

# 基于可分性的多类目标特征选择算法\*

任双桥\*\* 高勋章 黎湘 庄钊文

国防科技大学电子科学与工程学院空间电子信息技术研究所, 长沙 410073

**摘要** 对于多类目标识别, 首先通过扩展类别标示符编码, 给出了特征线性可分定义, 基于 Farkas 引理推导了多类目标线性可分的充要条件. 然后, 借鉴支撑向量机原理, 在特征可分条件的基础上, 定义了特征有效率并提出了一种新的多类目标特征选择方法. 实测数据实验结果表明, 与经典的 Relief 特征选择算法相比, 文中算法的推广性能优于前者. 特别地, 对于小样本数据效果更为明显.

**关键词** 多类目标识别 特征可分性 特征选择 特征有效率 支撑向量机

特征选择是模式识别中的关键技术之一. 一般情况下, 只有在特征矢量中包含了足够的类别信息, 才能通过分类器实现正确分类. 由于特征是否包含足够的信息很难确定, 为了提高识别率, 总是最大限度地提取特征信息, 结果不仅使特征维数增大, 而且其中可能存在较多的无效和冗余特征. 因此, 选择合适的特征来描述模式对模式识别精度、训练时间和存储空间等许多方面都影响较大<sup>[1]</sup>, 并且对分类器的构造也起着非常重要的作用.

特征选择的标准较多, 经典的选择算法大多采用概率度量、距离度量、熵度量、类内距离、类间距离和散布矩阵度量等衡量标准<sup>[1,2]</sup>. 在足够多的样本前提下, 基于大样本统计理论的上述准则和相应的分类算法是合理的, 而当训练样本为小样本时, 上述准则不一定总能取得好的效果<sup>[3]</sup>. 而且, 这些方法虽然应用广泛, 但是, 其识别算法大多是建立在经验模型的基础上<sup>[4]</sup>, 如神经网络等, 其模型参数和结构的确定受数据和识别算法的影响较大. 总之, 经典的特征算法和相应的分类器大多注重的是经验风险最小, 而对分类器的推广性能要求较低<sup>[3]</sup>. 分类器的分类能力和推广性能往往是一对相互制衡的量, 好的分类器和识别算法应能很好地

兼顾这两方面的性能要求<sup>[5,6]</sup>.

Kira 等提出的 Relief<sup>[7,8]</sup>算法是公认的性能较好的二类目标特征选择方法. 该方法通过借鉴最近邻学习的思想, 认为一个好的特征应该使最近邻的同类样本之间特征值相同或相近, 而使最近邻的不同类样本之间值不同或者差别很大. 因此, 若赋予每个特征相应的权值并进行特征排序, 并通过设定特征权阈值或者特征子集的数目即可进行相应的特征选择. Kononenko 对该方法进行了扩展<sup>[9]</sup>, 使得 Relief 方法可以处理多类目标识别.

对于多类目标问题, 本文首先通过扩展类别标示符编码, 给出了特征线性可分的定义及其判别条件. 然后, 基于支撑向量机原理(SVM)探讨了特征选择问题, 以特征有效率为切入点, 提出了一种新的特征选择算法. 该算法秉承了统计学习理论和支撑向量机原理的基础——结构风险最小, 也即要求选择出的特征子集能较好地兼顾分类器的分类能力和推广性能. 实测数据实验结果表明, 相比于经典的 Relief 特征选择算法, 本算法的推广性能要优于前者. 特别地, 对于小样本数据效果更为明显.

2006-03-27 收稿, 2006-09-26 收修改稿

\* 国家自然科学基金(批准号: 60402032)和国防预研项目(批准号: 41303040203)资助

\*\* E-mail: shuangqiaoatr@yahoo.com.cn

# 1 多类目标线性可分条件

## 1.1 线性可分性定义

给定输入空间  $l$  个训练样本  $D = \{(x_i, y_i), i = 1, \dots, l\}$ , 其中,  $x_i \in \mathbb{R}^n$ ,  $y_i \in \{1, \dots, c\}$  为类别标示符,  $c$  为类别总数. 将类别标示符用二进制加以扩展编码, 即令:

$$y_i = (y_{i1}, \dots, y_{id})^T \in \mathbb{R}^d, \quad i = 1, \dots, l$$

其中,  $d = \lceil \log_2 c \rceil$  或  $\lceil \log_2 c \rceil + 1$ ,  $y_{ij} \in \{1, -1\}$ ,  $j = 1, \dots, d$ . 标示符扩展编码方式如图 1 所示.

经过上述类别标示符扩展编码后, 就可以考察多类目标的特征可分性了.

定义 1.1: 设样本集  $D = \{(x_i, y_i), i = 1, \dots, l\}$ , 其中,  $x_i \in \mathbb{R}^n$ ,  $y_i = (y_{i1}, \dots, y_{id})^T \in \mathbb{R}^d$ ,  $y_{ij} \in \{1, -1\}$ , 则称  $D$  线性可分是指, 存在  $w_j \in \mathbb{R}^n$ ,  $b = (b_1, \dots, b_d)^T \in \mathbb{R}^d$ ,  $\rho = (\rho_1, \dots, \rho_d)^T \in \mathbb{R}^d$ ,  $\rho_j > 0$ ,  $j = 1, \dots, d$ , 使得下式成立:

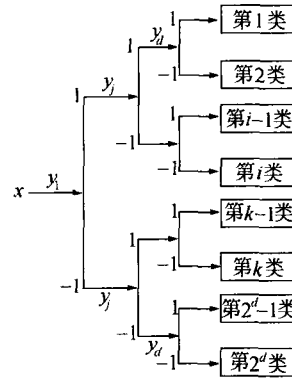


图 1 类别标示符扩展编码示意图

$$y_{ij} (w_j^T x_i + b_j) \geq \rho_j, \quad i = 1, \dots, l, \quad j = 1, \dots, d \quad (1)$$

若不存在满足不等式(1)的  $(w_1, \dots, w_d, b, \rho)$ , 则称多类目标样本集  $D$  线性不可分.

