PERGAMON

Pattern Recognition 36 (2003) 3031–3033

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and Brief Communication

Theory analysis on FSLDA and ULDA

Yong Xu*, Jing-yu Yang, Zhong Jin

Department of Computer Science, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, People's Republic of China

Received 6 March 2003; accepted 27 March 2003

Abstract

It is first revealed that the Fisher criterion ratio of each FSLDA discriminant vector must not be less than that of corresponding ULDA discriminant vector. So, the phenomenon in Yang et al. (Pattern Recognition 35 (2002) 2665) is not strange but certain, and must be available in all experiments! In addition, it is also first illustrated that in fact ULDA discriminant vectors are the S_t- orthogonal eigenvectors of a generalized eigenequation. As a result, the algorithms to obtain S_t- orthogonal eigenvectors of the generalized eigenequation are equivalent to the ULDA algorithm. Consequently, it is possible to work out ULDA discriminant vectors more efficiently.

Keywords: Fisher criterion; Foley–Sammon linear discriminant analysis; Uncorrelated linear discriminant analysis

1. Introduction

It is known that linear discriminant analysis (LDA) is an important approach for linear feature extraction. There are two very effective LDA methods. One is Foley–Sammon linear discriminant analysis (FSLDA) [1], proposed by Foley and Sammon, and the other is uncorrelated linear discriminant analysis (ULDA), recently proposed by Jin [2]. FSLDA and ULDA have been widely used. Compared to FSLDA, ULDA also tries to find discriminant vectors $\varphi_1, \dots, \varphi_m$, which maximize Fisher criterion

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

where, S_b and S_w are the between-class scatter matrix and the within-class scatter matrix, respectively. Suppose $\zeta_1, \dots, \zeta_m$ are FSLDA discriminant vectors and $\xi_1, \dots, \xi_m$ are ULDA discriminant vectors. FSLDA discriminant vectors are successively obtained. So do ULDA discriminant vectors. That is to say, we should start to work out the kth FSLDA (or ULDA) discriminant vector after the previous $k-1$ FSLDA (or ULDA) discriminant vectors have been worked out. In addition, we must note that orthogonal constraints

$$\zeta_i^T \zeta_j = 0 \quad \forall i \neq j \tag{2}$$

must be satisfied among FSLDA discriminant vectors while ULDA discriminant vectors must be subject to S_t- orthogonal constraints

$$\xi_i^T S_t \xi_j = 0 \quad \forall i \neq j. \tag{3}$$

In fact, this is the key difference between FSLDA and ULDA. S_t is the total scatter matrix.

According to one experiment about CENPARMI handwritten numeral database, Jian Yang found two interesting things. The first is that the Fisher criterion ratios of FSLDA discriminant vectors are much larger than those of ULDA discriminant vectors. But, the second is that the classifying results of FSLDA is not better than those of ULDA [3]. The discovery is attractive and seemingly contradictory. But, what are other experiments? It is doubtful whether the discovery is sure. Under what conditions are the Fisher criterion ratios of FSLDA discriminant vectors larger than those of ULDA discriminant vectors?

* Corresponding author. Tel.: +86-25-4315751-705; fax: +86-25-4315510.

E-mail addresses: laterfall@sohu.com, laterfall@sina.com (Y. Xu), yangjy@mail.njust.edu.cn (J. Yang), jinzhong@mail.njust.edu.cn (Z. Jin).

0031-3203/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/S0031-3203(03)00157-2

2. Theory and analysis

The following discussion will be based on the equation:

$$S_b\zeta = \lambda S_w\zeta. \quad (4)$$

For simplicity, suppose that S_w is positive definite and $S_b, S_w \in R^{n\times n}$. As to eigenequation (4), there must be n S_t- orthogonal eigenvectors, $X_1, X_2, \ldots, X_n$, corresponding to eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$. Moreover, the following formulas are satisfied:

$$X_i^{\mathrm{T}} S_b X_j = \begin{cases} 0 & i \neq j, \\ \lambda_i & i = j, \end{cases} \quad (5)$$

$$X_i^{\mathrm{T}} S_w X_j = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases} \quad (6)$$

Obviously, $X_1, X_2, \ldots, X_n$ are subject to S_t- orthogonal constraints (3), and it is the reason why $X_1, X_2, \ldots, X_n$ are called S_t- orthogonal eigenvectors. It is certain that the set consisting of $X_1, X_2, \ldots, X_n$ is a basis of R^n, because $X_1, X_2, \ldots, X_n$ are linearly independent [4,5]. If $L(X_k, X_{k+1}, \ldots, X_n)$ is called the S_t- orthogonal complement of $S_{k-1} = L(X_1, X_2, \ldots, X_{k-1})$ and denoted by $S_{k-1}^{\perp}$, the following theorems will be inferred.

