第32卷 第2期
1993年3月

厦门大学学报(自然科学版)
Journal of Xiamen University (Natural Science)

Vol. 32 No. 2
Mar. 1993

最优判别平面法改进及其在癌症诊断中的应用①

朱尔一 邓志威 王小如 杨芃原 黄本立

(化学系)

摘要 采用了一种最优判别平面法的改进方法.在该法中用判别矢量 P_1 和 P_2 的共轭约条件代替原来 P_1 和 P_2 的正交约条件,从而使投影矢量 t_1 与 t_2 正交.在模式判别分析中可得到更好的分类效果.用该方法对癌证诊断问题进行了分析,结果表明有较好的分类效果.

关键词 模式识别,最优判别平面法,癌症诊断

最优判别平面法[1]是统计模式识别中判别两类问题常用的方法.在该法中由类间方差投影与类内方差总和投影之比值 R 取极大时确定第一判别矢量或 Fisher 矢量 P_1,第二判别矢量 P_2 为满足与 P_1 正交的条件取 R 值为极大确定,而二维平面判别图是由投影矢量 $t_1=XP_1$, $t_2=XP_2$ 确定,其中 X 为原始数据矩阵.然而 P_1 与 P_2 正交并不能保证 t_1 与 t_2 正交,因而 t_1 与 t_2 存在一定相关性,t_2 中的信息有一部分与 t_1 中信息重迭,在实际应用中当 t_1 与 t_2 相关性较大时,在判别分析图上的点容易集中到一条线上,当点较多时,图上的点难以分辨.为了克服上述缺点,我们提出了改进最优判别平面法.

本工作中研究实例为癌症病诊断问题,根据人体血液中6种微量元素 Zn, Ba, Mg, Ca, Cu, Se 的含量,判断所测样本是否来自癌症病人.研究中所用数据为本实验室用等离子体原子发射光谱以及石墨炉原子吸收光谱仪测定的结果.以上分析方法的特点是采用了流动注入微量进样技术[2],待测样品的体积仅为 0.25 ml.实验中所用血清样品由厦门大学抗癌中心提供.

1 最优判别平面法

最优判别平面法是研究两类样本判别分析问题的方法.该法中第一判别矢量 P_1 为:

$$P_1=aW^{-1}(m_1-m_2)=aW^{-1}\Delta \tag{1}$$

式中 a 为使 P_1 变为单位矢量的规范常数,W 为类内方差矩阵,Δ 为两类样本均值矢量 m_1 和 m_2 之差.第二判别矢量 P_2 满足与 P_1 正交的条件,并取判别比值 R 为极大值而确定,即

$$P_2^TP_1=0 \tag{2}$$

$$R=P_2^TBP_2/(P_2^TWP_2) \tag{3}$$

上式中上标 T 为矢量的转置,B 为类间方差矩阵.P_2 的算式为

① 本文1992-06-16收到

$$P_2=\beta[W^{-1}-\frac{\Delta^T(W^{-1})^2\Delta}{\Delta^T(W^{-1})^3\Delta}(W^{-1})^2\Delta] \tag{4}$$

式中 β 为规范常数. 将样本投影到矢量 P_1 和 P_2 方向上,可得投影矢量 t_1 和 t_2 构成最优判别平面图.

2 最优判别平面法改进

在改进的最优判别平面法中第一判别矢量 P_1 仍由式(1)确定. 求第二判别矢量 P_2 用矢量 P_1 与 P_2 的共轭约束条件代替原来 P_1 与 P_2 正交约束条件,即用

$$P_2^TX^TXP_1=0 \tag{5}$$

代替式(2),上式为矢量 P_1 与 P_2 关于矩阵 X^TX 相互共轭,上式实际为投影矢量 t_1 与 t_2 正交,即 $t_1^Tt_2=0$,在共轭约束条件下,再取判别比值 R 为极大,以确定 P_2.