将(1)式写成矩阵形式, 记向量  $\beta = (w_1^T, b_1, \rho_1, \dots, w_d^T, b_d, \rho_d)^T \in \mathbb{R}^{d(n+2)}$ , 矩阵

$$A = \begin{pmatrix} 0 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix}_{d \times (d(n+2))}$$

$$B_i = \begin{pmatrix} y_{i1} x_i^T & y_{i1} & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i2} x_i^T & y_{i2} & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{id} x_i^T & y_{id} & -1 \end{pmatrix}_{d \times (d(n+2))}, \quad i = 1, \dots, l$$

$$B = (B_1^T, \dots, B_l^T)^T$$

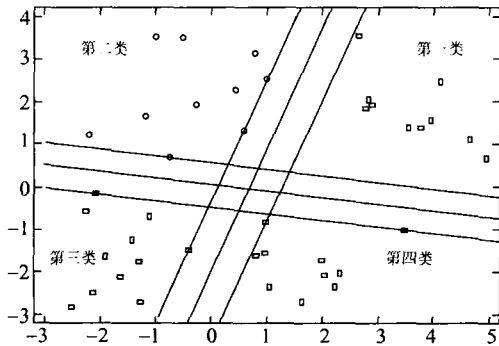


图 2 多类目标样本集线性可分示意图

则有:

$$A\beta > 0 \quad (2)$$

$$B\beta \geq 0 \quad (3)$$

## 1.2 线性可分性条件

根据 Farkas 引理<sup>[10]</sup>, 若  $D$  线性不可分, 则不等式组(2)和(3)无解, 其充要条件是存在  $u \in \mathbb{R}^n$ ,  $v \in \mathbb{R}^p$ ,  $u \geq 0$ ,  $v \geq 0$  且  $u \neq 0$ , 使得  $A^T u + B^T v = 0$ . 化简可得:

$$\sum_{i=1}^l \alpha_{ij} y_{ij} \mathbf{x}_i = \mathbf{0}, \quad \sum_{i=1}^l \alpha_{ij} y_{ij} = 0 \quad (4)$$

$$\sum_{i=1}^l \alpha_{ij} = 1, \alpha_{ij} \geq 0, \quad i = 1, \dots, l, \\ j = 1, \dots, d \quad (5)$$

$$\mathbf{H} = \begin{pmatrix} y_{11} \mathbf{x}_1 & & & y_{21} \mathbf{x}_2 & & \dots & & y_{l1} \mathbf{x}_l \\ & y_{12} \mathbf{x}_1 & & & y_{22} \mathbf{x}_2 & & \dots & & y_{l2} \mathbf{x}_l \\ & & \ddots & & & \ddots & & & \\ & & & y_{1d} \mathbf{x}_1 & & & y_{2d} \mathbf{x}_2 & \dots & & y_{ld} \mathbf{x}_l \end{pmatrix} \quad (6)$$

$$\mathbf{y} = (\text{diag}(y_{11}, \dots, y_{1d}) \dots \text{diag}(y_{l1}, \dots, y_{ld}))^T \in \mathbb{R}^{dl \times d} \quad (7)$$

$$\mathbf{e} = (\text{diag}(1, \dots, 1) \dots \text{diag}(1, \dots, 1))^T \in \mathbb{R}^{dl \times d} \quad (8)$$

则可将不等式组(4), (5)写成如下矩阵形式:

$$\mathbf{G} = \mathbf{Q}^{-1} (\mathbf{E}_{dl \times dl} \mathbf{0}) \mathbf{P}^{-1} \quad (16)$$

$$\mathbf{H}\boldsymbol{\alpha} = \mathbf{0} \quad (9)$$

$$\mathbf{y}^T \boldsymbol{\alpha} = 0 \quad (10)$$

$$\mathbf{e}^T \boldsymbol{\alpha} = 1 \quad (11)$$

$$\mathbf{0} \leq \boldsymbol{\alpha} \leq \mathbf{1} \quad (12)$$

从而, 可得如下样本集  $\mathbf{D}$  线性可分性理论形式的判  
据.

**定理 1.1** (可分性条件) 观测样本集  $\mathbf{D} = \{(\mathbf{x}_i, \mathbf{y}_i), i=1, \dots, l\}$  线性可分的充要条件是不  
等式组(4), (5)无解.

记非负矢量  $\boldsymbol{\alpha} = (\alpha_{11}, \dots, \alpha_{1d}, \dots, \alpha_{l1}, \dots, \alpha_{ld})^T \in \mathbb{R}^{dl}$ ,  
矩阵

由广义逆矩阵可知<sup>[6]</sup>, 不等式组(9)–(12)与下述  
不等式组同解:

$$\mathbf{y}(\mathbf{E}_{dl \times dl} - \mathbf{GH})\mathbf{Z} = \mathbf{0} \quad (13)$$

$$\mathbf{e}^T (\mathbf{E}_{dl \times dl} - \mathbf{GH})\mathbf{Z} = 1 \quad (14)$$

$$(\mathbf{E}_{dl \times dl} - \mathbf{GH})\mathbf{Z} \geq \mathbf{0} \quad (15)$$

其中,  $\mathbf{Z} \in \mathbb{R}^{dl}$ ,  $\mathbf{G}$  为任意给定的矩阵  $\mathbf{H}$  的广义逆.  
由此可得样本集  $\mathbf{D}$  线性可分充要条件是不等式组  
(13), (14), (15)无解.

**推论 1.1** 若矩阵  $\mathbf{H}$  列满秩, 即秩( $\mathbf{H}$ ) =  $dl$ ,  
则样本集  $\mathbf{D} = \{(\mathbf{x}_i, \mathbf{y}_i), i=1, \dots, l\}$  线性可分.

