Neurocomputing 122 (2013) 521–534

Contents lists available at ScienceDirect

Neurocomputing

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

Regularized least squares fisher linear discriminant with applications to image recognition

Xiaobo Chen [a,b,*], Jian Yang [b], Qirong Mao [a], Fei Han [a]

[a] School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, PR China
[b] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 26 November 2012
Received in revised form
27 March 2013
Accepted 22 May 2013
Communicated by Ran He
Available online 3 June 2013

Keywords:
Linear discriminant analysis (LDA)
Regularization technique
Concave–convex programming (CCP)
2-Norm loss function

ABSTRACT

Recursive concave-convex Fisher Linear Discriminant (RPFLD) is a novel efficient dimension reduction method and has been successfully applied to image recognition. However, RPFLD suffers from singularity problem and may lose some useful discriminant information when applied to high-dimensional data. Moreover, RPFLD is computationally expensive because it has to solve a series of quadratic programming (QP) problems to obtain optimal solution. In order to improve the generalization performance of RPFLD and at the same time reduce its training burden, we propose a novel method termed as regularized least squares Fisher linear discriminant (RLS-FLD) in this paper. The central idea is to introduce regularization into RPFLD and simultaneously use the 2-norm loss function. In doing so, the objective function of RLS-FLD turns out to be positive-definite, thus avoiding singularity problem. To solve RLS-FLD, the concave-convex programming (CCP) algorithm is employed to convert the original nonconvex problem to a series of equality-constrained convex QP problems. Each optimization problem in this series has a closed-form solution in its primal formulation via classic Lagrangian method. The resulting RLS-FLD thus leads to much fast training speed and does not need any optimization packages. Meanwhile, theoretical analysis is provided to uncover the connections between RLS-FLD and regularized linear discriminant analysis (RLDA), thus giving more insight into the principle of RLS-FLD. The effectiveness of the proposed RLS-FLD is demonstrated by experimental results on some real-world handwritten digit, face and object recognition datasets.

1. Introduction

With the rapid accumulation of high-dimensional data in numerous real-world applications such as face recognition, dimensionality reduction (DR) plays a more and more important role and attracts much attention during the last few decades. The key assumption of DR is that many classes of data, such as face images, possibly lie on or close to a subspace or manifold of much lower dimension than that of their ambient space. Performing DR by linearly combining the original features is of particular interest because it is simple to calculate and analysis. Two representative DR methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2,3] which have been extensively applied to face recognition [4–6] and other pattern recognition problems.

PCA seeks a set of mutually orthogonal basis vectors to represent the original data in a least squares sense. In fact, these basis vectors are exactly the eigenvectors associated with the largest eigenvalues of the covariance matrix of all the sample points. However, PCA is an unsupervised DR method because it does not take into account of the label information in the learning steps. As a result, the vectors found by PCA may be insufficient for pattern classification task. Unlike PCA, LDA searches the discriminant vectors that make the samples from the different classes far away from each other while keeping the samples from the same class to be closer to each other. The optimal discriminant vectors are obtained by maximizing the ratio of between-class scatter to within-class scatter. Recent studies have shown that LDA achieves better classification performance than PCA in many cases [5,7].

A major problem in LDA is known as singularity problem or small sample size (SSS) problem. Specifically, when the number of samples is small compared with the size of features, the within-class scatter matrix S_w in LDA is singular. To handle this problem, a simple yet effective strategy is based on regularization notion [8–10]. It adds a small perturbation to the diagonal elements of S_w so that it becomes nonsingular. Regularized LDA (RLDA) thus stabilizes the estimation of S_w and improves the performance of LDA. Another option to handle singularity problem is based on subspace analysis [5,11]. The resulting method is coined as Fisherfaces, in which PCA is first used as a pre-processing step to make S_w nonsingular before the application of LDA. Another problem of LDA is that it may fail to find a good discriminant vector for multi-class data especially when the classes distribute unevenly in original feature space [12]. This is because LDA tries to maximize the average distances among different classes. To deal with this

* Corresponding author at: School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, PR China. Tel.: +86 18 3628 80812.
E-mail addresses: xbchen82@gmail.com, xbchen82@yahoo.com.cn (X. Chen).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.05.006

so-called class separation problem, a heuristic method called aPAC (approximate Pairwise Accuracy Criterion) [13] is developed that assigns larger weights for similar classes in estimating between-class scatter. Another kind of recently emerging method tries to find discriminant features by maximizing the minimal distance among classes [14,15]. Local Fisher Discriminant Analysis (LFDA) [29], however, modifies the definition of within-class scatter and between-class scatter matrices of LDA, thus showing usefulness when samples in a class are multimodal.

Recently, some research works [16–18] have shown that casting an eigenvalue based method as a related SVM-type problem [19] can significantly improve the performance. Following this line, a novel discriminant analysis algorithm called recursive concave-convex Fisher linear discriminant (RPFLD) [16] is developed by converting the eigenvalue problem in LDA to a related SVM-type optimization problem which can be solved by quadratic programming (QP) iteratively. Although RPFLD empirically shows good discriminant performance compared with the original LDA, it still suffers from some drawbacks as described in the next section. In this paper, in order to enhance the performance of RPFLD and at the same time accelerate its training speed, we develop a novel DR method termed as regularized least squares Fisher linear discriminant (RLS-FLD). Specifically, to simplify the formulation of RPFLD, we replace the 1-norm loss function in RPFLD with the 2-norm loss function, borrowing the same idea as in LSSVM related literatures [20–23]. Meanwhile, taking into account that RPFLD only minimizes the empirical risk, we introduce regularization on the discriminant vectors so as to implement structural risk minimization [19]. Furthermore, aiming to generate more discriminant vectors, we directly impose orthogonal constraints on discriminant vectors. In such a way, we finally obtain an equality-constrained nonconvex optimization problem. To solve the resulting optimization problem, an iterative algorithm is developed based on both concave-convex procedure (CCP) [24] and classic Lagrangian multiplier method [25]. In what follows, we summarize some advantages of RLS-FLD compared with the original RPFLD.