下面为用 lagrange 乘子法求上述条件极值问题.

$$P^*=\frac{P_2^TBP_2}{P_2^TWP_2}+\lambda P_2^TX^TXP_1=\frac{(\Delta^TP_2)^2}{P_2^TWP_2}+\lambda P_2^TX^TXP_1$$

式中类间方差阵 $B=\Delta\Delta^T$,λ 为 lagrange 乘子. 令

$$\frac{\partial R^*}{\partial P_2}=\frac{2(\Delta^TP_2)(P_2^TWP_2)\Delta-2(\Delta^TP_2)^2WP_2}{(P_2^TWP_2)^2}+\lambda X^TXP_1=0$$

上式经整理后可得

$$P_2=\beta(P_1+\lambda^*W^{-1}X^TXP_1) \tag{6}$$

式中 β 为规范常数,λ^* 为一常数. 用 $P_1^TX^TX$ 左乘上式的两边,并注意式(5)条件,可得

$\lambda^*=-P_1X^TXP_1/(P_1^TX^TXW^{-1}X^TXP_1)$最后可得判别矢量 P_2 为

$$P_2=\beta(P_1-\frac{P_1^TX^TXP_1}{P_1^TX^TXW^{-1}X^TXP_1}W^{-1}X^TXP_1) \tag{7}$$

3 数据处理与讨论

本工作应用等离子体原子发射光谱仪及石墨炉原子吸收光谱仪共分析 76 个样本,其中癌症病患者样本 43 个,正常人样本 33 个,在数据处理中将正常人样本规定为 1 类样本,癌症病人样本为 2 类样本. 而每一个样本包含 6 个因子,分别为人体血清中微量元素 Zn,Ba,Mg,Ca,Cu,Se 的含量.

用最优判别平面法处理以上数据,得到判别矢量 P_1 和 P_2 为

$$P_1=(-0.016\,0,0.097\,6,-0.011\,1,0.006\,2,-0.044\,9,0.994\,0)^T$$

$$P_2=(-0.155\,5,0.834\,2,-0.090\,1,0.054\,9,-0.507\,0,-0.108\,6)^T$$

所得判别分析图示于图 1. 用改进后的方法对以上同一套数据进行处理,所得 P_1^* 与上面相同,P_2^* 为

$$P_2^*=(-0.283\,0,0.565\,2,-0.158\,6,0.057\,1,-0.706\,0,-0.271\,2)^T$$

所得判别分析图示于图 2.

由图 1 与 2 结果比较可见,用改进后最优判别平面法进行分类的效果明显优于经典最优判别平面法,图 2 中的点较分散,容易辨认,而图 1 中的点较集中于一条线上,其原因为投影矢量 t_1 与 t_2 不正交,有重合部分的信息.

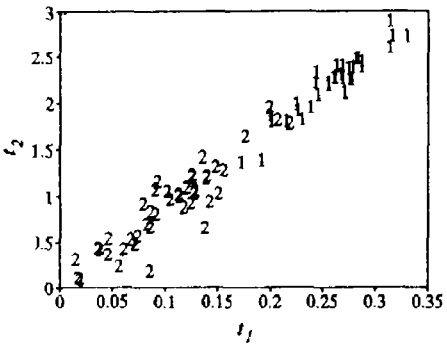

图 1　最优判别平面法映射图

Fig. 1　Mapping results by ODP method

1. normal people, 2. patient with cancer

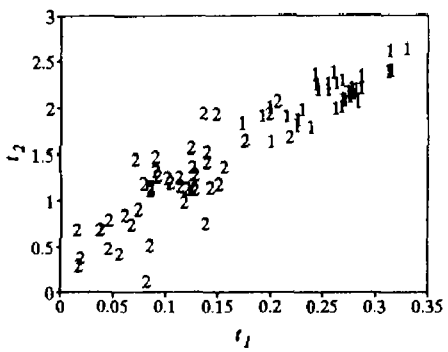

图 2　改进后的方法映射图

Fig. 2　Mapping results by improved ODP method

1. normal, 2. patient with cancer

参 考 文 献

1　李介谷,蔡国康.计算机模式识别技术.上海:上海交通大学出版社,1986

2　Riley C. *Talanta*, 1984,31:879

Improvement of Optimal Discriminant Plane Method and Application in Diagnosis to Cancer

Zhu Eryi　Deng Zhiwei　Wang Xiaoru　Yang Pengyuan　Huang Benli

(*Dept. of Chem.*)

Abstract　An improved method of optimal discriminant plane has been introduced in this paper. In the method, the othogonality between discriminant vectors P_1 and P_2 is replaced with the conjugation between P_1 and P_2, so that the score vectore t_1 and t_2 become othogonal. Therefore, a better classification could be achieved in the discriminant analysis of data. The comparison between the new method and the classical method has been made in the real data treatment for the diagnosis problem of the cancer. As expected, the results have been improved by use of the new method.

Key words　Pattern recognition, Optimal discriminant plane, Cancer diagnosis