第26卷, 第10期
2006年10月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 26, No. 10, pp1960-1964
October, 2006

# 基于广义判别分析的光谱分类

许 馨[1,2], 杨金福[1], 吴福朝[1], 赵永恒[2]

1. 中国科学院自动化研究所国家模式识别实验室, 北京 100080
2. 中国科学院国家天文台, 北京 100012

摘 要 提出了基于广义判别分析(generalized discriminant analysis, GDA)方法对恒星(Star)、星系(Galaxy)和类星体(Quasars)的光谱进行分类。广义判别分析将核技巧与Fisher判别分析结合起来, 通过非线性映射将样本集映射到高维特征空间$F$, 在$F$空间中进行线性判别分析。实验对比了LDA, GDA, PCA, KPCA算法对于恒星、星系和类星体的光谱分类性能。结果表明基于GDA的算法对于这3种类型光谱的分类正确率最高, LDA次之; 尽管KPCA也是一种基于核的方法, 但是选择主成分个数较少时效果较差, 甚至低于LDA; 基于PCA的分类效果最差。

主题词 光谱分类; 广义判别分析; 线性判别分析; 核主成分分析
中图分类号: TP29 文献标识码: A 文章编号: 1000-0593(2006)10-1960-05

## 引 言

天文分类技术具有相对特殊的背景, 目前使用的方法许多是基于PCA的线性的方法[1, 2]和基于神经网络的非线性方法[3]。近年来, 随着统计理论的发展, 基于核技术的学习方法如SVM[4]、核主成分分析(KPCA)[5]和核Fisher判别分析(kernel fisher discriminant, KFD)[6]在模式识别与机器学习领域的优越表现引起了各行业学者的广泛关注。因为核学习方法既可以提取数据的非线性特征, 又有较好的推广能力。KPCA由Scholkopf[5]首先提出, 随后Mika[6]提出了基于核的Fisher判别分析算法(KFD)。KPCA是一种用于描述数据的非线性的特征提取方法, KFD是用于分类的非线性的特征提取方法, 后者更适合分类问题。KFD的主要思想是将样本数据集非线性映射到高维特征空间, 以求数据能够线性可分或近似线性可分, 然后在核特征空间中进行线性Fisher判别分析, 找到最利于分类的投影方向, 从而导致输入空间的非线性判别能力。Baudat等[7]将KFD发展到多类问题, 称之为generalized discriminant analysis(GDA)。

KFD(或者GDA)能够抽取基于分类的非线性特征, 它不像神经网络那样依赖对模型的选择, 而且也不存在维数灾难和局部极小值问题, 因而它在许多实际分类问题中非常有效。在天文光谱的分类问题中, 面对的数据依然是具有非线性特征的, 天体红移和类型的耦合使得我们在做分类时遇到了很大的困难。由此, 我们尝试用基于核技术的广义判别分析(GDA)对于恒星、星系和类星体进行分类, 得到了较好的实验结果。

本文结构如下, 第1部分简单的介绍了多类的线性Fisher判别分析; 第2部分的广义判别分析为基于核技术的多类线性判别分析; 第3部分是KPCA与GDA的联系; 第4部分为实验结果及分析; 第5部分为全文总结。

## 1 多类的线性Fisher判别分析

Fisher线性判别分析的本质在于找到一个子空间, 使得各个类别在这个子空间中能较好地分离, 易于分类器的设计[8]。

假设有一组$N$个$d$维样本的集合$A = \{x_1, x_2, \cdots, x_N \in R^d\} = \bigcup_{j=1}^{c} A_j$, $d \geqslant c$, 其中, $c$为类别数; $A_j$是第$j$类的$n_j$个样本构成的样本子集, $N = \sum_{j=1}^{c} n_j$。

常用的标准Fisher准则函数如下

$$J(W) = |W^{\mathrm{T}} S_{\mathrm{B}} W| / |W^{\mathrm{T}} S_{\mathrm{W}} W| \quad (1)$$

$S_{\mathrm{B}}$, $S_{\mathrm{W}}$分别称为类间散度矩阵和类内散度矩阵

$$S_{\mathrm{B}} = \sum_{k=1}^{c} n_k (\mu_k - \mu)(\mu_k - \mu)^{\mathrm{T}}$$

$$S_{\mathrm{W}} = \sum_{k=1}^{c} \sum_{x_i \in A_k} (x_i - \mu_k)(x_i - \mu_k)^{\mathrm{T}}$$

