Neurocomputing 173 (2016) 290–297

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Graph Regularized Sparsity Discriminant Analysis for face recognition

Songjiang Lou [a,*], Xiaoming Zhao [a], Yuelong Chuang [a], Haitao Yu [b], Shiqing Zhang [a]

[a] Institute of Image Processing & Pattern Recognition, Tai Zhou University, Taizhou, Zhejiang 318000, China
[b] College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

ARTICLE INFO

Article history:
Received 28 July 2014
Received in revised form
11 April 2015
Accepted 16 April 2015
Chennai Guest Editor
Available online 1 September 2015

Keywords:
Sparse representation
Graph embedding
Sparsity preserving projection
Feature extraction
Face recognition

ABSTRACT

Manifold learning and Sparse Representation Classifier are two popular techniques for face recognition. Because manifold learning can find low-dimensional representations for high-dimensional data, it is widely applied in computer vision and pattern recognition. Most of the manifold learning algorithms can be unified in the graph embedding framework, where the first step is to determine the adjacent graphs. Traditional methods use k nearest neighbor or the ε-ball schemes. However, they are parametric and sensitive to noises. Moreover, it is hard to determine the size of appropriate neighborhoods. To deal with these problems, in this paper, Graph Regularized Sparsity Discriminant Analysis, GRSDA, for short, is proposed. Based on graph embedding and sparsity preserving projection, the weight matrices for intrinsic and penalty graphs are obtained through sparse representation. GRSDA seeks a subspace in which samples in intra-classes are as compact as possible while samples in inter-classes are as separable as possible. Specifically, samples in the low-dimensional space can preserve the sparse locality relationship in the same class, while enhancing the separability for samples in different classes. Hence, GRSDA can achieve better performance. Extensive experiments were carried out on ORL, YALE-B and AR face databases, and the results confirmed that the proposed algorithm outperformed LPP, UDP, SPP and DSNPE.

1. Introduction

Face recognition, as one popular application of pattern recognition and computer vision, has aroused great interest among researchers. The main steps for face recognition include preprocessing, feature extraction and classification. In order to make the subsequent tasks easier, many algorithms have been proposed for the preprocessing part, like detection [1,2], and a survey on face detection can be found in [3]. Classification for face recognition has developed from the simple yet elegant nearest neighbor (NN) [4] method to recently proposed regression-based classification algorithms such as Linear Regression Classifier (LRC) [5], Sparse Representation Classifier (SRC) [6] and Collaborative Representation Classifier (CRC) [7]. These three regression-based algorithms have achieved comparable results and they have shown to have a great potential in practical applications. Moreover, in [8–10] many extensions to the aforementioned classification algorithms have also been proposed. Besides these, Support Vector Machine (SVM) [11] and deep learning [12,13] are also very popular in face recognition. For example, when deep learning was applied to face recognition, one algorithm is called DeepFace [12], which can achieve very impressive results in the Labeled Faces in the Wild (LFW) dataset.

Among all the algorithms for face recognition, appearance-based subspace learning schemes attract considerable interest due to its simplicity and desirable performance. Because the dimensionality of face images is usually very high, dimensionality reduction, which is also called feature extraction, is a key issue for face recognition, and has received tremendous attention in the past 20 years.

Many applications in computer vision and pattern recognition fields, such as face recognition, content-based image retrieval, bioinformatics etc., often confront high-dimensional and nonlinear samples. Nevertheless, dimensionality reduction gives an effective way to avoid the curse of dimensionality [14]. A lot of algorithms have been proposed in the past decades, and the two widely-used classic techniques are Principal Component Analysis (PCA) [15] and Linear Discriminant Analysis (LDA) [16,17], which are both matrix-decomposition based approaches [18] and assume that the distribution of samples is globally linear. However, in many applications such as face images where high-dimensional data are considered, distribution of samples is often nonlinear. One way to handle this problem is to use the kernel trick where the data from the original space is mapped to a higher-dimensional space. In the kernel space data is assumed to be linearly separable. Kernel Principal Component Analysis (KPCA) [19] and Kernel Linear Discriminant Analysis (KLDA) [20] are two representatives and can find their effectiveness in pattern recognition. However, how to choose the appropriate kernel is not an easy task, as it often influences the success of the algorithms.

* Corresponding author. Tel.: +8613586117695.
E-mail address: lousongjiangac@163.com (S. Lou).

http://dx.doi.org/10.1016/j.neucom.2015.04.116
0925-2312/

Another category is manifold learning algorithms (for instance, ISOMAP [21], Locally Linear Embedding (LLE) [22], Laplacian Embedding (LE) [23], Local Tangent Space Alignment (LTSA) [24], Parallel vector Field Embedding (PFE) [25], Geodesic Distance Function Learning (GDL) [26], and Parallel Field Alignment for cross media Retrieval (PFAR) [27]), which have been proposed to discover the intrinsic low-dimensional presentations for high-dimensional and nonlinear data. However, these kinds of algorithms cannot map a new sample to the corresponding low-dimensional space, which is also called the Out-Of-Sample extension problem [28]. Thus numerous methods have been proposed to solve this problem, which can achieve an explicit mapping, like Locality Preserving Projection (LPP) [29,30], Neighborhood Preserving Embedding (NPE) [31], Unsupervised Discriminant Projection (UDP) [32], Marginal Fisher Analysis (MFA) [33], Linear Discriminant Embedding (LDE) [34], Orthogonal LPP [35], locality preserving discriminant projections (LPDP) [36], Discriminative multi-manifold analysis [37] and Iterative Subspace Analysis based on Feature Line Distance [38]. These algorithms considered the manifold structure or the discriminant information in one way or another, and showed to be more efficient than the traditional methods in some special scenarios. For example, in [37] the inter-manifold and intra-manifold graphs were defined according to the label information and then the optimal projection was searched, which yielded impressive results for a special case where only one sample was available for each person.

