PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Pattern Recognition 38 (2005) 613–616

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and brief communication

Uncorrelated heteroscedastic LDA based on the weighted pairwise Chernoff criterion

A.K. Qin[a], P.N. Suganthan[a,*], M. Loog[b]

[a]School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Block S2, Singapore 639798, Singapore

[b]Department of Innovation, Image Analysis Group, IT University of Copenhagen, Rued Langgards Vej 7, DK-2300 Copenhagen, Denmark

Received 7 September 2004; accepted 20 September 2004

Abstract

We propose an uncorrelated heteroscedastic LDA (UHLDA) technique, which extends the uncorrelated LDA (ULDA) technique by integrating the weighted pairwise Chernoff criterion. The UHLDA can extract discriminatory information present in both the differences between per class means and the differences between per class covariance matrices. Meanwhile, the extracted feature components are statistically uncorrelated the maximum number of which exceeds the limitation of the ULDA. Experimental results demonstrate the promising performance of our proposed technique compared with the ULDA.

Keywords: Uncorrelated linear discriminant analysis; Heteroscedastic; Weighted pairwise Chernoff criterion

1. Introduction

In statistical pattern recognition, linear dimensionality reduction (LDR) techniques are widely applied to reduce the complexity of the statistical model and often result in the improved classification accuracy in the transformed lower-dimensional space. Fisher's linear discriminant analysis (LDA) is one of the most popular supervized linear dimensionality reduction techniques, which tries to find an optimal set of discriminant vectors $\mathbf{W} = [\varphi_1, \ldots, \varphi_d]$ by maximizing the Fisher criterion: $J_F(\mathbf{W}) = |\mathbf{W}^{\mathrm{T}}\mathbf{S}_b\mathbf{W}|/|\mathbf{W}^{\mathrm{T}}\mathbf{S}_w\mathbf{W}|$. Here, $\mathbf{S}_b$ and $\mathbf{S}_w$ are the between-class scatter matrix and average within-class scatter matrix of the training sample group, respectively, which can be estimated as follows:

$$\mathbf{S}_b = \sum_{i=1}^{C} P_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^{\mathrm{T}} = \sum_{i=1}^{C-1}\sum_{j=i+1}^{C} P_i P_j(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^{\mathrm{T}} \quad \text{and}$$

$$\mathbf{S}_w = \sum_{i=1}^{C} P_i \mathbf{S}_i, \tag{1}$$

where $C, P_i, \mathbf{m}_i, \mathbf{m}$ and $\mathbf{S}_i$ represent the total number of pattern classes, a priori probability of pattern class ω_i, the mean vector of class ω_i, the mean vector of all training samples and the covariance matrix of class ω_i, respectively. The between-class scatter matrix $\mathbf{S}_b$ can be expressed by both the original definition and its equivalent pairwise decomposition form [1].

* Corresponding author. Tel.: +00 65 6790 5404; fax: +00 65 6792 0415.

E-mail addresses: qinkai@pmail.ntu.edu.sg (A.K. Qin), epnsugan@ntu.edu.sg (P.N. Suganthan), marco@isi.uu.nl (M. Loog).

0031-3203/$30.00 © 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.09.007

Uncorrelated features are usually desirable in pattern recognition tasks because an uncorrelated feature set is likely to contain more discriminatory information than a correlated one of the same dimension. Recently, Jin et al. [2] proposed the uncorrelated LDA technique (ULDA), which can obtain discriminant vectors by maximizing the Fisher criterion under the constraints that the extracted feature components are statistically uncorrelated, i.e. the derived discriminant vectors are subject to the $\mathbf{S}_t$-orthogonal constraints: $\varphi_i \mathbf{S}_t \varphi_j = 0, \forall i \neq j, i, j = 1, \ldots, d$. Yang et al. [4] also demonstrated that ideal discriminant vectors should not only correspond to maximal Fisher criterion values but also correspond to minimal correlations between the extracted feature components. Therefore, the ULDA can yield a set of discriminant vectors with better discriminating power as shown experimentally in Refs. [2,4].

However, the ULDA technique still suffers from some deficiencies: firstly, it is incapable of dealing with heteroscedastic data in a proper way due to the implicit assumption that the covariance matrices for all the classes are equal. Hence, the derived discriminant vectors by the ULDA can merely attempt to separate the class means as much as possible while ignoring the discriminatory information present in the differences between the per class covariance matrices. This fact leads to the upper bound of the number of discriminant vectors extracted by ULDA to be limited to $C - 1$ as proven in Ref. [2]. Secondly, from the equivalent pairwise decomposition expression of the $\mathbf{S}_b$ matrix, we can easily find that the class pair with large distance between them in the original feature space are overemphasized in the pairwise $\mathbf{S}_b$ formula, which results in the obtained transformation attempting to preserve the distances of already well separated classes while causing larger overlap between pairs of classes that are not well separated in the original feature space. Consequently, the discriminant directions that may well separate the neighboring classes in the original feature space cannot be obtained by the ULDA if there are some classes far away and well separated from some other classes. In this paper, we propose an uncorrelated heteroscedastic LDA (UHLDA) technique based on the weighted pairwise Chernoff criterion, which can successfully solve the above problems.

