

# 分片支撑矢量机

任双桥 杨德贵 黎 湘 庄钊文

(国防科学技术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** 文中借鉴了分段线性识别的基本思想,提出了分片支撑矢量机模型.该模型首先将特征空间剖分成若干子空间,在每个子空间中基于支撑矢量机构造一个最优分类面,然后,将各个分类面链接起来构成一个分片最优分类面以逼近理论上的最优分类超曲面.同时,文中还从理论上分析探讨了其推广能力的界,为分片支撑矢量机模型提供了坚实的基础.最后,经典双螺旋线数据实验结果表明,相对于传统支撑矢量机,分片支撑矢量机的计算速度、分类能力以及推广能力均有了明显提高.

**关键词** 分段线性;支撑矢量机;空间剖分;分片;识别

中图法分类号 TP18

DOI号: 10.3724/SP.J.1016.2009.00077

## Piecewise Support Vector Machines

REN Shuang-Qiao YANG De-Gui LI Xiang ZHUANG Zhao-Wen

(School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073)

**Abstract** A novel piecewise support vector machines (PSVM) model is provided in this paper, which used the traditional piecewise linear recognition method for reference. In this new model, the feature space was firstly partitioned into several subspaces, then the piecewise classification surface was developed by linking the optimal classification surfaces in each subspaces based on SVM. As the experimental results shown, validated with the Two Spiral Data and the actually measuring data, the performance of PSVM such as computational efficiency, classification capability and generalized capability are improved obviously in contrast to SVM.

**Keywords** piecewise linear; SVM; space partition; piecewise; recognition

## 1 引 言

线性支撑矢量机以其结构简单和计算量低而备受学者的关注,但遗憾的是线性支撑矢量机的分类能力和推广能力均为有限,这在一定程度上限制了其适应范围.特别地,对于比较复杂的识别问题,如两类样本分布区域凹凸交错,或其中一类的类域不是单连通的而另一类却嵌入其中,当观测样本并非线性可分时,用线性超平面分类识别就可能会导致

很大的误识率.为了解决复杂的线性不可分识别问题,支撑矢量机引入了核函数技巧,通过某种非线性映射将原空间映入到一个高维的特征空间中,使得在新的特征空间中目标线性可分,从而可以在特征空间中构造一个线性超平面分类识别.

核函数的引入虽然巧妙地解决了线性不可分而非线性可分的识别问题<sup>[1]</sup>,但核函数自适应选择与构造一直是一个难题<sup>[2-3]</sup>,而且非线性支撑矢量机本质上也是一个线性分类器,与线性支撑矢量机的差别就在于是在高维特征空间而非原空间中构造最优

分类超平面. 因此, 当样本分布的区域十分散乱和高度交错时, 非线性支撑矢量机的分类能力和推广能力均会受到很大的挑战<sup>[4]</sup>.

在传统模式识别理论中<sup>[5]</sup>, 分段线性识别是介于线性识别和非线性识别之间一个较好的折中方案 (见图 1). 分段线性识别函数是一种特殊的非线性识别函数, 其分类界面是由若干个超平面组成. 因此, 与一般的非线性超曲面相比, 分段线性分类面依然是比较简单的, 而且能逼近各种形状的超曲面, 柔性地通过两类之间的空间, 具有较强的适应性与分类能力.

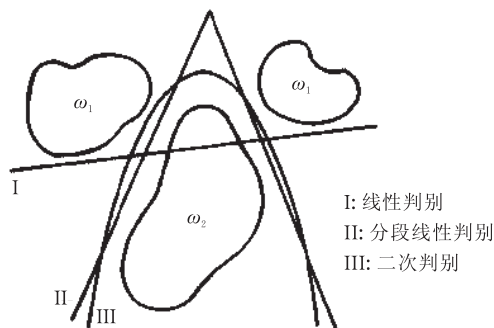


图 1 分段线性识别原理示意图

本文借鉴了分段线性模式识别的基本思想, 提出了一种新的支撑矢量机模型——分片支撑矢量机

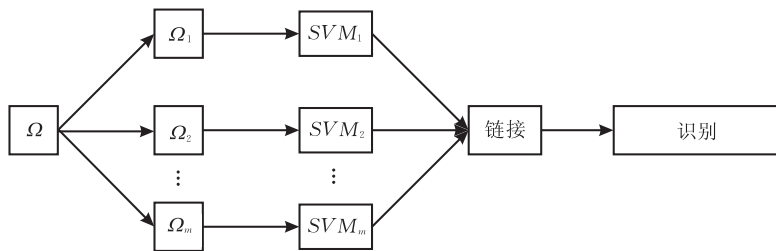


图 2 分片支撑矢量机原理图

由图 2 可知, 分片支撑矢量机是通过剖分原空间  $\Omega$ , 然后在每个子空间上基于传统的 SVM 构造局部最优分类超曲面, 进而对  $m$  个局部最优分类超曲面加以链接以形成最终的全空间分类超曲面. 显然, 紧空间  $\Omega$  的剖分情形直接影响到分片支撑矢量机的识别性能和推广能力, 这是十分重要的一个关键环节, 我们将在随后的内容中加以讨论.

