

一种适用于小样本问题的基于边界的特征提取算法

黄 睿 何明一 杨少军

(西北工业大学电子信息学院 西安 710072)

摘 要 特征提取技术是模式识别领域进行数据降维和强化判别信息的有效方法. 线性判别分析是监督特征提取方法的典型代表, 获得广泛应用, 但受到小样本问题的制约. 对此提出一种适用于小样本问题的基于边界的特征提取算法. 算法利用高维数据小样本情况下线性可分概率增加以及其低维投影趋于正态分布的特点, 定义了新的类别边界, 不但考虑了由线性判别分析提出的类内、类间离散度, 也兼顾各类别的方差差异性. 通过极大化该边界获得最优投影向量, 同时避免因类内离散度矩阵奇异导致的小样本问题. 进一步将算法推广到多类问题. 高光谱数据特征提取与分类实验表明, 算法在小样本情况下对于两类和多类问题均具有良好的推广性能, 优于多种线性判别分析的改进算法, 并且在样本较多时也取得了满意结果.

关键词 特征提取; 线性判别分析; 小样本问题; 模式分类; 最大化类别边界

中图法分类号 TP391

A Margin Based Feature Extraction Algorithm for the Small Sample Size Problem

HUANG Rui HE Ming-Yi YANG Shao-Jun

(School of Electrical and Information, Northwestern Polytechnical University, Xi'an 710072)

Abstract Feature extraction techniques are widely employed to reduce the dimensionality of data and to enhance the discriminatory information for the classification and recognition tasks. Linear Discriminant Analysis (LDA) is the most popular supervised method for feature extraction, but it often suffers the small sample size problem due to the singularity of the within-class scatter which arises if the number of samples is smaller than the dimensionality of samples. A margin based feature extraction algorithm is proposed for the problem. In view of the facts that for the high-dimensional data, the probability of linear separability may grow in case of small samples and the low-dimensional projection is approximately normal, the proposed algorithm introduces a new definition of the margin, which involves not only the between-class scatter and within-class scatter proposed by LDA criterion, but also the differences of the class variances. Through maximizing the margin, we can obtain the optimal projection vector, and avoid the small sample size problem. Through theoretical analysis, the algorithm is further extended to the multi-class case. The experiment results show that the algorithm outperforms several improved versions of LDA in the case of small samples. At the same time, a satisfying performance is also achieved for larger samples.

Keywords feature extraction; linear discriminant analysis; small sample size problem; pattern classification; maximum margin

收稿日期: 2006-04-18; 修改稿收到日期: 2007-01-11. 本课题得到国家自然科学基金(60572097)资助. 黄 睿, 女, 1976年生, 博士, 主要研究方向为特征约简、模式识别、神经网络与遥感图像分析. E-mail: huangr@263.net. 何明一, 男, 1958年生, 教授, 博士生导师, 主要研究领域为神经网络与智能信息处理、图像处理、三维信息获取与处理等. 杨少军, 男, 1976年生, 博士, 主要研究方向为图像处理、多媒体通信与智能信息处理、无线传感器网络.

1 引言

特征提取技术可广泛应用于数据降维和强化判别信息,能降低数据分类计算量,且在标记样本有限的情况下有效克服 Hughes 现象^[1]. 线性判别分析 (Linear Discriminant Analysis, LDA) 是监督特征提取方法的典型代表,通过定义类内离散度矩阵和类间离散度矩阵,将线性变换矩阵的求解转换为极大化 Fisher 准则(一种 Rayleigh 系数)问题,获得广泛应用. 但在小样本 (small sample size) 情况下,因类内离散度矩阵接近奇异或奇异, LDA 失效.

