Neurocomputing 151 (2015) 207–214

Contents lists available at ScienceDirect

Neurocomputing

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

A deterministic approach to regularized linear discriminant analysis

Alok Sharma [a,b,*], Kuldip K. Paliwal [a]

[a] School of Engineering, Griffith University, Brisbane, QLD, Australia
[b] School of Engineering and Physics, University of the South Pacific, Suva, Fiji

ARTICLE INFO

Article history:
Received 19 January 2014
Received in revised form
25 September 2014
Accepted 26 September 2014
Communicated by D. Tao
Available online 8 October 2014

Keywords:
Linear discriminant analysis (LDA)
Regularized LDA
Deterministic approach
Cross-validation
Classification

ABSTRACT

The regularized linear discriminant analysis (RLDA) technique is one of the popular methods for dimensionality reduction used for small sample size problems. In this technique, regularization parameter is conventionally computed using a cross-validation procedure. In this paper, we propose a deterministic way of computing the regularization parameter in RLDA for small sample size problem. The computational cost of the proposed deterministic RLDA is significantly less than the cross-validation based RLDA technique. The deterministic RLDA technique is also compared with other popular techniques on a number of datasets and favorable results are obtained.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Linear discriminant analysis (LDA) is a popular technique for dimensionality reduction and feature extraction. Dimensionality reduction is a pre-requisite for many statistical pattern recognition techniques. It is primarily applied for improving generalization capability and reducing computational complexity of a classifier. In LDA the dimensionality is reduced from d-dimensional space to h-dimensional space (where $h < d$) by using a transformation $\mathbf{W} \in \mathbb{R}^{d \times h}$. The transformation (or orientation) matrix $\mathbf{W}$ is found by maximizing the Fisher's criterion: $J(\mathbf{W}) = |\mathbf{W}^T\mathbf{S}_B\mathbf{W}|/|\mathbf{W}^T\mathbf{S}_W\mathbf{W}|$, where $\mathbf{S}_W \in \mathbb{R}^{d \times d}$ is within-class scatter matrix and $\mathbf{S}_B \in \mathbb{R}^{d \times d}$ is between-class scatter matrix. Under this criterion, the transformation of feature vectors from higher dimensional space to lower dimensional space is done in such a manner that the between-class scatter in the lower dimensional space is maximized and within-class scatter is minimized. The orientation matrix $\mathbf{W}$ is computed by the eigenvalue decomposition (EVD) of $\mathbf{S}_W^{-1}\mathbf{S}_B$ [1].

In many pattern classification applications, the matrix $\mathbf{S}_W$ becomes singular and its inverse computation becomes impossible. This is due to the large dimensionality of feature vectors compared to small number of vectors available for training. This is known as small sample size (SSS) problem [2]. There exist several techniques that can overcome this problem [3–11,19–34]. Among these techniques, regularized LDA (RLDA) technique [3] is one of the pioneering methods for solving SSS problem. The RLDA technique has been widely studied in the literature [12–14]. It has been applied in areas like face recognition [13,14] and bioinformatics [15].

In the RLDA technique, the $\mathbf{S}_W$ matrix is regularized to overcome the singularity problem of $\mathbf{S}_W$. This regularization can be done in various ways. For example, Zhao et al. [12,16,17] have done this by adding a small positive constant α (known as regularization parameter) to the diagonal elements of matrix $\mathbf{S}_W$; i.e., the matrix $\mathbf{S}_W$ is approximated by $\mathbf{S}_W + \alpha\mathbf{I}$ and the orientation matrix is computed by EVD of $(\mathbf{S}_W + \alpha\mathbf{I})^{-1}\mathbf{S}_B$. The performance of RLDA technique depends on the choice of the regularization parameter α. In the past studies [18], this parameter is chosen rather heuristically, for example, by applying cross-validation procedure on the training data. In the cross-validation based RLDA technique (denoted here as CV-RLDA), the training data is divided into two subsets: training subset and validation subset. The cross-validation procedure searches over a finite range of α values and finds an α value in this range that maximizes the classification accuracy over the validation subset. In the cross-validation procedure, the estimate of α depends on the range over which it is explored. For a given dataset, its classification accuracy can vary depending upon the range of α being explored. Since many values of α have to be searched in this range, the computational cost of this procedure is quite high. In addition, the cross-validation procedure used in the CV-RLDA technique is biased towards the classifier used.