在对观测样本测试时, 分片支撑矢量机的决策函数为

$$f(x) = f_j(x) = \sum_{i=1}^{l_j} \alpha_{ij} y_{ij} K(x, x_{ij}) + b_j, \quad x \in \Omega_j \quad (1)$$

其中,  $\alpha_{ij}$  为求解第  $j$  个支撑矢量机的非负 Langrange 乘子,  $b_j$  为第  $j$  个最优分类面的平移分量,  $K(\cdot)$  为

(Piecewise SVM, PSVM). 该模型首先将特征空间划分成若干子空间, 在每个子空间中基于支撑矢量机原理构造一个最优分类面, 然后, 将各个分类面链接起来构成一个分片最优分类超曲面以逼近理论上的最优分类超曲面. 此外, 本文还从理论上分析探讨了其推广能力的界, 为分片支撑矢量机模型提供了坚实的基础. 最后, 经典双螺旋线数据和实测数据实验结果表明, 相对于传统支撑矢量机, 分片支撑矢量机的计算速度、分类能力以及推广能力均有了明显提高.

## 2 分片支撑矢量机

### 2.1 分片支撑矢量机原理

本文主要探讨二类目标识别问题, 对于多类目标可以很方便地推广. 给定原输入空间中 1 个观测样本  $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$ , 其中,  $x_i \in \mathcal{R}^n$ ,  $y_i \in \{1, -1\}$  为类别标示符. 假设样本观测值  $x \in \Omega \subset \mathcal{R}^n$ ,  $\Omega$  为  $\mathcal{R}^n$  中紧子集, 在紧空间  $\Omega$  中对其实施空间剖分,

$$\Omega = \bigcup_{j=1}^m \Omega_j,$$

其中, 子空间  $\Omega_i \cap \Omega_j = \emptyset, i \neq j$ , 则分片支撑矢量机的基本原理可以用图 2 来表示.

选择的核函数,  $y_{ij}$  为观测样本  $x_{ij}$  的类别标示符,  $l_j$  为子空间  $\Omega_j$  中的训练样本数,  $x_{ij}$  为子空间  $\Omega_j$  中第  $i$  个训练样本. 由式(1)可知, 分片支撑矢量机测试时是分区域识别的, 也即在不同的区域采用相应的最优分类曲面来识别目标.

### 2.2 分片支撑矢量机推广能力

由于分片支撑矢量机的实质是通过剖分原空间, 增加决策函数的个数分区域识别, 也即, 令决策函数

$$f(x) = \sum_{j=1}^m I_{\Omega_j}(x) f_j(x) \quad (2)$$

其中,  $f_j(x)$  为子空间  $\Omega_j$  中由 SVM 构造的最优决策函数,  $I_{\Omega}(\cdot)$  为示性函数, 即

$$I_{\Omega}(x) = \begin{cases} 1, & x \in \Omega \\ 0, & x \notin \Omega \end{cases} \quad (3)$$

为了探讨更一般的分区域识别算法的推广性能,可假设训练样本集  $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$  是依据联合分布  $F(x, y) = F(x)F(y|x)$  随机独立地抽取出来的,对于任意一个决策函数

$$f(x, \alpha) = \sum_{j=1}^m I_{\Omega_j}(x) f_j(x, \alpha) \quad (4)$$

其中,参数  $\alpha \in \Lambda$ ,  $f_j(x, \alpha)$  为子空间  $\Omega_j$  中任意决策函数,若记  $z = (x, y)$ ,则定义  $f(x, \alpha)$  的决策风险函数为

$$\begin{aligned} Q(z, \alpha) &= \sum_{j=1}^m I_{\Omega_j}(z) Q_j(z, \alpha) \\ &= \sum_{j=1}^m I_{\Omega_j}(x) L(y_j, f_j(x, \alpha)) \end{aligned} \quad (5)$$

其中,  $Q_j(z, \alpha) = L(y_j, f_j(x, \alpha))$ ,  $L(\cdot, \cdot)$  为损失函数,对于目标识别问题,可以考虑最简单的 0-1 损失. 则期望风险与经验风险可分别定义为

$$R(\alpha) = \int Q(z, \alpha) dF(z) \quad (6)$$

$$R_{\text{emp}}(\alpha) = \frac{1}{l} \sum_{i=1}^l Q(z_i, \alpha) \quad (7)$$

假设期望风险  $R(\alpha)$  在函数  $Q(z, \alpha_0)$  上取得最小值,经验风险  $R_{\text{emp}}(\alpha)$  在函数  $Q(z, \alpha_l)$  上取得最小值. 为了探讨函数集  $Q(z, \alpha)$ ,  $\alpha \in \Lambda$  的推广性能,必须寻找期望风险  $R(\alpha)$  的界以及  $R(\alpha_l)$  和  $R(\alpha_0)$  之间的差. 在下述的讨论中,需要用到一个基本的引理.

**引理 1**<sup>[6]</sup>. 若事件  $\bigcap_{i=1}^n A_i \subset A$ , 则有  $P(A) \geq \sum_{i=1}^n P(A_i) - n + 1$ .

基于引理 1, 可以很方便地分析分片支撑矢量机识别算法的推广能力.

**定理 1.** 对于函数集  $Q(z, \alpha)$ ,  $\alpha \in \Lambda$ , 下述不等式至少以  $1 - 2m\eta$  的概率成立.

$$R(\alpha) \leq R_{\text{emp}}(\alpha) \left[ 1 + \sqrt{\frac{-l \ln \eta}{2}} \sum_{j=1}^m \frac{1}{l_j} \right] + \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \sqrt{\frac{h_j (\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} + \frac{1}{l},$$

$$R(\alpha) \leq R_{\text{emp}}(\alpha) \left[ 1 + \sqrt{\frac{-l \ln \eta}{2}} \sum_{j=1}^m \frac{1}{l_j} \right] + \frac{1}{2} \sum_{j=1}^m \epsilon_j(l) \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \left[ 1 + \sqrt{1 + \frac{4R_{\text{emp}}(\alpha)}{l_j \epsilon_j(l)}} \right],$$

其中,  $\Omega = \bigcup_{j=1}^m \Omega_j$ ,  $l_j$  为子空间  $\Omega_j$  中的样本数,  $h_j$  为函