收稿日期: 2005-08-08, 修订日期: 2005-11-08
基金项目: 国家"863"计划项目(2003AA133060)资助
作者简介: 许 馨, 女, 1974年生, 中国科学院自动化研究所国家模式识别实验室博士研究生

其中，$\mu_k$ 为第 $k$ 类样本的均值，$\mu_k = \frac{1}{n_k}\sum_{x_i \in A_k} x_i$；$\mu$ 为全体样本的均值，$\mu = \frac{1}{N}\sum_{k=1}^{c} n_k \mu_k$。求解使得公式(1)最大化的矩阵 W 的列向量是下列等式中的非零特征值对应的特征向量

$$S_B w_i = \lambda_i S_W w_i \Rightarrow (S_B - \lambda_i S_W) w_i = 0)$$

文献[9]提出扩展的 Fisher 准则函数

$$J(W) = | W^T S_B W | / | W^T S_T W | \qquad (2)$$

$S_T = S_W + S_B$，$S_T$ 称为总散度矩阵。在寻找最优投影方向时，公式(1)和(2)是等价的。

## 2　广义判别分析

广义判别分析(GDA)的思想是通过一个非线性映射，将输入空间的样本映射到高维特征空间，在这个特征空间中进行线性 Fisher 判决分析，从而实现相对于输入空间的非线性判决分析[6, 7]。

令 $\Phi$ 为输入空间到特征空间 $F$ 的非线性映射，即

$$\Phi: R^d \rightarrow F,\ x \Phi \rightarrow \Phi(x)$$

参照准则函数(2)，在 $F$ 空间中进行线性 Fisher 判决，其扩展的准则函数为

$$J(W^\Phi) = \arg\max_{W^\Phi} (| (W^\Phi)^T S_B^\Phi W^\Phi | / | (W^\Phi)^T S_T^\Phi W^\Phi |) \qquad (3)$$

其中，$W^\Phi \in F$；假设数据在 $F$ 空间已经中心化[5]，$S_B^\Phi$ 和 $S_T^\Phi$ 为相应的 $F$ 空间的类间散度矩阵和总散度矩阵

$$S_B^\Phi = \sum_{k=1}^{c} n_k (\mu_k^\Phi)(\mu_k^\Phi)^T$$

$$S_T^\Phi = \frac{1}{N}\sum_{i=1}^{N} \Phi(x_i)\Phi(x_i)^T \qquad (4)$$

其中，$\mu_k^\Phi = \frac{1}{n_k}\sum_{l=1}^{n_k} \Phi(x_{kl})$。

求最优的 $W^\Phi$ 等价于求解下述广义特征值和特征向量问题：

$$\lambda_i S_T^\Phi W_i^\Phi = S_B^\Phi W_i^\Phi,\ i = 1, \cdots, m,$$
$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_2 \geqslant \lambda_m,\ m \leqslant c - 1 \qquad (5)$$

注意到 $W^\Phi$ 的列向量 $W_i^\Phi$ 一定在 $\Phi(x_1), \Phi(x_2), \cdots, \Phi(x_N)$ 张成的空间 $F$ 中，因而存在系数 $\alpha_{ij}$ $(j = 1, \cdots, N)$，使得

$$W_i^\Phi = \sum_{j=1}^{N} \alpha_{ij} \Phi(x_j),\ i = 1, \cdots, m \qquad (6)$$

将公式(5)两边同乘 $\Phi(x_l)^T$，得到

$$\lambda_i \Phi(x_l)^T S_T^\Phi W_i^\Phi = \Phi(x_l)^T S_B^\Phi W_i^\Phi,\ l = 1, 2, \cdots, N \qquad (7)$$

公式(5)与公式(7)有相同的特征向量解。

$F$ 空间的维数非常高，加上 $\Phi$ 是一个隐式的映射，使得数值计算异常困难。为了便于数值计算，将上述公式表达为 $F$ 空间中的内积形式。为此引入核函数[4]

$$k(x, y) = (\Phi(x) \cdot \Phi(y)) \qquad (8)$$

综合(4)，(6)，(8)整理得[7]

$$\lambda K K \alpha = K M K \alpha \qquad (9)$$

其中，$K$ 为 $N \times N$ 的矩阵，$K_{ij} = (\Phi(x_i) \cdot \Phi(x_j))$；$M$ 为 $N \times N$ 的块对角阵，$M = (M_k)$，$k = 1, \cdots, c$。$M_k$ 是元素均为 $1/n_k$ 的 $n_k \times n_k$ 的矩阵。

