

文章编号: 0253-3782(2006)01-0029-08

# 我国大陆强震预测的支持向量机方法<sup>\*</sup>

王 炜<sup>1)</sup> 刘 悦<sup>2)</sup> 李国正<sup>2)</sup> 吴耿锋<sup>2)</sup>  
马钦忠<sup>1)</sup> 赵利飞<sup>1)</sup> 林命邇<sup>1)</sup>

1) 中国上海 200062 上海市地震局

2) 中国上海 200072 上海大学计算机工程与科学学院

**摘要** 统计学习理论是研究小样本情况下机器学习规律的理论. 支持向量机是基于统计学习理论框架下的一种新的通用机器学习方法. 它不但较好地解决了以往困扰很多学习方法的小样本、过学习、高维数、局部最小等实际难题, 而且具有很强的泛化(预测)能力, 其预测效果通常优于人工神经网络. 我国大陆强震与全球主要板块边界的强震活动之间具有一定的关系, 但是这种关系具有较强的非线性. 尽管这种关系还不清楚, 但是通过支持向量机可以很好地进行建模, 并对我国大陆强震进行预测.

**关键词** 统计学习理论 支持向量机 人工神经网络 地震形势

**中图分类号**: P315.3

**文献标识码**: A

## 引言

统计学习理论(SLT, statistical learning theory)(Vapnik, 1995)是研究小样本情况下机器学习规律的基本理论和数学构架, 也是小样本统计估计和预测学习的最佳理论. Vapnik等从上世纪 60—70 年代开始致力于该方面研究; 到 90 年代中期, 随着该理论的不断发展和成熟, 产生了基于统计学习理论体系的新的通用机器学习方法, 即支持向量机(SVM, support vector machines)(Cristianini, Shawe-Taylor, 2000). 它与神经网络完全不同, 神经网络学习算法的构造是受模拟生物启发, 而支持向量机的思想来源于最小化错误率的理论界限. 这些界限是通过对学习过程的形式化分析得到的. 基于这一思想产生的支持向量机算法, 不但具有良好的数学性质, 如解的唯一性、不依赖输入空间的维数等, 而且在应用中也表现出了良好的性能. 它所得到的最优解超过了传统的学习方法. 它避免了人工神经网络等方法的网络结构难于确定、过学习和欠学习, 以及局部极小等问题. 一些学者认为, SLT 和 SVM 正在成为继神经网络研究之后新的研究热点, 并将有力推动机器学习理论和技术的发展, 因此被认为是目前针对小样本的分类、回归等问题的最佳理论.

支持向量机是处理非线性分类问题的一种新方法. 研究发现支持向量机的各项性能尤其是泛化(预测)能力, 优于传统的统计方法和人工神经网络方法. 因此近几年(尤其是最近 1~2 年)来, 支持向量机方法已在一些领域, 如文字识别、人脸识别、医疗诊断、故障诊断、基因分类、遥感图像分析、函数逼近、勘察测井、气象、化学计量(陈念贻等, 2002)等

<sup>\*</sup> 地震科学联合基金(104090)资助.

2005-05-08 收到初稿, 2005-09-26 收到修改稿, 2005-10-24 决定采用.

领域得到了一些十分有意义的应用. 各方面的应用结果表明, 支持向量机比传统的分类和预测方法具有更佳的性能.

目前在地震活动趋势分析中, 主要是根据一些与地震活动趋势有关的相关因素进行预测. 例如, 根据全球或我国大陆的强震活动进行预测, 根据日本海沟、台湾等某些相关地区的地震活动进行预测, 根据东北深震或某些特定事件进行预测, 以及通过与历史上发生过的地震序列对比分析进行预测等. 而所用的方法则主要是根据某一个或多个相关因素进行相关分析, 或进行某种定性的分析描述. 但是目前的这类分析方法带有很大的不确定性, 使用各种预报因子进行预报或同一预报因子在不同时段进行预报所得到的结论往往是不一致的, 有时甚至是矛盾的. 这表明这些预报用的相关因子与实际结论之间有着很明显的非线性关系. 使用支持向量机则可以建立起它们之间的非线性关系, 从而对未来地震活动趋势作出科学的预测. 本研究将根据全球强震活动与我国大陆强震活动的关系, 使用支持向量机进行学习和预测. 结果表明, 该方法具有较好的预报效果, 其预测结果优于 BP 神经网络.

