Pattern Recognition 62 (2017) 87–98

Contents lists available at ScienceDirect

Pattern Recognition

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

Regularized coplanar discriminant analysis for dimensionality reduction ☆

Ke-Kun Huang [a], Dao-Qing Dai [b,*], Chuan-Xian Ren [b]

[a] School of Mathematics, JiaYing University, Meizhou, Guangdong, 514015, China
[b] Intelligent Data Center and Department of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China

ARTICLE INFO

Article history:
Received 13 January 2016
Received in revised form
26 May 2016
Accepted 22 August 2016
Available online 26 August 2016

Keywords:
Dimensionality reduction
Sparse representation classifier
Face recognition
Hyperspectral image classification
Coplanar discriminant analysis

ABSTRACT

The dimensionality reduction methods based on linear embedding, such as neighborhood preserving embedding (NPE), sparsity preserving projections (SPP) and collaborative representation based projections (CRP), try to preserve a certain kind of linear representation for each sample after projection. However, in the transformed low-dimensional space, the linear relationship between the samples may be changed, which cannot make the linear representation-based classifiers, such as sparse representation-based classifier (SRC), to achieve higher recognition accuracy. In this paper, we propose a new linear dimensionality reduction algorithm, called Regularized Coplanar Discriminant Analysis (RCDA) to address this problem. It simultaneously seeks a linear projection matrix and some linear representation coefficients that make the samples from the same class coplanar and the samples from different classes not coplanar. The proposed regularization term balances the bias from the optimal linear representation and that from the class mean to avoid overfitting the training data, and overcomes the matrix singularity in solving the linear representation coefficients. An alternative optimization approach is proposed to solve the RCDA model. Experiments are done on several benchmark face databases and hyperspectral image databases, and results show that RCDA can obtain better performance than other dimensionality reduction methods.

1. Introduction

In many fields of pattern classification, such as face recognition, hyperspectral image classification, object categorization, action recognition and target detection, the original data is usually provided in high-dimensional form, while the underlying structure in many cases can be characterized by a small number of features. Dimensionality reduction (DR) is necessary and helpful for classification. In the past decades, many useful techniques for DR have been developed, such as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], Fisherface [3], maximum margin criterion (MMC) [4], regularized discriminant analysis (RDA) [5], locality preserving projections (LPP) [6], and marginal fisher analysis (MFA) [7].

In recent years, researchers proposed a number of methods to improve the above traditional DR methods. Lu et al. proposed a parametric regularized LPP [8], which results in better locality preserving power. Deng et al. proposed a transform-invariant PCA [9], characterizing accurately the intrinsic structures of the human face that are invariant to the in-plane transformations. Lu et al. proposed a sparse exponential family PCA method [10], to achieve simultaneous dimension reduction and variable selection for better interpretation of the results. Oh et al. presented a generalized mean PCA [11], overcoming the problem that PCA is prone to outliers included in the training set. Abou-Moustafa et al. presented a Pareto LDA [12], to maximize each pairwise distance to maximally separate all class mean. Ghassabeh et al. proposed a fast incremental LDA [13], which accelerates the convergence rate of the incremental LDA algorithm. Ren et al. proposed the outlier Suppressing LDA [14], exploring the importance of the sample itself in building the optimal subspace. Wang et al. proposed a semi-supervised LDA [15], which can use limited number of labeled data and a quantity of the unlabeled ones for training. Ji et al. proposed a relevance MFA [16], to model the pairwise constraints of relevance-link and irrelevance-link into the relevance graph and irrelevance graph.

The above DR methods all consider the property about

☆This work is supported in part by National Science Foundation of China under Grants 61375033, 61403164, 61572536 and 11631015, in part by the Fundamental Research Funds for the Central Universities under Grant 161GZD16, and in part by the Foundation for Distinguished Young Talents in Higher Education of Guangdong, China (2013LYM_0085).

* Corresponding author.
E-mail addresses: kkcocoon@163.com (K.-K. Huang), stsddq@mail.sysu.edu.cn (D.-Q. Dai), rchuanx@mail.sysu.edu.cn (C.-X. Ren).

http://dx.doi.org/10.1016/j.patcog.2016.08.024
0031-3203/© 2016 Published by Elsevier Ltd.

Euclidean distance between pairwise points, while there are some DR methods which consider the property about the distance between a sample and the linear combination of other samples, called linear embedding property, such as neighborhood preserving embedding (NPE) [17], sparsity preserving projections (SPP) [18,19] and collaborative representation based projections (CRP) [20]. NPE [17] represents each data point as a linear combination of the neighboring points. Then it finds an optimal embedding such that the neighborhood structure can be preserved in the subspace. Recently, many discriminative versions of NPE have been proposed, such as neighbourhood preserving discriminant embedding (NPDE) [21] and double adjacency graphs-based discriminant neighborhood embedding (DAG-DNE) [22]. Different from NPE, where the nearest neighbors are manually chosen, SPP [18,19] automatically constructs a graph. It aims to preserve the sparse linear reconstructive relationship of the data, which is achieved by minimizing a l_1 regularization objective function. Because SPP is unsupervised too, recently, some works consider its discriminative version, such as LDSNPE [23], GDSNPE [24], discriminative SPP [25] and sparse discriminative multi-manifold embedding (SDMME) [26]. Like SPP, CRP [20] aims to preserve the collaborative representation based reconstruction relationship of data. CRP is much faster than SPP since CRP calculates the objective function with l_2 regularization while SPP calculates the objective function with l_1 regularization.

