第5卷第3期 空军工程大学学报(自然科学版) Vol.5 No.3
2004年6月 JOURNAL OF AIR FORCE ENGINEERING UNIVERSITY(NATURAL SCIENCE EDITION) Jun. 2004

支持向量机研究与应用

王晓丹，王积勤
(空军工程大学 导弹学院，陕西 三原 713800)

摘　要:支持向量机是在统计学习理论基础上发展起来的一种新的机器学习方法,同时也是到目前为止统计学习理论最成功的实现。支持向量机在模式识别、回归估计、函数逼近等领域有了广泛的应用。论述了支持向量机的研究、应用状况,指出了支持向量机研究和应用中待解决的一些问题和今后进一步的研究方向。

关键词:支持向量机;优化算法;训练;分类
中图分类号:TP391　**文献标识码**:A　**文章编号**:1009-3516(2004)02-0049-07

支持向量机SVM(Support Vector Machine)是在统计学习理论[1]基础上发展起来的一种新的机器学习方法。支持向量机又称为支持向量网络[2],具有理论完备、适应性强、全局优化、训练时间短、泛化性能好等优点,已经成为目前国际、国内研究的热点[3]。

支持向量机的核心内容从1992年才开始提出,是到目前为止统计学习理论最成功的实现,目前仍处于不断发展阶段。据统计,1999年以前国际上公开发表的有关统计学习理论的文章不足50篇,但1999年以后这方面的文章有数千篇之多,其应用范围和成果不断扩大。近年来专门或相关的国际、国内会议(如ICANN,ICONIP,ICML,CCSP等)也都列有统计学习理论和支持向量机的专题。

本文将在对支持向量机分类机理分析的基础上,对支持向量机目前的研究、应用作一综述,最后指出对支持向量机进一步研究和应用亟待解决的一些重要问题。

1　SVM分类机理分析

SVM是从线性可分情况下的最优分类超平面发展而来的。对两类分类问题,设训练样本集为$(\boldsymbol{x}_i, y_i)$，$i=1,2,\cdots,n$，n为训练样本的个数，$\boldsymbol{x}_i \in R^d$为训练样本，$y_i \in \{+1,-1\}$是输入样本$\boldsymbol{x}_i$的类标记(期望输出)。SVM算法的出发点是寻找最优分类超平面。最优分类超平面不但能将所有样本正确分开(训练错分率为0),而且能够使两类间的边际(margin)最大,边际定义为训练数据集到该分类超平面的最小距离之和。最优分类超平面意味着对测试数据平均分类误差最小。

若样本集线性可分,d维空间中线性判别函数$g(x)=\boldsymbol{w}\cdot x+b$，分类面方程为$\boldsymbol{w}\cdot x+b=0$。将判别函数归一化,使离分类面最近样本的$|g(x)|=1$。若分类面对所有样本都能够正确分类,则满足:

$$y_i(\boldsymbol{w}\cdot x_i+b)\geqslant 1\,,\quad i=1,2,\cdots,n \tag{1}$$

容易证明,两个超平面$H_1:\boldsymbol{w}\cdot x_1+b=1$和$H_2:\boldsymbol{w}\cdot x_1+b=-1$间的边际宽度为$d=2/\|w\|$。使边际最大等价于使$\|w\|$(或$\|w\|^2$)最小。所以,分类超平面$H:\boldsymbol{w}\cdot x+b=0$为最优,当且仅当$(\boldsymbol{w},b)$是下面优化问题的解:

min:　$\|\boldsymbol{w}\|^2/2$　　$1,2,\cdots,n$

subject to:　$y_i(\boldsymbol{w}\cdot \boldsymbol{x}_i+b)\geqslant 1$，　$i=1,2,\cdots,n$

用lagrange乘子法解上述QP问题。为此,可以定义如下的lagrange函数:

收稿日期:2003-07-18
作者简介:王晓丹(1966-),女,陕西汉中人,副教授,博士(后),主要从事智能信息处理研究.

