learning

1. Introduction

Recently, Sparse Representation (SR) has attracted a lot of attention in the machine learning, computer vision and image processing communities [1, 2, 3, 4, 5]. The idea of SR stems from the compression sensing that most signals have a sparse representation as a linear combination of a reduced subset of signals from the same space [3]. Naturally, the signals tend to have a representation biased towards their own class which endows SR with the strong discriminating power and feature selection ability. However, an important condition of SR is that the dictionary should be overcomplete. In the lack of training samples case which is very common in the real world applications, the dictionary constructed by training samples is too small to sparsely represent the query sample. Consequently, the small size of dictionary restricts the classification performance of SR.

Motivated by some recent successes in SR which utilize SR to construct a sparse graph for tackling the clustering and dimensionality tasks [3, 4, 5], we address this problem by constructing a sparse graph for introducing the sparse representation to the graph-based transduction [7, 8]. We name this novel SR-based classifier Sparse Graph-based Classifier (SGC). In SGC, SR is leveraged to measure the correlation between each two samples and then a graph Laplacian is constructed to plug into the conventional graph-based transduction framework for classification. Note, the graph Laplacian is constructed from both training samples and testing samples. Therefore, in the classification procedure, SGC can only use the sparse representation relationships between the query sample and training samples, which is as same as the traditional sparse representation-based classifier does, but also can use the sparse representation relationships between the query sample and the other test samples to infer the class label. We apply our work to image classification. Yale [9], AR [10] and Caltech256 [11] databases are employed for evaluation. The experimental results show that our method can get a promising result in comparison with SRC and some state-of-the-art transductive learning-based classifiers particular in the small training samples size case.

2. Methodology

2.1. Sparse Graph Laplacian

In the sparse representation, since the signals tend to have a representation biased towards their own class, it highlights the important correlations among the samples. In this section, we introduce a new graph-based transductive learning algorithm to utilize this sparse representation property for image classification. We name this algorithm Sparse Graph-based Classifier (SGC). The core of graph-based transductive learning is the construction of graph Laplacian which should encode the sparse representation relationships among samples in our case. We utilize the sparse representation to compute the weight between each two samples which is regarded as the element of the affinity matrix \( W \). Given a sample \( x_q \) as a query and considered the rest of samples \( X_{i\neq q} = \{x_1, \cdots, x_{q-1}, x_{q+1}, \cdots, x_n\} \) as the dictionary, we can present the following collaborative representation model to measure the correlations between the query sample and the other samples.

\[
\hat{C}_q = \arg\min_{C_q} ||C_q||_1, \text{ s.t. } \|x_q - X_{i\neq q}C_q^T\|^2_2 \leq \epsilon
\]  

(1)

*Corresponding author (Dan Yang): dyang@cqu.edu.cn

August 28, 2014
where the vector \( C_q = [c_{q1}, \cdots, c_{q(i-1)}, c_{q(i+1)}, \cdots, c_{qn}]^T \) is the regression weights of sample \( x_i \) and \( c_{qj} \) is the element of \( C_q \) corresponding to the sample \( x_j \). \( \epsilon \) is the measurement noise. This is a typical convex optimization problem. We adapt the SLEP method [12] to solve this problem. The regression coefficient \( c_{ij} \) indicates the correlation of the sample \( x_i \) and the sample \( x_j \). So, we compute the similarity (or the weight of edge) between sample \( x_i \) and \( x_j \) as follows

\[
w_{ij} = w_{ji} = \frac{|c_{ij}| + |c_{ji}|}{2} \tag{2}
\]

and we also provide a way to compute the self-similarity of the sample as follows

\[
w_{ii} = \sum_{j \neq i} w_{ij} \tag{3}
\]

Then we can obtain the affinity matrix of the sparse graph \( W \). According to the Laplacian Eigenmapping [13], the normalized graph Laplacian can be computed as follows

\[
L = D^{-1/2}(D - W)D^{-1/2} = I - D^{-1/2}WD^{-1/2} \tag{4}
\]

where \( D \) is a diagonal matrix and \( D_{ii} = \sum_j w_{ij} \). \( I \) is an identical matrix. The normalized graph Laplacian encodes the sparse representation property of data and has a strong discriminating power.

