第30卷 第10期 计 算 机 工 程 2004年5月
Vol.30 №10 Computer Engineering May 2004
·博士论文· 文章编号：1000—3428(2004)10—0006—04 文献标识码：A 中图分类号：TP181

支持向量机及其应用研究综述

祁亨年
(浙江林学院信息工程学院，临安 311300)

摘 要：在分析支持向量机原理的基础上，分别从人脸检测、验证和识别、说话人/语音识别、文字/手写体识别、图像处理及其他应用研究等方面对SVM的应用研究进行了综述，并讨论了SVM的优点和不足，展望了其应用研究的前景。

关键词：支持向量机；机器学习；统计学习理论

Support Vector Machines and Application Research Overview

QI Hengnian
(School of Information Engineering，Zhejiang Forestry College，Lin'an 311300)

【**Abstract**】The paper reviews the principles of SVM and then overviews its application research such as face detection, verification and identification, speaker/speech identification, character/script identification, image processing, and other applications. In the conclusion section, it discusses the advantages and shortcomings of SVM and looks forward to its attractive application research prospect.

【**Key words**】Support vector machine; Machine learning; Statistical learning theory

机器学习研究从观测数据出发寻找规律，利用这些规律对未来数据或无法观测的数据进行预测。其重要理论基础之一是统计学。统计学习理论(Statistical Learn-ing Theory，SLT)专门研究实际应用中有限样本情况的机器学习规律，并发展了支持向量机(Support Vector Machine，SVM)[1,2]这一新的通用学习方法，由于它基于结构风险最小化(SRM)原理，而不是传统统计学的经验风险最小化(ERM)，表现出很多优于已有方法的性能，迅速引起各领域的注意和研究兴趣，取得了大量的应用研究成果，推动了各领域的发展。

1 机器统计学习的原理

机器学习的目的是根据给定的训练样本 $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ 来估计某系统的输入和输出之间的依赖关系，即寻找最优函数f(x,w_0)，使它能对未知输出作尽可能准确的预测。评估的方法是使期望风险R(w)最小。

1.1 结构风险最小化

由于可以利用的信息只有有限样本，无法计算期望风险，因此传统的学习方法中采用了所谓经验风险最小化(Empirical Risk Minimization，ERM)准则，即用样本定义经验风险：

$$R_{emp}(w)=\frac{1}{n}\sum_{i=1}^{n}L(y_i,f(x_i,w)) \quad (1)$$

统计学习理论系统地研究了各种类型的函数集，经验风险和实际风险之间的关系，即推广性的界。关于两类分类问题的结论是：对指示函数(即两类分类情况的预测函数)集中的所有函数，经验风险 $R_{emp}(w)$ 和实际风险之间以至少1-η的概率满足如下关系[4]：

$$R(w)\le R_{emp}(w)+\Phi(h/n) \quad (2)$$

其中h是函数集的VC维，表征了复杂性高低；n是样本数。这一结论从理论上说明了学习机器的实际风险是由两部分组成的：一是经验风险(训练误差)，另一部分称作置信范围，它和学习机器的复杂性及训练样本数有关。它表明，在有限训练样本下，学习机器的VC维越高(复杂性越高)则置信范围越大，导致真实风险与经验风险之间可能的差别越大。这就是为什么会出现过学习现象的原因。机器学习过程不但要使经验风险最小，还要使VC维尽量小以缩小置信范围，才能取得较小的实际风险。这种思想称作结构风险最小化(Structural Risk Minimization[1]，SRM)即SRM准则。

1.2 支持向量机

SVM是从线性可分情况下的最优分类面发展而来的，基本思想可用两类线性可分情况说明。如图1所示，实心点和空心点代表两类样本。假如这两类样本(训练集)是线性可分的，则机器学习的结果是一个超平面(二维情况下是直线)或称为判别函数，该超平面可以将训练样本分为正负两类。