近年来,小样本问题得到人们的关注,已经提出了一系列修正方案. PCA+LDA^[2]方法是首先利用 PCA 降维,保证类内离散度矩阵非奇异;再用 LDA 提取特征. 由于 PCA 降维很可能丢失一些有关判别的关键信息,因此对接下来的 LDA 特征提取影响较大. 近年来出现了一些利用零空间改进 LDA 的方法. Chen 等^[3]提出的 LDA+PCA 首先将数据变换到类内离散度的零空间,接着在该空间最大化类间离散度,是在类内离散度矩阵的零空间寻找最优投影. Huang 等^[4]对该方法进行改进,在移除类内离散度矩阵与类间离散度矩阵的共同零空间后,再应用 LDA+PCA. Direct LDA^[5]是另一种零空间方法,它首先移除类间离散度矩阵的零空间,接着寻找极小化类内离散度矩阵的投影向量. Li 等^[6]将 LDA 准则中类间离散度矩阵与类内离散度矩阵的比值关系改为相减关系,提出基于最大化边界准则 (Maximum Margin Criterion, MMC) 的特征提取方法. 这些改进都基于原始 Fisher 准则,没有考虑类别间方差差异的影响.

本文提出一种适用于小样本问题的基于边界的特征提取算法 (Margin-based Feature Extraction, MFE). 首先讨论了两类分类问题,利用高维数据小样本情况下线性可分概率增加以及其低维投影趋于正态分布的特点,定义了新的类别边界,在考虑类内、类间离散度的同时也兼顾各类方差的差异性,通过极大化该边界获得最优投影向量. 接着将 MFE 算法推广到多类问题. 最后通过高光谱数据特征提取与分类实验证明算法的有效性.

2 高维数据的统计特性

研究表明,高维数据特征空间呈现出不同于低

维空间的特性. 随着维数的增长,有以下结论^[7-8]:

(1) 数据向超立方体的角落、超(椭)球体的外层聚集,特征空间几乎是空的. 数据在高维空间具有低维结构,可以将其投影到低维空间而不损失过多分类判别信息.

(2) 正态分布数据有向尾部聚集的趋势;均匀分布数据有向角落聚集的趋势. 因此在高维空间估计数据概率分布更为困难.

(3) 高维数据的低维投影有服从正态分布或正态分布组合的趋势,当维数无穷大时,低维投影数据以概率 1 逼近正态分布.

(4) 在有限样本情况下,高维空间的线性可分概率随维数的增加而增加. 当样本数接近或小于维数时,样本的线性可分概率趋于 1.

因此,利用高维数据在特征空间具有低维结构并且低维投影趋于正态分布的特点,在判别信息损失不大的情况下对其降维,将数据投影到低维子空间再进行后续处理,这时标记样本相对充裕,往往可以得到更好的结果.

3 MFE 算法描述

3.1 两类分类问题的 MFE 算法

根据高维数据的空间特性可知,其低维投影趋于正态分布,并且维数越高,趋势越明显. 因此,当把高维数据映射到低维空间后,可以认为数据服从正态分布而不会带来太大误差.

对于小样本问题,数据在高维空间几乎是线性可分的. 对于两类分类问题,存在一个决策超平面,使得分类在一定准则下最优,因此只需提取一个垂直于该超平面的特征即可实现分类. MFE 算法就是要寻找最优特征向量 \mathbf{p} ,使得数据投影到一维空间后类别边界最大. 图 1 说明了 MFE 算法的原理.

由于正态分布数据可以完全由均值 u 和方差 σ^2 来描述,因此,一维空间类别边界可直观表示为

$$D(c_1, c_2) = |u_1 - u_2| - (\sigma_1 + \sigma_2) \quad (1)$$

可以看到,理想的类别边界应满足均值差较大,并使各类数据尽可能集中在其均值附近,即方差要小. 另一方面,当类别方差不同时,也有必要将这一差异体现出来. 为此,定义类别边界为

$$D(c_1, c_2) = (1 - \beta)(u_1 - u_2)^2 + \beta|\sigma_1^2 - \sigma_2^2| - (\sigma_1^2 + \sigma_2^2) \quad (2)$$

其中,参数 $\beta \in [0, 1]$ 协调均值差异与方差差异对边界的影响.

