

用于一维图像识别的支撑向量机方法^{*}

张 莉 周伟达 焦李成

(西安电子科技大学雷达信号处理重点实验室, 陕西, 西安, 710071)

摘要 研究了支撑向量机的分类机理, 并利用支撑向量机对雷达目标一维像进行了识别, 识别的结果表明了该方法的优越性, 并显示它可以对残缺不全的样本进行识别。

关键词 雷达目标识别, 神经网络, 支撑向量机。

SUPPORT VECTOR MACHINE FOR 1-D IMAGE RECOGNITION^{*}

ZHANG Li ZHOU Wei-Da JIAO Li-Cheng

(National Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract The classification mechanism of support vector machine (SVM) was analyzed in detail. The one-dimensional image of radar target was recognized by SVM. The results of the simulation show the advantage of SVM over other methods.

Key words radar target recognition, neural network, support vector machine.

引言

当前模式识别已获得了广泛的应用, 几乎遍及各个科学领域. 现已提出了大量的模式分类方法, 包括贝叶斯决策分类^[1], 神经网络^[2]和进化算法^[3]等. 由于雷达目标的运动比较复杂, 包括平移、旋转等, 得到的识别信号仅是间接地大致地对目标的反映, 所以雷达目标的识别一直是模式识别问题中的难点. 尤其对于军事目标识别, 由于目标对象是非合作性的, 要获得敌方目标各种角度各个姿态的雷达一维像几乎是不可能的, 同时一般目标对象不仅具有伪装能力, 而且具有反雷达能力, 所以我们不仅缺乏样本数据, 即使已获得的样本数据也是残缺不全的. 针对上述情况, 传统的雷达一维像识别方法把重点放在目标特征提取上, 即提取目标对象一维像中有用的能真实反映目标的信息, 如引入 Garbor 变换^[4], 子波变换^[5]来进行时频域识别, 以及用分形来进行特征提取^[6]. 但是由于变换本身的冗余性, 数据量大大增加, 这无疑给数据量已经很大, 计算复杂的识别方法又增加了难度. 为了减少计算复杂度,

我们不得不进行特征再提取, 这需要冒由于提取方法不当而删除了有用信息的风险. 此项工作有待继续深入研究.

本文从另一个角度, 在分类器上展开工作. 目前用于雷达一维像识别的分类器存在许多缺点: (1) 训练时只考虑了逼近已有样本的分类曲面, 而未涉及推广误差, 即对未被采样到的样本的分类误差没有考虑. 这样的分类器对于样本获取代价较低的情况是适用的, 而对军事目标, 它的推广能力差的弱点就会显露无疑. (2) 计算复杂. 对于高维样本处理能力弱. (3) 收敛速度慢. 容易陷入局部极小. 支撑向量机^[7-9]是新近提出的一种学习机器, 它不仅具有较强的推广能力, 适用于小样本训练的情况, 而且非常适合用于雷达目标识别, 同时它又具有强大的处理能力, 对高维样本的处理复杂度与对低维样本近似, 并能巧妙地引入核函数来实现非线性映射, 从而完成非线性处理. 另外它的模型结构预先是不确定的, 可根据学习的对象在训练过程中确定. 本文用支撑向量机来进行雷达目标识别, 不仅希望它较强的推广能力能改善由于样本不全, 样本数据残缺带来

^{*} 国家自然科学基金(批准号 60073053, 60133010, 69831040)资助项目

稿件收到日期 2001-07-14, 修改稿收到日期 2001-09-10

^{*} The project supported by the National Natural Science Foundation of China (No. 60073053, 60133010, 69831040)

Received 2001-07-14, revised 2001-09-10

的性能降低,而且希望通过其强大的处理能力来提高训练和识别的计算速度.由于支撑向量机具有较强的高维样本处理能力,本文直接采用雷达目标原始一维像作为样本,而不进行变换,以免特征提取方法的不当而丢失有用信息.