数集  $Q_j(z, \alpha)$  的 VC 维,

$$\epsilon_j(l) = 4 \frac{h_j (\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}.$$

证明. 由于期望风险

$$R(\alpha) = \int Q(z, \alpha) dF(z) \quad (8)$$

将式(5)代入式(8), 可得

$$\begin{aligned} R(\alpha) &= \sum_{j=1}^m \int_{\Omega_j} Q_j(z, \alpha) dF(z) \\ &= \sum_{j=1}^m P(\Omega_j) \int_{\Omega_j} Q_j(z, \alpha) dF_j(z) \end{aligned} \quad (9)$$

其中, 联合分布  $F_j(z) = \frac{F(z)}{P(\Omega_j)} I_{\Omega_j}(z)$ . 若记

$$R_j(\alpha) = \int_{\Omega_j} Q_j(z, \alpha) dF_j(z) \quad (10)$$

则有,  $R(\alpha) = \sum_{j=1}^m P(\Omega_j) R_j(\alpha)$ , 也即总的期望风险为各子空间期望风险的加权平均. 根据统计学习理论可知<sup>[7]</sup>, 下述不等式至少以  $1 - \eta$  的概率成立.

$$R_j(\alpha) \leq R_{\text{emp}}^j(\alpha) + \sqrt{\frac{h_j (\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} + \frac{1}{l} \quad (11)$$

$$R_j(\alpha) \leq R_{\text{emp}}^j(\alpha) + \frac{\epsilon_j(l)}{2} \left[ 1 + \sqrt{1 + \frac{4R_{\text{emp}}^j(\alpha)}{\epsilon_j(l)}} \right] \quad (12)$$

$$P(\Omega_j) \leq \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \quad (13)$$

其中, 经验风险  $R_{\text{emp}}^j(\alpha) = \frac{1}{l_j} \sum_{i=1}^{l_j} Q_j(z_{ij}, \alpha)$ . 则由引理 1 和式(11)、(12)可知, 不等式

$$R(\alpha) \leq \sum_{j=1}^m P(\Omega_j) R_{\text{emp}}^j(\alpha) + \sum_{j=1}^m P(\Omega_j) \sqrt{\frac{h_j (\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} + \frac{1}{l},$$

$$R(\alpha) \leq \sum_{j=1}^m P(\Omega_j) R_{\text{emp}}^j(\alpha) + \frac{1}{2} \sum_{j=1}^m P(\Omega_j) \epsilon_j(l) \left[ 1 + \sqrt{1 + \frac{4R_{\text{emp}}^j(\alpha)}{\epsilon_j(l)}} \right],$$

至少以  $1 - m\eta$  的概率成立. 由于  $R_{\text{emp}}(\alpha) = \sum_{j=1}^m \frac{l_j}{l} R_{\text{emp}}^j(\alpha)$ , 则由引理 1 和式(13)可知,

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \sqrt{\frac{-\ln \eta}{2l}} \sum_{j=1}^m R_{\text{emp}}^j(\alpha) + \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \sqrt{\frac{h_j (\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} + \frac{1}{l},$$

$$R(\alpha) \leq R_{\text{emp}}(\alpha) + \sqrt{\frac{-\ln \eta}{2l}} \sum_{j=1}^m R_{\text{emp}}^j(\alpha) + \frac{1}{2} \sum_{j=1}^m \varepsilon_j(l) \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \left[ 1 + \sqrt{1 + \frac{4R_{\text{emp}}^j(\alpha)}{\varepsilon_j(l)}} \right],$$

至少以  $1 - 2m\eta$  的概率成立.

又由于  $R_{\text{emp}}(\alpha) = \sum_{j=1}^m \frac{l_j}{l} R_{\text{emp}}^j(\alpha)$ , 因此,

$$\frac{l_j}{l} R_{\text{emp}}^j(\alpha) \leq R_{\text{emp}}(\alpha), \quad j=1, 2, \dots, d \quad (14)$$

则有

$$\sum_{j=1}^m R_{\text{emp}}^j(\alpha) \leq \sum_{j=1}^m \frac{l}{l_j} R_{\text{emp}}(\alpha) \quad (15)$$

从而由式(14)、(15)可知, 不等式

$$R(\alpha) \leq R_{\text{emp}}(\alpha) \left[ 1 + \sqrt{\frac{-\ln \eta}{2}} \sum_{j=1}^m \frac{1}{l_j} \right] + \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \sqrt{\frac{h_j(\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} + \frac{1}{l},$$

$$R(\alpha) \leq R_{\text{emp}}(\alpha) \left[ 1 + \sqrt{\frac{-\ln \eta}{2}} \sum_{j=1}^m \frac{1}{l_j} \right] + \frac{1}{2} \sum_{j=1}^m \varepsilon_j(l) \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \left[ 1 + \sqrt{1 + \frac{4lR_{\text{emp}}(\alpha)}{l_j \varepsilon_j(l)}} \right],$$

至少以  $1 - 2m\eta$  的概率成立.

证毕.