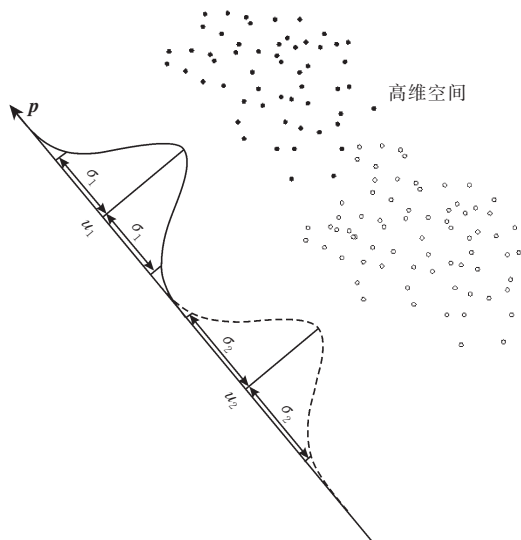


图1 MFE算法原理示意

原始高维空间数据沿投影向量 p 映射到一维空间, 则类别边界可表示为

$$D(c_1, c_2) = (1-\beta) p^T (m_1 - m_2) (m_1 - m_2)^T p + \beta |p^T (\Sigma_1 - \Sigma_2) p| - p^T (\Sigma_1 + \Sigma_2) p \quad (3)$$

最优特征向量 p 的求解可描述为如下优化问题:

$$\begin{aligned} \arg \max_p J_{\text{mfe}}(p) &= p^T ((1-\beta) \Sigma_b + \beta \Sigma^{(1-2)} - \Sigma_w) p \\ \text{subject to } p^T p &= 1 \end{aligned} \quad (4)$$

其中, Σ_b , Σ_w 分别为类间离散度矩阵与类内离散度矩阵;

$$\Sigma^{(1-2)} = \begin{cases} \Sigma_1 - \Sigma_2, & p^T (\Sigma_1 - \Sigma_2) p \geq 0 \\ \Sigma_2 - \Sigma_1, & p^T (\Sigma_1 - \Sigma_2) p < 0 \end{cases}$$

采用 Lagrange 乘子法, 定义 Lagrange 函数为

$$L(p, \lambda) = p^T ((1-\beta) \Sigma_b + \beta \Sigma^{(1-2)} - \Sigma_w) p - \lambda (p^T p - 1) \quad (5)$$

上式对 p 求偏导数, 并使其为零, 可得

$$((1-\beta) \Sigma_b + \beta \Sigma^{(1-2)} - \Sigma_w) p = \lambda p \quad (6)$$

优化问题转化为特征值求解, p 为相应特征向量, $J_{\text{mfe}}(p) = \lambda$.

当不考虑对投影向量模的要求时, 投影方向由下式确定:

$$\frac{\partial J_{\text{mfe}}}{\partial p} = ((1-\beta) \Sigma_b + \beta \Sigma^{(1-2)} - \Sigma_w) p = 0 \quad (7)$$

根据 Σ_b 的定义, 有

$$\Sigma_b p = (m_1 - m_2) (m_1 - m_2)^T p = (m_1 - m_2) R,$$

其中 $R = (m_1 - m_2)^T p$ 为标量. 则式(7)写为

$$(1-\beta) R (m_1 - m_2) = (\Sigma_w - \beta \Sigma^{(1-2)}) p \quad (8)$$

去除与方向无关的标量, 有

$$p_{\text{mfe}} \propto (\Sigma_w - \beta \Sigma^{(1-2)})^{-1} (m_1 - m_2) \quad (9)$$

