

广义统计不相关最优鉴别矢量集的一个理论结果

吴小俊^{1,2,3}, 杨静宇⁴, 王士同^{1,4}, Josef Kittler³

(1. 江苏科技大学计算机系, 江苏镇江 212003; 2. 中国科学院沈阳自动化所机器人学重点实验室, 辽宁沈阳 110015;

3. CVSSP, University of Surrey, Surrey GU2 7XH, UK; 4. 南京理工大学信息学院, 江苏南京 210094)

摘 要: 本文对统计不相关最优鉴别矢量集的理论问题进行研究, 提出了广义统计不相关最优鉴别准则, 并给出了广义统计不相关最佳鉴别矢量集的一个理论结果, 研究表明, 广义统计不相关最佳鉴别矢量集的计算公式与基于 Fisher 最优鉴别准则的统计不相关最佳鉴别矢量集的计算公式完全一样, 但是以前这一点没有被认识到. 本文的研究丰富了统计不相关最优鉴别分析的特征抽取理论.

关键词: 模式识别; 特征抽取; 鉴别分析; 广义最佳鉴别矢量集

中图分类号: TP391.4, O235 **文献标识码:** A **文章编号:** 0372-2112 (2004) 10-1720-03

A Theoretical Result on the Generalized Optimal Set of Statistically Uncorrelated Discriminant Vectors

WU Xiao-jun^{1,2,3}, YANG Jing-yu⁴, WANG Shi-tong^{1,4}, Josef Kittler³

(1. Dept. of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China;

2. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning 110015 China;

3. CVSSP, University of Surrey, Surrey GU2 7XH, UK;

4. School of Information, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094 China)

Abstract: This paper presents a research on the theoretical issue of optimal set of statistically uncorrelated discriminant vectors. First, the criterion of the generalized optimal set of statistically uncorrelated discriminant vectors is proposed, then a theoretical result on the generalized optimal statistically uncorrelated discriminant vectors is obtained. The research result indicates that the formula of the generalized optimal set of statistically uncorrelated discriminant vectors is the same as that of the conventional optimal set of statistically uncorrelated discriminant vectors, however, which was not realized. The research enriches the theory of feature extraction based on the statistically uncorrelated discriminant analysis.

Key words: pattern recognition; feature extraction; discriminant analysis; generalized optimal set of discriminant vectors

1 引言

在模式识别领域中, Fisher 线性判别方法有着重大的影响^[1~15]. Foley 和 Sammon 在 1970 年发展了 Fisher 线性判别方法, 提出了 Sammon 最佳鉴别平面的技术, 并将它用于解决两类问题^[2]. Sammon 最佳鉴别平面的技术在模式识别领域中得到广泛的应用与发展, Duchene and Leclercq 给出了对多类问题的 Foley-Sammon 最佳鉴别矢量集的计算公式^[3], Kittler 提出了基于 Fisher 鉴别准则的特征选择方法^[4], Turk 和 Pentland 提出了特征脸的方法^[11], Hong 和 Yang 提出了基于 SVD 的特征抽取方法^[5,6], Cheng 和 Yang 提出了一种新的相似鉴别准则^[7], Liu 提出了广义最佳鉴别平面和广义最佳鉴别矢量集的一系列方法^[8,9], 郭提出了广义最佳鉴别矢量的改进算法^[10~12], 我们最近又提出了广义最佳鉴别矢量集的近似解析算法^[13]. 在实际问题中, 模式原始特征的维数一般比较高, 特征分量可能是相关的, 因此, 具有统计不相关性的最佳鉴别矢量集的算法引起了人们的关注^[14~18]. 本文第 2 部分先介绍统计不相关

最佳鉴别准则和统计不相关最佳鉴别矢量集, 然后, 将统计不相关最佳鉴别准则推广, 提出了广义统计不相关最佳鉴别准则, 第 3 部分给出了广义统计不相关最佳鉴别矢量集的一个理论结果, 最后给出结论.