定理 1 给出了经验风险  $R_{\text{emp}}(\alpha)$  依概率一致收敛和相对一致收敛到期望风险  $R(\alpha)$  的收敛速率的上界. 下面将给出  $\Delta(\alpha_l) = R(\alpha_l) - R(\alpha_0)$  的上界, 其中, 期望风险  $R(\alpha)$  在函数  $Q(z, \alpha_0)$  上取得最小值, 经验风险  $R_{\text{emp}}(\alpha)$  在函数  $Q(z, \alpha_l)$  上取得最小值. 在此, 需要对参数  $\alpha \in \Lambda$  做一个必要的假设, 也即令决策函数

$$f(x, \alpha) = \sum_{j=1}^m I_{\Omega_j}(x) f_j(x, \alpha_j) \quad (16)$$

其中, 参数  $\alpha = (\alpha_1, \dots, \alpha_j) \in \Lambda$ . 另外, 在对期望风险  $R(\alpha)$  和经验风险  $R_{\text{emp}}(\alpha)$  求最小值的过程中, 假设对参数  $\alpha = (\alpha_1, \dots, \alpha_m)$  可以在  $m$  个子空间中分别解耦求解. 因此, 若假设子空间  $\Omega_j$  中期望风险  $R_j(\alpha) = \int_{\Omega_j} Q_j(z, \alpha) dF_j(z)$  在函数  $Q_j(z, \alpha_j^0)$  上取得最小值,

而其经验风险  $R_{\text{emp}}^j(\alpha) = \frac{1}{l_j} \sum_{i=1}^{l_j} Q_j(z_{ij}, \alpha)$  在函数  $Q_j(z, \alpha_j^l)$  上取得最小值,  $j=1, 2, \dots, m$ , 则全空间  $\Omega$  中总的期望风险  $R(\alpha) = \sum_{j=1}^m P(\Omega_j) R_j(\alpha)$  和经验风险

$R_{\text{emp}}(\alpha) = \sum_{j=1}^m \frac{l_j}{l} R_{\text{emp}}^j(\alpha)$  将分别在函数集  $Q(z, \alpha_0)$  和

$Q(z, \alpha_l)$  上取得最小值, 其中

$$\alpha_0 = (\alpha_1^0, \dots, \alpha_d^0) \in \Lambda \quad (17)$$

$$\alpha_l = (\alpha_1^l, \dots, \alpha_d^l) \in \Lambda \quad (18)$$

关于式(16)~(18)的假设, 一方面可以较为方便地讨论  $\Delta(\alpha_l) = R(\alpha_l) - R(\alpha_0)$  的上界, 而另一方面是因为考虑到识别算法的实用性. 对于一般的分区域识别模型式(3), 若其参数  $\alpha \in \Lambda$  不能在各子空间上解耦分离求解, 则不管是采用经验风险最小原则还是结构风险最小原则, 其实用性是较差的.

**定理 2.** 假设参数  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \Lambda$ , 且可以在各子空间中解耦求解, 则有, 对于函数集  $Q(z, \alpha)$ , 下述不等式至少以  $1 - 3m\eta$  的概率成立.

$$\Delta(\alpha_l) = R(\alpha_l^l) - R(\alpha_j^0) \leq \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \cdot \left[ \frac{1}{l} + \sqrt{\frac{-\ln \eta}{2l}} + \sqrt{\frac{h_j(\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} \right],$$

$$\Delta(\alpha_l) = R(\alpha_j^l) - R(\alpha_j^0) \leq \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \cdot \left[ \sqrt{\frac{-\ln \eta}{2l}} + \frac{\varepsilon_j(l)}{2} \left[ 1 + \sqrt{1 + \frac{4lR_{\text{emp}}(\alpha_l)}{l_j \varepsilon_j(l)}} \right] \right],$$

其中,  $R(\alpha)$  和  $R_{\text{emp}}(\alpha)$  分别在函数集  $Q(z, \alpha_0)$  和  $Q(z, \alpha_l)$  上取得最小值,  $\alpha_0 = (\alpha_1^0, \alpha_2^0, \dots, \alpha_d^0)$ ,  $\alpha_l = (\alpha_1^l, \alpha_2^l, \dots, \alpha_d^l)$ ,  $h_j$  为函数集  $Q_j(z, \alpha)$  的 VC 维,

$$\varepsilon_j(l) = 4 \times \frac{h_j(\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}.$$

证明. 根据假设式(16)~(18), 期望风险

$$R_j(\alpha) = \int_{\Omega_j} Q_j(z, \alpha) dF_j(z) \text{ 和经验风险 } R_{\text{emp}}^j(\alpha) =$$

$\frac{1}{l_j} \sum_{i=1}^{l_j} Q_j(z_{ij}, \alpha)$  分别在函数集  $Q_j(z, \alpha_j^0)$  和  $Q_j(z, \alpha_j^l)$  上取得最小值. 因此, 由统计学习理论可知<sup>[7]</sup>, 下述不等式至少以  $1 - 2\eta$  的概率成立

$$\Delta_j(\alpha_l) = R_j(\alpha_j^l) - R_j(\alpha_j^0) \leq \frac{1}{l} + \sqrt{\frac{-\ln \eta}{2l}} + \sqrt{\frac{h_j(\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} \quad (19)$$

$$\Delta_j(\alpha_l) = R_j(\alpha_j^l) - R_j(\alpha_j^0) \leq \sqrt{\frac{-\ln \eta}{2l}} + \frac{\varepsilon_j(l)}{2} \left[ 1 + \sqrt{1 + \frac{4R_{\text{emp}}^j(\alpha_j^l)}{\varepsilon_j(l)}} \right] \quad (20)$$

其中,  $j=1, 2, \dots, m$ . 由于  $R(\alpha) = \sum_{j=1}^m P(\Omega_j) R_j(\alpha)$ , 且不等式

$$P(\Omega_j) \leq \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \quad (21)$$

至少以  $1 - \eta$  的概率成立. 则由引理 1 和式 (19)、(21) 可知, 不等式

$$\Delta(\alpha_i) = R(\alpha_i^j) - R(\alpha_i^0) \leq \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \cdot \left[ \frac{1}{l} + \sqrt{\frac{-\ln \eta}{2l}} + \sqrt{\frac{h_j(\ln(2l/h_j) + 1) - \ln(\eta/4)}{l}} \right],$$

