Neurocomputing 87 (2012) 41–50

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Reconstructive discriminant analysis: A feature extraction method induced from linear regression classification

Yi Chen*, Zhong Jin

School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 2 June 2011
Received in revised form
15 January 2012
Accepted 5 February 2012
Communicated by Xiaofei He
Available online 23 February 2012

Keywords:
Reconstructive discriminant analysis
Feature extraction
Dimensionality reduction
Linear discriminant analysis
Linear regression classification
Face recognition
Finger Knuckle print recognition

ABSTRACT

Based on linear regression, a novel method called reconstructive discriminant analysis (RDA) is developed for feature extraction and dimensionality reduction (DR). RDA is induced from linear Regression classification (LRC). LRC assumes each class lies on a linear subspace and finds the nearest subspace for a given sample. But the original space cannot guarantee that the given sample matches its nearest subspace. RDA is designed to make the samples match their nearest subspaces. Concretely, RDA characterizes the intra-class reconstruction scatter as well as the inter-class reconstruction scatter, seeking to find the projections that simultaneously maximize the inter-class reconstruction scatter and minimize the intra-class reconstruction scatter. Actually, RDA can also be seen as another form of classical linear discriminant analysis (LDA) from the reconstructive view. The proposed method is applied to face and finger knuckle print recognition on the ORL, extended YALE-B, FERET face databases and the PolyU finger knuckle print database. The experimental results demonstrate the superiority of the proposed method.

1. Introduction

With the last several decades, dimensionality reduction (DR) has drawn considerable attention in the areas of image processing and pattern recognition. Generally, in practical applications, the raw data may contain variations of illumination and noises, which probably lead to misclassifications. And, it is time-consuming to perform classification directly in the high-dimensional space. For robust recognition and fast computation, DR techniques are usually performed first before the classification step. Although the DR step may cause information loss, recent literatures [2,24] indicate that the lost information has no substantial impact on the classification results. Even more, the researchers achieve much higher recognition rates in the reduced subspace [50,51]. As a fundamental problem in many scientific fields, DR plays an important role in scientific research. The goal of DR is to find a meaningful low dimensional representation of high dimensional data. With respect to pattern recognition, DR is an effective way to overcome the “curse of dimensionality” [1]. And more importantly, it reveals the distinctive features from the original data for pattern matching [2].

In the task of pattern recognition, discriminant analysis has shown its significant discriminability and becomes the fundamental tool in many areas. By far, numerous discriminant analysis methods have been developed. Among the proposed methods, the most well-known technique is linear discriminant analysis (LDA) [3]. Based on Euclidean distance, LDA searches for the project axes on which the inter-class data points are far away from each other while the intra-class data points are close to each other. Unfortunately, it has been pointed out that there are still some drawbacks existed in LDA. For example, (1) it usually suffers from the small sample size (SSS) problem [4] when the within-class scatter matrix is singular; (2) it is only optimal for the case where the distribution of the data in each class is a Gaussian with an identical covariance matrix [47]; (3) LDA can only extract at most $c-1$ features (c is the number of total classes), which is suboptimal for many applications. Numerous LDA variants [4–17,41–43] have been developed to solve the limitations mentioned above. Recently, motivated by manifold learning algorithms [18–20], researchers proposed a family of locality characterization based discriminant analysis techniques [21–26,34]. Different from LDA, these techniques extract local discriminative information. Despite the different motivations of these algorithms, they can be nicely interpreted in a general graph embedding framework [19,22,26]. The graph embedding view of subspace learning provides us a powerful platform to develop various kinds of dimensionality reduction algorithms. However, the high computational cost restricts these algorithms to be applied to large scale high dimensional data sets. To address this issue, a strong tool named spectral regression (SR) [52–56] was proposed for efficient subspace learning.

* Corresponding author. Tel.: +86 25 8431 7297x403 (lab).
E-mail address: cystory@qq.com (Y. Chen).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2012.02.001

Although the existing discriminant analysis techniques achieve remarkable performances, we notice that these methods were designed independently of classifiers. At the classification stage of the pattern recognition progress, the classifier is usually selected by experience. Obviously, the subspaces learned by different discriminant analysis methods have different characteristics that are invisible to the classifiers. However, one specific classifier just explores the subspace following the classification rule rather than the characteristic of the subspace. Therefore, the DR method may not match the random selected classifier perfectly, which potentially degrades the performance of the pattern recognition system. To connect DR methods with classifiers, one feasible way is to design the DR methods according to the classification rule of a specific classifier. In literatures, we find Yang et al. have designed discriminant analysis methods [28,29] according to the minimal local reconstruction error (MLRE) measure based classifier and the local mean based nearest neighbor classifier (LM-NNC) respectively. By combining the discriminant analysis methods with their corresponding optimal classifiers, the researchers demonstrated remarkable improvements against conventional discriminant analysis methods.