1 支撑向量机的分类机理分析

1.1 线性支撑向量机的分类机理

对两类模式的识别问题,假设已知一组训练样本 $\{(x_1, y_1), \dots, (x_l, y_l)\}$,其中 $x \in R^N, y \in \{-1, 1\}$, R^N 表示输入模式的特征空间.学习的目的就是寻找判决函数 $f(x, a)$ 中适当的参数 a ,使得对任意 $x, f(x, a)$ 的函数值与 y 相等,这样就能够得到一个正确的分类结果.假设有一超平面能够把期望值为正负1的样本分开,则位于超平面的矢量 x 满足 $W \cdot x + b = 0$,其中 W 是该超平面的法向量, $|b|/\|W\|$ 是从超平面的原点的距离,其中 $\|\cdot\|$ 表示取2-范数,即欧氏距离.令 $d_+(d_-)$ 是期望值为正(负)的样本矢量到超平面的最短距离.定义超平面的间隔(margin)为 $(d_+ d_-)$.支撑向量机算法就是要求得到具有最大间隔的超平面,即 $\max(d_+ + d_-)$.且所有的样本矢量满足下面的不等式

$$W \cdot x_i + b \geq 1; \quad (y_i = +1), \quad (1)$$

$$W \cdot x_i + b \leq -1; \quad (y_i = -1), \quad (2)$$

当式(1)等式成立,矢量位于超平面 $H_1: W \cdot x_i + b = +1$ 上,其法向量为 W ,到原点的距离为 $|1-b|/\|W\|$;当式(2)等式成立,矢量位于超平面 $H_2: W \cdot x_i + b = -1$ 上,法向量也是 W ,到原点的距离变为 $|1-b|/\|W\|$.这样, $d_+ = d_- = 1/\|W\|$,所以分离超平面的边界为 $d_+ = d_- = 2/\|W\|$.因此关键是使得边界最大,即是在式(1)和式(2)下,使得 $\|W\|^2$ 最小^[8].

图1是一个典型的线性可分的例子,说明了线性支撑向量机的分类机理.图1中用圆圈圈起来的点位于超平面 H_1 或 H_2 上,即我们所说的支撑矢量.

考虑到对式(1)和式(2)不成立的矢量,即不能线性分开的情况,文献[7]引入了松弛变量 ξ_i ,其实际物理意义是分类误差

$$\xi_i \geq 0, \quad i = 1, \dots, l \quad (3)$$

从而有下面的最小优化问题;式(4)中 $C > 0$ 是用户自定义的参数, C 越大表明惩罚性越大,因而可以认为它是惩罚因子.目标函

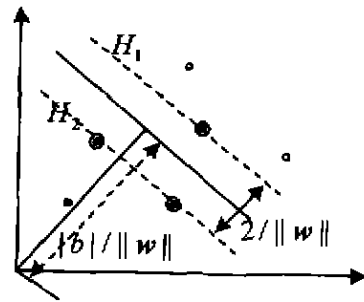


图1 线性SVM分类机理

Fig. 1 The classification mechanism of linear SVM

$$\text{最小化} \quad \frac{1}{2} \|W\|^2 + C \cdot \sum_{i=1}^l \xi_i, \quad (4)$$

$$\text{约束} \quad y_i((W \cdot x_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \quad i = 1, \dots, l$$

数中的第1项最小保证分类边界最大,第2项最小是指样本错分的总误差最小.

利用Lagrange乘子法,可以把式(4)变成拉格朗日方程:

$$L(w, \xi, b, a, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l (a_i (y_i ((w \cdot x_i) + b) - 1 + \xi_i) + \gamma_i \xi_i), \quad (5)$$

式(5)中 a_i 和 γ_i 为正的Lagrange乘子.为了求得式(5)的极小值,分别对 W, ξ 和 b 求偏导,并使之等于0

$$\frac{\partial L}{\partial W} = W - \sum_{i=1}^l a_i y_i x_i = 0, \Rightarrow W = \sum_{i=1}^l a_i y_i x_i, \quad (6)$$

$$\frac{\partial L}{\partial \xi_i} = C - (a_i + \gamma_i) = 0, \Rightarrow C = a_i + \gamma_i, \quad (7)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l a_i = 0, \Rightarrow \sum_{i=1}^l a_i = 0, \quad (8)$$

式中符号“ \Rightarrow ”表示左边的等式推出右边的等式.把式(6)、式(7)和式(8)代入到式(5)中,得到对偶规则:

最大化

$$W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j (x_i \cdot x_j), \quad (9)$$

约束

$$\sum_{i=1}^l a_i y_i = 0, a_i \in [0, C] \quad i = 1, \dots, l$$

利用式(7)消去了Lagrange乘子 γ_i ,在对偶规则中

没有该变量.采用线性判决函数

$$f(x) = \text{sgn}(W \cdot x + b), \quad (10)$$

由式(6)得

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i (x_i \cdot x) + b\right). \quad (11)$$

式中 w 能够完全用训练样本 x_i 的线性组合来描述. 如果 $\alpha_i \neq 0$, 则训练样本 x_i 就被称为是支撑矢量. 支撑矢量机的判决函数与支撑矢量的个数有关, 与样本的维数无关, 这有利于处理高维输入空间模式.

1.2 非线性支撑矢量机的分类机理

采用核方法把判决函数推广为非线性函数. 在线性支撑矢量训练算法中, 数据以点积形式 $(x_i \cdot x_j)$ 出现. 现在用非线性映射把输入空间映射到某一特征空间, 记为 $\Phi: R^N \rightarrow H$. 如果存在一种核函数 K , 使得

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j), \quad (12)$$

这样就可以在特征空间中进行许多计算, 而且并不需要知道具体的映射 Φ .

事实上一个函数只要满足 Mercer 条件, 就称为是支撑矢量核^[7-9]. 这样的核函数就可以分解成特征空间的点积形式. Mercer 条件可描述如下^[8,9]

$$\iint_{L_2 \supset L_2} K(x, x^*) g(x) g(x^*) dx dx^* \geq 0, \quad (13)$$

$$\forall g(x) \in L_2(R^N), K(x, x^*) \in L_2(R^N) \odot L_2(R^N)$$

如果式(13)满足, 可以找到 K 的特征函数和特征值 $(\Phi_i(x), \lambda_i)$. 核函数 K 可以写成点积的形式

$$K(x, x^*) = \sum_{i=1}^{N_H} \lambda_i \Phi_i(x) \Phi_i(x^*), \quad (14)$$

其中 N_H 是特征空间的维数. 那么非线性映射函数可以写成

$$\Phi(x) = (\sqrt{\lambda_1} \Phi_1(x), \sqrt{\lambda_2} \Phi_2(x), \dots, \sqrt{\lambda_{N_H}} \Phi_{N_H}(x))^T, \quad (15)$$

但通常在知道这些核函数 $K(x, x^*)$ 后, 很难求出映射 $\Phi(x)$ 或者映射 $\Phi(x)$ 是无穷维的.

文献[7]给出了一些核函数, 比如高斯核函数

$$K(x_i, x_j) = \exp(-\beta \|x_i - x_j\|^2) \quad (\beta > 0), \quad (16)$$

以及多项式核函数

$$K(x_i, x_j) = (x_i \cdot x_j)^d \quad (d \in N), \quad (17)$$

等等. 用核函数代替线性支撑矢量机中的点积形式, 对偶规则即为

最大化

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i y_i \alpha_j y_j K(x_i \cdot x_j), \quad (18)$$

$$\text{约束} \quad \sum_{i=1}^l \alpha_i y_i = 0, \\ \alpha_i \in [0, C], \quad i = 1, \dots, l$$

从而得到非线性支撑矢量机的判决函数

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right). \quad (19)$$

这样巧妙地利用核函数映射使支撑矢量机具有了非线性处理的能力.

1.3 KKT 条件及 b 的计算

b 的计算不是在优化时确定的, 而是利用优化后的 α 值, 根据 Karush-Kuhn-Tucker(KKT) 条件来计算的. KKT 条件表明, 当 $g(x_i, y_i) \geq 0$, 相应的 Lagrange 乘子(即对偶变量)为 α_i , 则对所有 i , 解满足 $\alpha_i g(x_i, y_i) = 0$. 在支撑矢量机中, 这就意味着

$$\alpha_i (y_i ((W \cdot x_i) + b) - 1 + \xi_i) = 0, \quad (20)$$

和

$$y_i \xi_i = (C - \alpha_i) \xi_i = 0, \quad (21)$$

当 $0 < \alpha_i < C$ 时, 有 $\xi_i = 0$, 可以得到

$$b = y_i - \sum_{i=1}^l \alpha_i y_i K(x_i, x_i). \quad (22)$$

将式(22)中得到的 b 代入式(19), 就可完成对模式的分类.