In order to address these drawbacks of CV-RLDA technique, we explore a deterministic way for finding the regularization parameter α. This would provide a unique value of the regularization

* Corresponding author.
E-mail address: alok.sharma@griffith.edu.au (A. Sharma).

http://dx.doi.org/10.1016/j.neucom.2014.09.051
0925-2312/© 2014 Elsevier B.V. All rights reserved.

parameter on a given training data. We call this approach as the deterministic RLDA (DRLDA) technique. This technique avoids the use of the heuristic (cross-validation) procedure for parameter estimation and improves the computational efficiency. We show that this deterministic approach computes the regularization parameter by maximizing the Fisher's criterion and its classification performance is quite promising compared to other LDA techniques.

2. Related work

In a SSS problem, the within-class scatter matrix $\mathbf{S}_W$ becomes singular and its inverse computation becomes impossible. In order to overcome this problem, generally inverse computation of $\mathbf{S}_W$ is avoided or approximated for the computation of orientation matrix $\mathbf{W}$. There are several techniques that can overcome this SSS problem. One way to solve this problem is by estimating the inverse of $\mathbf{S}_W$ by its pseudoinverse and then the conventional eigenvalue problem can be solved to compute the orientation matrix $\mathbf{W}$. This was the basis of pseudoinverse LDA (PILDA) technique [20]. Some improvements of PILDA have also been presented in [28,31]. In Fisherface (PCA+LDA) technique, d-dimensional features are firstly reduced to h-dimensional feature space by the application of PCA [2,52,53] and then LDA is applied to further reduce features to k dimensions. There are several ways for determining the value of h in PCA+LDA technique [4,5]. In the Direct LDA (DLDA) technique [7], the dimensionality is reduced in two stages. In the first stage, a transformation matrix is computed to transform the training samples to the range space of $\mathbf{S}_B$, and in the second stage, the dimensionality of this transformed samples is further transformed by some regulating matrices. The Improved DLDA technique [11], addresses drawbacks of DLDA technique. In the improved DLDA technique, first $\mathbf{S}_W$ is decomposed into its eigenvalues and eigenvectors instead of $\mathbf{S}_B$ matrix as of DLDA technique. Here, both its null space and range space information are utilized by approximating $\mathbf{S}_W$ by a well deterministic substitution. Then $\mathbf{S}_B$ is diagonalized using regulating matrices. For the Null LDA (NLDA) technique [6], the orientation $\mathbf{W}$ is computed in two stages. In the first stage, the data is projected on the null space of $\mathbf{S}_W$ and in the second stage it finds $\mathbf{W}$ that maximizes $|\mathbf{W}^{\mathrm{T}}\mathbf{S}_B\mathbf{W}|$. In orthogonal LDA (OLDA) technique [8], the orientation matrix $\mathbf{W}$ is obtained by simultaneously diagonalizing scatter matrices. It has shown that OLDA is equivalent to NLDA under a mild condition [8]. The Uncorrelated LDA (ULDA) technique [21], is a slight variation of OLDA technique. In ULDA, the orientation matrix $\mathbf{W}$ is made uncorrelated. The fast NLDA (FNLDA) technique [25], is an alternative method of NLDA. In this technique, the orientation matrix is obtained by using the relation $\mathbf{W} = \mathbf{S}_T^{+}\mathbf{S}_B\mathbf{Y}$, where $\mathbf{Y}$ is a random matrix of rank $c-1$, and c is the number of classes. This technique is so far the fastest technique of performing null LDA operation. In extrapolation LDA (ELDA) technique [32], the null space of $\mathbf{S}_W$ matrix is regularized by extrapolating eigenvalues of $\mathbf{S}_W$ using exponential fitting function. This technique utilizes range space information and null space information of $\mathbf{S}_W$ matrix. The two stage LDA (TSLDA) technique [34], exploits all four informative spaces of scatter matrices. These spaces are included in two separate discriminant analyses in parallel. In the first analysis, null space of $\mathbf{S}_W$ and range space of $\mathbf{S}_B$ are retained. In the second analysis, range space of $\mathbf{S}_W$ and null space of $\mathbf{S}_B$ are retained. In eigenfeature regularization (EFR) technique [10], $\mathbf{S}_W$ is regularized by extrapolating its eigenvalues in its null space. The lagging eigenvalues of $\mathbf{S}_W$ is considered as noisy or unreliable which are replaced by an estimation function. The general tensor discriminant analysis (GTDA) technique [48] has been developed for image recognition problems. This work focuses on the representation and preprocessing of appearance-based models for human gait sequences. Two models were presented: Gabor gait and tensor gait. In [49], authors proposed a constrained empirical risk minimization framework for distance metric learning (DML) to solve SSS problem. In double shrinking sparse dimension reduction technique [50], the SSS problem is solved by penalizing the parameter space. A detailed explanation regarding LDA is given in [51] and an overview regarding SSS based LDA techniques is given in [47]. There are other techniques which can solve SSS problem and applied in various fields of research [54–62]. In this paper, we focus on regularize LDA (RLDA) technique. This technique overcomes SSS problem by utilizing a small perturbation to the $\mathbf{S}_W$ matrix. The details of RLDA have been discussed in the next section.