**证明** 由于秩( $\mathbf{H}$ ) =  $dl$ , 故存在可逆矩阵  $\mathbf{P}$ 、  
 $\mathbf{Q}$ , 使得:

$$\mathbf{H} = \mathbf{P} \begin{pmatrix} \mathbf{E}_{dl \times dl} \\ \mathbf{0} \end{pmatrix} \mathbf{Q}$$

从而, 其广义逆矩阵:

齐次方程(9)的解为<sup>[11]</sup>:

$$\boldsymbol{\alpha} = (\mathbf{E}_{dl \times dl} - \mathbf{GH})\mathbf{Z} = \mathbf{0} \quad (17)$$

显然, (17)与(14)式矛盾, 故(9), (10), (11)式  
无解,  $\mathbf{D}$  线性可分. 证毕.

若记广义 Gram 矩阵:

$$\mathbf{K}_G = \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_1 & & \dots & \mathbf{x}_1^T \mathbf{x}_l \\ & \ddots & & \vdots & & \ddots \\ & & \mathbf{x}_l^T \mathbf{x}_1 & \dots & & \mathbf{x}_l^T \mathbf{x}_l \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_l^T \mathbf{x}_1 & & \dots & \mathbf{x}_l^T \mathbf{x}_l \\ & \ddots & & \vdots & & \ddots \\ & & \mathbf{x}_l^T \mathbf{x}_1 & \dots & & \mathbf{x}_l^T \mathbf{x}_l \end{pmatrix} \quad (18)$$

则有,  $\mathbf{H}^T \mathbf{H} = \mathbf{Y}_G^T \mathbf{K}_G \mathbf{Y}_G$ , 其中:

$$\mathbf{Y}_G = \text{diag}(y_{11}, \dots, y_{1d}, \dots, y_{l1}, \dots, y_{ld}) \quad (19)$$

由于  $\mathbf{Y}_G^T \mathbf{Y}_G = \mathbf{E}$ , 因此, 可得如下更实用的结论.

**推论 1.2** 若广义 Gram 矩阵  $\mathbf{K}_G$  正定, 则样本  
集  $\mathbf{D} = \{(\mathbf{x}_i, \mathbf{y}_i), i=1, \dots, l\}$  线性可分.

**证明:** 由于广义 Gram 矩阵  $\mathbf{K}_G$  正定, 且

$H^T H = Y_G^T K_G Y_G$ ,  $Y_G^T Y_G = E$ , 故矩阵  $H^T H$  也是正定矩阵, 从而秩  $(H^T H) = dl$ . 又由于:

$$dl = \text{秩}(H^T H) \leq \text{秩}(H) \leq dl$$

故秩  $(H) = dl$ , 由推论 1.1 可知, 样本集  $D = \{(x_i, y_i), i=1, \dots, l\}$  线性可分. 证毕.

**推论 1.3** 若 Gram 矩阵  $K$  正定, 则样本集  $D = \{(x_i, y_i), i=1, \dots, l\}$  线性可分, 其中, Gram 矩阵  $K = (k_{ij})_{l \times l}$ ,  $k_{ij} = x_i^T x_j$ .

**证明** 设向量  $\alpha = (\alpha_{11}, \dots, \alpha_{1d}, \dots, \alpha_{l1}, \dots, \alpha_{ld})^T \in \mathbb{R}^{dl}$ , 则有:

$$\alpha^T K_G \alpha = \sum_{j=1}^d \sum_{i_1, i_2=1}^l \alpha_{i_1 j} \alpha_{i_2 j} k_{i_1 i_2} = \sum_{j=1}^d \alpha_j^T K \alpha_j$$

其中,  $\alpha_j = (\alpha_{1j}, \dots, \alpha_{lj})^T \in \mathbb{R}^l$ . 由于 Gram 矩阵  $K$  正定, 因此,  $\alpha_j^T K \alpha_j > 0, j=1, \dots, d$ , 于是

$$\alpha^T K_G \alpha > 0$$

也即, 广义 Gram 矩阵  $K_G$  正定. 从而, 由推论 1.2 可知, 样本集  $D$  线性可分. 证毕.

## 2 多类目标特征选择

对于多类问题, 由定理 1.1 可知, 样本集  $D$  线性可分的充要条件是不等式组(4), (5)无解, 即下述不等式组无解:

$$H\alpha = 0 \quad (20)$$

$$y^T \alpha = 0, \quad e^T \alpha = 1 \quad (21)$$

$$0 \leq \alpha \leq 1 \quad (22)$$

由于  $H\alpha = 0$  的充要条件是  $\alpha^T S_G \alpha = 0$ , 这里,  $S_G = H^T H$ . 显然, 对称矩阵  $S_G$  半正定, 从而对任意的  $\alpha \in \mathbb{R}^l$ , 有  $\alpha^T S_G \alpha \geq 0$ . 因此, 构造如下二次凸规划问题:

$$\begin{aligned} \min & \frac{1}{2} \alpha^T S_G \alpha \\ \text{s. t.} & y^T \alpha = 0, \quad e^T \alpha = 1 \\ & 0 \leq \alpha \leq 1 \end{aligned} \quad (23)$$

由定理 1.1 和优化问题(23)可得.

**推论 2.1** 对于多类目标识别, 观测样本集  $D = \{(x_i, y_i), i=1, \dots, l\}$  线性可分的充要条件为二次凸规划问题(23)的目标函数值大于零.

**证明** 必要性. 假设样本集  $D = \{(x_i, y_i), i=1, \dots, l\}$  线性可分, 则不等式组(20), (21), (22)无解. 因此, 对任意满足  $y^T \alpha = 0$  和  $e^T \alpha = 1$  的非负向量  $\alpha$ , 必有  $H\alpha \neq 0$ , 从而可得  $\alpha^T S_G \alpha > 0$ , 故二次凸规划问题(23)的目标函数值大于零. 必要性得证.

充分性. 假设二次凸规划问题(23)的目标函数值大于零, 则对任意满足  $y^T \alpha = 0$  和  $e^T \alpha = 1$  的非负向量  $\alpha$ , 必有  $\alpha^T S_G \alpha > 0$ , 从而  $H\alpha \neq 0$ . 因此, 不等式组(20), (21), (22)无解, 即样本集  $D = \{(x_i, y_i), i=1, \dots, l\}$  线性可分. 充分性得证. 证毕.

