Neurocomputing 137 (2014) 185–191

Contents lists available at ScienceDirect

Neurocomputing

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

Regularized complete linear discriminant analysis

Wuyi Yang*, Houyuan Wu

Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Xiamen University, Ministry of Education, Xiamen, China

ARTICLE INFO

Article history:
Received 5 January 2013
Received in revised form
9 June 2013
Accepted 18 August 2013
Available online 18 February 2014

Keywords:
Linear discriminant analysis
Regularization
Small sample size problems

ABSTRACT

Linear discriminant analysis (LDA) searches for a linear transformation that maximizes class separability in a reduced dimensional space. Because LDA requires the within-class scatter matrix to be non-singular, it cannot be directly applied to small sample size (SSS) problems in which the number of available training samples is smaller than the dimensionality of the sample space. To solve SSS problems, this paper develops a system of regularized complete linear discriminant analysis (RCLDA). In RCLDA, two regularized criteria are used to derive discriminant vectors that include "regular" and "irregular" discriminant vectors in the range space and null space, respectively, of the within-class scatter matrix. Extensive experiments on the SSS problem of image recognition are performed to evaluate the proposed algorithm in terms of classification accuracy, and the experimental results demonstrate its effectiveness.

1. Introduction

Dimensionality reduction (DR), a common method of handling high-dimensional data, can be divided into the processes of feature selection and feature extraction. Feature selection tries to find a subset of the original variables, and feature extraction transforms the data from the high-dimensional space to a low-dimensional space. DR techniques have attracted much attention in the field of pattern recognition. Principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are the two most popular dimensionality reduction algorithms because of their relative simplicity and effectiveness. As an unsupervised method, PCA computes eigenvectors of the data covariance matrix and approximates the original data samples by linear combinations of the leading eigenvectors. In contrast, LDA takes the class labels into consideration and seeks a linear transformation that maximizes class separability in the reduced dimensional space. LDA requires the within-class scatter matrix to be non-singular, and the linear transformation can be found by applying an Eigen-decomposition on the scatter matrices of the given training dataset.

In many interesting machine learning and data mining problems, such as information retrieval, microarray data analysis, face recognition, and the number of available training samples are generally smaller than the dimensionality of the sample space. These problems are known as small sample size (SSS) problems, and have a significant influence on the design and performance of pattern recognition systems. For SSS problems, as the within-class scatter matrices are singular, LDA cannot be directly applied.

To address SSS problems, many extensions of LDA have been proposed. Linear algebra techniques were applied by pseudo-inverse LDA (PLDA) [2–5] to solve the numerical problem of inverting the singular within-class scatter matrix. PLDA finds the linear transformation by applying an Eigen-decomposition to the matrix $\mathbf{S}_w^+\mathbf{S}_b$, where $\mathbf{S}_w^+$ is the pseudoinverse of the within-class scatter matrix and $\mathbf{S}_b$ is the between-class scatter matrix. Based on the generalized singular value decomposition (GSVD) technique, LDA/GSVD [6,7] aims to find an optimal transformation $\mathbf{A}$ that would minimize the trace of the matrix $(\mathbf{A}^T\mathbf{S}_b\mathbf{A})^+(\mathbf{A}^T\mathbf{S}_w\mathbf{A})$. Regularized LDA (RLDA) [8–10] regularizes the within-class scatter matrix so that the regularized within-class scatter matrix is non-singular and LDA can be applied. RLDA is both space and time consuming, and efficient algorithms for RLDA were proposed [11,12]. The two-stage PCA plus LDA framework [13–15] is another technique for dealing with SSS problems. In this framework, an intermediate dimension reduction stage applies PCA to reduce the dimension of the original data before performing LDA. The dimension of the subspace transformed by PCA is chosen such that the total scatter matrix or the within-class scatter matrix of the mapped data points in the subspace is non-singular. Then, LDA can be applied in the subspace. Because the rank of the within-class scatter matrix is at most $n-c$, where n is the number of training samples and c is the number of classes, Fisherface [13] uses PCA to reduce the dimension of the feature space to $n-c$ to overcome the singularity problem. A limitation of the PCA plus LDA framework is that it is difficult to determine the optimal value of the reduced dimension for PCA. Since the null space of the within-class scatter matrix contains significant discriminatory

* Corresponding author. Tel.: +86 15280217168.
E-mail addresses: wyyang@xmu.edu.cn (W. Yang), why-08@163.com (H. Wu).

