PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition 38 (2005) 441–443

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and brief communications

Modified linear discriminant analysis

Songcan Chen*, Daohong Li

Department of Computer Science & Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China

Received 22 July 2004; accepted 5 August 2004

Abstract

In this paper, a modified Fisher linear discriminant analysis (FLDA) is proposed and aims to not only overcome the rank limitation of FLDA, that is, at most only finding a discriminant vector for 2-class problem based on Fisher discriminant criterion, but also relax singularity of the within-class scatter matrix and finally improves classification performance of FLDA. Experiments on nine publicly available datasets show that the proposed method has better or comparable performance on all the datasets than FLDA.

Keywords: Fisher linear discriminant analysis (FLDA); Modified FLDA (ModLDA); Rank limitation; Singularity; Pattern recognition

1. Introduction

Fisher linear discriminant analysis (FLDA) is a very popular and effective feature extraction and discriminant analysis approach [1] for 2-class problem in pattern recognition and data analysis and even becomes a comparison standard in face recognition [2]. Formally it can briefly be formulated as follows: Given two pattern classes $X^{(p)} = [x_{p1}, x_{p2}, \ldots, x_{pN_p}]$, p=1, 2 with N_p D-dimensional patterns in the pth class, respectively. FLDA attempts to seek an optimal discriminating vector φ by maximizing the Fisher criterion:

$$J(\varphi) = \frac{\varphi^{\mathrm{T}} S_b \varphi}{\varphi^{\mathrm{T}} S_w \varphi}, \tag{1}$$

where S_b is the between-class scatter matrix and denoted by $S_b=(m_1-m_2)(m_1-m_2)^{\mathrm{T}}$ and S_w is the within-class scatter matrix and denoted by $S_w=\sum_{p=1}^{2}\sum_{j=1}^{N_p}(x_{pj}-m_p)(x_{pj}-m_p)^{\mathrm{T}}$, m_1 and m_2 denote two corresponding class means, respectively. On one hand, by maximizing criterion (1), we can only get one optimal discriminating vector φ denoted by $S_w^{-1}(m_1-m_2)$ due to that the rank limitation of S_b. On the other hand, for the S_w, there is a possibility of nonexistence of its inverse and thus a singularity problem arises. The former shortcoming limits us to seek more discriminant vectors and thus may hinder further promotion for classification performance while the latter needs some indirect technique such as the regularization trick [3] to eliminate it. The goal of this paper is to (1) overcome the rank limitation so that we can seek as many discriminant vectors as possible and thus desire improving classification performance; (2) directly relax the singularity by newly defining a within-class scatter matrix. Finally, we perform experiments on nine publicly available datasets and obtain the experimental results of that ModLDA is better or comparable performance than FLDA.

2. Modified linear discriminant analysis

2.1. Definition of the new criterion

Focusing on the aforesaid problems with FLDA, in this section, we propose a new linear discriminant criterion

* Corresponding author. Tel.: +86 25 84892805; fax: +86 25 84498069.

E-mail addresses: s.chen@nuaa.edu.cn (S. Chen), lidaohong_2002@sohu.com (D. Li).

0031-3203/$30.00 © 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.08.008

(ModLDA) and aim to make it overcome the rank limitation and relax the singularity. The criterion of ModLDA can be formulated as follows:

$$J(\Phi) = \frac{tr(\Phi^{\mathrm{T}}\tilde{S}_b\Phi)}{tr(\Phi^{\mathrm{T}}\tilde{S}_w\Phi)}, \tag{2}$$

where

$$\tilde{S}_b = \sum_{p=1}^{c} \frac{J_p}{N} \sum_{j=1}^{J_p} \sum_{\substack{k=1 \\ k\notin I_p}}^{N} (x_{pj} - x_k)(x_{pj} - x_k)^{\mathrm{T}}, \tag{3}$$

a new c-class between-class scatter matrix and

$$\tilde{S}_w = \sum_{p=1}^{c}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} (x_{pk} - x_{pl})(x_{pk} - x_{pl})^{\mathrm{T}} + \eta \sum_{p=1}^{c} \left(\frac{N_p}{N}\right)^2 \sum_{k=1}^{N}\sum_{l=1}^{N} (x_k - x_l)(x_k - x_l)^{\mathrm{T}}, \tag{4}$$

a new c-class within-class scatter matrix, $N = \sum_{p=1}^{c} N_p$ denotes the total pattern number, N_p the number of patterns in the pth class C_p, x_{pj} is the jth pattern in C_p and x_k the kth pattern in $C = \bigcup_{p=1}^{c} C_p$, $I_p = \{k \mid x_k \in C_p\}$ is an index set of C_p, η is a tunable non-negative coefficient, controls a tradeoff between the two terms in the $\tilde{S}_w$ and plays a role in relaxing singularity of $\tilde{S}_w$, $\mathbf{\Phi}_d = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_d]$ consists of the optimal discriminant eigen-vectors obtained by maximizing Eq. (2) or equivalently, solving the following eigen-system equation

$$\tilde{S}_w^{-1}\tilde{S}_b\Phi_d = \Phi_d \Lambda_d, \tag{5}$$

where Λ_d is diagonal and consists of the eigenvalues corresponding to the d vectors making (2) maximization. Although criterion (2) is defined for multi-class discriminant analysis, here, we only consider 2-class problem mainly for a comparison with the 2-class FLDA.