由于最优化模型(23)待求参数过多, 若对样本直接求解则计算量过大. 因此, 为使该模型具有实用意义, 必须寻找更为有效的求解途径.

若令  $\alpha_j = (\alpha_{1j}, \dots, \alpha_{lj})^T \in \mathbb{R}^l, y_j = (y_{1j}, \dots, y_{lj})^T \in \mathbb{R}^l, e_j = (1, \dots, 1)^T \in \mathbb{R}^l$ , 则有

$$\frac{1}{2} \alpha^T S_G \alpha = \sum_{j=1}^d \frac{1}{2} \alpha_j^T S_j \alpha_j \quad (24)$$

其中,  $S_j = Y_j^T K Y_j, Y_j = \text{diag}(y_{1j}, \dots, y_{lj})$ . 由此, 二次凸规划问题(23)可等效为

$$\begin{aligned} \min & \sum_{j=1}^d \frac{1}{2} \alpha_j^T S_j \alpha_j \\ \text{s. t.} & y_j^T \alpha_j = 0, \quad e_j^T \alpha_j = 1 \\ & 0 \leq \alpha_j \leq C_j, \quad j=1, \dots, d \end{aligned} \quad (25)$$

由上可知, 原规划问题(23)是对非负乘子  $\alpha = (\alpha_{11}, \dots, \alpha_{1d}, \dots, \alpha_{l1}, \dots, \alpha_{ld})^T$  求最优值, 然而, 由于其目标函数和约束条件均可分离为只对向量  $\alpha_j, j=1, \dots, d$  的求解. 因此, 可以采用序贯方法来寻优, 也即每次寻优仅对  $\alpha_j$  实施, 依次遍历  $d$  次即可完成对  $\alpha$  的求解. 采用序贯法求解不仅可以大幅度降低计算量和节省存储空间. 又由支撑向量机原理<sup>[3]</sup>可知, (25)式中第  $j$  个子规划问题即为二类情形下  $v$ -SVM 模型的对偶形式. 因此, 满足线性方程(20)和(21)的非负权值  $\alpha$ , 其实质就是  $d$  个二类  $v$ -SVM 对偶问题中的非负 Lagrange 乘子.

下面来考察凸规划问题(25)目标函数值的性质. 假设每个观测样本由  $n$  个特征构成, 即  $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T \in \mathbb{R}^n, i=1, \dots, l$ , 其中  $x_{ij}$  为第  $j$  个特征. 由于

$$\mathbf{S}_n = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \cdots & \mathbf{S}_{1l} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \cdots & \mathbf{S}_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{l1} & \mathbf{S}_{l2} & \cdots & \mathbf{S}_{ll} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1 \mathbf{X}_{11} \mathbf{Y}_1 & \mathbf{Y}_1 \mathbf{X}_{12} \mathbf{Y}_2 & \cdots & \mathbf{Y}_1 \mathbf{X}_{1l} \mathbf{Y}_l \\ \mathbf{Y}_2 \mathbf{X}_{21} \mathbf{Y}_1 & \mathbf{Y}_2 \mathbf{X}_{22} \mathbf{Y}_2 & \cdots & \mathbf{Y}_2 \mathbf{X}_{2l} \mathbf{Y}_l \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}_l \mathbf{X}_{l1} \mathbf{Y}_1 & \mathbf{Y}_l \mathbf{X}_{l2} \mathbf{Y}_2 & \cdots & \mathbf{Y}_l \mathbf{X}_{ll} \mathbf{Y}_l \end{pmatrix} \quad (26)$$

其中,  $\mathbf{S}_{ij} = \mathbf{Y}_i \mathbf{X}_{ij} \mathbf{Y}_j$ ,

$$\mathbf{X}_{ij} = \begin{pmatrix} \mathbf{x}_i^T \mathbf{x}_j & & \\ & \ddots & \\ & & \mathbf{x}_i^T \mathbf{x}_j \end{pmatrix}_{d \times d} \quad (27)$$

令

$$\mathbf{S}_n = \sum_{k=1}^n \mathbf{S}_n^k \quad (28)$$

其中,  $\mathbf{S}_n^k = (\mathbf{S}_{ij}^k)_{d \times d}, \mathbf{S}_{ij}^k = \mathbf{Y}_i \mathbf{X}_{ij}^k \mathbf{Y}_j$ ,

$$\mathbf{X}_{ij}^k = \begin{pmatrix} x_{ik} x_{jk} & & \\ & \ddots & \\ & & x_{ik} x_{jk} \end{pmatrix}_{d \times d} \quad (29)$$

则有,  $\mathbf{S}_n^k$  半正定,  $k=1, \dots, n$ . 令

$$J(n) = \frac{1}{2} \mathbf{a}_n^T \mathbf{S}_n \mathbf{a}_n = \frac{1}{2} \sum_{k=1}^n \mathbf{a}_n^T \mathbf{S}_n^k \mathbf{a}_n \quad (30)$$

其中,  $\mathbf{a}_n$  为(25)式的解.