Theorem 1. *If S_{k-1} is a $k-1$-dimensional subspace of R^n and denotes $S_{k-1} = L(X_1, X_2, \ldots, X_{k-1})$, it will be sure that $J(cX_k) = \max_{\zeta\in S_{k-1}^{\perp}} J(\zeta) = \lambda_k$, $c =$ constant and $c \neq 0$.*

Proof.

$$S_{k-1} = L(X_1, X_2, \ldots, X_{k-1}), \quad (7)$$

$$S_{k-1}^{\perp} = L(X_k, X_{k+1}, \ldots, X_n). \quad (8)$$

Suppose $\zeta = l_k X_k + l_{k+1} X_{k+1} + \cdots + l_n X_n \quad \forall \zeta \in S_{k-1}^{\perp}$. Then, we have

$$J(\zeta) = \frac{l_k^2\lambda_k + l_{k+1}^2\lambda_{k+1} + \cdots + l_n^2\lambda_n}{l_k^2 + l_{k+1}^2 + \cdots + l_n^2}. \quad (9)$$

Obviously, if $\zeta^0 = l_k X_k$ and $l_k \neq 0$, then $J(\zeta^0) = \max_{\zeta\in S_{k-1}^{\perp}} J(\zeta) = \lambda_k$, i.e.

$$J(cX_k) = \max_{\zeta\in S_{k-1}^{\perp}} J(\zeta) = \lambda_k, \quad c = consant$$

$$\text{and} \quad c \neq 0. \quad \square \quad (10)$$

Theorem 2. *If S is an arbitrary subspace of R^n, $\dim S = n - k + 1$ and $1 \leqslant k \leqslant n$, it will be certain that $\max_{\zeta\in S} J(\zeta) \geqslant \lambda_k$.*

Proof. Suppose $S_k = L(X_1, X_2, \ldots, X_k)$.

Because of $\dim S = n - k + 1$ and $\dim S_k = k$, we have $\dim S + \dim S_k > n$.

Obviously, $S \cap S_k$ is not null, even though S is arbitrary.

Suppose $\zeta^0 = l_1 X_1 + l_2 X_2 + \cdots + l_k X_k \quad \forall 0 \neq \zeta^0 \in S \cap S_k$, so,

$$J(\zeta^0) = \frac{l_1^2\lambda_1 + l_2^2\lambda_2 + \cdots + l_k^2\lambda_k}{l_1^2 + l_2^2 + \cdots + l_k^2} \geqslant \lambda_k. \quad (11)$$

It is sure that $\max_{\zeta\in S} J(\zeta) \geqslant J(\zeta^0)$. As a result,

$$\max_{\zeta\in S} J(\zeta) \geqslant \lambda_k \quad \forall \dim S = n - k + 1. \quad \square \quad (12)$$

It is notable that in practice X_1 is simultaneously taken as the first vector of FSLDA discriminant vectors and ULDA discriminant vectors. But, according to the algorithms, FSLDA obtains the second vector in the orthogonal complement of the first vector, whereas ULDA obtains the second vector in the S_t- orthogonal complement of the first vector. Successively, FSLDA obtains the kth vector in the orthogonal complement of the previous $k-1$ FSLDA discriminant vectors, whereas ULDA obtains the kth vector in the S_t- orthogonal complement of the previous $k-1$ ULDA discriminant vectors [1,2].

Actually, the S_t- orthogonal complement of the previous $k-1$ ULDA discriminant vectors is an $n-k+1$-dimensional subspace. In other words, in fact the kth vector of ULDA discriminant vectors is obtained in the $n-k+1$-dimensional subspace and the vector, corresponding to the maximal Fisher criterion ratio among vectors in the $n-k+1$-dimensional subspace, will be taken as the kth ULDA discriminant vector. Based on inductive method, the following statement can be illustrated.

Corollary 1. *If there is $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d > 0$, $c_1X_1, c_2X_2, \ldots, c_dX_d$ will be ULDA discriminant vectors, where, $c_1, c_2, \ldots, c_d$ are constants and each is not zero.*

Proof. Firstly, for simplicity, the case that $\lambda_i \neq \lambda_j, \forall i \neq j$, $i = 1, 2, \ldots, d$, $j = 1, 2, \ldots, d$ is discussed. Since the second discriminant vector of ULDA must come from the S_t- orthogonal complement of X_1 and the vector c_2X_2 satisfies the condition $J(c_2X_2) = \max_{\zeta\in L(X_2, X_3, \ldots, X_n)} J(\zeta)$, according to the algorithm of ULDA, c_2X_2 should be the second vector of ULDA discriminant vectors. Moreover, if $c_1X_1, c_2X_2, \ldots, c_{k-1}X_{k-1}$ $(k-1 < d)$ have been taken as the previous $k-1$ ULDA discriminant vectors, the algorithm of ULDA will work out the kth discriminant vector in $L(X_k, X_{k+1}, \ldots, X_n)$. It is known that $J(c_kX_k) = \max_{\zeta\in L(X_k, X_{k+1}, \ldots, X_n)} J(\zeta)$. So $\zeta_k = c_kX_k$ should be the kth discriminant vector of ULDA discriminant vectors. According to inductive method, it is convinced that if there is $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_d > 0$, $c_1X_1, c_2X_2, \ldots, c_dX_d$ will be ULDA discriminant vectors, under the condition $\lambda_i \neq \lambda_j \quad \forall i \neq j$.