3. Regularized linear discriminant techniques for SSS problem

In the RLDA technique, the within-class scatter matrix $\mathbf{S}_W$ is approximated by adding a regularization parameter to make it a non-singular matrix [3]. There are, however, different ways to perform regularization (see for details, [3,12–14,16,17,30,33]). In this paper we adopted Zhao's model [12,16,17] to approximate $\mathbf{S}_W$ by adding a positive constant in the following way $\hat{\mathbf{S}}_W = \mathbf{S}_W + \alpha\mathbf{I}$.[1] This will make within-class scatter matrix a non-singular matrix and then its inverse computation would be possible. The RLDA technique computes the orientation matrix $\mathbf{W}$ by EVD of $\hat{\mathbf{S}}_W^{-1}\mathbf{S}_B$. Thus, this technique uses null space of $\mathbf{S}_W$, range space of $\mathbf{S}_W$ and range space of $\mathbf{S}_B$ in one step (i.e., simultaneously).

In the RLDA technique, a fixed value of regularization parameter can be used, but it may not give the best classification performance as shown in Appendix B. Therefore, the regularization parameter α is normally computed by the cross-validation procedure. The cross-validation procedure (e.g. leave-one-out or k-fold) employs a particular classifier to estimate α and is conducted on the training set (which is different from the test set). We briefly describe below the leave-one out cross-validation procedure used in the CV-RLDA technique. Let $[a, b]$ be the range of α to be explored and α_0 be any value in this range. Consider a case when n training samples are available. The training set is first subdivided into training subset (consisting of $n-1$ samples) and validation subset (consisting of 1 sample). For this particular subdivision of training set, the following operations are required: (1) computation of scatter matrices $\mathbf{S}_B$, $\mathbf{S}_W$ and $\hat{\mathbf{S}}_W = \mathbf{S}_W + \alpha_0\mathbf{I}$ for $n-1$ samples in the training subset; (2) EVD of $\hat{\mathbf{S}}_W^{-1}\mathbf{S}_B$ to compute orientation matrix $\mathbf{W}$; and (3) classification of the left out sample (from the validation subset) by the classifier to obtain the classification accuracy. These computational operations are carried out for $n-1$ subdivisions of the training set and the average classification accuracy over the $n-1$ runs is computed. This average classification accuracy is obtained for a particular value of α (namely α_0). All the above operations will be repeated for other values of α in the range $[a, b]$ to get the highest average classification accuracy. From this description, it is obvious that the cross-validation procedure used in the CV-RLDA technique has the following drawbacks:

- Since the cross-validation procedure repeats the above-mentioned computational operations many times for different values of α, its computation complexity is extremely large.

[1] In the Friedman's model [3], $\mathbf{S}_W$ is estimated as $\hat{\mathbf{S}}_W = (1-\alpha)\mathbf{S}_W + \alpha\mathbf{I}$. We have compared Zhao's model and Friedman's model of CV-RLDA and found that Zhao's model exhibits comparatively better generalization capability (see Appendix A for details). Furthermore, we have considered Zhao's model because it is relatively simpler for establishing deterministic approach of computing α (in DRLDA).

Download English Version:

https://daneshyari.com/en/article/409706

Download Persian Version:

https://daneshyari.com/article/409706

Daneshyari.com