(i) Due to regularization, RLS-FLD can reduce model complexity and is expected to produce better generalization ability. Interestingly, it overcomes the singularity problem in RPFLD at the same time.
(ii) Rather than solving QP problems, RLS-FLD boils down to solving a system of linear equations iteratively. Therefore, RLS-FLD does not require any specialized optimization package and produces much faster training speed.
(iii) The relation between RLS-FLD and regularized LDA is revealed based on the power method for eigenequations, thus giving a deep insight into the principle of RLS-FLD.

The rest of this paper is organized as follows. In Section 2, we briefly review LDA and the recently proposed RPFLD algorithm. Then, we present the formulation of RLS-FLD, develop an effective algorithm, and discuss its relation to RLDA in Section 3. Subsequently, we report the experimental results on some publicly available databases in Section 4. Finally, we conclude this paper in Section 5.

2. Brief reviews of LDA and RPFLD

2.1. Linear discriminant analysis (LDA)

Suppose $\omega_1, \omega_2, \ldots, \omega_K$ are K known pattern classes in D-dimensional input space. The number of samples in class ω_i is L_i $(i=1,2,\ldots,K)$ and let $L=\sum_{i=1}^{K} L_i$ be the total number of samples. Let $X=[x_1, x_2, \ldots, x_L]$ where x_i is a D-dimensional sample and $y_i \in \{\omega_1, \omega_2, \ldots, \omega_K\}$ be the associated class labels. The mean vectors for each class and for all of the classes can be defined as

$$m_i = \frac{1}{L_i} \sum_{y_j=\omega_i} x_j \text{ and } m = \frac{1}{L} \sum_{j=1}^{L} x_j \tag{1}$$

Then, the within-class scatter matrix S_w, between-class scatter matrix S_b and total scatter matrix S_t are defined, respectively, by the following formula

$$S_w = \sum_{i=1}^{K} \sum_{y_j=\omega_i}^{L_i} (x_j - m_i)(x_j - m_i)^T \tag{2}$$

$$S_b = \sum_{i=1}^{K} L_i (m_i - m)(m_i - m)^T = \frac{1}{2L} \sum_{i,j=1}^{K} L_i L_j (m_i - m_j)(m_i - m_j)^T \tag{3}$$

$$S_t = S_w + S_b = \sum_{j=1}^{L} (x_j - m)(x_j - m)^T \tag{4}$$

Note that two equivalent expressions for S_b are given in (3). According to the Fisher's criterion, LDA tries to find a discriminant vector such that the ratio of between-class scatter to within-class scatter is maximized after the projection of samples. Therefore, LDA seeks the optimal discriminant vector w by maximizing

$$J(w) = \frac{w^T S_b w}{w^T S_w w} \tag{5}$$

The objective function of (5) is known as Rayleigh quotient. Thus, when S_w is of full rank, the set of optimal discriminant vectors can be calculated by performing an eigenvalue decomposition of $S_w^{-1} S_b$ and taking the discriminant vectors to equal the eigenvectors corresponding to the d largest eigenvalues. However, when S_w is singular, an extra regularization term λI can be added to S_w which allows $S_w + \lambda I$ to be positive-definite, where λ is a small positive number and I is an identity matrix of appropriate dimensions. This method is called regularized LDA (RLDA) that has been proved to produce better performance than the original LDA.

2.2. Recursive concave–convex fisher discriminant analysis

RPFLD incorporates the notion of SVM [18,26] into the formulation of LDA by constraining the distance between mean vector of class i and total mean vector to be at least one. Subjecting to these constraints, RPFLD minimizes within-class scatter at the same time. To this end, RPFLD can be formulated as the following constrained minimization problem

$$\min_{w,\xi_i} \tfrac{1}{2} w^T S_w w + C \sum_i \xi_i$$
$$\text{s.t.} |w^T m_i - w^T m| \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \ldots, K \tag{6}$$

where the parameter C controls the trade-off between the within-class compactness and between-class separability whileas ξ_is $(i = 1, 2, \ldots, K)$ are nonnegative slack variables for relaxing the constraints.

Some properties of RPFLD should be pointed out as follows. First, to handle the nonconvexity of constraint $|w^T m_i - w^T m| \geq 1 - \xi_i$, RPFLD follows the idea of Concave–Convex Procedure [24] and converts those constraints to linear ones by replacing $|w^T m_i - w^T m|$ with its first-order Taylor expansion at current approximate solution. In other words, we can solve (6) approximately by solving a series of convex QP problems. Therefore, computing the solution for RPFLD is very time consuming and requires elaborated optimization toolbox. Second, when the number of features is much larger than that of samples, the SSS problem occurs and within-class matrix S_w tends to be singular or ill-conditioned. In such a case, RPFLD applies PCA on the raw data to reduce dimension such that S_w is invertible. Although it does overcome SSS problem, this method may discard some useful