Secondly, we discuss the problem on $\lambda_i = \lambda_j$ for some $i \neq j$. If there is $\lambda_{k+1} = \lambda_k$, i.e. $J(c_k X_k) = J(c_{k+1} X_{k+1})$, $k+1 \leqslant d$ and the algorithm is searching the kth vector of ULDA discriminant vectors, $c_k X_k$ and $c_{k+1} X_{k+1}$ will be equivalent to be taken as the kth vector of ULDA discriminant vectors, because of the same discriminatory capability and the same S_t- orthogonal attribute. Now Corollary 1 has been inferred. Specially, $c_1 X_1, c_2 X_2, \ldots, c_d X_d$ are also called S_t- orthogonal eigenvectors of eigenequation (4).

As a result, naturally, the algorithms to obtain $X_1, X_2, \ldots, X_d$ ($d \leqslant n$) are alternative algorithms of ULDA. Moreover, $X_1, X_2, \ldots, X_d$ can be directly worked out from eigenequation (4) and the corresponding algorithm is much more efficient than the original algorithm of ULDA, which was proposed by Zhong Jin. Therefore, the above analysis and illustration are very significant and they make us obtain ULDA discriminant vectors easier.

It is obviously that FSLDA also obtains the kth discriminant vector in an $n-k+1$-dimensional subspace. According to Theorem 2, the kth vector of FSLDA discriminant vectors is subject to $J(\zeta_k) \geqslant \lambda_k$. Because ULDA discriminant vectors are the S_t- orthogonal eigenvectors of eigenequation (4), it is sure that $J(\xi_k) = \lambda_k$, where, ξ_k is the kth vector of ULDA discriminant vectors. As a result, $J(\zeta_k) \geqslant J(\xi_k)$ is inferred. Up to now, it is undoubting that the Fisher criterion ratio of each FSLDA discriminant vector is not less than that of corresponding ULDA discriminant vector! In any case, it is certain. In other words, the phenomenon in Ref. [3] is not haphazard but certifiable and must be available in all experiments!

Acknowledgements

We wish to thank National Science Foundation of China under Grant No. 60072034 for supporting.

References

[1] D.H. Foley, J.W. Sammon Jr., An optimal set of discriminant vectors, IEEE Trans. Comput. 24 (3) (1975) 281–289.
[2] Z. Jin, J.Y. Yang, Z.S. Hu, Z. Lou, Face recognition based on the uncorrelated discriminant transformation, Pattern Recognition 34 (7) (2001) 1405–1416.
[3] J. Yang, J.Y. Yang, David Zhang, What's wrong with Fisher criterion? Pattern Recognition 35 (2002) 2665–2668.
[4] X.R. Ding, G.T. Cai, Matrix Theory, Tianjin University Press, Tianjin, 1985 (in Chinese).
[5] Y.P. Cheng, K.Y. Zhang, Z. Xu, Matrix Theory, Xi'an University of Technology Press, Xi'an, 2002 (in Chinese).

About the Author—YONG XU was born in Sichuan, China, in 1972. He received his B.S. degree and M.S. degree in 1994 and 1997, respectively. Now, he is working for his Ph.D. degree in pattern recognition and intelligence system. He is the author of eight scientific papers in pattern recognition and image processing. His current interests include face recognition and detection, handwritten character recognition and image processing.

About the Author—JING-YU YANG received the B.S. degree in computer science from NUST, Nanjing, China. From 1982 to 1984 he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994 he was a visiting professor at the Department of Computer Science, Missuria University. And in 1998, he acted as a visiting professor at Concordia University in Canada. He is the author of over 100 scientific papers in computer vision, pattern recognition, and artificial intelligence. His current research interests are in the areas of pattern recognition, image processing and artificial intelligence, and expert system.

About the Author—ZHONG JIN was born in Jiangsu, China, in 1961. He received B.S. degree, M.S. degree and Ph.D. degree in 1982, 1984 and 1999, respectively. He is now a professor in the Department of Computer Science at NUST. He is the author of over 10 scientific papers in pattern recognition, image processing, and artificial intelligence. His current interests are in the areas of pattern recognition, video image processing, face recognition.