Very recently, an important work called linear regression classification (LRC) [27] is reported by Naseem et al., where linear regression is applied to estimate the reconstruction error. Then the label of the probe image will be assigned as the class with a minimum reconstruction error. In Naseem et al.'s pioneer work, the downsampled images are directly used for classification. However, neither the original space nor the downsampled image space can guarantee that the intra-class reconstruction error is minimal. To obtain the best performance, the original space should have smaller intra-class reconstruction errors and larger inter-class reconstruction errors. Due to the variations of illumination and noises, the inter-class reconstruction error is probably smaller than the intra-class reconstruction error in the original space. Under this circumstance, the performance of LRC will degrade. In order to strengthen the performance of LRC, we first inherit the assumption and the classification rule of LRC. Based on the inherited assumption and classification rule, we aim to find a subspace that has smaller intra-class reconstruction errors and larger inter-class reconstruction errors. Then we present a new method called reconstructive discriminant analysis (RDA) for feature extraction and DR.

To have an intuitive impression, we show the characteristics of RDA, LDA and the MLRE-based feature extractor (MLREF). Based on Euclidean distance, LDA searches for the directions that are most discriminative to separate the samples belonging to different classes. Different from LDA, MLREF and RDA are representation-based methods. MLREF finds the projections on which samples can be best represented by their local intra-class neighbors. Motivated by the classification rule of LRC, RDA finds the projections on which samples can be best expressed by all of their intra-class samples.

The rest of the paper is organized as follows. Related works are reviewed in Section 2. In Section 3, RDA is described in detail. Connections with some related works are analyzed in Section 4. In Section 5, the experiments are presented on the well-known databases to demonstrate the effectiveness of the proposed method. Finally, conclusions are drawn in Section 6.

2. Related works

2.1. Linear discriminant analysis (LDA)

Assume we have n samples from c classes. Let n_i represents the training number of the ith class and $\mathbf{x}_i^j \in R^d$ denotes the jth sample of the ith class, $i=1,2,\ldots,c$, $j=1,2,\ldots,n_i$. The objective function of LDA is as follows:

$$\mathbf{w}_{opt} = \arg\max_{\mathbf{w}} \frac{|\mathbf{w}^T\mathbf{S}_b\mathbf{w}|}{|\mathbf{w}^T\mathbf{S}_w\mathbf{w}|} \tag{1}$$

where

$$\mathbf{S}_b = \sum_{i=1}^{c} n_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^T \tag{2}$$

$$\mathbf{S}_w = \sum_{i=1}^{c}\left(\sum_{j=1}^{n_i}(\mathbf{x}_i^j-\mathbf{m}_i)(\mathbf{x}_i^j-\mathbf{m}_i)^T\right) \tag{3}$$

$\mathbf{m}_i$ is the average vector of the ith class, and $\mathbf{m}$ is the average vector of all samples. The optimal projections are the generalized eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$ corresponding to the largest generalized eigenvalues.

2.2. Linear regression classification (LRC)

LRC is based on the assumption that samples from a specific object class lie on a linear subspace. Using this concept, a linear model is developed. In this model, a probe image is represented as a linear combination of class-specific samples. Thereby the task of recognition is defined as a problem of linear regression. Least-squares estimation (LSE) [31–33] is used to estimate the reconstruction coefficients for a given probe image against all class models. Finally, the label is signed as the class with the most precise estimation.

Assume $\mathbf{X}_i$ is a class-specific model generated by stacking the n-dimensional image vectors

$$\mathbf{X}_i = [\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^{n_i}] \in R^{n\times n_i}, \quad i=1,2,\ldots,c \tag{4}$$

Suppose $\mathbf{y}$ is a probe sample from the ith class, it should be represented as a linear combination of the images from the same class (lying on the same subspace), i.e.,

$$\mathbf{y} = \mathbf{X}_i\boldsymbol{\beta}_i, \quad i=1,2,\ldots,c \tag{5}$$

where $\boldsymbol{\beta}_i \in R^{n_i\times 1}$ is the reconstruction coefficients. Given that $n \geq n_i$, the system of equations in Eq. (5) is well conditioned and can be estimated by LSE:

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\mathbf{y} \tag{6}$$

The probe sample can be reconstructed by Eq. (7):

$$\begin{aligned}\hat{\mathbf{y}}_i &= \mathbf{X}_i\hat{\boldsymbol{\beta}}_i, \quad i=1,2,\ldots,c \\ &= \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\mathbf{y}\end{aligned} \tag{7}$$

Then the distance measure between the probe sample $\mathbf{y}$ and reconstructed sample $\hat{\mathbf{y}}_i$, $i=1,2,\ldots,c$ can be computed, and the label is signed as the class with the minimum distance, i.e.,

$$\min_i \|\mathbf{y}-\mathbf{X}_i\hat{\boldsymbol{\beta}}_i\|^2, \quad i=1,2,\ldots,c \tag{8}$$

2.3. Minimal local reconstruction error measure based discriminant feature extraction

MLREF [28] is induced from the MLRE measure based Classifier (MLREC).The MLRE-based feature extractor aims to find the projections $\mathbf{P}$ that maximize the following criterion:

$$J(\mathbf{P}) = \frac{\mathrm{tr}(\mathbf{P}^T\mathbf{S}_b^L\mathbf{P})}{\mathrm{tr}(\mathbf{P}^T\mathbf{S}_w^L\mathbf{P})} \tag{9}$$

where

$$\mathbf{S}_w^L = \sum_{\substack{i,j \\ s=i}}\left(\mathbf{x}_i^j-\sum_t w_{st}^{ij}\mathbf{x}_s^t\right)\left(\mathbf{x}_i^j-\sum_t w_{st}^{ij}\mathbf{x}_s^t\right)^T \tag{10}$$