将 $W_i^\Phi$ 归一化得到：$\alpha_i^T K \alpha_i = 1$。

对于测试数据 $z$，它在 $F$ 空间中的像 $\Phi(z)$ 在 $W_i^\Phi$ 上的投影为

$$(W_i^\Phi \cdot \Phi(z)) = \sum_{j=1}^{N} \alpha_{ij} (\Phi(x_j) \cdot \Phi(z)) = \sum_{j=1}^{N} \alpha_{ij} k(x_j, z) \qquad (10)$$

总结上述讨论，GDA 的主要计算步骤如下。

(1)在 $F$ 空间要找到 Fisher 判决的最优投影方向等价于求公式(5)的特征向量解；

(2)将求解公式(5)的 $W^\Phi$ 转化为求解公式(9)的 $\alpha$；

(3)对于任意测试数据 $z$，根据公式(10)求其在 $F$ 空间中 $W_i^\Phi$ 方向上的投影。

可以看出，不必给出显示的非线性映射 $\Phi$，所有的运算都是通过输入空间中定义的内积核函数来完成的，这就是所谓的核技巧。不同的核函数表示不同的非线性映射 $\Phi$。常用的核函数有高斯核，多项式核和 Sigmoid 核[4]。

## 3　广义判别分析与核主成份分析(KPCA)的联系

文献[10]提出了 KFD(kernel fisher discriminant)的框架：先基于 KPCA 的 $m$ 个主成分得到空间 $R^m$，然后在此空间进行线性判别分析。该文使用标准 Fisher 准则函数

$$J(W^\Phi) = \arg\max_{W^\Phi} (| (W^\Phi)^T S_B^\Phi W^\Phi | / | (W^\Phi)^T S_T^\Phi W^\Phi |)$$

在本小节中，我们使用扩展的准则函数(2)，作为上述框架的补充。

由公式(5)知：若 $S_T^\Phi$ 可逆，则 $W_i^\Phi$ 就是 $(S_T^\Phi)^{-1} S_B^\Phi$ 关于特征值 $\lambda_i$ 的特征向量。实际问题中 $S_T^\Phi$ 经常不可逆。为此有许多正则化方法来解决这个问题。

核 PCA 方法则是找到一组最优的投影方向最大化总散度矩阵，即

$$J_{PCA}(W_{opt}^\Phi) = \arg\max_{W^\Phi} (| (W^\Phi)^T S_B^\Phi W^\Phi | = \arg\max_{W^\Phi} \frac{| (W^\Phi)^T S_T^\Phi W^\Phi |}{| (W^\Phi)^T W^\Phi |}$$

其中 $| (W^\Phi)^T W^\Phi | = I$。$W^\Phi$ 的列向量是互相正交的投影向量，它是 $S_T^\Phi$ 关于非零特征值 $\lambda$ 的特征向量。由此看出，核 PCA 找到的投影方向也去除了 $S_T^\Phi$ 的零空间，并且特征值 $\lambda$ 的大小代表了样本集方差的大小。设 $(\beta_1, \beta_2, \cdots, \beta_m)$ 是 $S_T^\Phi$ 相应的 $m$ 个最大特征值对应的特征向量，$m = \mathrm{rank}(S_T^\Phi)$。对于输入空间的任意 $x$，KPCA 变换后可以得到：$y = P^T \Phi(x)$，$y = (y_1, y_2, \cdots, y_m)^T$，$P = (\beta_1, \beta_2, \cdots, \beta_m)$。定义特征空间 $F$ 中的子空间 $\Theta = span\{\beta_1, \beta_2, \cdots, \beta_m\}$，它的正交补空间记为 $\Theta^\perp$。实际上，$\Theta^\perp$ 就是 $S_T^\Phi$ 的零空间。因此 $S_T^\Phi W_\Theta^{\Phi\perp} = 0$，$F = \Theta + \Theta^\perp$。于是，$F$ 空间中的扩展的 Fisher 准则函数为

$$J(W^\Phi) = \frac{(W^\Phi)^T S_B^\Phi W^\Phi}{(W^\Phi)^T S_T^\Phi W^\Phi} =$$

$$\frac{(W_{\Theta}^{\Phi}+\ W_{\Theta}^{\Phi\perp})^{\mathrm{T}} S_{\mathrm{B}}^{\Phi}(W_{\Theta}^{\Phi}+\ W_{\Theta}^{\Phi\perp})}{(W_{\Theta}^{\Phi}+\ W_{\Theta}^{\Phi\perp})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi}(W_{\Theta}^{\Phi}+\ W_{\Theta}^{\Phi\perp})} \tag{11}$$