至少以  $1 - 3m\eta$  的概率成立. 又由于  $R_{\text{emp}}(\alpha) = \sum_{j=1}^m \frac{l_j}{l} R_{\text{emp}}^j(\alpha)$ , 因此,

$$\frac{l_j}{l} R_{\text{emp}}^j(\alpha) \leq R_{\text{emp}}(\alpha), \quad j = 1, 2, \dots, m \quad (22)$$

从而由引理 1 和式 (20)~(22) 可知, 不等式

$$\Delta(\alpha_i) = R(\alpha_i^j) - R(\alpha_i^0) \leq \sum_{j=1}^m \left[ \frac{l_j}{l} + \sqrt{\frac{-\ln \eta}{2l}} \right] \cdot \left[ \sqrt{\frac{-\ln \eta}{2l}} + \frac{\epsilon_j(l)}{2} \left[ 1 + \sqrt{1 + \frac{4l R_{\text{emp}}(\alpha_i)}{l_j \epsilon_j(l)}} \right] \right],$$

至少以  $1 - 3m\eta$  的概率成立. 证毕.

定理 2 描述了在经验风险最小化原则下, 经验风险最小值  $R_{\text{emp}}(\alpha_i)$  关于期望风险最小值  $R(\alpha_0)$  的接近程度. 由统计学习理论可知<sup>[7]</sup>, 假设  $x \in \mathcal{R}^n$  含有一个半径为  $r$  的超球中, 则分类间隔为  $\Delta$  的超平面集合的 VC 维  $h$  以下述不等式为界

$$h \leq \min \left( \left\lceil \frac{r^2}{\Delta^2} \right\rceil, n \right) + 1 \quad (23)$$

显然, 覆盖每个子空间  $\Omega_i$  中样本的超球半径均会小于覆盖全空间  $\Omega$  中样本的超球半径, 而且, 每个子空间  $\Omega_i$  中的样本分类间隔至少不小于全空间  $\Omega$  中的样本分类间隔. 因此, 通过剖分原空间  $\Omega$  可以有效地降低决策函数集  $Q_j(z, \alpha)$  的 VC 维  $h_j$ , 从而由定理 1 和定理 2 可知, 可以有效地提高分区识别模型的推广能力和识别性能. 此外, 由于全空间上的期望风险和经验风险均为各子空间上的期望风险和经验风险的加权平均, 因此, 相对于传统的支撑矢量机而言, 分片支撑矢量机在分类性能和推广能力上均有了一定的增强.

### 2.3 分片支撑矢量机计算量

经理论分析和仿真实验证明, 二类目标 SVM 的训练时间  $t$  大致满足如下关系<sup>[8]</sup>.

$$t = al^u \quad (24)$$

其中,  $a$  为一常数,  $l$  为训练样本的个数,  $u$  的大小与训练算法有关, 当采用 SMO<sup>[8]</sup> 时,  $u \approx 2$ .

由于分片支撑矢量机在训练过程中, 可以分区域在各子空间  $\Omega_j$  上解耦求解最优化问题. 因此, 若假设每个子空间的训练样本数相同, 则每个 SVM

的训练样本有  $\frac{l}{m}$  个, 从而, 其总的训练时间为

$$t = ma \left( \frac{l}{m} \right)^u = am^{1-u} l^u \quad (25)$$

特别地, 当采用 SMO 算法时, 也即令  $u = 2$ , 则有

$$t = \frac{al^2}{m} \quad (26)$$

由式 (26) 可知, 当采用 SMO 时, 分片支撑矢量机相对于传统的支撑矢量机, 其训练速度可以提高  $m$  倍. 因此, 当剖分的子空间数越多时, 计算量将越

小, 训练速度越快, 但是, 由定理 1 可知, 由于  $l_j = \frac{l}{m}$ ,  $j = 1, 2, \dots, m$ , 则有

$$\sqrt{\frac{-\ln \eta}{2}} \sum_{j=1}^m \frac{1}{l_j} = m^2 \sqrt{\frac{-\ln \eta}{2l}} \quad (27)$$

因此, 随着子空间数  $m$  越大, 期望风险的界也越大, 而且经验风险与期望风险的差距也越大, 这就有可能在很大程度上降低分片支撑矢量机的分类性能和推广能力. 因此, 关于子空间的剖分, 必须兼顾训练速度与分类器的性能. 一般地, 可以通过求解下届最优化问题确定剖分子空间数目.

$$\min \gamma m^2 \sqrt{\frac{-\ln \eta}{2l}} + \frac{al^2}{m} \quad (28)$$

其中,  $\gamma$  为正正则化参数. 则求解最优化问题式 (28), 可得

$$m = \left( \frac{a}{\gamma \sqrt{-2\ln \eta}} \right)^{\frac{1}{3}} l^{\frac{5}{6}} \quad (29)$$

特别地, 若令正则化参数  $\gamma = \frac{al}{\sqrt{-2\ln \eta}}$ , 则有

$$m = \sqrt{l} \quad (30)$$

将式 (30) 代入式 (26), 可得分片支撑矢量机的训练时间为

$$t = al^{\frac{3}{2}} \quad (31)$$

为了更形象地对比 SVM 和 PSVM 的训练时间, 可假定  $a = 1, u = 2$ , 图 3 为实验结果.

由图 3 可知, 随着训练样本数的增加, 分片支撑矢量机相对于传统的支撑矢量机, 其训练时间急剧下降, 较好地解决了支撑矢量机计算量过大这一瓶颈问题.