1 支持向量机的分类算法

SVM 方法的基本思想是: 基于 Mercer 核展开定理(Cristianini, Shawe-Taylor, 2000; Chen *et al*, 2004), 可以通过非线性映射  $\varphi$  把样本空间映射到一个高维乃至无穷维的特征空间(Hilbert 空间), 使得在特征空间中可以应用线性学习机的方法解决样本空间中的高度非线性分类等问题. 降维(即把样本空间向低维空间作投影)是人们处理复杂问题常用的简化方法之一, 这样做可以降低计算的复杂性; 而升维(即向高维空间作映射)一般只会增加计算的复杂性, 甚至会引起“维数灾难”, 因而人们很少问津. 但是作为分类问题来说, 很可能在低维样本空间无法线性处理的样本集, 在高维特征空间却可以通过一个线性超平面实现线性划分, 而与特征空间的线性划分相对应的则是样本空间的非线性分类. 这里, 自然发生的两个问题是如何求得非线性映射  $\varphi$  和解决算法的复杂性. SVM 方法巧妙地解决了这两个难题: 由于应用了核函数的展开定理, 所以根本不需要知道非线性映射的显式表达式; 由于是在高维特征空间中应用线性学习机的方法, 所以与线性模型相比几乎不增加计算的复杂性, 这在某种程度上避免了“维数灾难”.

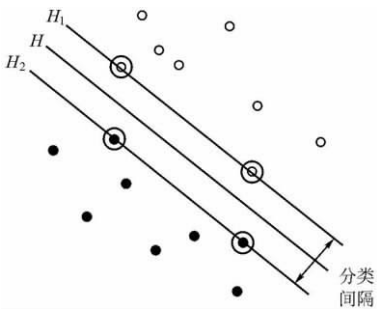


图 1 线性可分情况下的最优分类面  
○ 和 ● 为两类不同的样本,  $H_1$  和  $H_2$  平面上的样本点为支持向量

1.1 线性支持向量机

支持向量机是从线性可分情况下的最优分类面发展而来的, 其基本思想可用图 1 的两维情况说明. 图中, 实心圆和空心圆代表两类样本;  $H$  为分类超平面,  $H_1$  和  $H_2$  分别为各类中离分类超平面最近的样本且平行于分类超平面的平面, 它们之间的距离叫做分类间隔(margin). 所谓最优分类面就是要求分类面不但能将两类正确分开(训练错误率为 0), 而且能使其分类间隔最大.

最大间隔和最优超平面只由落在边界上的样本点完全确定, 我们称这样的样本点为支持向量. 只由少数训练样本点(支持向量)就可把最大间隔和最优超平面完全确定, 其余非支持向量的样本

点均不起作用, 这具有重要的意义. 它说明间隔最大化原则下的最优划分不是依赖于所有点, 而只是由支持向量决定. 求最优超平面和最大间隔等同于确定各个样本点是否为支持向量, 这预示着该方法具有好的鲁棒性(robust). 由支持向量确定的线性分类机称为线性支持向量机.

设样本为  $n$  维向量, 样本集为  $(x_i, y_i), i = 1, 2, \dots, k; x_i \in R^n (R \text{ 为实数域}); y_i \in \{-1, +1\}$ . 则  $n$  维空间中线性判别函数的一般形式为  $f(x) = \mathbf{w} \cdot \mathbf{x} + b$ , 其分类超平面  $H$  的方程为

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (1)$$

这里,  $\cdot$  表示内积;  $\mathbf{w}, b$  为待定的参数;  $\mathbf{w}$  为超平面  $H$  的法方向向量. 将判别函数归一化, 使两类样本都满足  $|f(x)| \geq 1$ , 即使离分类面最近样本的  $|f(x)| = 1$ , 则对所有样本  $x_i$  应满足

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq 1 & \text{若 } y_i = 1 \\ \mathbf{w} \cdot \mathbf{x}_i + b &\leq -1 & \text{若 } y_i = -1 \end{aligned}$$