由于

$$(W_{\Theta}^{\Phi}+\ W_{\Theta}^{\Phi\perp})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi}(W_{\Theta}^{\Phi}+\ W_{\Theta}^{\Phi\perp}) =$$

$$(W_{\Theta}^{\Phi})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi} W_{\Theta}^{\Phi}+\ 2(W_{\Theta}^{\Phi\perp})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi} W_{\Theta}^{\Phi}+\ (W_{\Theta}^{\Phi\perp})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi} W_{\Theta}^{\Phi\perp} =$$

$$(W_{\Theta}^{\Phi})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi} W_{\Theta}^{\Phi}$$

并且可以证明 $S_{\mathrm{T}}^{\Phi}$ 的零空间必然是 $S_{\mathrm{B}}^{\Phi}$ 的零空间[11]，从而(11)式的分子简化为：$(W_{\Theta}^{\Phi})^{\mathrm{T}} S_{\mathrm{B}}^{\Phi} W_{\Theta}^{\Phi}$。所以公式(11)可以写成

$$J(W^{\Phi}) = (W_{\Theta}^{\Phi})^{\mathrm{T}} S_{\mathrm{B}}^{\Phi}(W_{\Theta}^{\Phi}) / (W_{\Theta}^{\Phi})^{\mathrm{T}} S_{\mathrm{T}}^{\Phi}(W_{\Theta}^{\Phi}) \tag{12}$$

此时可以得出，基于扩展的 Fisher 准则函数的广义判别分析依然可以纳入到文献[10]所提的框架中：在核 PCA 张成的子空间中进行 LDA 可以得到核 Fisher 的投影向量，前提是去掉了 $S_{\mathrm{T}}^{\Phi}$ 的零空间。

## 4 实验结果与分析

将 GDA 算法用于恒星、星系和类星体的自动分类。实验数据来自 SLOAN 的 DR2 的光谱数据库，其中恒星有 1 878条，星系 2 529 条，类星体 3 026 条。光谱波长范围：380~ 900 nm，步长取为 1 nm，插值后每条光谱 521 个点。将整条光谱作为输入参与 GDA 的运算。

为了确定核函数的参数，恒星、星系和类星体光谱各取 1 000 条作为交叉验证样本集(训练样本集)，剩下的作为测试数据集。实验分别对比了线性判别分析(LDA)和 GDA 算法的结果。其中 GDA 的核函数取为高斯核函数：$k(x, Y) = \exp(-\ \| x - y \|^2/\sigma^2)$。因为有三类数据，所以投影方向有 2 个。

图 1 所示的是三类数据基于 LDA 在第 1、第 2 个投影方向上的投影结果。图 2 所示的为三类数据基于 GDA 算法在第 1、第 2 个投影方向上的投影结果。可以看到，对于 LDA，三类数据基本也能分开，但是数据的散度较大；对于 GDA，三类数据的分布更加内敛，散度相对较小。可见 GDA 使得不同类别样本的特征差别加大，而同类样本的特征差别减小，因而 GDA 所提取的特征比 LDA 的特征更接易于分类。

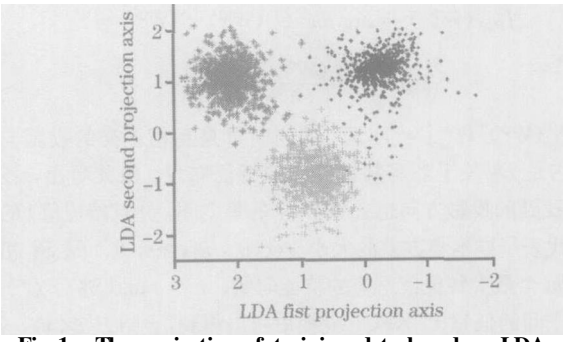


**Fig 1 The projection of training data based on LDA**

用 LDA 和 GDA 分别提取特征后，用最近邻分类器(KNN)完成分类。GDA 对于核宽有一定的敏感度，核宽不同，分类的正确率相差较大。图 3 所示是在训练样本集上 GDA 在不同的核宽下的分类表现。实验按照 5 倍交叉验证完成，核宽范围：0 01~ 10。可以看出，核宽较小时，分类正确率高；核宽增大，分类正确率降低。最好的结果在 $\sigma$= 0 2 时出现，为0 96。相比 LDA 的0 909 可以看出，选择合适的核宽，GDA 的分类性能远高于 LDA。