总之, 通过借鉴传统模式识别中的分段线性识别方法, 分片支撑矢量机首先将原紧空间  $\Omega$  剖分成  $m$  个子空间, 然后在每个子空间上基于传统的 SVM 构造局部最优分类超曲面, 进而对  $m$  个局部最优分

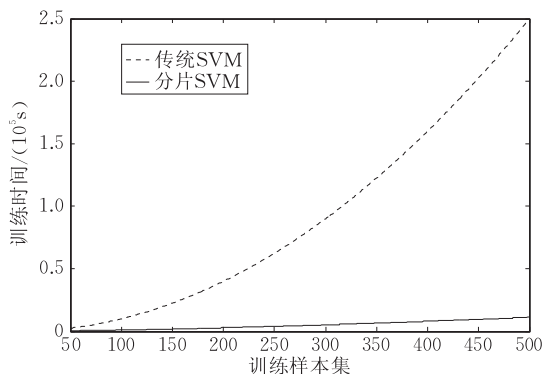


图3 传统 SVM 与分片支撑矢量机训练速度对比

类超曲面加以链接以形成最终的全空间分类超曲面. 分片支撑矢量机可以在各子空间上解耦求解, 可以很方便地应用于实际工程背景. 另外, 在训练速度、分类性能和推广能力上, 分片支撑矢量机相对于传统的支撑矢量机都有了一定的提高.

### 3 特征空间剖分

由图1可知, 对紧空间  $\Omega$  的剖分情形直接影响到分片支撑矢量机的识别性能和推广能力, 不恰当的空间剖分准则可能会导致分片支撑矢量机完全失效. 本文通过借鉴仿生模式识别原理<sup>[9]</sup>, 提出了如下特征连续性空间剖分准则.

传统模式识别理论与方法的基本出发点都是考察若干类目标之间的最佳分类划分, 其根本的原因或许在于采用这样的数学描述与其相应的处理方法最具有—般性和通用性. 对此, 文献[9]另辟蹊径提出了一种仿生模式识别理论, 该识别理论的基本出发点就在于把模式识别问题看成是模式的“认识”, 而不是分类划分, 是一类样本的“认识”, 而不是多类样本的划分.

仿生模式识别方法指出<sup>[9]</sup>, 自然界任何欲被认识的事物若存在两个同源同类而不完全相等的事物, 且这两个事物的差别是可以渐变的或非量子化的, 则这两个事物之间必定至少存在一个渐变过程, 在这个渐变过程中间的各事物都是属于同一类的. 在特征空间中, 同类样本点之间所存在的这个连续性规律超出了传统模式识别与学习的基本假定, 该假定认为“可用的信息都包含在训练集中”. 由于特征连续性规律是客观世界人类直观认识范围存在的客观规律, 因而可以作为仿生模式识别方法中关于样本点分布的先验信息加以利用, 用来提高对事物的认识能力. 因此, 仿生模式识别是把特征空间训练

样本之间的关系作为基点, 而特征空间中样本分布的连续性规律为此提供了可靠的基石.

由于同类事物的特征具有的连续性, 而不同类事物的特征之间存在一个渐变过程, 因此, 对于观测样本  $D = \{(x_i, y_i), i = 1, 2, \dots, l\}$ , 构造某个特征

$$t_x = \varphi(x) \quad (32)$$

比如可令  $t_x = \varphi(x) = \|x\|$ , 则特征  $t_x$  的连续性为全空间  $\Omega$  的剖分提供了一条新的思路. 假设  $t_0 \leq t_x \leq t_m$ , 则给出区间  $[t_0, t_m]$  上  $m-1$  个节点,

$$t_0 < t_1 < \dots < t_{m-1} < t_m \quad (33)$$

区间节点可以采用等距划分, 即令

$$t_{j+1} = t_j + \Delta t \quad (34)$$

其中, 步长  $\Delta t = \frac{t_m - t_0}{m}$ , 也可以采用等样本数划分, 即要求  $m$  个区间中的观测样本数大致相等. 给出区间  $[t_0, t_m]$  上各节点后, 可令子空间

$$\Omega_j = \{x \in \Omega | t_{j-1} < t_x \leq t_j\}, j = 1, 2, \dots, m \quad (35)$$

子空间  $\Omega_j$  即为依据特征连续性准则对全空间  $\Omega$  的剖分结果.

依据特征连续性准则剖分全空间  $\Omega$  时, 由于特征空间中样本分布的连续性, 有可能导致某个子空间  $\Omega_j$  中仅包含一类样本. 此时, 可以减少子空间个数重新划分空间, 也可以认为在该子空间  $\Omega_j$  中只分布一类样本, 可以不必训练该样本子集.

### 4 仿真实验与结果分析

本节将利用经典的双螺旋线数据和微波暗室数据对 SVM 和 PSVM 的分类性能与计算速度进行多组对比实验, 通过实验结果来验证本文算法的有效性和优越性.

双螺旋线(图4)分类识别一直是模式识别领域公认的一个相当有难度的问题, 该问题因其难度而经常被用作检验模式识别算法性能的试金石<sup>[10-12]</sup>. 螺旋线圈数的多少代表了问题难易程度, 圈数越多问题的可分性越差. 本文分别对 2 圈、4 圈、8 圈和 16 圈的双螺旋线, 在分类性能和计算速度上, 将传统支撑矢量机和分片支撑矢量机进行了对比. 实验中, 所选择的核函数为 Gaussian 核函数, 以样本能量大小依次剖分全空间, 在所有观测样本中每隔一个选择一个样本作为训练样本, 其余观测样本作为测试样本, 子空间个数为  $m = \sqrt{l}$ .