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\alpha_i[y_i(w \cdot x_i + b) - 1] \tag{2}$$

其中 $\alpha_i(i = 1,2,\cdots,n)$ 为正 lagrange 乘子,需要对 w 和 b 求 lagrange 函数的极小值。

因上述问题是一凸 QP 问题,等价于解对偶问题:

$$\max \cdot \quad L_D = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\|w\|^2 - \sum_{i,i=1}^{n}\alpha_i\alpha_j y_i y_j x_i \cdot x_j \tag{3}$$

$$\text{subject to} \cdot \quad \sum_{i=1}^{n}\alpha_i y_i = 0, \quad \alpha_i \geqslant 0, \quad \imath = 1,2,\cdots,n \tag{4}$$

对每个训练样本 x_i都有一个 lagrange 乘子 α_i。解 $\alpha_i > 0$ 所对应的 x_i称为支持向量,它满足式(1) 中的等式约束。支持向量距最优超平面最近,通常只是全体样本中的很少一部分,是对分割两类非常重要的样本点。

若 α_i为最优解,则 $w = \sum_{i=1}^{n}\alpha_i^* y_i x_i$ (5)

所以,得到的分类决策函数为:$g(x) = \text{sgn}(w^* \cdot x + b^*) = \text{sgn}(\sum_{i\in SV}\alpha_i^* y_i x_i \cdot x + b^*)$ (6)

其中 x 为待分类样本,SV 为支持向量集, b^* 为分类阈值,可用任一支持向量求得。

若样本集线性不可分,可以引入正松弛变量 ξ_i,$i = 1,2,\cdots,n$。约束条件式(1) 变为

$$y_i(w \cdot x_i + b) \geqslant 1 - \xi_i, \quad \imath = 1,2,\cdots,n \tag{7}$$

当训练样本 x_i不能满足式(1) 时,$\xi_i > 0$,表示 x_i被误分类,此样本为造成线性不可分的点;否则 $\xi_\iota = 0$。可以看作训练样本关于最优分类超平面的偏差, $\sum_{i=1}^{n}\xi_i$ 是训练误差的上界。

为求得线性不可分情况下的最优分类超平面,称为广义最优分类超平面,需要在条件式(7) 的约束下,$\min:\frac{1}{2}\|w\|^2 - C\sum_{i=1}^{n}\xi_i$。其中 C 为某个指定的常数,它实际上起控制对错分样本惩罚程度的作用。

用 lagrange 乘子法及对偶原理,得到线性不可分情况下的对偶问题:

$$\max: L_D = \sum_{i=1}^{n}\alpha_\iota - \frac{1}{2}\|w\|^2 - \sum_{i\ i=1}^{n}\alpha_i\alpha_j y_i y_j x_i \cdot x_j \tag{8}$$

$$\text{subject to:} \sum_{\iota=1}^{n}\alpha_i y_i = 0, C \geqslant \alpha_i \geqslant 0, i = 1,2,\cdots,n \tag{9}$$

求解得到的分类决策函数同式(6)。

对非线性可分数据集,超平面的分类能力毕竟有限,为此 SVM 通过对待分类数据进行非线性特征映射 $\phi(x):R^d \to E$,将数据映射到某一更高维特征空间 E 中,从而能够线性可分,然后在 E 中构造(广义) 最优分类超平面。由于优化函数和分类函数都只涉及样本间的内积运算$(x_\iota \cdot x)$,因此,在变换后的高维空间 E 中也只需进行内积运算$(\phi(x_i) \cdot \phi(x))$ 。根据泛函有关理论,如果核函数 $K(x_i,x)$ 满足 Mercer 条件,则它对应某变换空间中的内积,$\phi(x_i) \cdot \phi(x) = K(x_i,x)$ 。因此采用适当核函数 $K(x_i,x)$ 就可代替向高维空间的非线性映射,实现某一非线性变换后的线性分类。此时优化问题为

$$\max: L_D = \sum_{\iota=1}^{n}\alpha_i - \frac{1}{2}\|w\|^2 - \sum_{i,i=1}^{n}\alpha_i\alpha_j y_i y_j x_i \cdot x_j K(x_i,x_j) \tag{10}$$

$$\text{subject to:} \sum_{i=1}^{n}\alpha_i y_i = 0, \quad C \geqslant \alpha_i \geqslant 0, \quad \imath = 1,2,\cdots,n \tag{11}$$

