Pattern Recognition 46 (2013) 1579–1594

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Complete large margin linear discriminant analysis using mathematical programming approach

Xiaobo Chen [a,b,*], Jian Yang [b], David Zhang [c], Jun Liang [d]

[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
[c] Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University, Hong Kong
[d] Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013, PR China

ARTICLE INFO

Article history:
Received 19 December 2011
Received in revised form
17 August 2012
Accepted 19 November 2012
Available online 30 November 2012

Keywords:
Feature extraction
Linear discriminant analysis
Large margin classifier
Constrained concave–convex procedure

ABSTRACT

In this paper, we develop a novel dimensionality reduction (DR) framework coined complete large margin linear discriminant analysis (CLMLDA). Inspired by several recently proposed DR methods, CLMLDA constructs two mathematical programming models by maximizing the minimum distance between each class center and the total class center respectively in the null space of within-class scatter matrix and its orthogonal complementary space. In this way, CLMLDA not only makes full use of the discriminative information contained in the whole feature space but also overcome the weakness of linear discriminant analysis (LDA) in dealing with the class separation problem. The solutions of CLMLDA follow from solving two nonconvex optimization problems, each of which is transformed to a series of convex quadratic programming problems by using the constrained concave–convex procedure first, and then solved by off-the-shelf optimization toolbox. Experiments on both toy and several publicly available databases demonstrate its feasibility and effectiveness.

1. Introduction

In many real-world applications, e.g. bioinformatics, face identification, we are often faced with high-dimensional data. Directly applying classic pattern recognition methods such as nearest neighborhood classifier, to deal with these data may bring about many problems, e.g. (1) it is generally time-consuming to carry out classification based on all of the original high-dimensional features, (2) large quantities of features may likely to deteriorate the classification performance since many features are irrelevant and redundant for predicting the desired output. By first discovering a low-dimensional subspace where a small number of good features are extracted from the original high-dimensional data and then performing classification in such subspace, we can address these problems by speeding up learning process and improving generalization ability. Therefore, dimensionality reduction (DR) techniques have attracted much attention in pattern recognition and machine learning community during the last few decades. By far, numerous DR methods have been developed, among which Principle component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are two representative linear subspace learning methods.

In contrast to PCA which does not take into account the class label information and thus is not reliable enough for classification task [3], LDA looks for a low-dimensional subspace by maximizing Fisher criterion, i.e. the ratio of between-class scatter to within-class scatter. Empirical study [4] showed that LDA owns better performance than PCA in face recognition applications and the corresponding method is known as Fisherfaces. Due to its effectiveness, LDA receives intense attention and has been widely used in many real-world applications. However, when the number of samples is much smaller than the dimensionality of sample space, the within-class scatter matrix S_w becomes singular preventing direct application of LDA. This is the well-known small sample size (SSS) problem and may occur in many circumstances. Besides, for multi-class problems, LDA may fail to find good projecting direction since it overemphasizes the class pairs with larger distance in the original sample space and tends to merge the neighboring class pairs after projecting onto a low-dimensional subspace [5,6]. This so-called class separation problem [7] roots in the fact that LDA tries to maximize the average distance between the centers of different classes [8] as indicated by the definition of between-class scatter matrix S_b. To handle the aforementioned problems of LDA, extensive methods were put forward in the literatures and demonstrated better performance.

To overcome small sample size problem, the conventional methods apply PCA first to generate a low-dimensional subspace in which S_w is no longer singular and then perform LDA in that

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

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.11.019

reduced space to yield discriminant vector, as it was done for example in Fisherfaces [4]. However, the PCA plus LDA strategy is imperfect in that some small principle components are thrown away in the PCA step and hence potentially loses some useful discriminative information. Considering the null space of S_w contains the most important discriminative information for classification if a SSS problem takes place, Chen et al. [9] proposed null space LDA (NLDA) algorithm aiming to seek discriminant vector by directly maximizing the between-class scatter in the null space of S_w. However, when the dimensionality D of sample space is too large, NLDA suffers from high computational complexity since it needs to decompose S_w with size of $D \times D$. Taking into account that the null space of the total scatter matrix S_t usually contains no discriminative information, Huang et al. [10] seek first a rather small discriminative subspace in the whole null space of S_w by removing the null space of S_t and then perform NLDA. In doing so, Huang′s method only needs to decompose a matrix whose size is bounded by $L-1$, where L denotes the total number of samples thus cutting down the computation load of NLDA. Zheng et al. [11] presented a similar approach which further incorporates statistically uncorrelated constraint on the discriminant vector into NLDA such that the extracted features are uncorrelated [12]. However, these null space-based methods [9–12] only consider discriminative information from the null space of S_w, whereas discarding that contained in its orthogonal complement space. To make full use of the information residing in both subspaces, Yang et al. [13,14] proposed a complete LDA (CLDA) framework which extracts the irregular discriminant vectors from the null space of S_w as well as regular discriminant vectors from its orthogonal complementary space, respectively. Then these two kinds of discriminant vectors are used jointly for better feature extraction.