**定理 2.1**  $J(n)$  具有单调递增性, 即  $J(n) \leq J(n+1)$ .

**证明** 假设观测样本  $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T \in \mathbb{R}^n$  时规划问题(25)的解为  $\mathbf{a}_n$ , 当样本增加一个特征时, 即  $\mathbf{x}_i = (x_{i1}, \dots, x_{in}, x_{in+1})^T \in \mathbb{R}^{n+1}$ , 规划问题(25)的解为  $\mathbf{a}_{n+1}$ . 则有

$$\begin{aligned} J(n+1) &= \frac{1}{2} \mathbf{a}_{n+1}^T \mathbf{S}_{n+1} \mathbf{a}_{n+1} = \frac{1}{2} \sum_{k=1}^{n+1} \mathbf{a}_{n+1}^T \mathbf{S}_{n+1}^k \mathbf{a}_{n+1} \\ &\geq \frac{1}{2} \sum_{k=1}^n \mathbf{a}_{n+1}^T \mathbf{S}_{n+1}^k \mathbf{a}_{n+1} = \frac{1}{2} \sum_{k=1}^n \mathbf{a}_n^T \mathbf{S}_n^k \mathbf{a}_n \\ &\geq \frac{1}{2} \sum_{k=1}^n \mathbf{a}_n^T \mathbf{S}_n^k \mathbf{a}_n = J(n) \end{aligned}$$

其中, 第一个不等式是由于  $\mathbf{S}_{n+1}^k \geq 0$ , 第二个不等式是由于  $J(n)$  为规划问题(25)的最优目标函数值. 证毕.

**定理 2.2** 假设两个观测样本集  $\mathbf{D}_n = \{(\mathbf{u}_i, \mathbf{y}_i), i=1, \dots, l\}$ ,  $\mathbf{D}_m = \{(\mathbf{v}_i, \mathbf{y}_i), i=1, \dots, l\}$ , 其中,  $\mathbf{u}_i \in \mathbb{R}^n, \mathbf{v}_i \in \mathbb{R}^m$ . 令样本集  $\mathbf{D}_{n+m} = \{(\mathbf{x}_i, \mathbf{y}_i), i=1, \dots, l\}$ ,  $\mathbf{x}_i = (\mathbf{u}_i^T, \mathbf{v}_i^T)^T \in \mathbb{R}^{n+m}$ , 则有

$$J(n) + J(m) \leq J(n+m)$$

其中,  $J(n), J(m), J(n+m)$  分别为样本集  $\mathbf{D}_n, \mathbf{D}_m, \mathbf{D}_{n+m}$  所对应的规划问题(25)的最优目标函数值.

**证明** 由于  $\mathbf{x}_i = (\mathbf{u}_i^T, \mathbf{v}_i^T)^T \in \mathbb{R}^{n+m}$ , 则有

$$\begin{aligned} \mathbf{S}_{n+m} &= \mathbf{H}_{n+m}^T \mathbf{H}_{n+m} = \begin{pmatrix} \mathbf{Y}_1 \mathbf{X}_{11} \mathbf{Y}_1 & \cdots & \mathbf{Y}_1 \mathbf{X}_{1l} \mathbf{Y}_l \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_l \mathbf{X}_{l1} \mathbf{Y}_1 & \cdots & \mathbf{Y}_l \mathbf{X}_{ll} \mathbf{Y}_l \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Y}_1 (\mathbf{U}_{11} + \mathbf{V}_{11}) \mathbf{Y}_1 & \cdots & \mathbf{Y}_1 (\mathbf{U}_{1l} + \mathbf{V}_{1l}) \mathbf{Y}_l \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_l (\mathbf{U}_{l1} + \mathbf{V}_{l1}) \mathbf{Y}_1 & \cdots & \mathbf{Y}_l (\mathbf{U}_{ll} + \mathbf{V}_{ll}) \mathbf{Y}_l \end{pmatrix} \end{aligned}$$

其中,

$$\begin{aligned} \mathbf{U}_{ij} &= \begin{pmatrix} \mathbf{u}_i^T \mathbf{u}_j & & \\ & \ddots & \\ & & \mathbf{u}_i^T \mathbf{u}_j \end{pmatrix}_{d \times d}, \\ \mathbf{V}_{ij} &= \begin{pmatrix} \mathbf{v}_i^T \mathbf{v}_j & & \\ & \ddots & \\ & & \mathbf{v}_i^T \mathbf{v}_j \end{pmatrix}_{d \times d}. \end{aligned}$$

因此,  $\mathbf{S}_{n+m} = \mathbf{S}_n + \mathbf{S}_m$ . 从而

$$\begin{aligned}
 J(n+m) &= \frac{1}{2} \alpha_{n+m}^T S_{n+m} \alpha_{n+m} = \\
 &\frac{1}{2} \alpha_{n+m}^T S_n \alpha_{n+m} + \frac{1}{2} \alpha_{n+m}^T S_m \alpha_{n+m} \geq \\
 &\frac{1}{2} \alpha_n^T S_n \alpha_n + \frac{1}{2} \alpha_m^T S_m \alpha_m = \\
 &J(n) + J(m)
 \end{aligned}$$

其中,  $\alpha_n, \alpha_m, \alpha_{n+m}$  分别为  $D_n, D_m, D_{n+m}$  所对应的规划问题(25)的最优解. 证毕.

**定理 2.3** 假设两个观测样本集  $D_n = \{(u_i, y_i), i=1, \dots, l\}$ ,  $D_m = \{(v_i, y_i), i=1, \dots, l\}$ , 其中,  $u_i \in \mathbb{R}^n, v_i \in \mathbb{R}^m$ . 若  $J(n) \leq J(m)$ , 则有  $\Delta_n^* \leq \Delta_m^*$ , 其中,  $J(n), J(m)$  分别为样本集  $D_n, D_m$  所对应的规划问题(25)的最优目标函数值

$$\Delta_n^* = \sqrt{\sum_{j=1}^d (\Delta_j^n)^2}, \quad \Delta_m^* = \sqrt{\sum_{j=1}^d (\Delta_j^m)^2}$$

$\Delta_j^n, \Delta_j^m$  分别为样本集  $D_n$  和  $D_m$  的第  $j$  个最优分类间隔.