### 2.2. Sparse Graph-based Transduction

After we get the graph Laplacian of samples, the encoded sparse representation relationships of samples in the sparse graph are utilized to transductively infer the unlabel samples. At first, we take the simple binary classification as an example to show how to perform the sparse graph-based transductive inference. Let a graph cut \( f \) as a classification function. An optimal cut should not only minimize the sparse representation relationship change but also the classification loss over the sparse graph partition. So this transduction problem can be formulated as following minimization issue.

\[
\hat{f} = \arg \min_f \frac{1}{2} |f|^T L f + \lambda ||f - y||^2 \tag{5}
\]

where the vector \( y \) is regard as a function encodes the labels. \( y(i) = 1 \) or \(-1\) if the sample \( x_i \) has been labeled as positive or negative respectively, and 0 if it is unlabeled. \( \lambda \) is a positive to reconcile the sparse representation relationship change and the classification loss. Note, the graph Laplacian \( L \) is constructed from both training samples and testing samples, since the testing samples is transductively inferred from the correlations among all the samples, which are learned by the sparse representation. We adapt the one-versus-all strategy to extend the binary classification to multi-class classification. The multi-class version can be denoted as follows

\[
\hat{F} = \arg \min_\hat{F} \sum_i (\hat{f}_i^T L f_i + \lambda ||\hat{f}_i - y_i||^2) \tag{6}
\]

\[
= \arg \min_\hat{F} \hat{F}^T L \hat{F} + \lambda \|\hat{F} - Y\|^2
\]

where \( F = [f_1, \cdots, f_i, \cdots, f_n] \) and \( Y = [y_1, \cdots, y_i, \cdots, y_n] \) are the collection of classification functions and the collection of label functions respectively corresponding to the different classes. In the function \( y_i \), only the samples from class \( i \) are considered as positive while the samples from other classes are considered as negative. Since \( L \) is a positive semi-definite matrix, Equation 6 can be efficiently solved by Regularized Least Square. We obtain the partial derivative of Equation 6 with respect to \( F \), and let it equal to zero.

\[
\frac{\partial}{\partial F} \left[ F^T L F + \lambda ||Y - F||^2 \right] = 0
\]

\[
\Rightarrow 2(LF + \lambda F - \lambda Y) = 0
\]

\[
\Rightarrow F = \frac{\lambda Y}{L + \lambda I}
\]

Finally, the classification of \( i \)-th sample can be accomplished by assigning it to the \( j \)-th class that satisfies

\[
\hat{y}_i = \arg \max_j F_{ij} \tag{8}
\]

### 3. Experimental Results

Yale, AR and Caltech256 databases are used to evaluate our work. The Yale face database totally has 15 subjects and 11 samples per subject [9]. The size of image is 32×32 pixels. The AR database consists of more than 4,000 images of 126 subjects [10]. The database characterizes divergence from ideal conditions by incorporating various facial expressions, luminance alterations, and occlusion modes. Following paper [14], a subset contains 1680 images with 120 subjects are constructed in our experiment. All these images are 50×40 pixels. Similarly, we follow the paper [7] and select a subset from Caltech256 database [11]. In this subset, there are 20 classes and 100 images per class. Since Caltech256 is more challenging than the other two databases. We adapt the Picodes feature [15] to represent the images.