2. Uncorrelated heteroscedastic LDA technique

2.1. ULDA technique

Suppose that $\mathbf{S}_b$ and $\mathbf{U}$ are positive semi-definite matrices and $\mathbf{S}_w$ is a positive definite matrix. The first ULDA discriminant vector, denoted by φ_1, is calculated as the eigenvector corresponding to the maximal eigenvalue of the eigenequation $\mathbf{S}_b \varphi = \lambda \mathbf{S}_w \varphi$. Suppose that i eigenvectors $\varphi_1, \varphi_2, \ldots, \varphi_i, i \geqslant 1$, have been obtained. The $(i+1)$th ULDA discriminant vector φ_{i+1}, which maximizes the Fisher criterion function $J_F(\mathbf{W})$ with $\mathbf{S}_t$-orthogonal constraints, is the eigenvector corresponding to the maximum eigenvalue of the eigenequation: $\mathbf{U}\mathbf{S}_b \varphi = \lambda \mathbf{S}_w \varphi$, where

$$\mathbf{U} = \mathbf{I} - \mathbf{S}_t \mathbf{D}^{\mathrm{T}} (\mathbf{D}\mathbf{S}_t \mathbf{S}_w^{-1} \mathbf{S}_t \mathbf{D}^{\mathrm{T}})^{-1} \mathbf{D}\mathbf{S}_t \mathbf{S}_w^{-1}, \quad \mathbf{D} = [\varphi_1 \varphi_2 \ldots \varphi_i]$$

and $\mathbf{I}$ is the identity matrix.

2.2. UHLDA based on the weighted pairwise Chernoff criterion

Although the set of discriminant vectors obtained by the ULDA technique can ensure the transformed feature components to be uncorrelated, which may significantly favor the subsequent pattern recognition tasks, these discriminant vectors set are actually not optimal due to the deficiencies described in introduction. Here, we introduce the multi-class Chernoff criterion function $J_C(\mathbf{W})$ into the ULDA, which can be regarded as the heteroscedastic extension of the Fisher criterion $J_F(\mathbf{W})$. The Chernoff criterion-based LDA solution was derived by Loog [1] and its efficiency has been experimentally demonstrated. However, the incorporation of the Chernoff criterion within the ULDA framework can ensure the discriminatory information in both class means' differences and class covariance matrices' differences to be extracted while the transformed feature components are statistically uncorrelated. The multi-class Chernoff criterion is defined as: $J_C(\mathbf{W}) = |\mathbf{W}^{\mathrm{T}} \mathbf{S}_C \mathbf{W}| / |\mathbf{W}^{\mathrm{T}} \mathbf{S}_w \mathbf{W}|$, where positive semi-definite $\mathbf{S}_c$ is the multi-class directed distance matrix that captures the summation of Chernoff distances between different class pairs and is defined by

$$\begin{aligned}
\mathbf{S}_c &= \sum_{i=1}^{C} \sum_{j=i+1}^{C} P_i P_j \mathbf{S}_C^{ij} \\
&= \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} P_i P_j \mathbf{S}_w^{1/2} ((\mathbf{S}_w^{-1/2} \mathbf{S}_{ij} \mathbf{S}_w^{-1/2})^{-1/2} \\
&\quad \times \mathbf{S}_w^{-1/2} (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^{\mathrm{T}} \mathbf{S}_w^{-1/2} (\mathbf{S}_w^{-1/2} \\
&\quad \times \mathbf{S}_{ij} \mathbf{S}_w^{-1/2})^{-1/2} \\
&\quad + \frac{1}{\pi_i \pi_j} (\log(\mathbf{S}_w^{-1/2} \mathbf{S}_{ij} \mathbf{S}_w^{-1/2}) - \pi_i \log(\mathbf{S}_w^{-1/2} \mathbf{S}_i \mathbf{S}_w^{-1/2}) \\
&\quad - \pi_j \log(\mathbf{S}_w^{-1/2} \mathbf{S}_j \mathbf{S}_w^{-1/2}))) \mathbf{S}_w^{1/2},
\end{aligned} \tag{2}$$

where $\pi_i = P_i/(P_i + P_j)$ and $\pi_j = P_j/(P_i + P_j)$ are relative a priori taking into account two classes that define the particular pairwise term. $\mathbf{S}_C^{ij}$ and $\mathbf{S}_{ij}$ are the pairwise directed distance matrix between classes i and j and the average pairwise within-class scatter matrix defined as $\pi_i \mathbf{S}_i + \pi_j \mathbf{S}_j$. $\mathbf{S}_i$ and $\mathbf{S}_j$ are the covariance matrices of classes i and j, respectively. The detailed derivation is presented in Ref. [1].

From Eq. (2), we notice that, in $\mathbf{S}_c$, all class pairs have same weights irrespective of their separability in the original space, which may possibly yield bad discriminant directions favoring originally well-separated class pairs. In order to avoid this problem, we introduce a weighting factor $w_{ij}(d_{ij}^c)$