可将上述不等式的规范形式合并为如下紧凑形式:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad i = 1, 2, \dots, k \quad (2)$$

此时分类间隔为  $2/\|\mathbf{w}\|$ , 使间隔最大等价于使  $\|\mathbf{w}\|^2$  最小. 满足条件(2)且使  $\|\mathbf{w}\|^2/2$  最小的分类面就叫做最优分类面,  $H_1$  和  $H_2$  上的训练样本点就称作支持向量. 因此构造最优分类超平面的问题可转化为在满足式(2)条件下对式(3)求最小化的问题.

$$\Phi(\mathbf{w}, b) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} \quad (3)$$

在线性条件下式(3)的优化问题是典型的二次规划问题, 其最优解为下面拉格朗日(Lagrange)函数的鞍点

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^k \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \quad (4)$$

式中,  $\alpha$  为非负 Lagrange 乘子. 在鞍点处, 由于  $\mathbf{w}$  和  $b$  的梯度均为零, 则可得到  $\sum_{i=1}^k \alpha_i y_i = 0$ ,

$$\mathbf{w} = \sum_{i=0}^k \alpha_i y_i \mathbf{x}_i \quad (5)$$

因为只有支持向量可以在  $\mathbf{w}$  的展开式中具有非零系数  $\alpha_i$ , 这时的支持向量就是使式(2)成立的向量, 即只有支持向量影响最终的分类结果. 根据 KKT(Karush-Kuhn-Tucher)(Cristianini, Shawe-Taylor, 2000)条件, 最后可以得到最优超平面方程为

$$\sum_{i=1}^k \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b = 0 \quad (6)$$

由此可得最优超平面的分类函数为式(7), 这里  $\text{sgn}$  为符号函数.

$$f(x) = \text{sgn}\left(\sum_{i=1}^k \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b\right) \quad (7)$$

显然, 以上介绍的方法是在样本能全部正确分类的情况下得到的. 考虑到可能存在一些样本不能被超平面正确分类, 因此引入松弛变量  $\xi_i$  来允许错分样本的存在

$$\xi_i \geq 0 \quad i = 1, 2, \dots, k \quad (8)$$

显然, 当分类出现错误时,  $\xi_i$  大于零,  $\sum_{i=1}^k \xi_i$  是分类错误数量的一个上界. 这时约束条件式

(2)变为

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, k \quad (9)$$

为此引入错误惩罚分量  $C$ , 因此构造广义最优分类超平面问题就转化为在式(8)、(9)约束条件下的最小化函数

$$\Phi(\mathbf{w}, b) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^k \xi_i \quad (10)$$

的问题. 上式中  $C$  为一正的常数,  $C$  越大, 对错误的惩罚越重. 其中第一项是样本到超平面的距离尽量大, 从而提高泛化能力; 第二项使误差尽量小, 即折衷考虑最少错分样本和最大分类间隔, 得到广义最优分类面.

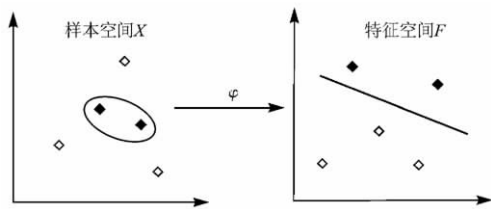


图2 样本空间到特征空间的非线性映射示意

## 1.2 非线性支持向量机

SVM方法真正有价值的应用是用来解决非线性问题. 其方法是通过一个非线性映射  $\varphi$ , 把样本空间映射到一个高维乃至无穷维的特征空间, 使得在特征空间中可以应用线性支持向量机的方法解决样本空间中的高度非线性分类和回归等问题. 图2对此给出了二维样本数据的直观示意图.