http://dx.doi.org/10.1016/j.neucom.2013.08.048
0925-2312

information [16], the PCA stage may lose some useful information for discrimination. Yu and Yang [17] suggested extracting the optimal discriminant vectors within the range space of the between-class scatter matrix, and proposed the direct LDA (DLDA) algorithm. DLDA is suboptimal in theory, and may discard some discriminatory information in the null space of the within-class scatter matrix. Chen et al. [16] projected the original space to the null space of the within-class scatter matrix, and then obtained the optimal discriminant vectors by finding the eigenvectors corresponding to the largest nonzero eigenvalues of the "reduced" between-class scatter matrix. However, Chen et al.'s algorithm (NLDA) is not efficient. Cevikalp et al. [18] proposed the discriminative common vectors (DCV) method for face recognition. Indeed, NLDA and DCV find the same projection subspace and obtain the optimal discriminant vectors in the null space of the within-class scatter matrix. Both NLDA and DCV overlook the discriminant information in the range space of the within-class scatter matrix. To find the discriminant vectors, maximal margin criterion (MMC) [19,20] maximizes the average margin between classes and is reported to be very effective for face recognition. Yang and Yang [21] built a theoretical foundation for the PCA plus LDA framework by proving that all optimal discriminant vectors can be derived from the range space of the total scatter matrix, and they suggested a complete linear discriminant analysis (CLDA) algorithm. CLDA finds irregular discriminant vectors in the null space of the within-class scatter matrix and regular discriminant vectors in the range space of the within-class scatter matrix. For classification, two kinds of discriminant feature vectors are fused in the decision level, and the fusion coefficient determines the weight of regular discriminant information [21,22].

In this paper, we propose a method of regularized complete linear discriminant analysis (RCLDA) to deal with SSS problems. RCLDA uses two regularized criteria to derive regular discriminant vectors in the range space of the within-class scatter matrix and irregular discriminant vectors in the null space of the within-class scatter matrix. For classification, the weight of regular discriminant information is determined while finding the discriminant vectors. The rest of the paper is organized as follows. Since RCLDA is based on CLDA, in Section 2, we briefly review the CLDA algorithm. Following that, RCLDA is introduced and analyzed in Section 3. In Section 4, we report on experiments conducted to demonstrate the effectiveness of the proposed algorithm. Our conclusions are given in Section 5.

2. Complete linear discriminant analysis

Given a training set composed of c classes, n_i denotes the number of samples in the ith class, $\mathbf{x}_j^i \in \mathbf{R}^m$ is the jth sample from the ith class, and a total of $n = \sum_{i=1}^{c} n_i$ samples are available in the set. The between-class scatter matrix $\mathbf{S}_b$ and the total scatter matrix $\mathbf{S}_w$ can be written as

$$\mathbf{S}_w = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(\mathbf{x}_j^i - \mathbf{m}_i)(\mathbf{x}_j^i - \mathbf{m}_i)^T,$$

$$\mathbf{S}_b = \frac{1}{n}\sum_{i=1}^{c}(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T,$$

where $\mathbf{m}_i$ is the centroid of the ith class and $\mathbf{m}$ is the centroid of the training set. LDA focuses on finding a linear transformation $\mathbf{A}=(\mathbf{a}_1,\ldots,\mathbf{a}_d)$ that maximizes the criterion

$$J_1(\mathbf{A}) = \underset{\mathbf{A}}{\arg\max}\frac{|\mathbf{A}^T\mathbf{S}_b\mathbf{A}|}{|\mathbf{A}^T\mathbf{S}_w\mathbf{A}|}.$$

when $\mathbf{S}_w$ is non-singular, the transformation matrix $\mathbf{A}$ is obtained by finding the eigenvectors corresponding to the d largest generalized eigenvalues of $\mathbf{S}_b\mathbf{a}=\lambda\mathbf{S}_w\mathbf{a}$. The linear transformation $\mathbf{A}$ maps each data sample in the m-dimensional space to a vector in the reduced d-dimensional space ($d \ll m$). For a sample $\mathbf{x}$ in the m-dimensional space, the embedded sample $\mathbf{y}$ in the d-dimensional space is given by $\mathbf{y}=\mathbf{A}^T\mathbf{x}$.

The within-class scatter matrices are singular in SSS problems, meaning that LDA cannot be directly applied. All optimal discriminant vectors can be derived from the range space of the total scatter matrix without any loss of the discriminatory information [21,22]. So, for SSS problems, all discriminant vectors can be extracted from the PCA transformed space, in which LDA can be performed. Suppose $\mathbf{p}_1,\ldots,\mathbf{p}_m$ are m orthonormal eigenvectors of $\mathbf{S}_t$, and the first s ($s=\text{rank}(\mathbf{S}_t)$) correspond to the nonzero eigenvalues, where $\mathbf{S}_t$ is the total scatter matrix. In the PCA transformed space, CLDA derives regular discriminant vectors in the range space of the within-class scatter matrix and derives irregular discriminant vectors in the null space of the within-class scatter matrix. Suppose $\mathbf{q}_1,\ldots,\mathbf{q}_s$ are s orthonormal eigenvectors of $\mathbf{P}^T\mathbf{S}_w\mathbf{P}$, and the first k ($k=\text{rank}(\mathbf{P}^T\mathbf{S}_w\mathbf{P})$) correspond to nonzero eigenvalues, where $\mathbf{P}=(\mathbf{p}_1,\ldots,\mathbf{p}_s)$. Let $\mathbf{Q}_1=(\mathbf{q}_1,\ldots,\mathbf{q}_k)$ and $\mathbf{Q}_2=(\mathbf{q}_{k+1},\ldots,\mathbf{q}_s)$. The column vectors of $\mathbf{PQ}_1$ form an orthonormal basis for $R(\mathbf{S}_w)$, where $R(\mathbf{S}_w)$ denotes the range space of $\mathbf{S}_w$. $(\mathbf{PQ}_1)^T\mathbf{S}_b\mathbf{PQ}_1$ is the "reduced" between-class scatter matrix and $(\mathbf{PQ}_1)^T\mathbf{S}_w\mathbf{PQ}_1$ is the "reduced" within-class scatter matrix in the range space of $\mathbf{S}_w$. Regular discriminant vectors are found with respect to the criterion