Sparse Representation-based Classifier (SRC) [1], Graph-based Classifier (GC), Normalized Hypergraph-based Classifier (NHC) [8] and Adaptive Hypergraph-based Classifier (AHC) [7] are employed as the compared methods. The last three algorithms are all transductive learning-based methods and their affinity matrices are computed based on Euclidean distance (Heat Kernel Weighting).

| Databases | Classification Error (Mean±STD%) |
|-----------|---------------------------------|
| SRC [1]   | ZHC [8]                         |
| AR        | 2.78±2.36                       |
| Caltech256| 23.49±2.36                      |
| Ours      | 23.49±2.36                      |

Table 1 shows the experimental results of different methods under two-fold cross validation while Figure 1 presents the performances of different classification methods under different sample percents for training. Clearly, we can know from the results that our proposed classifier outperforms all compared methods particularly compared with the traditional Euclidean
distance-based graph transduction algorithms. For example, the classification accuracy gains of SGC over NHP, AHP and GC are 5.48%, 5.60% and 15.95% respectively. This shows that the sparse representation can better measure the similarity among samples than the naive Euclidean distance-based methods. Moreover, from the observations of Figure 1, we find that the improvement of SGC over SRC is more significant when the training percent is low. We attribute this to the utilization of the sparse representation relationships between the query sample and the other testing samples in SGC. More specifically speaking, in the small training sample size case, the test samples are considered as a complement of the dictionary which can benefit the classification.

We empirically find that the performance of SGC is quite insensitive to the regularization parameter $\lambda$ when $\lambda > 100$. In our experiments, we let $\lambda = 1000$.

4. Conclusion

We proposed a new sparse representation-based classifier named Sparse Graph-based Classifier (SGC) via introducing the graph-based transduction to the sparse representation. In GSC, a graph Laplacian which encoded the correlations of samples is constructed by the sparse representation. And then this graph Laplacian is plugged into the graph-based transductive learning framework for image classification. The advantage of SGC over SRC is that it can not only use the sparse representation relationships between the query sample and the training samples but also utilize the ones between the query sample and the other testing samples to infer the class labels of testing samples. In such case, SGC can alleviate the problem that the dictionary is not overcomplete in the image classification task. There are a lot of interesting works can be done based on our method. For example, we can generalize the sparse graph into the sparse hypergraph to solve the multi-label classification tasks.

Acknowledgement

The work described in this paper was partially supported by National Natural Science Foundations of China (NO. 60975015 and 61173131), Fundamental Research Funds for the Central Universities (No. CDJXS11181162). The authors would like to thank useful comments of the anonymous reviewers and editors.

References

[1] J. Wright, A. Y. Yang, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.
[2] P. Ma, D. Yang, Y. Ge, X. Zhang, Y. Qu, S. Huang, J. Lu, Robust face recognition via gradient-based sparse representation, Journal of Electronic Imaging 22 (1) (2013) 013018-013018.
[3] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2790–2797.
[4] R. Timofte, L. Van Gool, Sparse representation based projections, in: British machine vision conference (BMVC), 2011, pp. 61–1.
[5] S. Gao, I. W.-H. Tsang, L.-T. Chia, Kernel sparse representation for image classification and face recognition, in: European Conference on Computer Vision (ECCV), 2010, pp. 1–14.
[6] D. L. Donoho, For most large underdetermined systems of linear equations the minimal $1$-norm solution is also the sparsest solution, Communications on pure and applied mathematics 59 (6) (2006) 797–829.
[7] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Transactions on Image Processing 21 (7) (2012) 3262–3272.
[8] D. Zhou, J. Huang, B. Schölkopf, Learning with hypergraphs: Clustering, classification, and embedding, in: Advances in neural information processing systems (NIPS), 2006, pp. 1601–1608.
[9] A. M. Martinez, A. C. Kak, Pca versus lda, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 228–233.
[10] A. Martinez, R. Benavente, The ar face database (Jun 1998).
[11] G. Griffin, A. Holub, P. Perona, Caltech-256 object category dataset.
[12] J. Liu, S. Ji, J. Ye, Slep: Sparse learning with efficient projections, Arizona State University 6.
[13] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural computation 15 (6) (2003) 1373–1396.
[14] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (11) (2010) 2106–2112.
[15] A. Bergamo, L. Torresani, A. W. Fitzgibbon, Picodes: Learning a compact code for novel-category recognition, in: Advances in Neural Information Processing Systems (NIPS), 2011, pp. 2088–2096.