**证明** 不妨假设  $J(n) > 0$ , 若不然, 由推论 2.1 可知, 结论显然成立. 因此, 样本集  $D_n$  和  $D_m$  均线性可分. 由于(25)式中第  $j$  个子规划问题即为二类情形下  $\nu$ -SVM 模型的对偶形式, 则其原规划模型为<sup>[3]</sup>

$$\begin{aligned}
 \min & \frac{1}{2} \|w_j\|^2 - \rho_j \\
 \text{s. t.} & y_i (w_j^T x_i + b_j) \geq \rho_j \\
 & \rho_j \geq 0, \quad i = 1, \dots, l
 \end{aligned} \quad (31)$$

其中,  $w_j$  为特征空间中第  $j$  个最优分类超平面的法矢,  $b_j$  为第  $j$  个最优分类面的平移分量. 从而, (25)式中第  $j$  个子规划问题目标函数值<sup>[3]</sup>

$$\frac{1}{2} \alpha_j^T S_j \alpha_j = \frac{1}{2} \|w_j\|^2, \quad j = 1, \dots, d$$

且, 样本集  $D_n$  和  $D_m$  的第  $j$  个最优分类间隔分别为<sup>[3]</sup>

$$\Delta_j^n = \|w_j^n\|, \quad \Delta_j^m = \|w_j^m\| \quad (32)$$

其中,  $w_j^n, w_j^m$  分别为样本集  $D_n$  和  $D_m$  关于  $\nu$ -SVM

模型(31)的最优解. 设  $\alpha_n$  为样本集  $D_n$  关于规划问题(25)的最优解, 则有

$$J(n) = \frac{1}{2} \alpha_n^T S_n \alpha_n = \frac{1}{2} \sum_{j=1}^d \alpha_j^T S_j \alpha_j = \frac{1}{2} \sum_{j=1}^d \|w_j^n\|^2 \quad (33)$$

因此, 由(32)式和(33)式可知, 样本集  $D_n$  的最优分类间隔为

$$\Delta_n^* = \sqrt{2J(n)} \quad (34)$$

同理, 线性可分样本集  $D_m = \{(v_i, y_i), i=1, \dots, l\}$  的最优分类间隔为

$$\Delta_m^* = \sqrt{2J(m)} \quad (35)$$

由于  $J(n) \leq J(m)$ , 因此  $\Delta_n^* \leq \Delta_m^*$ . 证毕.

**定理 2.4** 若样本集  $D_n = \{(u_i, y_i), i=1, \dots, l\}$  线性不可分, 且设  $u_i = (u_{i1}, \dots, u_{in})^T \in \mathbb{R}^n$ ,  $u_{ij}$  为第  $j$  个特征, 则任意剔除  $m$  个特征后组成新的样本集仍然线性不可分. 若样本集  $D_n$  线性可分, 则增加任意  $m$  个新特征后组成新的样本集仍然线性可分, 且其分类间隔不会下降.

**证明** 假设样本集  $D_n = \{(u_i, y_i), i=1, \dots, l\}$  线性不可分, 则由推论 2.1 可知:

$$J(n) = \frac{1}{2} \alpha_n^T S_n \alpha_n = 0$$

则任意剔除  $m > 0$  个特征后, 由定理 2.1 可知:

$$0 \leq J(n-m) \leq J(n) = 0$$

从而,  $J(n-m) = 0$ , 也即新的样本集线性不可分. 同理, 假设样本集  $D_n = \{(u_i, y_i), i=1, \dots, l\}$  线性可分, 则有  $J(n) > 0$ , 增加任意  $m > 0$  个新特征后, 则有  $0 < J(n) \leq J(n+m)$ , 也即新的样本集线性可分.

当新增  $m > 0$  个特征后, 假设样本集  $D_{n+m} = \{(x_i, y_i), i=1, \dots, l\}$ ,  $x_i = (u_i^T, v_i^T)^T \in \mathbb{R}^{n+m}$ , 则由  $J(n) \leq J(n+m)$  和定理 2.3 可知,  $\Delta_n^* \leq \Delta_{n+m}^*$ . 因此, 当样本新增  $m$  个特征后, 其最优分类间隔随特征维数的增加而递增. 证毕.

综上所述,对于观测样本集  $D = \{(x_i, y_i), i=1, \dots, l\}$ , 二次凸规划问题(25)的目标函数值  $J(n)$  具备了作为类别可分性判据的下述基本条件.

(1)  $J(n)$  与误判概率的界有单调关系.

根据统计学习理论<sup>[3]</sup>, 当采用超平面进行分类识别时, 期望风险和测试样本误判概率的界均为分类间隔的单调减函数. 又由于最优分类间隔  $\Delta_n^* = \sqrt{2J(n)}$ , 从而, 期望风险和测试样本误判概率的界均为  $J(n)$  的单调减函数, 也即  $J(n)$  与误判概率的界有单调关系.

(2)  $J(n)$  具有不完全可加性.

(3) 由于  $J(n) = \frac{1}{2} \sum_{j=1}^d \|w_j^n\|^2$ , 因此,  $J(n)$  具有“距离”的某些共性.

(4)  $J(n)$  关于特征维数是单调递增的.

由于  $J(n)$  以上4个基本条件, 因此, 可以将其作为类别可分性判据<sup>[1]</sup>. 假设样本集  $D$  线性可分, 则  $J(n) = \frac{1}{2} \alpha_n^T S_n \alpha_n = \frac{1}{2} \sum_{k=1}^n \alpha_n^T S_n^k \alpha_n > 0$ , 从而, 可定义第  $k$  个特征的有效率为:

$$\gamma_k = \frac{\alpha_n^T S_n^k \alpha_n}{\alpha_n^T S_n \alpha_n}, \quad k = 1, \dots, n$$

根据 SVM 原理<sup>[3]</sup>,  $w_j^n = \sum_{i=1}^l \alpha_{ij}^n y_{ij} x_i$ ,  $\sum_{j=1}^d \|w_j^n\|^2 = \alpha_n^T S_n \alpha_n$ . 因此:

$$\gamma_k = \frac{\sum_{j=1}^d |\tau w_j^n(k)|^2}{\sum_{j=1}^d \|w_j^n\|^2}, \quad k = 1, \dots, n \quad (36)$$

其中,  $w_j^n = (w_j^n(1), \dots, w_j^n(n))^T$ . 由(36)式可知, 第  $k$  个特征的有效率  $\gamma_k$  的实质就是  $d$  个最优分类超平面法矢  $\{w_j^n\}_{j=1}^d$  的第  $k$  个分量  $\{w_j^n(k)\}_{j=1}^d$  占其总能量  $\sum_{j=1}^d \|w_j^n\|^2$  的百分比. 又由于最优分类间隔  $\Delta_n^* = \sqrt{2J(n)} = \sqrt{\sum_{j=1}^d \|w_j^n\|^2}$ , 因此,  $\gamma_k$  很好的刻画了第  $k$  个特征对最优分类间隔  $\Delta_n^*$  的贡献, 也即有效率.