相应的 SVM 的决策函数为:$g(x) = \text{sgn}(\sum_{\forall x_i \in SV}\alpha_i y_i K(x_i,x) + b)$ (12)

由优化目标函数式(10) 可以看出,SVM 方法的训练复杂度与特征维数无关,但却受到训练集规模 n 的制约。它需要计算所有训练样本两两之间的核函数,产生一个 $n \times n$ 的核函数矩阵,当样本点数目很大时,存储该核函数矩阵需要大量内存。例如,当样本点数目超过 4 000 时,存储核函数矩阵需要多达 128MB 内存。同时,SVM 在二次型寻优过程中要进行大量矩阵运算,多数情况下,寻优算法是占用 SVM 训练算法时间的主要部分。由式(12) 可以看出,SVM 方法构造的分类函数的复杂程度取决于支持向量的个数。

以下几点可以概括 SVM 方法的特点:① 非线性映射是 SVM 方法的理论基础,SVM 利用内积核函数代替向高维空间的非线性映射;② 对特征空间划分的最优超平面是 SVM 的目标,最大化分类边际的思想是 SVM

方法的核心;③ 支持向量是 SVM 的训练结果,在 SVM 分类决策中起决定作用的是支持向量。

目前,对支持向量机的研究主要集中在对支持向量机本身机理的研究和完善以及加大支持向量机应用研究的深度和广度两方面。

2 SVM 的研究

对 SVM 本身性质的研究是 SVM 的一个重要研究内容。本文以下就 SVM 的训练算法、分类算法、多类算法、核函数及选择等热点问题分别加以讨论。

2.1 SVM 训练算法

传统的利用标准二次型优化技术解决对偶问题的方法,是 SVM 训练算法慢及受到训练样本集规模制约的主要原因。目前已提出了许多解决方法和改进算法,主要是从如何处理大规模样本集的训练问题、提高训练算法收敛速度等方面改进。以下分为分解方法、修改优化问题法、增量学习法、几何方法等分别讨论。

2.1.1 分解方法

分解方法是 SVM 训练一般采用的途径。块算法、固定工作变量集方法、顺序最小优化方法等最为常见。考虑到去掉 Lagrange 乘子等于零的训练样本不会影响原问题的解,块算法(Chunking algorithm)的出发点就是在迭代过程中按照某种准则逐步排除非支持向量[4]。当支持向量数目远小于训练样本数目时,块算法的效率较高。

固定工作变量集方法思想是在迭代过程中,当前求解子问题的优化变量数目不变,即参与训练的样本集(工作变量集)规模固定。工作样本集大小固定在算法速度可以容忍的限度内,迭代过程选择一种合适的换入换出策略,将剩余样本中的一部分与工作样本集中的样本进行等量交换。Osuna[5] 针对 SVM 训练速度慢及时间空间复杂度大的问题,最早提出了该分解算法,并用于了人脸检测。文献[5] 中将训练样本分为工作样本集和非工作样本集,工作样本集中的样本个数为 q 个,q 远小于训练样本总数。

工作样本集大小的确定、如何确定工作样本集、如何确定合适的迭代策略是固定工作样本集方法的主要问题。SVMlight[6] 中做了以下改进:在工作样本集的选择上,SVMlight 中是沿着最速下降可行方向 d,由非零元素对应的 q 个优化变量构成工作样本集。已经证明了只要最速下降可行方向 d 存在,则用相应子集构成的子问题可以进一步优化,而子问题的可行解也是原问题的可行解。这就解决了工作样本集不能包括所有支持向量的问题。在实现细节上,文献[6] 中提出了连续收缩(Shrinking) 策略来缩小优化问题的规模,从而使算法能较好地处理大规模的训练集问题。此外,该方法对常用的参数进行缓存,以提高训练速度。SVMlight 常被用作各种算法比较的标准。