算法 MFE 与 Malina 准则^[9]相似, 考虑了影响判别质量的三个因素, 并通过参数 β 来控制均值差异与方差差异的影响; 但 Σ_w 以相减的关系引入, 不需计算 Σ_w^{-1} , 因而不受小样本问题影响. 当类别方差相同时, MFE 就是 MMC 准则; 类别方差不同时, 算法通过 β 调节类别均值差异和方差差异对投影向量的影响; 类别均值相同时, p 由以下方程确定:

$$(\beta \Sigma^{(1-2)} - \Sigma_w) p = 0 \quad (10)$$

3.2 MFE 算法向多类分类问题的推广

多类分类问题的解决一般采取一对多或一对一的策略. 我们采用后者, 将多类问题分解为若干两类问题. 设类别数为 C , 数据维数为 L , p^{c_i, c_j} 是 MFE 对类 c_i 与 c_j ($1 \leq c_i, c_j \leq C, c_i \neq c_j$) 分类提取的特征向量, 则共提取了 $C(C-1)/2$ 个特征向量, 有投影矩阵 $P = [p^{1,2}, p^{1,3}, \dots, p^{C-1,C}]$, 简记为 $P = [p_1, p_2, \dots, p_{C(C-1)/2}]$.

为便于表述, 首先引入分类有效特征向量和 MFE 特征矩阵的概念^[10-11]:

(1) 若特征向量与分类决策面垂直, 我们称其为有效特征向量.

(2) 定义 MFE 特征矩阵为

$$\Sigma_{\text{mfe}} = PP^T \quad (11)$$

其中, Σ_{mfe} 是 $L \times L$ 的矩阵.

在所研究的高维数据小样本情况下, 分类决策面可由各两类问题的决策超平面组合而成. 我们有如下结论.

定理 1. Σ_{mfe} 非零特征值对应的特征向量是有效特征向量. 对于 C 类问题, 最多可提取 $C(C-1)/2$ 个特征.

证明. 设 $(\varphi_1, \varphi_2, \dots, \varphi_M)$ 为 Σ_{mfe} 非零特征值对应特征向量, 因此有

$$p_i = \sum_{j=1}^M \beta_j^{(i)} \varphi_j, \quad 1 \leq i \leq C(C-1)/2 \quad (12)$$

由于多类问题的分类决策面由各两类问题的决策超平面组合而成, 因此垂直于决策面的任意向量 v_{\perp} 可表示为

$$v_{\perp} = \sum_{i=1}^{C(C-1)/2} \gamma_i p_i \quad (13)$$

结合式(12), 令 $\alpha_j = \sum_{i=1}^{C(C-1)/2} \gamma_i \beta_j^{(i)}$, 则

$$v_{\perp} = \sum_{j=1}^M \alpha_j \varphi_j \quad (14)$$

由于任意线性独立向量集合可从有限维向量空间扩展为该空间的基, 因此, 可将 $(\varphi_1, \varphi_2, \dots, \varphi_M)$ 扩展为 L

维空间的一组正交基 $(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \cdots, \boldsymbol{\varphi}_M, \boldsymbol{\varphi}_{M+1}, \cdots, \boldsymbol{\varphi}_L)$. 对于向量 $(\boldsymbol{\varphi}_{M+1}, \boldsymbol{\varphi}_{M+2}, \cdots, \boldsymbol{\varphi}_L)$, 因 $\boldsymbol{\varphi}_i^T \boldsymbol{\varphi}_j = 0 (i \neq j)$, 有

$$\boldsymbol{\varphi}_i^T \mathbf{v}_\perp = \boldsymbol{\varphi}_i^T \sum_{j=1}^M \alpha_j \boldsymbol{\varphi}_j = \sum_{j=1}^M \alpha_j \boldsymbol{\varphi}_i^T \boldsymbol{\varphi}_j = 0, \quad M+1 \leq i \leq L \tag{15}$$

可见, 向量 $(\boldsymbol{\varphi}_{M+1}, \boldsymbol{\varphi}_{M+2}, \cdots, \boldsymbol{\varphi}_L)$ 正交于 \mathbf{v}_\perp . 由于它们与任意垂直于决策面的向量正交, 因此是与分类无关的冗余向量; 而 $(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \cdots, \boldsymbol{\varphi}_M)$ 是有效特征向量. 另一方面, 由于 $\text{rank}(\boldsymbol{\Sigma}_{\text{mfc}}) = M \leq C(C-1)/2$, 因此, 对于 C 类问题, 最多可提取特征数为 $C(C-1)/2$. 证毕.