All these methods can be unified in the graph embedding framework [39,40]. In this framework the first step is to construct the graphs, that is, the intrinsic and penalty graphs. However, the performance of the algorithms is heavily dependent on how to construct the graphs. Traditional schemes employ k nearest neighbor or the ε-ball method, but how to choose the appropriate neighborhood size or the ball radius remains unclear. Moreover, for these two methods, the graph construction and the weight assignment are independent. One ideal model is that there is no parameter, and the graph construction and the weight assignment can be finished in one step [41].

Sparse representation [6,42] has received considerable interest in the last few years. The main idea in it is that the given test sample can be represented as a linear combination of the training samples, and the classification is achieved by evaluating which class leads to the minimum reconstructive deviation. The coefficients obtained by sparse representation can reflect the contributions of the samples to reconstruct the given test sample. It is reported in [43] that the bigger coefficients were the more likely these samples belonged to the same class. Hence, the reconstruction coefficients can be considered as a measurement of similarity.

Motivated by this idea, some researchers attempted to construct the adjacent graphs in a nonparametric way in which the graph construction and weight assignment can be finished in one step, and this technique has been applied to a wide range of applications due to the fact that it was parameter-free and robust to noises.

Yan et al. [44,45] proposed l_1 graph for image analysis, and constructed the graphs by sparse representation. In [46] a graph regularized sparse coding method was proposed, which combined local manifold structure into sparse representation. However, it was unsupervised and the performance was limited to some extent. Similar to l_1 graph, Qiao et al. [47] proposed Sparsity Preserving Projection (SPP), in which every sample was presented as a linear combination of the remaining samples. SPP tried to find a projection which can preserve the sparse reconstructive relationship. There was no need to choose the parameter of neighborhood size, and the authors pointed out that it had natural discriminative power and was robust to noises to some extent. However, SPP took the whole training set as the dictionary, and it was an unsupervised method. Zhang et al. [48] introduced a graph optimization for dimensionality reduction with sparsity constraints (GODRSC) which attempted to learn the sparse representation coefficients and the embedding simultaneously. In [49] Sparse Representation Classifier Steered Discriminant Projection (SRCSDP) was proposed, which tried to find a projection by maximizing the inter-class reconstruction error while minimizing the intra-class reconstruction error. Therefore it had more discriminative power, but it neglected the manifold structure and was time-consuming due to the fact that the projection matrix and sparse presentation coefficients were obtained iteratively. Chen and Jin [50] proposed a new feature extraction method called Reconstructive Discriminant Analysis (RDA) from the viewpoint of linear regression classification. Gui et al. [51] designed a new scheme called Discriminant Sparse Neighborhood Preserving Embedding (DSNPE), which represented the data as a linear combination of samples from the same class and preserved the sparse reconstructive relationship in the same class. However, it ignored the inherent manifold structure of training samples, especially the inter-class manifold structure, as it only integrated SPP and maximum margin criterion (MMC) [52]. Similar works can also be found in [53,54]

To exploit the merits of manifold learning and robustness of sparse representation, this paper presents a new algorithm called Graph Regularized Sparsity Discriminant Analysis (GRSDA), which utilizes sparse representation as a way to graph construction and weight assignment. In GRSDA, the intrinsic and penalty graphs are constructed via sparse representation and the weights are obtained subsequently, so it avoids the difficulty of determining the neighborhood size. On the one hand, it inherits the property of preserving the manifold structure like LPP; on the other hand, it derives from LDA which has good discriminative power. Under the graph embedding framework, GRSDA seeks a subspace, where samples from the same class are as compact as possible, while samples from different classes are as separable as possible.

The rest of this paper is organized as follows: Section 2 presents an overview of the related works like sparse representation, sparsity preserving projection and graph embedding. Graph Regularized Sparsity Discriminant Analysis is proposed in Section 3. Experiment results for the proposed algorithm and the related algorithms are shown in Section 4. Section 5 gives the conclusion.

2. The related work

Suppose that we have a training set $X = \{X_1, X_2, \dots, X_C\} = \{x_1, x_2, \dots, x_N\}$ of n samples, where $x_i \in R^D$ $(i = 1, 2, \dots, N)$ and D is the dimensionality. There are C classes, and there are N_k $(k = 1, 2, \dots, C)$ samples in the kth class. The aim of dimensionality reduction is to seek a projection A, so that every sample in the original space can be mapped to a low-dimensional space by $y_i = A^T x_i \in R^d$, where $d \ll D$.

2.1. Sparse representation

If a given test sample y belongs to the ith class, sparse representation assumes that y can be represented as a linear combination of the training samples in the ith class $X_i = \{x_i^1, x_i^2, \cdots, x_i^{N_i}\}$. In other words, we can present y as follows:

$$y = w_i^1 x_i^1 + w_i^2 x_i^2 + \cdots + w_i^{N_i} x_i^{N_i} = X_i W_i \tag{1}$$

where W_i denotes the representation coefficient of y over X_i. Ideally, the representation coefficients of other classes are zero, that is, $W_j = 0, \forall j \neq i$. Thus y can be represented as a linear