顺序最小优化方法 SMO(Sequential Minimal Optimization)[7] 可以说是 Osuna [5]分解算法的极端特例,其工作样本集中只有两个样本。它把二次型寻优算法简化为线性寻优问题。SMO 特别适合稀疏样本。其工作集的选择采用启发式,而不是传统的最陡下降法。算法主要耗时是在最优条件的判断上。文献[8] 中对 SMO 进行了改进,在迭代过程中的判优条件和循环策略上做了一定的修改,加快了算法的速度。文献[9] 中通过实验分析提出了一种简单的工作集选择方法,对复杂情形,采用该方法的 SVM 能较快地收敛。

文献[10] 在 SVM 分解方法的框架下提出了一种快速 SVM 训练算法。对大规模手写数字数据库 MNIST 的实验表明,所提出的算法较改进的 SMO[11] 快约 9 倍。

2.1.2 修改优化问题法

通过修改目标函数、约束条件来简化优化问题本身,是提高 SVM 算法效率的途径之一。

由 SVM 原理可知,对于错分样本,SVM 的惩罚项是对松弛因子的累加,但这种累加不必一定是线性的[2]。因此,最近点算法 NPA(Nearest Point Algorithm)[12] 的基本思想是将 SVM 原问题的惩罚项由线性累加 $C\sum_{i=1}^{n}\xi_i$ 改为二次累加 $C\sum_{i=1}^{n}\xi_i^{\,2}$。通过这样修改,可以将 $\sqrt{C}\xi_i$ 作为 w 的一个分量,同时将样本维数增加一维并规定新的一维为常数 $y_i/\sqrt{C}$,改写后的目标函数中将不再包含惩罚项,而约束条件中也没有了松弛因子,从而使问题转化为无错分情况下求最大边际的问题。

连续超松弛方法 SOR(Successive Overrelaxation)[13] 通过在原目标函数中加一项 b^2,从而使其对偶问题

多出一项($\sum_{i=1}^{n}\alpha_i^2)/2$,而约束条件则少了一项等式约束。修改后的对偶问题变为边界约束条件下的二次规划问题,适合迭代求解。同时应用矩阵分解技术,每次只更新 Lagrange 乘子的一个分量,从而不需将所有样本放入内存,大大提高了收敛速度。

最小二乘支持向量机(Least Squares)[14] 将最小二乘引入到 SVM 中,其目标函数为 $\min\limits_{w,b,e}J_{LS}(w,b,e)=\frac{1}{2}w^{T}W+\gamma\frac{1}{2}\sum_{i=1}^{n}e_i$,约束条件为:$y_i[w^{T}\phi(x_i)+b]=1-e_i,\ i=1,2,\cdots,n$。定义相应的 Lagrange 函数,并运用 KT 最优条件,可得到一组线性方程。通过解线性方程组可得到问题的解。该方法显示出了较低的计算代价。

2.1.3 增量学习方法

上述两种方法均假设训练集大小是固定的,但现实问题中这一要求在很多情况下是不能满足的。因此,希望学习机的学习精度应随应用过程中样本集的积累而逐步提高,即学习机应具有增量(Incremental Learning) 学习能力。经典 SVM 学习算法并不直接支持增量学习。

文献[15] 中给出了 SVM 增量训练算法。它们直接通过支持向量实现增量 SVM 学习,即每次只选一小批常规二次规划算法能处理的训练样本,然后保留支持向量,抛弃非支持向量,和新进来的样本混合进行训练,直到训练样本用完。这种方法可以实现近似的增量学习。

为解决加入新样本后的 SVM 训练问题,文献[16] 中用统计力学上的 Adatron 方法训练 SVM 中的系数,它将系数的求解看成系统由不稳定态到稳定态的变化过程。由 Adatron 算法改进得出的 Kernel - Adatron 算法[17],通过在线学习构建了大边际超平面,该算法实现简单,但只对于可分数据集有效。