To tackle the class separation problem, an intuitive idea is to incorporate a weighting function into Fisher criterion to ensure that the neighboring class pairs in the original sample space have higher weights since they are more likely to be misclassified in the projected space [5,6,15,16]. Among these weighting-based methods, Loog et al. [5] presented a simple but effective criterion named approximate pairwise accuracy criterion (aPAC) that adds weights which are used to approximate the Bayes error for class pairs in the estimation of S_b. However, it is not clear how to select an appropriate function to set the weights. More recently, several new methods [8,17,18] have been developed which can avoid the difficult of selecting weighting function. Our previous work [19] demonstrated classification performance can be largely improved by casting an eigen-decomposition based problem as a SVM-type problem [20–22]. Following this line, RPFDA [17] is further developed by converting the maximum average between-class scatter criterion in LDA into a set of SVM-type constraints. Unfortunately, RPFDA suffers from the singularity problem and resorts to the PCA plus LDA strategy. Xu et al. [8] proposed a method called minimum distance maximization (MDM). MDM applies first a whitening transformation on the samples to yield a subspace wherein the projected S_w becomes an identity matrix. Then, it searches for discriminant vectors by maximizing the minimum pairwise distance between the centers of different classes. A similar method is also independently developed in [18]. Although these methods [8,17,18] can handle the class separation problem well, they require performing PCA a prior to reduce the computational complexity which may loss some useful discriminative information.

In this paper, motivated by the success of CLDA [13,14] and RPFDA [17], we propose systematically a novel supervised DR framework, coined complete large margin linear discriminant analysis or CLMLDA for short. We call our method CLMLDA because of the following two reasons. Firstly, CLMLDA is capable to make full use of the discriminative information in and out of the null space of S_w and thus does not lead to any loss of useful information. Secondly, we introduce the idea of large margin to construct CLMLDA model. In such a way, the conventional eigen-decomposition based methods are reformulated as associated SVM-type optimization problems, which in turn bring about numerous advantages. It is interesting of our method from the following perspectives:

(1) Comparing with those null space based discriminant analysis methods [5,9–11] which maximize the (weighted) average distance between the centers of different classes in the null space of S_w, CLMLDA attempts to maximize the minimum distance between the centers of each class and total classes. This is consistent with the large margin principle which exhibits impressive generalization performance in the well-known SVM [21,22]. By doing so, the derived discriminant vectors may own better discriminative power and are expected to be insensitive to the statistical distribution of data.
(2) Comparing with RPFDA [17] which directly enlightens our work, the formulation of CLMLDA is more straightforward than RPFDA. To remove the limitation on the number of discriminant vectors obtained, RPFDA needs to generate new features via discarding the information represented by the old features in advance. It is like to result in singularity problem and hence applies PCA to tackle it. Different from RPFDA, we impose directly the orthogonality constraint on the set of existing discriminant vectors which is essentially similar to the classic orthonormal FLD [23,24]. As a result, CLMLDA avoids the singularity problem and does not require performing PCA in each iteration.
(3) Comparing with those existing mathematical programming based DR methods [8,17,18] which do not give sufficient consideration to the discriminative information contained in the null space of S_w, our proposed CLMLDA can make full use of two kinds of discriminative information, e.g. irregular and regular ones, which respectively lie in and out of the null space of S_w. It makes CLMLDA a more powerful discriminant analysis method. In addition, the optimization models in Ref. [8,18] are generally based on semidefinite program (SDP) [39] (except MDM which can be solved by quadratic programming) whereas CLMLDA is based on quadratic programming problem (QPP) which is simple and easy to implement.
(4) Last but not least, the new formulation of CLMLDA demonstrates the feasibility for the introduction of large margin principle into the field of DR as well as the advantages of mathematical programming based DR method. For example, although there exist some methods which can transform the generalized eigenvalue problems of DR into equivalent least squares problems [25,26] and facilitate the introduction of regularization techniques [27], the proposed method provides another possibility and calculation procedure from a new point of view.

The paper is organized as follows. In Section 2, we briefly review the basic LDA and NLDA methods. Then, we introduce the idea of large margin and propose CLMLDA framework in Section 3. Section 4 reports the experimental results on both artificial and benchmark databases. Finally, we draw a conclusion to this paper in Section 5.

2. Fundaments of Fisher LDA and null space based LDA

Suppose there are K pattern classes $\omega_1, \omega_2, \cdots, \omega_K$ in D-dimensional input space. The number of samples in class ω_i is L_i $(i = 1,\ 2, \cdots, K)$ and let $L = \sum_{i=1}^{K} L_i$ be the total number of