对于 $C(C > 2)$ 类分类问题, MFE 算法可提取多于 $C-1$ 的特征, 可达到 LDA 准则的 $C/2$ 倍.

4 实验结果与分析

实验数据来自美国 AVIRIS 多光谱扫描仪 1992 年在印第安纳北部一块农业区获取的 220 波段高光谱数据. 从中筛除掉受大气污染严重、低信噪比的波段, 最终实验采用 190 波段高光谱数据.

分别研究 MFE 算法在样本数不同时解决两类问题和多类问题的性能, 并与 PCA+LDA、3 种零空间方法、MMC 算法进行比较. 所有特征提取方法均采用极大似然分类器进行分类. 选择 4 种地物进行多类实验, 并以地物 Soybeans-min 分别与 Corn-notill 和 Corn-min 构成两类实验的数据. 表 1 列出了实验样本情况.

表 1 样本描述				
类别	小样本实验		较大样本实验	
	训练 样本	测试 样本	训练 样本	测试 样本
Corn-notill(C1)	43	341	257	256
Corn-min(C2)	36	288	216	216
Soybeans-notill(C3)	34	272	204	204
Soybeans-min(C4)	51	408	307	306
总计	164	1309	984	982

4.1 两类分类实验

我们主要研究小样本情况下特征提取方法的性能, 表 2、表 3 是对两个数据集分别采用 6 种方法提取特征后的分类精度情况比较. 其中, 数据集 C1~C4 由 Corn-notill 和 Soybeans-min 构成; C2~C4 由 Corn-min 和 Soybeans-min 构成. 3 种零空间方法分别记为 C_LDA+PCA, H_LDA+PCA 和 DLDA. 实验中, 除 MMC 外的 6 种方法均提取一个特征. MMC 根据其提取规则, 可获得几十个特征, 本实验规定最多提取 10 个特征, 最终特征数由分类性能确定.

表 2 6 种特征提取方法用于数据集 C1~C4 的性能比较		
	训练精度/%	测试精度/%
PCA+LDA	100.00	80.77
C_LDA+PCA	100.00	82.51
H_LDA+PCA	100.00	82.38
DLDA	53.19	58.34
MMC	93.62	90.12
MFE	94.68	85.58

表 3 6 种特征提取方法用于数据集 C2~C4 的性能比较		
	训练精度/%	测试精度/%
PCA+LDA	100.00	80.03
C_LDA+PCA	100.00	78.88
H_LDA+PCA	100.00	79.89
DLDA	68.97	67.67
MMC	87.36	82.04
MFE	98.85	83.33

从表 2、表 3 可以看到, 利用 MFE 算法提取的特征取得了最好的推广性能; MMC 在两个实验中分别提取了 10 和 7 个特征, 虽然测试精度较高但训练精度偏低; DLDA 性能最差, 这是由于算法需首先移除 $\boldsymbol{\Sigma}_0$ 的零空间, 对于两类问题, 其零空间维数为 $L - (C-1) = L-1$, 因此造成大量有用判别信息的丢失; 其它 3 种方法的性能较为接近.

4.2 多类分类实验

我们考察了小样本情况下, PCA+LDA, C_LDA+PCA, H_LDA+PCA, DLDA, MMC 和 MFE 的性能. 表 4 为 6 种特征提取方法的分类精度比较. 其中, MMC 提取了 12 个特征.