文献[18] 提出了增量式求解全局优化问题精确解的方法,增加一个样本或减少一个样本对 Lagrange 系数和支持向量的影响实验表明算法是有效的。

文献[19] 提出了一种基于遗忘因子的 SVM 增量学习算法 $\boldsymbol{\alpha}$ - ISVM。该算法通过在增量学习中逐步积累样本的空间分布知识,使得对样本进行有选择地遗忘成为可能。文献[20] 中引入了增量线性近似支持向量机 PSVM(Proximal Support Vector Machine),PSVM 的分类精度与常规 SVM 相同,但训练速度显著提高。

2.1.4 几何方法

该类方法利用了训练集中的几何信息,从 SVM 的几何意义出发求解问题。

文献[21] 中利用了训练向量的结构信息,提出了用几何方法提取卫向量集,并使用卫向量集构建 SVM 的优化决策面的方法。

文献[22] 把 SVM 原理建立在距离空间上,设计出基于邻域原理的计算海量数据支持向量的算法,并进行了实验分析。研究表明,在大规模样本情况下,用邻域原理方法求解支持向量速度极快,同时对计算机资源要求很低。使用邻域原理方法求出的是一组近似最优的支持向量。邻域原理求支持向量的过程本质上是简化 SVM 中二次归划目标函数的 Hessian 矩阵的过程。该方法不但几何意义明确,而且计算速度快,每次可以消掉内积矩阵的多行多列,所需内存开销小。

2.2 SVM 分类算法

训练好 SVM 分类器后,得到的支持向量被用来构成决策分类面。对于大规模样本集问题,SVM 训练得到的支持向量数目很大,则进行分类决策时的计算代价就是一个值得考虑的问题[23]。

缩减集(Reduced Set)SVM 方法[23] 采用缩减集代替支持向量集,缩减集中的向量不是支持向量,数目比支持向量少,但它们在分类决策函数中的形式与支持向量相同。该方法的目的是用缩减集构造的决策分类面来逼近支持向量决策分类面,逼近准则是最小平方误差,即 $\min\left\|\sum_{i=1}^{N_Z}\gamma_i\phi(z_i)-\sum_{i=1}^{N_S}\alpha_i y_i\phi(x_i)\right\|$,其中 ϕ 是非线性映射,z_i 是缩减集向量,N_Z 是缩减集向量个数,γ_i 是相应权系数,x_i 是支持向量,N_S 是支持向量个数,α_i 是相应的 lagrange 乘子,y_i 是类别标号。该方法虽然缩减了分类器的计算量,但用最小二乘法构造近似分类器的计算量却是非常大的。

2.3 多类 SVM 算法

设现需要将 n 个样本分 k 类,x_i 为样本,y_i 为样本所属类别,即有(x_i,y_i),$i=1,2,\cdots,n$,$x_i\in R^n$,$y_i\in\{1,2,\cdots,k\}$。SVM 本质上是两类分类器,常用的 SVM 多值分类器构造方法有:

一对多方法(One - Against - The - Rest):在第 k 类和其它 $k-1$ 类之间构建超平面。在这种方式下,系

统仅构建 k 个SVM,每一个SVM分别将某一分类的数据从其它分类的数据中鉴别出来。对第 i 个SVM用第 i 个类中的训练样本作为正训练样本,而将其它的样本作为负训练样本。

一对一方法(One - Against - One):为任意两个类构建超平面,共需训练 $k\times(k-1)/2$ 个2值SVM分类器。在这种方式下,是对 k 个分类的训练集进行两两区分。测试时,常用投票法,得票最多(Max Wins)的类为测试样本所属的类。该方法的缺点是分类器的数目随分类数的增加而迅速增加,导致在决策时速度很慢。

SVM决策树(SVM Decision Tree)方法[24]:将SVM和二叉决策树结合起来,构成多类分类器。该方法的缺点是如果在某个节点上发生了分类错误,则会把分类错误延续到该节点的后续下一级节点上。