下面考察在一些特殊情形下特征的有效率, 以验证(36)式可以很好地吻合直观感知. 假设第  $k_1$  个

特征和第  $k_2$  个特征满足如下线性关系:

$$x_{ik_1} = \theta x_{ik_2} + \sigma, \quad i = 1, \dots, l \quad (37)$$

不妨假设,  $|x_{ik_1}| \leq |x_{ik_2}|$ , 也即  $|\theta| \leq 1$ . 则有

$$\begin{aligned} w_j^n(k_1) &= \sum_{i=1}^l \alpha_{ij}^n y_{ij} x_{ik_1} = \\ \theta \sum_{i=1}^l \alpha_{ij}^n y_{ij} x_{ik_2} &= \theta w_j^n(k_2) \end{aligned} \quad (38)$$

从而可得

$$\gamma_{k_1} = \theta^2 \gamma_{k_2} \quad (39)$$

因此, 由(39)式可知, 当第  $k_1$  和第  $k_2$  个特征完全线性相关时, 其特征的有效率是以  $\theta^2$  下降的. 由于大多数分类器均采用距离度量, 因此, 数值较小的特征提供的有效信息应不如数值较大的特征. 特别是当  $\theta=0$  时, 则其有效率  $\gamma_{k_1}=0$ , 这一结论也符合客观事实. 总之, 由这些结论可知, (36)式定义的特征有效率与直观感知的吻合度是很好的.

假设二次凸规划问题(25)的目标函数值  $J(n)=0$ , 也即观测样本集  $D$  线性不可分. 此时, 由线性不可分情形下  $\nu$ -SVM 原理<sup>[3]</sup>可知, 只须对 Lagrange 乘子  $\alpha$  的边界约束条件进行进一步限制, 即要求,  $0 \leq \alpha \leq C$ . 其中,  $C$  为正则化参数, 可设定  $0 < C < 1$ , 也即在分类间隔和误判概率之间做一个折中以确保  $J(n) > 0$ . 总之, 无论样本集  $D$  是否线性可分, 均可以根据特征有效率的大小进行特征的剔除和增加.

### 3 仿真实验及结果分析

#### 例1 给定样本集

$$D = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 & -1 \\ 1 & 1 & 2 & 3 & 3 \end{pmatrix}$$

其中, 矩阵  $D$  的第一和第二行为5个二维样本点, 第三行为各自的类别标示符. 证明, 样本集  $D$  在原空间中线性不可分, 而选用二次多项式核函数  $K(u, v) = (1 + u^T v)^2$  后, 则在此核函数导出的高维特征空间中线性可分.

**证明** 首先给类别标示符扩展编码, 由于有三

类目标, 取  $d=[lbc]+1=2$ , 令第一类的标识符为  $(1, 1)$ , 第二类的标识符为  $(-1, 1)$ , 第三类的标识符为  $(-1, -1)$ . 根据定理 1.1, 当取  $\alpha=(0.25 \ 0.25 \ 0.25 \ 0.25 \ 0 \ 0 \ 0.25 \ 0.25 \ 0.25 \ 0.25)^T$  时, 则不等式组 (4), (5) 成立, 因此, 样本集  $D$  在原空间中线性不可分.

如果选用二次多项式  $K(u, v)=(1+u^T v)^2$  作为核函数, 则其 Gram 矩阵为

$$K = \begin{pmatrix} 9 & 1 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 1 & 9 \end{pmatrix}$$

因为 Gram 矩阵  $K$  的秩为 5, 故  $K$  为正定矩阵. 由推论 1.3 可知, 样本集在二次多项式核函数导出的高维特征空间中线性可分. 图 3 是样本集  $D$  在二次多项式核函数下基于 SVM 的实验结果, 由图可知, 样本集  $D$  已被完全正确地分类识别.

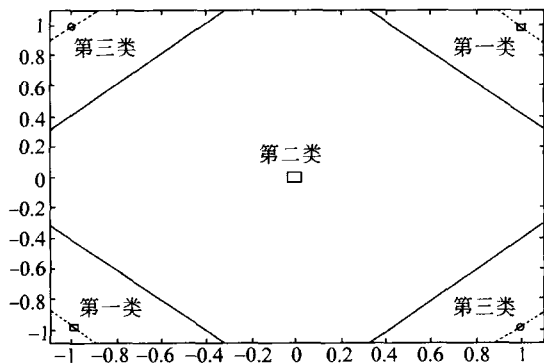


图 3 二阶多项式核函数分类结果

**例 2** 本组数据为五类目标缩微模型的微波暗室测量数据, 五类目标分别为 F-117、幻影-2000、J-8 机、无人驾驶飞机和运输机模型, 共有 155 个样本, 样本维数为 500. 雷达信号为阶梯变频形式, 起始频段为 34.7 GHz, 带宽为 1 GHz, 频率间隔为 2 MHz, 极化方式采用水平-垂直极化. 测量范围为  $0^\circ-30^\circ$ , 方位偏转角间隔为  $1^\circ$ . 实验中每类目标均间隔  $2^\circ$  选取一个训练样本, 共 55 个, 其余的为测试样本共 100 个. 特征的选择是依据特征有效率  $\gamma$  的大小从 15 开始依次递增特征维数, 并采用线性

v-SVM 方法对样本集进行分类识别, 实验结果如图 4 所示.