2.2. Breakdown of the rank limitation

In this subsection, we will give a brief proof about breaking down the rank limitation, i.e., proving $Rank(\tilde{S}_b) \geqslant Rank(S_b)$, where $Rank(\mathrm{M})$ denotes the rank of matrix M. First we will rewrite Eq. (3) as $\tilde{S}_b = XEX^{\mathrm{T}}$, where $X = (X^{(1)}, X^{(2)}), X^{(p)} = (x_{p1}, \ldots, x_{pN_p})$ is a matrix composed of the patterns in $C_p, p = 1, 2,$ and $E = \begin{pmatrix} N_2 I_{N_1} & -1_{N_1\times N_2} \\ -1_{N_2\times N_1} & N_1 I_{N_2} \end{pmatrix}$. Then it is not difficult to verify that E can be transformed into a diagonal matrix $diag(N_2 I_{N_1}, N_1 I_{N_2-1}, 0_1)$ through a series of elementary matrix transform. Where I_n ($n = N_1, N_2$ and $N_2 - 1$) is an $n \times n$ identity matrix and 0_1 a matrix only containing 1 zero. Therefore, E is a positive semi-definite and has a rank of $N - 1$. Now performing a spectrum decomposition for the E, we have $E = U\Lambda U^{\mathrm{T}} = U\Lambda^{1/2}\Lambda^{1/2}U^{\mathrm{T}}$, where U is orthogonal and $\Lambda = diag(\lambda_1, \ldots, \lambda_N)$with $\lambda_i \geqslant 0\ (i = 1, 2, \ldots, N)$ is a eigenvalue matrix of the E. Since $Rank(E) = N - 1$, thus we can obtain

$$\begin{aligned} Rank(\tilde{S}_b) &= Rank(XEX^{\mathrm{T}}) = Rank(XU\Lambda^{1/2}\Lambda^{1/2}U^{\mathrm{T}}X^{\mathrm{T}}) \\ &= Rank(XU\Lambda^{1/2}) \\ &\geqslant Rank(XU) + Rank(\Lambda^{1/2}) - N \\ &= Rank(X) - 1. \end{aligned}$$

For 2-class problem, generally, there exist at least two linearly uncorrelated vectors in X and hence we have $Rank(X) \geqslant 2$, which states the rank of $\tilde{S}_b$ is greater than or equal to that of S_b. As a result, the rank limitation is broken down, meaning that more discriminant vectors can be found and applied to subsequent classification.

2.3. Relaxation of the singularity

Using similar technique to the above subsection, we can analyze the relaxation of singularity of $\tilde{S}_w$. In order to fulfill the analysis, we just prove that the 2nd term in $\tilde{S}_w$ is positive semi-definite. Now first set $R = \sum_{k=1}^{N}\sum_{l=1}^{N} (x_k - x_l)(x_k - x_l)^{\mathrm{T}}$ and thus have the 2nd term rewritten as $\eta\sum_{p=1}^{c}\left(\frac{N_p}{N}\right)^2 R$, now it is easy to see that R is at least a positive semi-definite matrix and thus so is the whole 2nd term, as a consequence, the singularity of $\tilde{S}_w$ can be relaxed due to that the sum of two positive semi-definite matrices is still positive semi-definite and the rank of the summed matrix will obtain an increase [4].

2.4. Relationship with FLDA

In fact, we can prove that FLDA is a special case of the ModLDA. The analysis refers to as follows: Simplifying $\tilde{S}_b$ in (3) and $\tilde{S}_w$ in (4), respectively, we have $\tilde{S}_b = N_1 N_2 (S_b + \frac{1}{N_1} S_1 + \frac{1}{N_2} S_2)$ and $\tilde{S}_w = 2(N_1 S_1 + N_2 S_2)$, where S_i is the within-class scatter of the ith class $(i = 1, 2)$ and $S_w = S_1 + S_2$. Obviously, when particularly setting $\eta = 0$ and $N_1 = N_2$, $\tilde{S}_b = \frac{N^2}{4} S_b + N S_w$ and $\tilde{S}_w = N S_w$, hence

$$\tilde{S}_w^{-1}\tilde{S}_b = \frac{1}{N} S_w^{-1} \left(\frac{N^2}{4} S_b + N S_w\right) = \frac{N^2}{4} S_w^{-1} S_b + I. \tag{6}$$

Eq. (6) indicates that $\tilde{S}_w^{-1}\tilde{S}_b$ has the same eigenvectors as the matrix $S_w^{-1}S_b$, thus both have the same discriminant vector. Note that in this paper, we select unequal number training patterns in each class to construct all the scatter matrices mainly according to the scales of given sample sets.