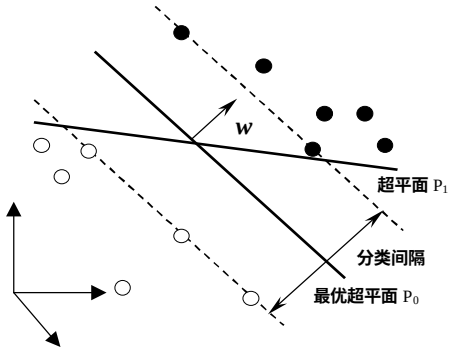

图1 线性可分情况下的分类超平面

显然，按ERM的要求，这样的超平面有无穷多个，但有的超平面对训练样本来说，其分类非常好(经验风险 $R_{emp}(w)$ 最小，为0)，但其预测推广能力却非常差，如图1中的超平面 P_1。而按照SRM的要求，学习的结果应是最优的超平面P_0，即该平面不仅能将两类训练样本正确分开，而且要使分类间隔(Margin)最大。实际上就是对推广能力的控制，这是SVM的核心思想之一。所谓分类间隔是指两类中离分类超平面最近的样本且平行于分类超平面的两个超平面间的距离，或者

基金项目：浙江省教育厅资助项目(20020980)
作者简介：祁亨年(1975－)，男，博士生、讲师，研究方向为神经网络、机器学习、模式识别等
收稿日期：2004-01-15 E-mail：qhn@zjfc.edu.cn

说是从分类超平面到两类样本中最近样本的距离的和，这些最近样本可能不止2个，正是它们决定了分类超平面，也就是确定了最优分类超平面，这些样本就是所谓的支持向量(Support Vectors)[3]。假设一个m维超平面由以下方程描述：

$$\boldsymbol{w}\cdot\boldsymbol{x}+b=0 \quad \boldsymbol{w}\in R^m, b\in R \tag{3}$$

则可以通过求 $\|w\|^2/2$ 的极小值获得分类间隔最大的最优超平面，这里的约束条件为

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)-1\geq 0 \quad i=1,\cdots,n \tag{4}$$

该约束优化问题可以用Lagrange方法求解，令

$$L(\boldsymbol{w},b,\boldsymbol{\acute{a}})=\frac{1}{2}\|\boldsymbol{w}\|^2-\sum_{i=1}^{m}a_i(y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)-1) \tag{5}$$

其中 $a_i\geq 0$ 为每个样本的拉氏乘子，由 L分别对b 和 w 导数为0，可以导出：

$$\sum_{i=1}^{m}a_i y_i=0 \tag{6}$$

$$\boldsymbol{w}=\sum_{i=1}^{m}a_i y_i \boldsymbol{x}_i \tag{7}$$

因此，解向量有一个由训练样本集的一个子集样本向量构成的展开式，该子集样本的拉氏乘子均不为0，即支持向量。拉氏乘子为0的样本向量的贡献为0，对选择分类超平面是无意义的。于是，就从训练集中得到了描述最优分类超平面的决策函数即支持向量机，它的分类功能由支持向量决定。这样决策函数可以表示为

$$f(\boldsymbol{x})=\mathrm{sgn}(\sum_{i=1}^{m}a_i y_i(\boldsymbol{x}\cdot\boldsymbol{x}_i)+b) \tag{8}$$

在线性不可分的情况下，比如存在噪声数据的情况，可以在式(4)中增加一个松弛项 $\xi_i\geq 0$，成为

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)\geq 1-x_i \quad i=1,\cdots,n \tag{9}$$

将目标改为求下式最小：

$$y(\boldsymbol{w},x)=\|\boldsymbol{w}\|^2/2+C\sum_{i=1}^{n}x_i \tag{10}$$

回顾决策函数式(8)，$(\boldsymbol{x}\cdot\boldsymbol{x}_i)$ 实际上相当于就是 x 和 x_i 的相似度。对更一般的情况，需要这样的函数 K ，对任意两个样本向量 x 和 x_i ，它的返回值 $K(x,x_i)$ 就是描述两者的相似度的一个数值，这样的一个函数就是所谓的核函数[4] (kernel)。对于实际上难以线性分类的问题，待分类样本可以通过选择适当的非线性变换映射到某个高维的特征空间(feature space)，使得在目标高维空间这些样本线性可分，从而转化为线性可分问题。Cover定理表明，通过这种非线性转换将非线性可分样本映射到足够高维的特征空间，非线性可分的样本将以极大的可能性变为线性可分[4]。如果这个非线性转换为 $f(x)$, 则超平面决策函数式(8)可重写为