NPE, SPP and CRP try to preserve a certain kind of linear representation for each sample after projection. However, in the transformed low-dimensional space, the linear relationship between the samples may be changed. In other words, the linear representation coefficients in the original high-dimensionality space for a sample may be different from those in the transformed low-dimensional space. But the classifiers work in the transformed space. Recently, sparse representation based classifier (SRC) has been successfully used in pattern classification [27,28], and there are many works to extend SRC [29–33]. SRC first codes a testing sample as a sparse linear representation of all the training samples, and then classifies the testing sample by evaluating which class leads to the minimum representation error. SRC supposes that the sparse nonzero representation coefficients concentrate on the training samples with the same class label as the testing sample. In other words, a testing sample should be linearly represented by the samples from the same class better than those from other classes to achieve higher recognition accuracy. The dimensionality reduction methods based on linear embedding meet the characteristic of SRC at a certain degree, but their performances are still limited, because the linear relationship between the samples may be changed in the transformed low-dimensional space.

To achieve better performance based on SRC, Yang et al. proposed a method called SRC steered discriminative projection (SRCDP) [34]. Observing that SRC adopts a class reconstruction residual-based decision rule, SRCDP maximizes the ratio of between-class reconstruction residual to within-class reconstruction residual in the transformed low-dimensional space. But the within-class reconstruction residual is too big to be minimized by projection, and it is influenced by the samples from different classes, which limits SRCDP to achieve better performance. To design a more effective DR method for SRC, we propose a new method called regularized coplanar discriminant analysis (RCDA).

The contributions of this paper are listed as follows:

1. We propose the coplanar projection model, which simultaneously finds a projection matrix and some linear representation coefficients such that the sample from the same class in the same hyperplane as much as possible.
2. We propose the model of regularized coplanar discriminant analysis (RCDA). RCDA makes the samples from the same class coplanar and the samples from different classes not coplanar. The linear representation coefficients are regularized by the proposed mean l_2 norm, which can balance the bias from the optimal linear representation and that from the class mean to avoid overfitting the training data, and overcomes the matrix singularity in solving the linear representation coefficients.
3. We propose an optimization algorithm to solve the model of RCDA, which alternatively calculates the projection matrix and the linear representation coefficients.

The remainder of the paper is organized as follows. In Section 2, we describe the proposed method in detail. In Section 3, we give some analyses of the proposed method. The experimental results are given in Section 4. Finally, the conclusion is provided in Section 5.

2. The proposed method

In this section, we describe the proposed method in detail. First, we discuss the basic idea of the proposed method in Section 2.1, only concerning the within-class representation for convenience. Then we propose the RCDA model by considering a supervised version of the model in Section 2.2. The corresponding optimization algorithm is proposed in Section 2.3.

2.1. RCDA: basic idea

The objective of the proposed method is to find a linear projection matrix that reduces the error of the within-class linear representation while preserves the error of the between-class linear representation in the transformed space.

2.1.1. Coplanar projection

To facilitate describing our method, we first discuss the within-class representation of ith training sample $\mathbf{x}_i$. Let

$$\mathbf{X}_i = \left\{ \mathbf{x}_j | c_j = c_i \quad \text{and} \quad j \neq i \right\} \in \mathbf{R}^{m \times n_i^w}$$

be the n_i^w training samples from the same class as $\mathbf{x}_i$ except $\mathbf{x}_i$. We want to simultaneously find a projection matrix $\mathbf{W} \in \mathbf{R}^{m \times d}$ and within-class linear representation coefficients $\boldsymbol{\beta}_i^w$ for each training sample such that the error of within-class linear representation is minimized after linear transformation, i.e:

$$\min_{\mathbf{W}, \boldsymbol{\beta}^w} \sum_{i=1}^{n} \left\| \mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{X}_i \boldsymbol{\beta}_i^w \right\|_2^2 \tag{1}$$

where $\mathbf{W}^T\mathbf{W} = \mathbf{I}$, $\mathbf{I}$ is the identity matrix, and $\boldsymbol{\beta}^w$ denotes all the linear representation coefficients $\left\{ \boldsymbol{\beta}_i^w | i = 1, \ldots, n \right\}$.

Note that the linear representation coefficients are calculated in the transformed lower-dimension space where the classifier actually works, instead of original space. Because the linear relationship between the samples may be changed after linear transformation, it will be more accurate to calculate the linear representation coefficients in the transformed space. If each sample can be linearly represented by the samples from the same class, then the samples from the same class are in the same hyperplane. So the coplanarity can be measured by the error of within-class linear representation. Model (1) finds some linear projection directions that make the training samples from the same class as much as possible in the same hyperplane after projection, so we call it *coplanar projection* model.

It should be mentioned that model (1) is a new model. The most similar work is NPE [17]. The linear representation