2.5. Design of classifier based on the new criterion

After obtaining the discriminating features for each pattern by Φ_d, we will use them to classify new patterns. Let x be a pattern to be classified, then it will be decided as the

1st class if $(m_2 - m_1)^{\mathrm{T}} \Phi_d \Phi_d^{\mathrm{T}} (x - b) < 0$, the 2nd class otherwise. Where b is taken as $\frac{N_1 m_1 + N_2 m_2}{N_1 + N_2}$ for two unequal training classes and m_i $(i = 1, 2)$ is defined as before.

3. Experimental results and conclusions

In order to compare the two approaches given in this paper, we perform the experiments on nine publicly available datasets from UCI database [5], Here is given a brief description for the experimental datasets: (1) Ionosphere (Ion): 34-dimension (34D), 225 patterns in the 1st class and 126 in the 2nd one, respectively. For short, (34D, 225, 126); (2) Musk clean2 (Mu2): (166D, 5581, 1017); (3) Water treatment (Wat): (38D, 65, 51); (4) Wisconsin diagnosis breast (WDB): (30D, 212, 357); (5) Wine: (12D, 107 (in original Class 1&2), 71 (in original Class 3)); (6) Breast cancer (Can): (10D, 444, 239); (7) Iris23(Iris): (4D, 50, 50); (8) Pima Indian diabetes (Diab): (8D, 500, 268) and (9) Bupa liver disorder (Liv): (6D, 145, 200).

All the dataset contain two classes from which we choose randomly half from each class to form the so-needed training set and then repeatedly perform the two algorithms 100 times on the separate training sets, finally the classification accuracies of 100 time runs on each dataset are averaged to give final classification performance as shown in Table 1.

Table 1
Classification performance comparison between FLD and ModLDA

Dataset	FLDA (%)	ModLDA (%)	*Rank* $(\tilde{S}_b)$	*Rank* $(\tilde{S}_w)$	*Rank* (S_w)
Ion	83.86	**85.42** (4)[a]	34	**.33**	**.33**
		85.41 (1)			
Mu2	79.86	**83.00** (3)	166	166	166
		82.84 (1)			
Wat	92.82	**93.04** (16)	38	38	38
		92.95 (1)			
WDB	95.30	**95.81** (1)	30	30	30
Wine	92.08	**92.30** (12)	12	12	12
		92.24 (1)			
Can	100.00	100.00 (1)	10	**.10**	**.9**
Iris	93.72	93.72 (1)	2	2	2
Diab	**73.57**	73.46 (1)	8	8	8
Liv	**62.80**	62.67 (2)	6	6	6
		61.68 (1)			

[a]The numbers in all the parentheses of this column denote the numbers of chosen discriminating vectors.

From the table, we observe that (1) ModLDA achieves better performance on the former five datasets, the same performance on Can and Iris, and comparable performance on the rest datasets compared to FLDA. In particular, on the Ion dataset, classification performance is increased by about 3 percentages. (2) On most of the datasets, the more the discriminating vectors are obtained, the more the performance is raised, but this is not necessarily always the case, for example, on the Liv dataset, the performance decreases conversely, which tells us how many discriminating vectors are selected at last is still problem-dependent. (3) The column marked as *Rank*$(\tilde{S}_b)$ indicates that the rank limitation is indeed broken down and thus more discriminating vectors are found, which is consistent with our theoretical analysis and (4) the last two columns state that our within-class scatter matrix also indeed relaxes singularity of FLDA corresponding matrix. On Can dataset, the rank is increased by 1 but this does also not necessarily always hold as in Ion dataset. In sum, ModLDA is effective and flexible in determining performance by appropriately selecting discriminating vectors and thus worthwhile to be investigated in future.

Acknowledgements

We thank National Natural Science Foundation of China and Jiangsu for the supports under Grant No. 60473035 and BK2004001, respectively.

References

[1] G.J. Mclachlan, Discriminant Analysis and Statistical Pattern Recognition, Wiley, New York, 1992.
[2] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE TPAMI 19 (1997) 711–720.
[3] J.H. Friedman, Regularized discriminant analysis, J. Am. Stat. Assoc. 84 (1989) 165–175.
[4] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
[5] UCI Repository available at http://www.kernel-machines.org/datasets/.