多类SVM(Multi - class objective functions)[25]:通过改写SVM2值分类中优化的目标函数,使得其满足多值分类的需要,在构造决策函数时同时考虑所有的类,则可将原始优化问题推广为 $\min:\Phi(w,\xi)=\frac{1}{2}\sum_{m=1}^{k}(w_m\cdot w_m)+C\sum_{i=1}^{n}\sum_{m\neq y_i}\xi_i^m$;subject to:$(\boldsymbol{w}_i\cdot\boldsymbol{x}_i)+b_{y_i}\geqslant(w_m\cdot\boldsymbol{x}_i)+b_m+2-\xi_i^m,\xi_i^m\geqslant0,i=1,2,\cdots,n$,$m\in\{1,2,\cdots,k\}$,相应的决策函数变为 $f(x)=\arg\max_k[(\boldsymbol{w}_i\cdot x)+b_i],i=1,2,\cdots,k$。多类SVM方法的二次规划形式复杂,求解计算量大,一般较少采用。

现有的多值分类算法一般直接采用一对一或其改进方法实现,如决策导向无环图算法[26]。

2.4　核函数及选择

SVM由训练样本集和核函数完全描述,因此采用不同核函数 $K(x,\boldsymbol{x}_i)$,就可以构造实现输入空间中不同类型的非线性决策面的学习机,导致不同的支持向量算法。在实际问题中,通常是直接给出核函数。目前研究最多、最常用的核函数有:线性核函数 $K(x,\boldsymbol{x}_i)=(x\cdot\boldsymbol{x}_i)$;多项式核函数 $K(x,\boldsymbol{x}_i)=((x\cdot\boldsymbol{x}_i)+1)^q$,$q$ 为参数;高斯核函数:$K(x,\boldsymbol{x}_i)=\exp(1-(x-\boldsymbol{x}i)^2/\sigma^2)$;Sigmoid核函数 $K(x,\boldsymbol{x}_i)=\tanh(v(x\cdot\boldsymbol{x}_i)+c)$;实多项式核函数 $K(x,\boldsymbol{x}_i)=(1-(x\cdot\boldsymbol{x}_i)^q)/(1-(x\cdot\boldsymbol{x}_i))$,$q$ 为参数,$-1<(x\cdot\boldsymbol{x}_i)<1$;完全多项式核函数 $K(x,\boldsymbol{x}_i)=((x\cdot\boldsymbol{x}_i)/a+b)^q$,$a$、$b$、$q$ 为参数。

由于核函数的重要性,如何去构造、选择核函数及参数成为人们关注的问题。一旦核函数确定,对于SVM分类器则只有一个参数可调整(误差惩罚参数 C)。通常的做法是找出样本集分布特点与最优分类器之间可能的对应关系,根据待训练样本的一些先验知识选择分类器类型和参数;或直接构造新的类型,可以预先确定或在训练过程中逐步优化。文献[27]中对将先验知识用于核函数的选择进行了研究。

3　SVM应用

加大SVM应用的深度和广度,是SVM的另一个重要研究内容。到目前,SVM已用于数据分类、回归估计、函数逼近等领域。

SVM应用最为广泛的当属在模式识别领域,已成功地用于许多模式识别问题,包括孤立手写字符识别[2,23]、计算机视觉[28]、网页或文本自动分类[29]、说话人识别[30]、人脸检测[5]、人脸识别[31]、头部姿态识别[32]、性别分类[33]、基因分类[34]、目标识别[35]、遥感图象分析[36]等。在函数回归、估计、函数逼近[37]、密度估计[38]、时间序列预测[39]、非线性系统控制[11]等问题中,都有SVM的成功应用。

此外,在解决数据压缩、文本过滤、数据挖掘[4]、股市预测[40]、计算机入侵检测[41]等应用问题中,SVM都显示出了良好的性能。

4　需要解决的问题和进一步的研究方向

理论和实践结果表明,SVM在存在全局优化、训练时间短、泛化性能好、算法复杂度与特征空间维数无关等优点的同时,存在着以下困难和问题:①核函数和参数的选择缺乏理论指导。SVM核函数的选择影响着分类器的性能,如何根据待解决问题的先验知识、选择和构造合适的核函数,如何根据实际样本数据、确定核函数的参数等问题都缺乏相应的理论指导;②训练大规模数据集的问题。训练集的规模和训练速度是一对矛盾。如何训练和分类大数据集,解决训练样本规模和训练速度间的矛盾、解决支持向量数目与分类速度间的矛盾,仍是需进一步解决的问题;③尽管训练多类SVM的算法已被提出,但SVM用于多类分类问题时的有效算法、多类SVM的优化设计仍是一个需进一步研究的问题;④具有增量学习能力是许多在线训练、实