表 4 小样本情况 6 种特征提取方法的特征值和分类精度比较												
特征数	PCA+LDA		C_LDA+PCA		H_LDA+PCA		DLDA		MMC		MFE	
	训练	测试	训练	测试	训练	测试	训练	测试	训练	测试	训练	测试
	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%	精度/%
1	100.00	41.02	100.00	39.80	100.00	39.80	46.95	49.66	32.32	35.14	49.39	50.27
2	100.00	43.70	100.00	43.39	100.00	42.93	47.56	52.33	37.80	40.26	67.68	66.77
3	100.00	44.77	100.00	48.74	100.00	48.43	51.22	55.69	50.00	49.20	83.54	79.98
4	—	—	—	—	—	—	—	—	57.93	57.98	89.02	79.22
5	—	—	—	—	—	—	—	—	67.07	62.34	94.51	85.49
6	—	—	—	—	—	—	—	—	68.90	64.17	96.95	84.80
7	—	—	—	—	—	—	—	—	70.73	64.32	—	—

(续 表)

特征数	PCA+LDA		C_LDA+PCA		H_LDA+PCA		DLDA		MMC		MFE	
	训练精度/%	测试精度/%	训练精度/%	测试精度/%	训练精度/%	测试精度/%	训练精度/%	测试精度/%	训练精度/%	测试精度/%	训练精度/%	测试精度/%
8	—	—	—	—	—	—	—	—	81.10	76.39	—	—
9	—	—	—	—	—	—	—	—	82.93	76.55	—	—
10	—	—	—	—	—	—	—	—	86.59	76.47	—	—
11	—	—	—	—	—	—	—	—	89.63	76.85	—	—
12	—	—	—	—	—	—	—	—	89.02	77.01	—	—

从表 4 中可以看到,算法 PCA+LDA,C_LDA+PCA 和 H_LDA+PCA 性能表现比较一致,它们虽然在提取一个特征时训练精度就达到 100%,但测试误差都很大;DLDA 在测试精度上有所提高,但对训练样本的学习并不充分;MMC 的性能随提取特征的增多而上升,但在 12 个特征时测试精度也仅为 77.01%;MFE 在较好地学习了训练样本类别信息的同时,提取 5 个特征时取得最好的测试精度 85.49%,体现出令人满意的学习能力.

此外,对 MFE 算法在较大样本情况下的性能也作了验证,并同 LDA 和 MMC 算法作了比较.表 5 是 3 种方法的分类精度比较.其中,MMC 提取 6 个特征.可以看到,虽然 MFE 是针对小样本问题设计的,但在大样本情况下也能取得较好的性能,与 LDA 相比,它可提取 6 个特征,获得了最好的测试精度.

表 5 较大样本情况 3 种特征提取方法的特征值和分类精度比较

特征数	LDA		MMC		MFE	
	训练精度/%	测试精度/%	训练精度/%	测试精度/%	训练精度/%	测试精度/%
1	68.09	62.22	51.42	50.92	54.98	56.21
2	88.72	82.99	78.05	77.60	75.30	75.66
3	92.89	87.27	87.09	82.99	85.06	83.40
4	—	—	87.09	82.69	88.11	85.54
5	—	—	86.89	82.69	89.74	86.56
6	—	—	86.89	82.69	92.38	89.92

5 结束语

本文提出了一种适用于小样本情况的 MFE 算法.算法利用高维数据小样本情况下线性可分概率增加以及其低维投影趋于正态分布的特点,定义了新的类别边界,不但考虑类内、类间离散度,也兼顾各类方差的差异性,同时又避免了因类内离散度矩阵奇异导致的小样本问题.并进一步将 MFE 算法推广到多类问题,指出对于 C 类数据分类,算法可最多提取 $C(C-1)/2$ 个特征,是基于 LDA 准则算

法的 $C/2$ 倍.高光谱数据特征提取实验表明小样本情况下 MFE 算法具有良好的推广性能,优于多种 LDA 改进算法;并在样本较多时也取得较好结果.