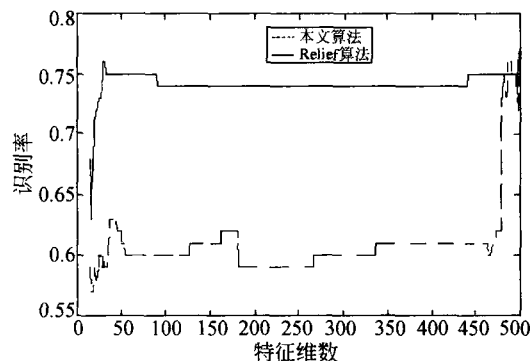


图 4 特征维数与测试样本分类正确率

由图 4 可以看出, 当特征维数等于 32 时, 测试分类结果最好, 测试率为 76%, 选出的特征为 [232 233 234 235 237 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 265 266]. 当全部特征都选用时, 测试率为 75%, 相对于最优特征集的测试率下降了 1%. 而由经典的 Relief 算法选择的最优特征子集维数为 499, 其相应的最高测试率为 77%. 由于经典的 Relief 算法考察的是特征之间的相关性而没有太多地注重算法的推广性能, 使得对于同样的特征维数, 本文算法的推广性能较经典的 Relief 算法要优, 实验结果也验证了这个结论.

#### 4 结束语

对于多类目标识别, 本文首先讨论了样本空间中特征可分性, 通过扩展类别标示符编码, 给出了特征线性可分的定义及其判别条件. 然后, 基于 SVM 原理探讨了特征选择问题, 以特征有效率为切入点, 提出了一种新的特征选择算法. 最后, 通过实测数据实验进一步验证了该特征选择算法比经典的 Relief 算法具有更好的推广性能.

#### 参 考 文 献

- 1 孙即祥著. 现代模式识别. 长沙: 国防科技大学出版社, 2002
- 2 范劲松, 方廷健. 特征选择和提取要素的分析及其评价. 计算机工程与应用, 2001, 31(13): 95--99
- 3 Vladimir N. Vapnik 著, 统计学习理论. 许建华, 张学工译. 北京: 电子工业出版社, 2004



- 4 Deon Garrett, David AP, Charles WA, et al. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans on Neural Systems and Rehabilitation Engineering*, 2003, 11(2): 141—144
- 5 Nojun K, Chong HC. Input feature selection for classification problems. *IEEE Transactions on Neural Networks*, 2002, 13(1): 143—159
- 6 Huan Liu, Lei Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(4): 491—502
- 7 Kira K, Rendell L. The feature selection problem: Traditional methods and a new algorithm. *Proceedings of the Ninth National Conference on Artificial Intelligence*. New Orleans: AAAI Press, 1992, 129—134
- 8 张丽新, 王家, 赵雁南, 等. 基于 Relief 的组合式特征选择. *复旦学报(自然科学版)*, 2004, 43(5): 893—898
- 9 Kononenko. Estimating attributes: Analysis and extension of Relief. *Proceedings of European Conference on Machine Learning*, 1994, 171—182
- 10 粟塔山著. 最优化计算原理与算法程序设计. 长沙: 国防科技大学出版社, 2001
- 11 北京大学数学系几何与代数教研室代数小组编. 高等代数. 北京: 北京高等教育出版社, 1995

(上接第 505 页)

4. 从工程科学角度研究: (1) 工程质量与工程安全; (2) 故障自愈率原理; (3) 矿山工程与安全。

通过本期论坛, 与会专家学者在如下方面形成共识:

(1) 突发公共事件应急管理中的科学问题涉及管理科学、信息科学、工程科学与生命科学等多学科领域, 是典型的跨层次、跨部门和综合性很强的问题, 需要不同学科间开拓、交叉、渗透与融合, 为解决此类问题的关键技术提供新的思路、理论和方法。

(2) 虽然我国在突发公共事件应急管理实践和应用层次的部分领域开展了研究工作, 已经有相当的研究基础, 但学界对应急/危机形成机理与演化规律还缺乏基础理论层次的深刻认知, 尚未形成系统的知识体系, 无法对应急科技的可持续发展和提升提供有效支撑, 急需开展有针对性的基础研究工作。

(3) 突发公共事件种类较多, 应重点关注各类突发公共事件应急管理中的共性和综合性需求, 兼顾现实, 虚实结合, 抽取、提炼其中基础性的科学问题, 开展前瞻性、开放性和原始创新性的基础科学研究, 可以结合典型和迫切需要解决的实际问题开展研究, 然后再研究跨领域和跨部门的问题。

(4) 突发公共事件应急管理中科学问题的基础研究, 应涵盖应急管理全过程, 在关注应急响应阶段的基础科学问题的同时, 也要重视突发公共事件应急预防、准备、恢复重建和后评估方面的研究。

(5) 突发公共事件应急管理中科学问题的解决是一项长期工作, 应注重与其他部门在类似的科技项目加强协调, 单个计划项目不可能解决所有问题. 需通过与其他项目的沟通与协调, 实现整体、系统、科学地解决相关问题。

与会专家学者也对今后如何开展针对性的工作提出建议和期望。

(1) 建议国家自然科学基金委员会针对“国家突发公共事件应急管理中的科学问题”设立一个重大研究计划. 与会专家学者一致认为国内已有必要的相关工作基础和研究积累, 对解决其中重大科学前沿问题充满信心, 期望管理科学部协调其他科学部继续围绕这个主题, 组织相关领域的研究人员, 提炼出交叉、综合和关键性的科学问题, 争取重大研究计划尽快在国家自然科学基金中成功立项。

(2) 建议在组织重大研究计划的过程中, 立项框架要体现出科学问题的结构性和层次性, 从而能够体现出总体设计的特性. 建议设计出整个工作的路线图, 对流程的层层设计, 保证各项目能形成一个有机的整体。

(3) 建议管理科学部以及其他相关科学部在正常的项目资助中, 加强对有关突发公共事件应急管理科学问题的面上项目和重点项目的资助。

(供稿: 杨列勋 邓云峰)