在特征空间中应用线性支持向量机的方法将  $\varphi(\mathbf{x})$  和  $\varphi(\mathbf{x}_i)$  代替了  $\mathbf{x}$  和  $\mathbf{x}_i$ , 分类决策函数式(7)变为

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^k \alpha_i y_i (\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})) + b\right) \quad (11)$$

直接确定非线性映射  $\varphi$  的形式是较困难的, 且计算量随特征空间维数增加呈指数递增. 根据 Hilbert-Schmidt 原理(Cristianini, Shawe-Taylor, 2000), 处理高维特征空间的计算问题可以避免求解空间映射  $\varphi$  的显式形式, 即通过引入所谓核函数  $K(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})$ , 将变换空间中的内积转化为原空间中某个函数的计算, 从而间接求解输入空间向高维特征空间的映射  $\varphi$ , 即分类决策函数为

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^k \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (12)$$

任意满足泛函 Mercer 条件(Cristianini, Shawe-Taylor, 2000)的对称函数均可作为核函数, 常用的核函数有: ① 多项式核函数  $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d$ ,  $d = 1, 2, \dots$ , 所得到的是  $d$  阶多项式; ② 径向基核函数(RBF)  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\}$ ; ③ Sigmoid 核函数  $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh[b(\mathbf{x}_i \cdot \mathbf{x}_j) + c]$  等等. 其中 RBF 函数因其优秀的局部逼近特性在 SVM 中应用最为广泛, 它利用局部接收域完成函数映射: 只有当输入落入输入空间的一个局部区域时, 基函数才产生一个重要的非零响应, 而其它情况下的响应近似为零. 本文采用 RBF 核函数.

式(12)为非线性支持向量学习机的最终分类决策函数. 用支持向量机求得的分类决策函数, 在形式上类似于一个神经网络(图3), 其输出为若干个中间节点的线性组合. 每个中间节点对应一个支持向量, 因此也被称为支持向量网络. 虽然用到了特征空间及非线性映射, 但实际计算中并不需要知道它们的显式表达. 通过核函数的计算即可得到原来样本

空间的非线性划分输出值. 这样我们就通过核函数和线性 SVM 方法解决了非线性 SVM 问题. 而线性 SVM 的算法归结为一个二次凸规划问题, 对此已有很多成熟的算法和应用软件可供使用.

2 我国大陆强震预测的支持向量机方法

众所周知, 全球强震主要分布在环太平洋地震带和欧亚地震带. 而我国地震活动是与全球地震活动密切联系的, 尤其是板块边界的强震活动, 它反映了板块运动的情况. 李献智等(1996)研究表明, 中国大陆强震与全球某些地震带的强震活动呈一定程度的正相关或负相关, 且某些地震带的强震超前于中国大陆强震的活动或平静. 一些学者也曾研究西太平洋地震带、欧亚地震带的地震活动与我国大陆强震的关系; 一些学者还研究日本海沟地震、台湾地震或者某些特定地区的地震与我国大陆强震的关系. 这些研究均表明, 我国大陆强震与全球主要板块边界的强震活动之间具有一定的关系, 但是这种关系具有较强的非线性.

为此, 王炜等(2001)根据全球的强震活动与板块边界的分布, 以及丁国瑜院士对中国及其邻近地区的活动边界的划分<sup>①</sup>, 将全球分为 16 个强震活动区(图 4), 并使用 BP 人工神经网络研究了这 16 个区域的 7 级以上强震活动频次与次年中国大陆地震活动的关系. 将这 16 个区域在一年中的  $M_s \geq 7.0$  地震次数  $N_i$  作为神经网络的输入元, 而将次年中国