$$f(\boldsymbol{x})=\mathrm{sgn}(\sum_{i=1}^{m}a_i y_i f(\boldsymbol{x})\cdot f(\boldsymbol{x}_i)+b) \tag{11}$$

在上面的问题中只涉及训练样本之间的内积运算 ，这样，在高维空间实际上只需进行内积运算，可以用原空间中的函数实现的，甚至没有必要知道变换的形式。根据泛函的有关理论，只要一种核函数 $K(x,x_i)$ 满足Mercer条件，它就对应某一变换空间中的内积。因此，在最优分类面中采用适当的内积函数 $K(x,x_i)$ 就可以实现某一非线性变换后的线性分类，而计算复杂度却没有增加。张铃证明了核函数存在性定理[5]，并提出了寻找核函数的算法。核函数存在性定理表明：给定一个训练样本集，就一定存在一个相应的函数，训练样本通过该函数映射到高维特征空间的相是线性可分的。

1.3 核函数

张铃进一步研究了支持向量机的支持向量集和核函数的关系[6]，研究表明对非线性可分情况，对一个特定的核函数，给定的样本集中的任意一个样本都可能成为一个支持向量。这意味着在一个支持向量机下观察到的特征在其它支持向量机下(其它核函数)并不能保持。因此，对解决具体问题来说，选择合适的核函数是很重要的。常见的核函数有3类，一种为多项式核函数：

$$K(\boldsymbol{x}\cdot\boldsymbol{x}_i)=[(\boldsymbol{x}\cdot\boldsymbol{x}_i)+1]^q \tag{12}$$

所得到的是q阶多项式分类器；第2种为径向基函数(RBF)：

$$K(\boldsymbol{x}\cdot\boldsymbol{x}_i)=Exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{x}_i|^2}{s^2}\right) \tag{13}$$

所得分类器与传统RBF方法的重要区别是，这里每个基函数中心对应一个支持向量，它们及输出权值都是由算法自动确定的；第3种采用Sigmoid函数作为内积，即

$$K(\boldsymbol{x}\cdot\boldsymbol{x}_i)=\tanh[v(\boldsymbol{x}\cdot\boldsymbol{x}_i)+c] \tag{14}$$

这时SVM实现的就是包含1个隐层的多层感知器。张铃的研究也证明了基于核函数的SVM与3层前向神经网络的等价性[5]。但采用Sigmoid函数为核函数的SVM中的多层感知器隐层节点数是由算法自动确定的，且不存在局部极小点问题，这是后者所不能比拟的。

1.4 多类分类问题

基本的支持向量机仅能解决两类分类问题，一些学者从两个方向研究用支持向量机解决实际的多类分类问题：一个方向就是将基本的两类支持向量机(Binary-class SVM，BSVM)扩展为多类分类支持向量机(Multi-Class SVM，MSVM)，使支持向量机本身成为解决多类分类问题的多类分类器;另一方向则相反,将多类分类问题逐步转化为两类分类问题,即用多个两类分类支持向量机组成的多类分类器。

1.4.1 多类分类支持向量机MSVM

实际应用研究中多类分类问题更加常见，只要将式(10)由两类改为多类(k类)情况，式(15)就可以很自然地将BSVM扩展为多类分类支持向量机MSVM[7]：

$$y(\boldsymbol{w},x)=\frac{1}{2}\sum_{j=1}^{k}\|\boldsymbol{w}_j\|^2+C\sum_{i=1}^{n}\sum_{j\neq y_i}x_i^j \tag{15}$$