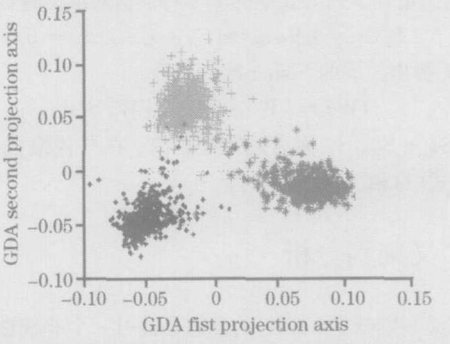


**Fig. 2 The projection of training data based on GDA**

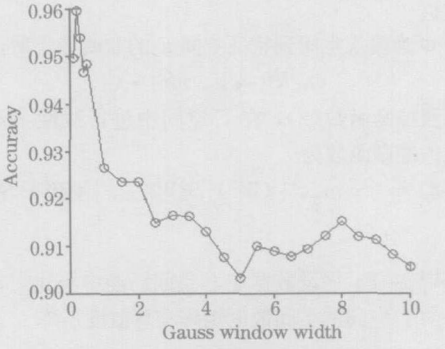


**Fig. 3 Correct classification rate based with GDA on different kernel width**

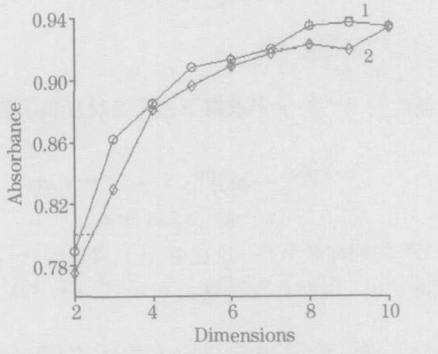


**Fig. 4 Correct classification rate based on PCA and KPCA**

1: KPCA; 2: PCA

同时比较了分别以 PCA 和 KPCA 作为特征提取(KNN 为分类器)在交叉验证样本集(训练样本集)上的分类表现，如图 4 所示。PCA 和 KPCA 分别取了 2~ 10 个主分量(其中 KPCA 的核宽为 2 0)。可以看出，当取 2 个主分量时，基于 PCA 的分类正确率为 0 773，基于 KPCA 的分类正确率为 0 778。这种分类结果比基于 LDA 和 GDA 的方法差很多，但是随着维数的增大，基于 PCA 和 KPCA 方法的分类正确率提升很快，PCA 取六维、KPCA 取五维时，分类正确率已超过了 0 9。基于 KPCA 的分类效果比 PCA 要稍好一些。

对于含有 4 433 条光谱的测试数据集，LDA 的测试光谱

分类的正确率为 0 852, 其中恒星的分类正确率为 0 952, 类星体为 0 815, 星系为 0 85。GDA(σ= 0 2) 的测试正确率为 0. 936, 其中: 恒星的分类正确率为 0. 975, 类星体为 0 94, 星系为 0 906 5(见表 1)。

**Table 1　The correct classification rate based on LDA and GDA**

| 输入类别 | LDA 分类正确率 | | | GDA 分类正确率 | | |
|---|---|---|---|---|---|---|
| | Star | Quasar | Galaxy | Star | Quasar | Galaxy |
| Star | **0 952** | 0 032 | 0 016 | **0 975** | 0. 025 | 0 |
| Quasar | 0 113 | **0 815** | 0 077 | 0 028 | **0. 94** | 0 032 |
| Galaxy | 0 03 | 0 12 | **0 85** | 0 009 8 | 0 083 7 | **0 906 5** |