由表1可知, 随着双螺旋线圈数和观测样本集个数的增加, 传统支撑矢量机的计算时间(训练与测

试时间总和)以非线性速率迅速增加,而分片支撑矢量机则以近似线性速率缓慢增加.在上述实验中,分片支撑矢量机的计算时间,相对于传统支撑矢量机而言可以降低 1 至 2 个数量级.另外,对于识别性能,由于分片支撑矢量机通过剖分全空间,在各子空间上大幅度降低了问题的复杂度,使得在训练集和测试集上均达到了 100% 的识别率,而传统支撑矢量机最好的识别率则分别只有 62.70% 和 63.10%.因此,分片支撑矢量机不仅大幅度降低了计算量,而且其识别性能相对于传统支撑矢量机已有了明显的提高.

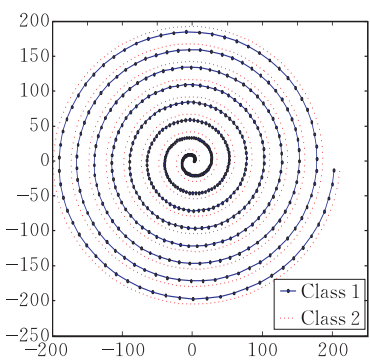


图 4 8 圈双螺旋线示意图

表 1 双螺旋线实验结果

圈数	训练样本	测试样本	参数设定	识别算法	时间/s	训练集识别率/%	测试集识别率/%
2	126	126	$\sigma=80$	SVM	0.2367	85.68	88.32
				PSVM	0.1029	100	100
			$\sigma=100$	SVM	0.2144	86.16	89.12
				PSVM	0.1029	100	100
4	252	252	$\sigma=80$	SVM	1.1720	51.98	51.98
				PSVM	0.3120	100	100
			$\sigma=100$	SVM	1.1560	62.70	63.10
				PSVM	0.3120	100	100
8	504	502	$\sigma=80$	SVM	8.4370	48.81	49.40
				PSVM	0.4840	100	100
			$\sigma=100$	SVM	8.2810	62.10	62.35
				PSVM	0.4840	100	100
16	1006	1006	$\sigma=80$	SVM	78.2810	51.29	50.60
				PSVM	0.7500	100	100
			$\sigma=100$	SVM	75.2500	52.88	52.88
				PSVM	0.7340	100	100

本实验数据为微波暗室测量数据,观测目标为无缝球锥和带缝球锥,两种观测目标的实物图如图 5 所示.雷达的工作频段为 8.75~10.75GHz,步进频率的采样间隔  $\Delta f=20\text{MHz}$ ,频率采样点数  $N=101$ .雷达极化方式包括了  $HH$ 、 $HV$ 、 $VH$  和  $VV$  4 种极化,雷达观测视线的俯仰角  $\theta=0$ ,方位角范围  $\phi=0\sim\pi$ ,方位角的采样间隔为  $\Delta\phi=\pi/900$ ,方位角

采样点数  $K=901$ ,观测样本个数为  $901\times4\times2=7208$ .实验中采用特征连续性准则,以样本能量大小依次剖分全空间.在所有观测样本中每隔一个选择一个样本作为训练样本,其大小为  $l=3604$ ,其余观测样本作为测试样本,共有 3604 个测试样本,子空间个数为  $m=14$ .



图 5 微波暗室测量实物图

由表 2 可知,分片支撑矢量机的计算时间(训练与测试时间总和),相对于传统支撑矢量机而言可以

降低 1 至 2 个数量级而且较为稳定.对于识别性能,由于分片支撑矢量机通过剖分全空间,降低了各子

空间上问题的复杂度,使得在训练集均达到了 100% 的识别率,而在测试集上略高于传统支撑矢量机的识别率. 由这个大样本实验可以有效地说明

PSVM 较好的解决了 SVM 计算量过大这一瓶颈问题,极大地增强了工程实用性.

表 2 微波暗室数据识别结果(训练样本 3604 个,测试样本 3604 个)

核函数类型	核参数设定	识别算法	时间/s	训练集识别率/%	测试集识别率/%
多项式	$p=5$	SVM	358.6870	100	97.92
		PSVM	5.5470	100	98.36
	$p=8$	SVM	449.5620	100	97.72
		PSVM	5.3590	100	98.58
Gaussian	$\sigma=1$	SVM	211.0780	100	98.06
		PSVM	5.0620	100	98.36
	$\sigma=1.5$	SVM	272.1870	100	98.36
		PSVM	5.2190	100	98.42

5 结束语

线性支撑矢量机以其结构简单和计算量低而备受学者的关注,但遗憾的是线性支撑矢量机的分类能力和推广能力均为有限,这在一定程度上限制了其适应范围. 核函数的引入虽然巧妙地解决了线性不可分而非线性可分的识别问题,但非线性支撑矢量机本质上也是一个线性分类器,与线性支撑矢量机的差别就在于是在高维特征空间而非原空间中构造最优分类超平面. 当样本分布的区域十分散乱和高度交错时,传统支撑矢量机的分类能力和推广能力均会受到很大的挑战.

本文借鉴了分段线性模式识别的基本思想,提出了一种新的支撑矢量机模型——分片支撑矢量机. 该模型通过某种特征空间划分算法,在每个子空间中基于 SVM 原理构造一个最优分类超曲面,然后,将各个分类超曲面链接起来构成一个分片最优分类超曲面以逼近理论上的最优分类超曲面. 由于 PSVM 可以有效地分解原问题的复杂度,因此, PSVM 不仅可以有效地处理高度复杂分布区域的识别问题,提高识别率和推广能力,而且可以大大降低支撑矢量机的计算量和存储容量,提高了 SVM 的实用能力. 此外,本文还从理论上分析探讨了其推广能力的界,为分片支撑矢量机模型提供了坚实的理论基础.