在以下约束条件下最小化式(15)

$$\begin{aligned}&\boldsymbol{w}_{y_i}\cdot\boldsymbol{x}_i+b_{y_i}\geq\boldsymbol{w}_j\cdot\boldsymbol{x}_i+b_j+2-x_i^j\\&x_i^j\geq 0, i=1,\cdots,n \quad j\in\{1,\cdots,k\}\setminus y_i\end{aligned} \tag{16}$$

以相似的方式可得到决策函数

$$\begin{aligned}f(x)=\arg\max_j[&\sum_{i:y_i=n}\sum_{i=1}^{k}a^i f(x)\cdot f(x^i)\\&-\sum_{i:y_i=n}a^i f(x)\cdot f(x^i)+b^n]\end{aligned} \tag{17}$$

1.4.2 基于BSVM的多类分类器

用多个两类分类支持向量机组成多类分类支持向量机结构的多类分类方案主要有3种类型：1-a-r 分类器，1-a-1分类器和多级BSVM分类器[8]。

(1) 1-a-r 分类器(One-against-rest classifiers)

这种方案是为每个类构建一个BSVM，如图2，对每个类的BSVM，其训练样本集的构成是：属该类的样本为正样本，而不属于该类的其他所有样本都是负样本，即该BSVM分类器就是将该类样本和其他样本分开。所以在训练1-a-r分类器过程中训练样本需要重新标注，因为一个样本只有在对应类别的BSVM分类器是正样本，对其他的BSVM分类器都是负样本。

(2) 1-a-1分类器(One-against-one classifiers)

对1-a-1分类器，解决K类分类问题就需要k(k-1)/2 个BSVM，因为这种方案是为每两个类别训练一个BSVM分类器，如图3,最后一个待识别样本的类别是由所有k(k-1)/2 个BSVM“投票”决定的。

图2 1-a-r 分类器

图3 1-a-1 分类器

(3)多级BSVM分类器

这种方案是把多类分类问题分解为多级的两类分类子问题,图4 是两种典型方案,其中 A、B、C、D、E和F分别是 7个不同的类。

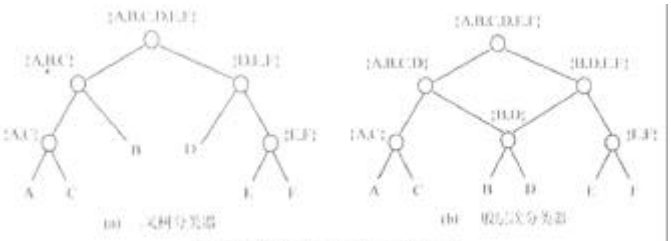

图4 两种多级BSVM分类器

2 支持向量机的应用研究现状

SVM方法在理论上具有突出的优势，贝尔实验室率先对美国邮政手写数字库识别研究方面应用了SVM方法[1]，取得了较大的成功。在随后的近几年内，有关SVM的应用研究得到了很多领域的学者的重视，在人脸检测、验证和识别、说话人/语音识别、文字/手写体识别、图像处理、及其他应用研究等方面取得了大量的研究成果，从最初的简单模式输入的直接的SVM方法研究，进入到多种方法取长补短的联合应用研究，对SVM方法也有了很多改进。

2.1 人脸检测、验证和识别

Osuna[9]最早将SVM应用于人脸检测，并取得了较好的效果。其方法是直接训练非线性SVM分类器完成人脸与非人脸的分类。由于SVM 的训练需要大量的存储空间，并且非线性SVM分类器需要较多的支持向量，速度很慢。为此，马勇等[10]提出了一种层次型结构的SVM分类器，它由一个线性SVM 组合和一个非线性SVM 组成。检测时，由前者快速排除掉图像中绝大部分背景窗口，而后者只需对少量的候选区域做出确认；训练时，在线性SVM 组合的限定下，与“自举(bootstrapping)”方法相结合可收集到训练非线性SVM的更有效的非人脸样本，简化SVM训练的难度，大量实验结果表明这种方法不仅具有较高的检测率和较低的误检率，而且具有较快的速度。