1.4 支撑矢量机推广能力分析

在统计学理论中, 最小风险的准则函数(或风险函数)^[8,9]如下

$$R(z, f) = \int Q(z, f) p(x) dx, \quad (23)$$

式(23)中 $z = (x, y)$, f 为假设空间的函数类, $p(x)$ 为样本的概率密度, $Q(z, f)$ 为损失函数, 一般定义为决策误差. 但是在一般情况下, 样本的概率密度是未知的, 所以在实际运用中认为样本均匀分布, 采用如下的风险函数

$$R(z, f) = \frac{1}{l} \sum_{i=1}^l Q(z, f) = \frac{1}{l} \sum_{i=1}^l \|\epsilon_i\|. \quad (24)$$

其中 $\|\epsilon_i\|$ 为判决误差的 1-范数. 在实际运用中我们经常碰到这样的情况, 训练过程非常好, 风险函数几乎为 0, 但检验效果很差; 有时训练结果不是很好, 但检验效果倒不错. 这实际上就是分类器推广能力的问题. 分类器训练得到判决函数 f^* 一般不止一个能使风险函数式(24)达到最小, 即有许多函数都能把样本分开, 说明此时仅考虑训练样本已经不够.

一般模式识别中首先要假设一个学习模型, 如一个函数类, 它的许多参数是要通过学习来确定的. 这样的一组假设构成我们的假设空间, 假设空间是否合适与分类器的性能直接相关. 假设空间太大, 收敛的速度就会很慢, 甚至不收敛. 假设空间太小, 就会导致学习误差和推广误差很大. 如何既能得到对于我们样本最优的分类函数包括在内, 又不能使假设空间太大的函数集, 这已经成为研究模式识别人们非常关注的问题. 支撑向量机正是通过在风险函数式(24)中引入权值项, 得到了一组很好的假设. 前面已经说明在风险函数中引入 $\frac{1}{2} \|W\|_2$ 的几何意义是使分类边界达到最大, 即它对假设空间做了一个限制, 不仅仅要寻找一个能正确识别训练样本的判决函数, 还要求该判决函数的分类边界是最大的, 所以支撑向量机得到判决函数相对于以前的模式识别方法来说是优中挑优. 同时控制假设空间的大小, 减小函数集的 VC 维数, 从而使得算法的收敛速度更快.

2 仿真实验

本实验所用的数据是在微波暗室里产生的仿真数据, 只涉及到角度的变换. 这里用到了 3 种飞机 B-52, Q-6 和 Q-7 从 $0 \sim 160^\circ$ 范围内的数据. 图

2 是 Q-6 飞机和 Q-7 飞机在 0° 时的一维像, 可以看出飞机在外形结构上的相似性也反映到它们的一维像上. 图 2 的横坐标有 64 个点, 表示输入样本的维数为 64 维; 纵坐标表示相应点的幅值. 对于这样高的输入维数, 如果不进行特征压缩, 许多模式识别方法是无能为力的.

在进行训练之前, 我们先对数据进行归一化处理, 这是由于同一类飞机样本点的最大值与最小值间至少差了 3 个数量级. 在实验中分别采用径向基函数网络和支撑向量机来进行识别, 识别结果如表 1 所列. 结果表明对高输入维的一维像问题, 神经网络分类器的推广能力很差, 而支撑向量机在雷达目标一维像识别中是一种有效且可行的方法.

3 结语

支撑向量机是一种新的学习机器. 与传统的模式识别的方法相比, 它是一种小样本学习方法, 推广能力强, 而且其算法不受到输入模式维数的影响, 通过引入核函数映射可方便地实现非线性处理. 本文针对雷达目标样本缺乏, 数据残缺不全, 同时对于一维像, 二维像这种样本数据大, 处理困难这一特点, 提出用支撑向量机进行雷达目标一维像识别, 得到了良好的效果.