2 统计不相关最优鉴别准则、统计不相关最优鉴别矢量集与广义统计不相关最优鉴别准则

设 $w_1, w_2, \dots, w_m$ 为 m 个模式类, $X = \{x_i\}$, $i = 1, 2, \dots, N$ 为 n 维训练样本集, X 中的每一个 x_i 属于 w_j 类, 即 $x_i \in w_j$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, m$. 设 w_i 类的平均矢量、协方差矩阵与先验概率分别为 $m_i, c_i, P(w_i)$, 类间散布矩阵为 S_b 、类内散布矩阵为 S_w 和总散布矩阵为 S_t 分别为:

$$S_b = \sum_{i=1}^m P(w_i) (m_i - m_0) (m_i - m_0)^T,$$

$$S_w = \sum_{i=1}^m P(w_i) C_i$$

$$C_i = E\{(x - m_i)(x - m_i)^T / w_i\}$$

收稿日期: 2003-06-09; 修回日期: 2004-04-16

基金项目: 国家自然科学基金课题 (No. 60072034); 中国科学院沈阳自动化研究所机器人学重点实验室基金 (No. RL200108); 江苏省高校自然科学基金研究计划项目 (No. 01KJB52002); 江苏省自然科学基金 (No. BK2002001, BK2004058); 图象处理与图象通信实验室开放基金 (No. KJS03038)

$$S_t = S_b + S_w = E\{(x - m_0)(x - m_0)^T\}$$

$$m_0 = E\{x\} = \sum_{i=1}^m P(w_i) m_i$$

其中 m_0 为全体训练样本的平均矢量, E 表示数学期望.

由散布矩阵, Fisher 鉴别函数可定义为:

$$J(\Phi) = (\Phi^T S_b \Phi) / (\Phi^T S_w \Phi) \quad (1)$$

其中, Φ 为任一 n 维列矢量. 使函数 $J(\Phi)$ 达到最大值的矢量 Φ_1^* 为 Fisher 最佳鉴别方向, 训练样本在方向 Φ_1^* 上的投影集在一维子空间 $\text{Span}\{\Phi_1^*\}$ 中有最小的类内距离和最大的类间距离.

引理 1^[15] 样本矢量投影到两个最佳鉴别矢量 Φ_i, Φ_j 得到的两个分量统计不相关的充要条件为: $\Phi_i^T S_w \Phi_j = 0$. 则称 Φ_i, Φ_j 为关于 S_t 共轭正交.

定义 1^[15] 若两个最佳鉴别矢量 Φ_i, Φ_j 满足 $\Phi_i^T S_w \Phi_j = 0$, 即 Φ_i, Φ_j 关于 S_t 共轭正交, 则称 Φ_i, Φ_j 为统计不相关的.

首先求出 Fisher 最佳鉴别方向 Φ_1 . 在求出 $r(r \geq 1)$ 个最佳鉴别方向 $\Phi_1, \Phi_2, \dots, \Phi_r$ 后, 第 $r+1$ 个最佳鉴别方向是在满足共轭正交条件(2)式下的使 Fisher 鉴别准则函数(1)式取到最大值的向量 Φ_{r+1} :

$$\Phi_{r+1}^T S_t \Phi_i = 0, \quad i = 1, 2, \dots, r \quad (2)$$

引理 2^[14,15] 具有统计不相关性的最佳鉴别矢量集的第 $r+1$ 个最佳鉴别方向 Φ_{r+1} 是下列广义特征方程中最大的特征值对应的特征向量:

$$P S_b \Phi_{r+1} = \lambda S_w \Phi_{r+1} \quad (3)$$

$$\text{式中: } P = I - S_t D^T (D S_t S_w^{-1} S_t D^T)^{-1} D S_t S_w^{-1} \quad (4)$$

这里, I 为单位矩阵.

$$D = [\Phi_1 \Phi_2 \dots \Phi_r]^T = [\Phi_1^T \Phi_2^T \dots \Phi_r^T]^T \quad (5)$$

作为统计不相关最佳鉴别准则的推广, 我们提出以下广义统计不相关最佳鉴别准则:

定义 3 设

$$J(\Phi) = \left(\sum_{i=1}^r \Phi_i^T S_b \Phi_i \right) / \left(\sum_{i=1}^r \Phi_i^T S_w \Phi_i \right) \quad (6)$$

$$J(\Phi) = \max_{\Phi} J(\Phi) \quad (7)$$

其中 $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_r)$, $\tilde{\Phi} = (\tilde{\Phi}_1, \tilde{\Phi}_2, \dots, \tilde{\Phi}_r)$, $\Phi_1, \Phi_2, \dots, \Phi_r$ 与 $\tilde{\Phi}_1, \tilde{\Phi}_2, \dots, \tilde{\Phi}_r$ 为统计不相关单位矢量集. 则称式(6)为广义统计不相关鉴别函数, 式(7)称为广义统计不相关最佳鉴别准则, $\tilde{\Phi}_1, \tilde{\Phi}_2, \dots, \tilde{\Phi}_r$ 为广义统计不相关最佳鉴别矢量.

3 广义统计不相关最佳鉴别矢量集的一个理论结果

在求出 $r(r \geq 1)$ 个最佳鉴别方向 $\Phi_1, \Phi_2, \dots, \Phi_r$ 后, 第 $r+1$ 个最佳鉴别方向是在满足共轭正交条件式(2)下的使广义鉴别准则函数式(6)取到最大值的向量 Φ_{r+1} .

$$\text{令 } \alpha_r = \sum_{i=1}^r \Phi_i^T S_b \Phi_i, \quad \beta_r = \sum_{i=1}^r \Phi_i^T S_w \Phi_i,$$

为了使广义鉴别准则函数 $J(\Phi_{k+1}) = \left(\sum_{i=1}^{r+1} \Phi_i^T S_b \Phi_i \right) / \left(\sum_{i=1}^{r+1} \Phi_i^T S_w \Phi_i \right) = \frac{\alpha_r + \Phi_{r+1}^T S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}}$ 最大, 构造如下 Lagrange 函数:

$$L(\Phi_{r+1}) = \frac{\sum_{i=1}^{r+1} \Phi_i^T S_b \Phi_i}{\sum_{i=1}^{r+1} \Phi_i^T S_w \Phi_i} = \frac{\alpha_r + \Phi_{r+1}^T S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - \sum_{i=1}^r \mu_i \Phi_{r+1}^T S_t \Phi_i \quad (8)$$

对 $L(\Phi_{r+1})$ 关于 Φ_{r+1} 求偏导数, 并令其等于 0, 得

$$\begin{aligned} & 2 \frac{S_b \Phi_{r+1} (\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}) - S_w \Phi_{r+1} (\alpha_r + \Phi_{r+1}^T S_b \Phi_{r+1})}{(\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1})^2} \\ & - \sum_{i=1}^r \mu_i S_t \Phi_i = 0 \\ & \frac{S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - \frac{S_w \Phi_{r+1} (\alpha_r + \Phi_{r+1}^T S_b \Phi_{r+1})}{(\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1})^2} \\ & - \frac{1}{2} \sum_{i=1}^r \mu_i S_t \Phi_i = 0 \end{aligned} \quad (9)$$

将式(9)两边同左乘 Φ_{r+1}^T , 由统计不相关性知式(9)第三项为 0.

令

$$\lambda = (\alpha_r + \Phi_{r+1}^T S_b \Phi_{r+1}) / (\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}) \quad (10)$$

则式(9)为:

$$2 \frac{\Phi_{r+1}^T S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - 2\lambda \frac{\Phi_{r+1}^T S_w \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} = 0 \quad (11)$$

整理上式得,

$$\lambda = (\Phi_{r+1}^T S_b \Phi_{r+1}) / (\Phi_{r+1}^T S_w \Phi_{r+1}) \quad (12)$$

式(12)表明, 要使广义鉴别准则函数 $J(\Phi_{k+1}) = \left(\sum_{i=1}^{r+1} \Phi_i^T S_b \Phi_i \right) / \left(\sum_{i=1}^{r+1} \Phi_i^T S_w \Phi_i \right) = \frac{\alpha_r + \Phi_{r+1}^T S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}}$ 最大, 只要使 $\frac{\Phi_{r+1}^T S_b \Phi_{r+1}}{\Phi_{r+1}^T S_w \Phi_{r+1}}$ 最大即可.