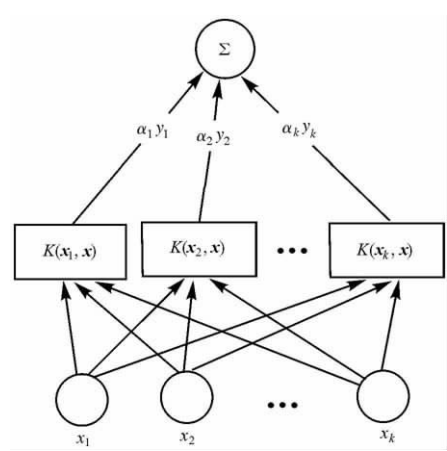


图 3 支持向量机示意图

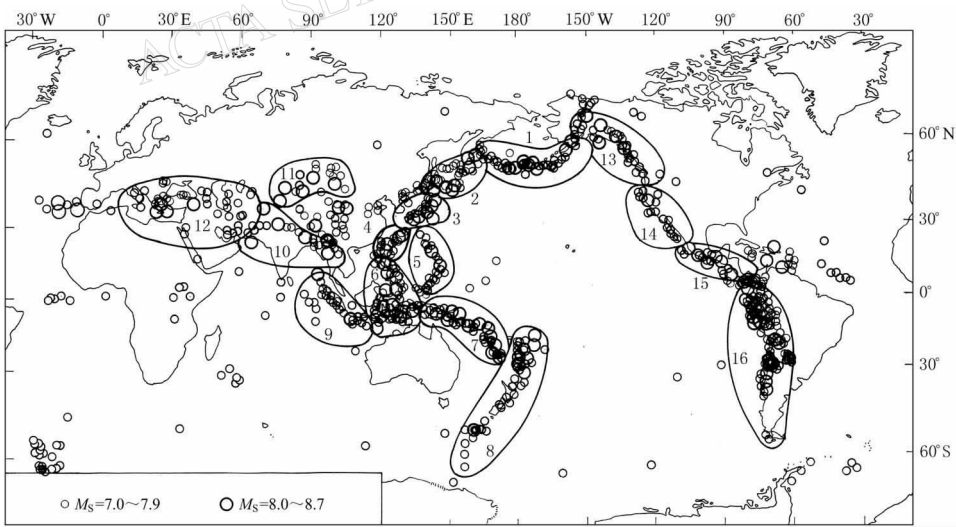


图 4 全球 7 级以上强震分布以及作为支持向量机输入项的 16 个强震活动区(各区域中的数字为区域编号)

① 丁国瑜. 2000. 中国地震活动的新构造背景, 地震分析预报技术骨干培训教材, 1~16.

大陆是否发生 7 级以上强震作为输出,得到了较好的预测效果. 由于上世纪早期的全球地震纪录不全,王炜等(2001)选取了 1925—1999 年的全球强震共 75 个样本进行神经网络的学习与检验. 为保证神经网络学习具有较多的样本,该文随机选取了其中的 10 个样本作为待检验样本,而另外的 65 个样本为学习样本. 该文的结果表明,对学习样本的内符检验全部正确;而对 10 个待检验样本的检验结果表明(表 1),有 8 个样本的检验结果正确,两个报错(1981 年和 1982 年),其报准率为 0.8.

为比较支持向量机方法与 BP 人工神经网络的预测效果,本文使用与王炜等(2001)相同的资料、方法以及相同的学习样本和检验样本应用支持向量机进行预测. 所不同的是,本文选取了 1925—2003 年(增加了 2000—2003 年 4 个检验样本,共 14 个检验样本)的资料使用支持向量机对我国大陆 1926—2004 年是否发生 7 级以上强震进行学习和预测.

所用资料取自中国地震局编辑的《中国历史强震目录》(国家地震局震害防御司,1995)、《中国近代地震目录》(国家地震局震害防御司,1999)和中国地震局分析预报中心编辑的《全球强震目录》. 本文同样取这 16 个区域在一年中的  $M_s \geq 7.0$  地震次数  $N_i$  为支持向量机输入项. 为使得输入数据归一化,其值取为  $N_i/10$ . 当  $N_i \geq 10$  时,取  $N_i = 10$ . 输出项为次年我国大陆是否发生 7 级以上强震. 如果发生,取值为 1; 否则为 0.

通过支持向量机对上述 65 个样本的学习,据此对学习样本的内符检验全部正确;对 14 个待检验样本的检验结果(表 1)表明,有 12 个样本的检验结果正确,两个报错(1969 年和 2003 年),其报准率为  $12/14 = 0.86$ .

表 1 使用支持向量机和 BP 神经网络对 14 个待检验样本的检验结果

预 测 年 份	1968	1969	1982	1983	1984	1985	1991	1992	1993	1994	2001	2002	2003	2004
中国大陆最大地震震级	5.7	7.4	6.6	6.8	6.2	7.4	6.5	6.9	6.6	7.3	8.1	5.9	6.8	6.7
检验结果(SVM)	0	0	0	0	0	1	0	0	0	1	1	0	1	0
检验结果(BP)	0	1	1	1	0	1	0	0	0	1	1	0	1	0

由表 1 中数据可见,全球强震与未来中国大陆强震之间具有很强的非线性关系. 但通过支持向量机的学习和检验,表明全球强震活动与我国大陆强震活动之间确实具有一定的联系,使用支持向量机可以较好地确立全球强震活动与我国大陆强震活动之间的非线性关系.