由表 1 看出, 经过线性 LDA 变换, Quasar 容易错分为 Star; 而 Quasar 和 Galaxy 也容易互相错分; Star 和 Galaxy 不容易互相错分。经过非线性的 GDA 变换后, Star, Quasar 和 Galaxy 的识别率都有增大, 以 Quasar 的识别率提高最多, 比 LDA 的识别率增加了近 14%, 说明 Quasar 对于基于核的线性判别方法非常敏感; Galaxy 的识别率有所上升, 它还是容易被错分为 Quasar; Star 的识别率上升最小; Quasar 和 Star 在非线性变换下不容易错分了; Star 和 Galaxy 的可分性也进一步增强。这充分体现了基于核的判别分析的方法的优越性。

## 5　总　结

近年来核方法在众多学科都得到广泛应用。本节主要讨论了基于核的广义判别分析(GDA) 算法在光谱分类中的应用, 并将基于扩展的 Fisher 准则函数也纳入到 KFD(kernel fisher discriminant) 算法的框架中, 即: 在核 PCA 张成的子空间中进行线性判别分析(LDA) 可以得到核 Fisher 的投影向量, 前提是去掉 $S_t^{\phi}$ 的零空间。实验对比了 LDA, GDA, PCA, KPCA 算法对于恒星、星系和类星体的光谱分类表现。结果是基于 GDA 的算法对于这三种类型光谱的分类正确率最高, LDA 次之; 尽管 KPCA 也是基于核的方法, 但是取低维时其效果不好, 甚至低于 LDA, 基于 PCA 的分类效果最差(低维时)。这也说明有监督的方法在光谱分类问题上比无监督的方法具有优势。

## 参 考 文 献


[1] QIN Dong-mei, HU Zhan-yi, ZHAO Yong-heng(覃冬梅, 胡占义, 赵永恒). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(1): 182.

[2] Cabanac R A, de Lapparent V, Hickson P. Astronomy and Astrophysics, 2002, 389: 1090.

[3] Weaver W B, et al. Astrophysical Journal, 1997, 487: 847.

[4] Vladimir N Vapnik. Translated by ZHANG Xue-gong(张学工译). Nature of Statistical Learning Theory(统计学习理论的本质). Beijing: Tsinghua University Press(北京: 清华大学出版社), 2000.

[5] Scholkopf B, Smola A, Muller K R. Neural Computation, 1998, 10(5): 1299.

[6] Mika S, Ratsch G, Weston J, et al. IEEE International Workshop Neural Networks for Signal Processing IX, 1999, (Aug.): 41.

[7] Baudat G, Anouar F. Neural Computation, 2000, 12(10): 2385.

[8] Richard O Duda, Peter E Hart, David G Stork. Transted by LI Hong-dong, YAO Tian-xiang, et al(李宏东, 姚天翔, 等译). Pattern Classification(Second Edition)(模式识别, 第 2 版). Beijing: China Machine Press(北京: 机械工业出版社), 2003.

[9] Chen Li-fen, et al. Pattern Recognition, 2000, 33(10): 1713.

[10] Yang Jian, Frangi A F, et al. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(2): 230.

[11] Huang R, Liu Q S, Lu H Q, et al. Solving the Small Sample Size Problem of LDA. Proceedings of 16th International Conference on Pattern Recognition, 2002, 3: 29C32.


# Spectra Classification Based on Generalized Discriminant Analysis


XU Xin[1,2], YANG Jin-fu[1], WU Fu-chao[1], ZHAO Yong-heng[2]

1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
2. National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China



**Abstract** A kernel based generalized discriminant analysis (GDA) technique is proposed for the classification of stars, galaxies, and quasars. GDA combines the LDA algorithm with kernel trick, and samples are projected by nonlinear mapping onto the feature space F with high dimensions, and then LDA is conducted in F. Also, it could be inferred that GDA which combines the extension of Fisher's criterion with kernel trick is complementary to kernel Fisher discriminant framework. LDA, GDA, PCA and KPCA were experimentally compared with these three different kinds of spectra. Among these four techniques, GDA obtains the best result, followed by LDA, and PCA is the worst. Although KPCA is also a kernel based technique, its performance is not satisfactory if the selected number of the principal components is small, and in some cases, it appears even worse than LDA, a non-kernel based technique.



**Keywords** Spectra classification; Generalized discriminant analysis; Linear discriminant analysis; Kernel principal component analysis

(Received Aug. 8, 2005; accepted Nov. 8, 2005)