在式(9)两边同时左乘 $\Phi_{r+1}^T S_t S_w^{-1}$, 由统计不相关性知第二项为 0, 则

$$2 \frac{\Phi_{r+1}^T S_t S_w^{-1} S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - \Phi_{r+1}^T S_t S_w^{-1} \left(\sum_{i=1}^r \mu_i S_t \Phi_i \right) = 0 \quad (13)$$

$$\text{令: } D = [\Phi_1, \Phi_2, \dots, \Phi_r]^T = [\Phi_1^T \Phi_2^T \dots \Phi_r^T]^T, \quad \mu = [\mu_1 \dots \mu_r]^T$$

则式(13)变成

$$2 \frac{D S_t S_w^{-1} S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - D S_t S_w^{-1} S_t D^T \mu = 0 \quad (14)$$

故有

$$\mu = \frac{2(D S_t S_w^{-1} S_t D)^{-1} D S_t S_w^{-1} S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} \quad (15)$$

将式(15)代入式(9), 并注意到 $\sum_{i=1}^r \mu_i S_t \Phi_i = S_t D^T \mu$, 则有

$$\frac{2 S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - \frac{2 \lambda S_w \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} - \frac{2 S_t D^T (D S_t S_w^{-1} S_t D)^{-1} D S_t S_w^{-1} S_b \Phi_{r+1}}{\beta_r + \Phi_{r+1}^T S_w \Phi_{r+1}} = 0$$

整理得:

$$S_b \Phi_{r+1} - S_t D^T (D S_t S_w^{-1} S_t D)^{-1} D S_t S_w^{-1} S_b \Phi_{r+1} = \lambda S_w \Phi_{r+1} \quad (16)$$

即

$$P S_b \Phi_{r+1} = \lambda S_w \Phi_{r+1} \quad (17)$$

其中,

$$P = I - S_t D^T (D S_t S_w^{-1} S_t D)^{-1} D S_t S_w^{-1} \quad (18)$$

因此, 广义统计不相关最佳鉴别矢量集的第 $r+1$ 个最佳鉴别

方向 φ_{r+1} 是广义特征方程 (17) 中最大的特征值对应的特征向量。

比较 (3) 和 (17) 两式, 我们惊讶地发现它们完全一样, 这说明广义统计不相关最佳鉴别矢量集的计算公式与基于 Fisher 最优鉴别准则的统计不相关最佳鉴别矢量集的计算公式完全一样, 只是人们以前没有注意到这一点。

4 结论

本文对统计不相关最佳鉴别矢量求解算法进行研究, 获得如下结论:

(1) 文献 [9] 将 Fisher 最优鉴别准则推广到广义最优鉴别准则情形, 这启发我们将统计不相关最佳鉴别准则推广到广义统计不相关最优鉴别准则。这是本文的一个基本出发点。

(2) 从理论上得到广义统计不相关最佳鉴别矢量集的一个理论结果, 研究发现, 广义统计不相关最佳鉴别矢量集的计算公式与基于 Fisher 最优鉴别准则的统计不相关最佳鉴别矢量集的计算公式完全一样, 只是以前我们没有注意到这一点。

(3) 本文的研究结果对揭示广义统计不相关最佳鉴别矢量集的本质以及它与一般统计不相关最佳鉴别矢量集的关系有重要的作用。

参考文献:

- [1] Turk M, Pentland A. Eigenfaces for face recognition[J]. J. Cognitive Neuroscience, 1991, 3(1): 71 - 86.
- [2] Foley D H, J W Sammon. An optimal set of discriminant vectors[J]. IEEE Trans Computers, 1975, 24(3): 281 - 289.
- [3] Duchene J, Leclercq S. An optimal transformation for discriminant and principal component analysis[J]. IEEE Trans PAMI, 1988, 10(6): 978 - 983.
- [4] J Kittler. On the discriminant vector method of feature selection[J]. IEEE Trans Computers, 1977, 26(6): 604 - 606.
- [5] Hong Z Q. Algebraic feature extraction of image for recognition[J]. Pattern Recognition, 1991, 24(3): 211 - 219.
- [6] 洪子泉, 杨静宇. 用于图像识别的图像代数特征抽取[J]. 自动化学报, 1992, 18(2): 232 - 238.
- [7] Cheng Y Q, Yang J Y, Liu K, et al. A novel feature extraction method for image recognition based on similar discriminant function[J]. Pattern Recognition, 1993, 26(1): 115 - 125.
- [8] Liu K, Cheng Yong-Qing, Yang J Y, et al. Algebraic feature extraction for image recognition based on an optimal discriminant criterion[J]. Pattern Recognition, 1993, 26(6): 903 - 911.
- [9] Liu K, Yang J Y. Generalized optimal set of discriminant vectors[J]. Pattern Recognition, 1992, 25(7): 731 - 739.
- [10] 郭跃飞, 杨静宇. 求解广义最佳鉴别矢量的一种迭代算法及人脸识别[J]. 计算机学报, 2000, 23(11): 1189 - 1195.
- [11] 郭跃飞. 人脸图像代数特征提取与最佳鉴别矢量的研究[D]. 南京: 南京理工大学, 2000.
- [12] Yue-Fei Guo, Jing-Yu Yang. Feature extraction method based on the generalized fisher discriminant criterion and facial recognition[J]. Pattern Analysis & Application. 2001, 4(1): 61 - 66.
- [13] 吴小俊, 杨静宇, 王士同, 等. A new algorithm for solving optimal discriminant vectors [J]. Journal of Computer Science and Technology, 2001, (17) 3: 324 - 331.
- [14] 金忠. 人脸图像特征抽取与维数研究[D]. 南京: 南京理工大学, 1999.
- [15] 金忠, 杨静宇. 一种具有统计不相关性的最佳鉴别矢量集[J]. 计算机学报, 1999, 22(10): 1105 - 1108.
- [16] 吴小俊. 图像特征抽取与识别理论及其在人脸识别中的应用研究[D]. 南京: 南京理工大学, 2002.
- [17] 吴小俊, 杨静宇, 王士同, 等. A new optimal set of uncorrelated discriminant vectors and it's application[A]. Proceedings of WCICA '02 (IEEE 02EX527) [C]. Shanghai, P. R. China: IEEE, WCICA, 2002.
- [18] 吴小俊, 杨静宇, 王士同等. 广义 DKL 变换及其在人脸识别中的应用研究[J]. 计算机科学, 2003, 30(1): 87 - 89.

作者简介:



吴小俊 男, 1967 年生于江苏省丹阳市, 博士, 副教授, 先后在国内外发表论文 50 余篇, 出版论著与教材 2 本, 先后承担过 UNESCO 国际合作项目, 国家自然科学基金, 国防预研重点项目, 国家重点实验室基金及其它省部级项目, 曾获省市科技进步奖, 目前主要从事神经网络、模式识别和人工智能的研究。



杨静宇 男, 1941 年生于吉林, 教授, 博士生导师, 先后在国内外发表论文 300 余篇, 先后承担过国家 863, 973 重点项目, 国家自然科学基金及其它省部级项目, 曾获省部级以上科技进步奖 20 多次, 主要研究领域为计算机视觉、信息融合、模式识别和智能机器人等。



王士同 男, 1964 年生于江苏邳江, 教授, 博士生导师, 先后在国内外发表论文 200 余篇, 出版论著与教材 8 本, 先后承担过多项国家自然科学基金, 国家教育部优秀教师基金及其它省部级项目, 曾获省部级科技进步奖多次, 目前主要从事神经网络、模糊系统和模糊人工智能的研究。



Josef Kittler 男, 1946 年生, 博士, 教授, 模式识别和计算机视觉领域知名学者, 多种国际知名刊物编委, 国际模式识别协会 (IAPR) 主席, 先后发表论文 500 余篇, 目前主要从事模式识别、计算机视觉和信号处理的研究。