人脸检测研究中更复杂的情况是姿态的变化。叶航军等[11]提出了利用支持向量机方法进行人脸姿态的判定，将人脸姿态划分成6个类别，从一个多姿态人脸库中手工标定训练样本集和测试样本集，训练基于支持向量机姿态分类器，分类错误率降低到1.67%，明显优于在传统方法中效果最好的人工神经元网络方法。

在人脸识别中，面部特征的提取和识别可看作是对3D物体的2D投影图像进行匹配的问题。由于许多不确定性因素的影响，特征的选取与识别就成为一个难点。凌旭峰等[12]及张燕昆等[13]分别提出基于PCA与SVM相结合的人脸识别算法，充分利用了PCA在特征提取方面的有效性以及SVM在处理小样本问题和泛化能力强等方面的优势，通过SVM与最近邻距离分类器相结合，使得所提出的算法具有比传统最近邻分类器和BP网络分类器更高的识别率。王宏漫等[14]在PCA基础上进一步做ICA，提取更加有利于分类的面部特征的主要独立成分；然后采用分阶段淘汰的支持向量机分类机制进行识别。对两组人脸图像库的测试结果表明，基于SVM的方法在识别率和识别时间等方面都取得了较好的效果。

2.2 说话人/语音识别

说话人识别属于连续输入信号的分类问题，SVM是一个很好的分类器，但不适合处理连续输入样本。为此，忻栋等[15]引入隐式马尔可夫模型HMM，建立了SVM和HMM的混合模型。HMM适合处理连续信号，而SVM适合于分类问题；HMM的结果反映了同类样本的相似度，而SVM的输出结果则体现了异类样本间的差异。为了方便与HMM组成混合模型，首先将SVM的输出形式改为概率输出。实验中使用YOHO数据库，特征提取采用12阶的线性预测系数分析及其微分，组成24维的特征向量。实验表明HMM和SVM的结合达到了很好的效果。

2.3 文字/手写体识别

贝尔实验室对美国邮政手写数字库进行的实验[1]，人工识别平均错误率是2.5%，专门针对该特定问题设计的5层神经网络错误率为5.1%(其中利用了大量先验知识)，而用3种SVM方法(采用3种核函数)得到的错误率分别为4.0%、4.1%和4.2%，且是直接采用16×16的字符点阵作为输入，表明了SVM的优越性能。

手写体数字0～9的特征可以分为结构特征、统计特征等。柳回春等[16]在UK心理测试自动分析系统中组合SVM和其他方法成功地进行了手写数字的识别实验。另外，在手写汉字识别方面，高学等[17]提出了一种基于SVM的手写汉字的识别方法，表明了SVM对手写汉字识别的有效性。

2.4 图像处理

(1)图像过滤。一般的互联网色情图像过滤软件主要采用网址库的形式来封锁色情网址或采用人工智能方法对接收到的中、英文信息进行分析甄别。段立娟等[18]提出一种多层次特定类型图像过滤法，即以综合肤色模型检验，支持向量机分类和最近邻方法校验的多层次图像处理框架，达到85%以上的准确率。

(2)视频字幕提取。视频字幕蕴含了丰富语义，可用于对相应视频流进行高级语义标注。庄越挺等[19]提出并实践了基于SVM的视频字幕自动定位和提取的方法。该方法首先将原始图像帧分割为N*N的子块，提取每个子块的灰度特征；然后使用预先训练好的SVM分类机进行字幕子块和非字幕子块的分类；最后结合金字塔模型和后期处理过程，实现视频图像字幕区域的自动定位提取。实验表明该方法取得了良好的效果。

(3)图像分类和检索。由于计算机自动抽取的图像特征和人所理解的语义间存在巨大的差距，图像检索结果难以令人满意。近年来出现了相关反馈方法，张磊等[20]以SVM为分类器，在每次反馈中对用户标记的正例和反例样本进行学习，并根据学习所得的模型进行检索，使用由9 918幅图像组成的图像库进行实验，结果表明，在有限训练样本情况下具有良好的泛化能力。