参 考 文 献

[1] Hughes G F. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory, 1968, 14(1): 55-63

[2] Belhumeur P N, Hespanha J, Kriegman D J. Eigenfaces vs. fisherfaces; Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720

[3] Chen L-F, Liao H-Y M, Ko M-T, Lin J-C, Yu G-J. A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognition, 2000, 33(10): 1713-1726

[4] Huang R, Liu Q, Lu H, Ma S. Solving the small sample size problem of LDA//Proceedings of the 16th International Conference on Pattern Recognition. Quebec, Canada, 2002: 29-32

[5] Yu H, Yang J. A direct LDA algorithm for high-dimensional data—with application to face recognition. Pattern Recognition, 2001, 34(10): 2067-2070

[6] Li H, Jiang T, Zhang K. Efficient and robust feature extraction by maximum margin criterion//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2004: 97-104

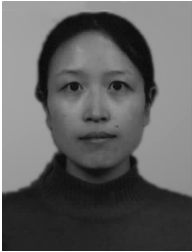
[7] Jimenez L O, Landgrebe D A. High dimensional feature reduction via projection pursuit[Ph. D. dissertation]. Purdue University, Indiana USA, 1995

[8] Bian Zhao-Qi, Zhang Xue-Gong et al. Pattern Recognition (2nd Edition). Beijing: Tsinghua University Press, 2000(in Chinese)
(边肇祺,张学工等.模式识别(第二版).北京:清华大学出版社,2000)

[9] Malina W. On an extended fisher criterion for feature selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1981, 3(5): 611-614

[10] Lee C, Landgrebe D A. Feature extraction based on decision boundaries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(4): 388-400

[11] Xia Jian-Tao, He Ming-Yi. Feature extraction in kernel space using Bhattacharyya distance as criterion function. Chinese Journal of Computers, 2004, 27(5): 683-689(in Chinese)



HE Ming-Yi, born in 1958, professor, Ph. D. supervi-

HUANG Rui, born in 1976, Ph.D.. Her research interests include feature reduction, pattern recognition, neural networks and remote images analysis, etc.

Background

Advances in hyperspectral remote sensing have provided an important means for monitoring the world. The resulting high-dimensional data collected at hundreds of adjoining and narrow wavelengths benefits better discrimination among similar spectral signatures or fingerprints than the traditional multispectral data with low spectral resolution, and have been widely used in aerospace, earth observing, lunar and mars exploration, biomedical engineering etc. However, the vast amount of data volume presents challenging problems for the subsequent information processing. Task-oriented feature extraction has become one of the most important research tasks and attracted more and more attentions.

Feature extraction, transforming the original data from a high dimension into a lower dimension with most of the desired information content preserved, has been widely used for dimensionality reduction and discriminatory information enhancement. However, how to extract useful features is still an open issue.

Linear Discriminant Analysis (LDA) is one of the most

nese)
(夏建涛,何明一. 基于 Bhattacharyya 距离准则的核空间特征提取算法. 计算机学报, 2004, 27(5): 683-689)

sor. His main research interests include neural network and intelligent information processing, image processing, 3D data acquisition and processing.

YANG Shao-Jun, born in 1976, Ph. D.. His research interests include image processing, multimedia communication and intelligent information processing, wireless sensor networks.

popular supervised techniques for feature extraction. It finds the set of the projection vectors which maximize the ratio of between-class scatter against within-class scatter (Fisher's criterion). However, LDA may encounter the so-called small sample size (SSS) problem which arises whenever the number of samples is smaller than the dimensionality of samples. In the case of the SSS problem, the within-class scatter becomes singular and thus LDA fails. In recent years, researchers have proposed different schemes to solve this problem. But most of them ignore the difference of class variances as the LDA criterion does. In view of these, the method proposed for the SSS problem in the paper defines a new margin which involves not only the between-class scatter and within-class scatter, but also the differences of the class variances. Besides, it is further extended to the multi-class case through theoretical analysis. This work is supported by a grant from the National Natural Science Foundation of China (No. 60572097).