本文也将新增的 2000—2003 年 4 个样本使用王炜等(2001)中同样的 BP 神经网络模型进行外推检验,预测 2001—2004 年我国大陆的强震活动. 王炜等(2001)的 10 个检验样本结果以及本文新增样本的检验结果表明,这 14 个检验样本中有 11 个样本的检验结果正确,3 个报错(1982 年、1983 年和 2003 年),其报准率为  $11/14 = 0.79$ (表 1). 本文结果与其它一些文献(陈念盼等, 2002; Cristianini, Shawe-Taylor, 2004)的结果相一致,支持向量机的预测检验结果要优于 BP 神经网络.

3 讨论

本文的结果表明,支持向量机在解决地震预测的分类问题时具有较好的效能. 目前在地震预测中,常常使用人工神经网络和各种模式识别方法进行分类. 但是在实践中,人们经常发现使用上述方法对已知数据(即学习样本)拟和得较好,而在预测未知样本时偏差则

往往较大. 当学习样本较少, 影响因子(子变量)较多时, 在小样本问题中此问题尤其严重. 在数学上将这种现象称为“泛化能力”(generalization ability)不足的问题. 如何提高算法和模型的泛化能力, 以增强预测结果的可靠性, 显然是地震预测领域中的一个十分重要的问题.

传统的模式识别算法通常认为, 在分类问题中样本数必须较空间维数大 3 倍以上, 最好大 10 倍以上. 但在许多实际问题中, 往往样本数少, 而影响因数多, 常常达不到上述要求. 因此传统的模式识别算法非常强调特征筛选或通过投影进行降维. 统计学习理论证明: 对于多维空间中线性可分的分类问题, 其误报率的上限  $E_{\text{error}}$  主要由以下 3 个因素中的最小因素决定(陈念贻等, 2002):

$$E_{\text{error}} \leq \min[(m/l), (R^2 \|\omega\|^2/l), (n/l)]$$

这里,  $n$  是空间维数,  $l$  是学习样本数,  $m$  是支持向量数. 此处支持向量是指在多维空间中距离最佳分类面最近的那一部分学习样本(图 1, 图 5).  $R^2 \|\omega\|^2/l$  中  $R$  是包括全体样本点的超球半径,  $l/\|\omega\|^2$  与两类点分布区中间空白间隔宽度成正比.

由此可见, 在支持向量机分类方法中, 学习样本数并不是控制分类效果的唯一因素. 如果能够充分利用另外两个因素, 即设法使两类点中距离分界面最近的点在全体样本中所占比例  $m/l$  降低, 或设法加大两类点分布区中间空白区的宽度  $l/\|\omega\|^2$ . 因此, SVM 算法与传统的模式识别算法相反, 它不是强调降维而是采用升维技术. 利用升维办法使得高维空间中两类样本线性可分, 并能有较小的支持向量数, 两类点分布区中间空白区较宽. 从而得到距离两类点最远的分类“最佳超平面”作为分类判据, 建立起预测能力较好的数学模型(图 5).

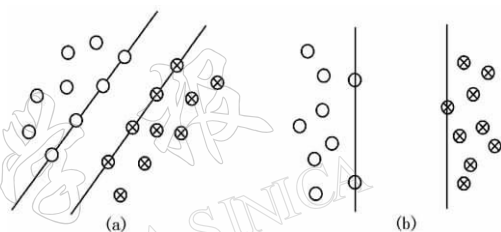


图 5 支持向量在总样本中的比例和两类样本( $\circ$ 与 $\otimes$ )分布区空白间隔宽度对预测能力的影响(示意图)