参 考 文 献

[1] Chapelle O, Vapnik V, Bacsquest et al. Choosing multiple parameters for support vector machines. Machine Learning, 2002, 46(1): 131-159

[2] Wu Tao, He Han-Gen, He Ming-Ke. Interpolation based kernel function's construction. Chinese Journal of Computers, 2003, 26(8): 990-996(in Chinese)  
(吴涛, 贺汉根, 贺明科. 基于插值的核函数构造. 计算机学报, 2003, 26(8): 990-996)

[3] Ayat N E, Cheriet M, Remaki L et al. KOMD—A new support vector machine kernel with moderate decreasing for pattern recognition, application to digit image recognition//Proceedings of the 6th International Conferemce on Document Analysis and Recognition. Seattle, USA, IEEE, 2001: 1215-1221

[4] Anguita D, Boni A. Improved neural network for SVM learning. IEEE Transactions on Neural Networks, 2002, 13(5): 1243-1244

[5] Sun Ji-Xiang et al. Modern Pattern Recognition. Changsha: National University of Defense Technology Press, 2002 (in Chinese)  
(孙即祥等. 现代模式识别. 长沙: 国防科学技术大学出版社, 2002)

[6] Liang Zhi-Shun, Deng Ji-Xian, Yang Wei-Quan, Situ Rong, Deng Yong-Lu. The Principle of Probability and Statistics. 2nd Edition. Beijing: Higher Education Press, 1988(in Chinese)  
(梁之舜, 邓集贤, 杨维权, 司徒荣, 邓永录. 概率论及数理统计. 第 2 版. 北京: 高等教育出版社, 1988)

[7] Vapnik V N. Statistical Learning Theory. Beijing: Publishing House of Electronics Industry, 2004

[8] Platt J. Fast Training of support vector machines using sequential minimal optimization//Scholkopf B, Burges C et al eds. Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 1999: 185-208

[9] Wang Shou-Jue. Bionic(topological) pattern recognition——A new model of pattern recognition theory and its application. Acta Electronica Sinica, 2002, 30(10): 1417-1420(in Chinese)  
(王守觉. 仿生模式识别(拓扑模式识别)——一种模式识别新模型的理论与应用. 电子学报, 2002, 30(10): 1417-1420)

[10] Zhou Wei-Da, Zhang Li, Jiao Li-Cheng. Linear programming support vector machine. *Acta Electronica Sinica*, 2001, 29 (11): 1507-1511(in Chinese)  
(周伟达, 张莉, 焦李成. 线性规划支撑向量机. *电子学报*, 2001, 29(11): 1507-1511)

[11] WU You-Shou, Zhao Ming-Sheng, Ding Xiao-Qing. A new artificial neural network with adjustable activation function and its application. *Science in China (Series E)*, 1997, 27

(1): 55-60(in Chinese)  
(吴佑寿, 赵明生, 丁晓青. 一种激励函数可调的新人工神经网络及应用. *中国科学(E辑)*, 1997, 27(1): 55-60)

[12] Zhang Ling, Zhang Bo, Yin Hai-Feng. An alternative covering design algorithm of multi-layer neural network. *Journal of Software*, 1999, 10(7): 737-742(in Chinese)  
(张铃, 张钊, 殷海风. 多层前向网络的交叉覆盖设计算法. *软件学报*, 1999, 10(7): 737-742)



**REN Shuang-Qiao**, born in 1977, Ph. D., assistant researcher. His research interests include machine learning, pattern recognition and radar signal processing.

**YANG De-Gui**, born in 1976, Ph. D., associate profes-

sor. His main research interests include pattern recognition and information fusion.

**LI Xiang**, born in 1967, Ph. D., professor. His main research interests include pattern recognition, image processing and information fusion.

**ZHUANG Zhao-Wen**, born in 1958, Ph. D., professor. His main research interests include pattern recognition, signal processing and information fusion.

### Background

In recent years, many scholars have carefully studied the linear classifier because of little complex form and computational time. When the samples can not be correctly classified by the linear classification hyper plane, the linear classifier will be not competent for recognition the samples. In the interesting of correctly recognizing the samples with complex distribution, Vladimir N. Vapnik introduced the kernel function technique and support vector machines. The kernel functions in SVM allow the classifier to be extremely flexible to suit the requirements of the classification problem, but it is an on-going issue that the kernel function can be adaptively selected and constructed. Additional, the nonlinear SVM essentially is a linear classifier in the high dimension feature space which is educed through the nonlinear kernel function. When the samples complexly distributing, the performance of nonlinear SVM, such as the classification capability, generalized capability and the computation time, will encounter the hard challenge.

In classical pattern classification theory and methods, the piecewise linear recognition technique is a compromise between linear and nonlinear recognition method. Its classification surface is a special nonlinear function, which consists of

several classification hyper planes. Comparing to the general nonlinear classification hyper surface, the piecewise linear classification surface is more simple and has the ability to approach a great variety of optimal classification nonlinear hyper surface.

A novel piecewise support vector machine (PSVM) model based on piecewise linear classification concept is provided in this paper. In this model, the feature space was firstly partitioned into several subspaces, and the optimal classification surface was constructed based on SVM in each subspace, then the piecewise classification surface in feature space was developed by linking the optimal classification surfaces in subspaces. As the experimental results shown, the performance of PSVM such as computational efficiency, classification capability and generalized capability are improved obviously in contrast to classical SVM.

This work is supported by the National Science Foundation of China, under grant No. 60402032. The aim of this project is to solve the automatic recognition for radar observing objects. This paper offeres the kernel function adaptively construction method, which is an important portion to the automatic objection recognition using SVM.