表 1 三种飞机的识别结果

Table 1 The recognition results of three airplanes

飞机模型	训练样本数(个)	检验样本数(个)	径向基网络(识别率%)		支撑向量机(识别率%)	
			训练样本	检验样本	训练样本	检验样本
B-52	33	322	100	70.93	100	95.50
歼-6	32	311	100	57.35	100	88.89
歼-7	46	451	100	79.51	100	88.40

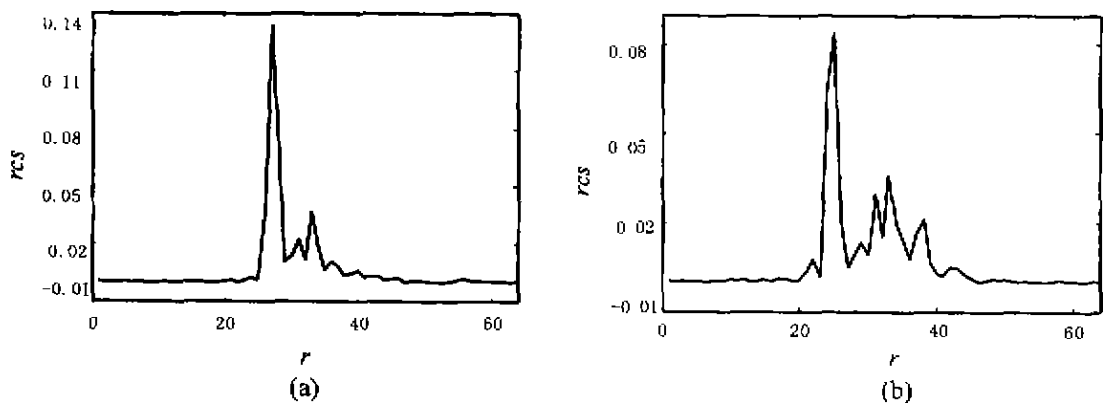


图 2 两种外形相似的飞机在 0° 下的一维像 (a) Q-6 飞机 (b) Q-7 飞机
Fig. 2 The one-dimensional image under zero degree of two airplanes with similar forms
(a) Q-6 airplane (b) Q-7 airplane

REFERENCES

- [1] BIAN Zhao-Qi, YAN Ping-Fan, YANG Cun-Rong, *et al.* *Pattern Recognition*. Beijing: Tsinghua Publishing House (边肇祺, 阎平凡, 杨存荣, 等. 模式识别. 北京: 清华大学出版社), 1998: 9—38
- [2] JIAO Li-Cheng. *The Application and Realization of Neural Networks*. Xi'an: Xidian University Press (焦李成. 神经网络的应用与实现. 西安: 西安电子科技大学出版社), 1993: 1—97
- [3] CHEN Guo-Liang, WANG Xi-Fa, Zhuang Zhen-Quan, *et al.* *Genetic Algorithm and Application*. Beijing: People's Post and Telecommunication Publishing House (陈国良, 王熙法, 庄镇泉, 等. 遗传算法及其应用. 北京: 人民邮电出版社), 1996: 262—266, 345—354
- [4] ZHANG Xian-Da, BAO Zheng. *Non-stable Signal Analysis and Processing*. Beijing: National Defence Industry Press (张贤达, 保铮. 非平稳信号分析与处理. 北京: 国防工业出版社), 1998: 109—152
- [5] LIU Fang, WU Ming. The target recognition based on wavelet and genetic network. *Xidian University Transaction* (刘芳, 吴明. 基于子波和进化网络的目标识别. 西安电子科技大学学报), 1999, 26(3): 328—331
- [6] XIAN Ming, ZHUANG Zhao-Wen, XIAO Shun-Ping, *et al.* Radar target recognition based on fractal compression characteristic. *J Infrared Millim. Waves* (鲜明, 庄钊文, 肖顺平, 等. 基于分形压缩特征的雷达目标识别研究. 红外与毫米波学报), 1999, 18(4): 289—294
- [7] Cortes C, Vapnik V. Support vector networks. *Machine Learning*, 1995, 20: 273—297
- [8] Burges C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, 2(2): 1—47
- [9] Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 1995: 1—156