(a) 支持向量在样本中所占比例大, 空白间隔窄, 预测能力差; (b) 支持向量在样本中所占比例小, 空白间隔宽, 预测能力强

本文结果与其它一些文献(陈念贻等, 2002; Cristianini, Shawe-Taylor, 2000)的结果表明, 应用 SVM 方法比 BP 神经网络法有更高的预测精度和鲁棒性. SVM 方法比 BP 神经网络法之所以性能更优, 原因在于 BP 网络容易陷入的局部极值问题, 因为它采用梯度下降法优化权值, 这一优化过程只能保证收敛到其中的一个点; 而在 SVM 算法中, 训练 SVM 就相当于解决一个线性约束的二次规划问题, 因而 SVM 的解是唯一的、全局的和最优的. 同时通过非线性变换和核函数巧妙地解决了高维数问题, 使得其算法复杂度与样本维数无关, 加快了训练学习速度. 另外, 它能根据有限的样本信息在模型的复杂型和学习能力之间寻求最佳折衷, 保证其有较好的泛化性能.

总之, 支持向量机有如下一些特点: ① 基于结构风险最小化(SRM, structural risk minimization)原则, 保证学习机器具有良好的泛化能力; ② 巧妙地解决了算法复杂度与输入向量维数密切相关的问题; ③ 应用核函数, 将输入空间中的非线性问题, 通过非线性函数映射到高维特征空间中, 在高维空间中构造线性判别函数; ④ 与传统统计学不同, 支持向量机专门针对小样本情况, 它的最优解基于已有样本信息, 而不是样本数趋于无穷大时

的最优解；⑤ 算法最终转化为一个凸二次规划的优化问题，保证了算法的全局最优性，解决了神经网络无法避免的可能出现局部最小的问题；⑥ 神经网络、遗传算法等传统方法的实现中带有有很大的经验成分，而支持向量机具有更严格的理论和数学基础。

### 参 考 文 献

- 陈念贻, 陆文聪, 叶晨周, 等. 2002. 支持向量机及其他核函数算法在化学计量学中的应用[J]. 计算机与应用化学, 19(6): 82~106
- 国家地震局震害防御司. 1995. 中国历史强震目录[M]. 北京: 地震出版社, 475~504
- 国家地震局震害防御司. 1999. 中国近代地震目录[M]. 北京: 中国科技出版社, 1~443
- 李献智, 吕梅梅, 刘杰. 1996. 中国大陆地震活跃期和全球部分地震带强震活动的相关性研究[J]. 中国地震, 12(2): 155~162
- 王炜, 曹雪峰, 宋先月. 2001. 使用人工神经网络判断未来我国地震形势[J]. 地震学刊, 3(21): 10~14
- Chen Nianyi, Lu Wencong, Yang Jie, et al. 2004. *Support Vector Machines in Chemistry*[M]. Singapore: World Scientific Publishing Company, 1~32
- Cristianini N, Shawe-Taylor J 著. 2000. 李国正, 王猛, 曾华军译. 2004. 支持向量机导论[M]. 北京: 电子工业出版社, 82~106
- Vapnik V 著. 1995. 张学工译. 2000. 统计学习理论的本质[M]. 北京: 清华大学出版社, 1~50

## SUPPORT VECTOR MACHINE METHOD FOR FORECASTING FUTURE STRONG EARTH- QUAKES IN CHINESE MAINLAND

Wang Wei<sup>1)</sup> Liu Yue<sup>2)</sup> Li Guozheng<sup>2)</sup> Wu Gengfeng<sup>2)</sup>  
Ma Qinzong<sup>1)</sup> Zhao Lifei<sup>1)</sup> Lin Mingzhou<sup>1)</sup>

1) *Earthquake Administration of Shanghai Municipality, Shanghai 200062, China*

2) *School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China*

**Abstract:** Statistical learning theory is for small-sample statistics. And support vector machine is a new machine learning method based on the statistical learning theory. The support vector machine not only has solved certain problems in many learning methods, such as small sample, over fitting, high dimension and local minimum, but also has a high generalization (forecasting) ability than that of artificial neural networks. The strong earthquakes in Chinese mainland are related to a certain extent to the intensive seismicity along the main plate boundaries in the world; however, the relation is nonlinear. In the paper, we have studied this unclear relation by the support vector machine method for the purpose of forecasting strong earthquakes in the Chinese mainland.

**Key words:** statistical learning theory; support vector machine; artificial neural networks; earthquake situation