$$J_2(\mathbf{u}) = \frac{\mathbf{u}^T(\mathbf{PQ}_1)^T\mathbf{S}_b\mathbf{PQ}_1\mathbf{u}}{\mathbf{u}^T(\mathbf{PQ}_1)^T\mathbf{S}_w\mathbf{PQ}_1\mathbf{u}}. \tag{1}$$

Suppose $\mathbf{u}_1,\ldots,\mathbf{u}_d$ ($d \le c-1$) are the generalized eigenvectors of $(\mathbf{PQ}_1)^T\mathbf{S}_b\mathbf{PQ}_1\mathbf{u}=\lambda(\mathbf{PQ}_1)^T\mathbf{S}_w\mathbf{PQ}_1\mathbf{u}$ corresponding to the d largest positive eigenvalues. Then, $\mathbf{PQ}_1\mathbf{u}_1,\ldots,\mathbf{PQ}_1\mathbf{u}_d$ are the optimal regular discriminant vectors. Since $(\mathbf{PQ}_2)^T\mathbf{S}_w\mathbf{PQ}_2=0$, the column vectors of $\mathbf{PQ}_2$ form an orthonormal basis for $R(\mathbf{S}_t)\cap N(\mathbf{S}_w)$, where $R(\mathbf{S}_t)$ denotes the range space of $\mathbf{S}_t$ and $N(\mathbf{S}_w)$ denotes the null space of $\mathbf{S}_w$. Thus, irregular discriminant vectors can be derived from the range space of $\mathbf{PQ}_2$. $(\mathbf{PQ}_2)^T\mathbf{S}_b\mathbf{PQ}_2\mathbf{u}$ is the "reduced" between-class scatter matrix in the null space of $\mathbf{S}_w$. Irregular discriminant vectors are found with respect to the criterion

$$J_3(\mathbf{v}) = \frac{\mathbf{v}^T(\mathbf{PQ}_2)^T\mathbf{S}_b\mathbf{PQ}_2\mathbf{v}}{\mathbf{v}^T\mathbf{v}}. \tag{2}$$

Suppose $\mathbf{v}_1,\ldots,\mathbf{v}_l$ ($l \le c-1$) are the eigenvectors of $(\mathbf{PQ}_2)^T\mathbf{S}_b\mathbf{PQ}_2\mathbf{v}=\lambda\mathbf{v}$ corresponding to the l largest positive eigenvalues. Then, $\mathbf{PQ}_2\mathbf{v}_1,\ldots,\mathbf{PQ}_2\mathbf{v}_l$ are the optimal irregular discriminant vectors.

Projecting a sample $\mathbf{x}$ onto the regular discriminant vectors, we can obtain the regular discriminant feature vector

$$\mathbf{y}^1 = \mathbf{U}^T(\mathbf{PQ}_1)^T\mathbf{x},$$

where $\mathbf{U}=(\mathbf{u}_1,\ldots,\mathbf{u}_d)$. Similarly, projecting the sample $\mathbf{x}$ onto the irregular discriminant vectors, we can obtain the irregular discriminant feature vector

$$\mathbf{y}^2 = \mathbf{V}^T(\mathbf{PQ}_2)^T\mathbf{x},$$

where $\mathbf{V}=(\mathbf{v}_1,\ldots,\mathbf{v}_l)$. CLDA fuses the regular and irregular discriminant feature vectors in the decision level for classification. Denote a pattern $\mathbf{y}=[\mathbf{y}^1, \mathbf{y}^2]$, where $\mathbf{y}^1$ and $\mathbf{y}^2$ are the regular and irregular discriminant feature vectors, respectively, of the same sample. The summed normalized-distance [10] between sample y and the training sample $\mathbf{y}_i=[\mathbf{y}_i^1, \mathbf{y}_i^2]$ ($i=1,\ldots,n$) is defined by

$$g(\mathbf{y},\mathbf{y}_i) = \theta\frac{||\mathbf{y}^1-\mathbf{y}_i^1||}{\sum_{j=1}^{n}||\mathbf{y}^1-\mathbf{y}_j^1||} + \frac{||\mathbf{y}^2-\mathbf{y}_i^2||}{\sum_{j=1}^{n}||\mathbf{y}^2-\mathbf{y}_j^2||}, \tag{3}$$