时应用的关键。需要找到有效的算法,同时满足在线学习和期望风险控制的要求。

因此,在对 SVM 方法进行进一步的机理分析和实验分析、拓宽 SVM 的应用领域的同时,针对上述几个方面将是 SVM 今后研究的方向,随着研究者们的努力和实际应用需求的不断增长,对 SVM 研究和应用必将获得更深入和广泛的成果。

参考文献:

[1] Vapnik V N. The Nature of Statistical Learning Theory[M]. New York: Springer - Verlag, 1995.

[2] Cortes C, Vapnik V. Support Vector Networks[J]. Machine Learning, 1995, 20:273 - 297.

[3] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.

[4] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2):121 - 167.

[5] Osuna E, Freund R, Girosi F. Training Support Vector Machines: An Application to Face Detection[A]. Proceedings of IEEE Conference on CVPR97[C]. Puerto Rico, 1997. 130 - 136.

[6] Joachims T. Making Large - Scale SVM Learning Practical[A]. Advances in Kernel Methods: Support Vector Learning[C]. Cambridge, MA: MIT Press, 1999.

[7] Platt J C. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines[A]. Advances in Kernel Methods: Support Vector Learning[C]. Cambridge, MA: MIT Press, 1999. 185 - 208.

[8] Keerthi S S. Improvements to Platt's SMO Algorithm for SVM Classifier Design[J]. Neural Computaion, 2001, 13(3):637 - 649.

[9] Chih Wei Hsu, Chih Jen Lin. A Simple Decomposition Method for Support Vector Machines[J]. Machine Learning, 2002, 46(1):291 - 314.

[10] Jain Xiong Dong, Ching Y S, Adam K. A Fast SVM Training Algorithm[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2003, 17(3):367 - 384.

[11] Suykens J A K. Optimal Control By Least Squares Support Vector Machines[J]. Neural Networks, 2001, 14(1):23 - 25.

[12] Keerthi S S. A Fast Iterative Nearest Point Algorithm for Support Vector Machine Classifier Design[J]. IEEE Trans. on Neural Networks, 2000, 11(1):124 - 136.

[13] Mangasarian Olvi L, Musicant David R. Successive Overrelaxation for Support Vector Machines[J]. IEEE Trans. on Neural Networks, 1999, 10(5):1032 - 1037.

[14] Suykens J A K, Vandewalle J. Least Squares Support Vector Machine Classifiers[J]. Neural Processing Letters, 1999, 9(3):293 - 300.

[15] Syed A N. Incremental learning With Support Vector Machines[A]. Proceedings of IJCAIGG[C]. Sweden: Stockhdm, 1999.

[16] Mattera D, Palmieri F, Haykin S. An Explicit Algorithm for Training Support Vector Machines[J]. IEEE Signal Processing Letters, 1999, 6(9):243 - 244.

[17] Frieβ T T, Cristianini N, Campbell C. The Kernel Adatron Algorithm: A Fast And Simple Learning Procedure for Support Vector Machines[A]. Proceedings of 15th Intl. Conf. Machine Learning, ICML98. [C]. Morgan Kaufmann Publishers, 1998. 188 - 196.

[18] Cauwenberghs G, Poggio T. Incremental and Decremental Support Vector Machine Learning[J]. Machine Learning, 2001, 44(13):409 - 415.

[19] 萧 嵘. 一种 SVM 增量学习算法 α - ISVM[J]. 软件学报, 2001, 12(12):1818 - 1824.

[20] Fung G, Mangasarian O L. Incremental Support Vector Machine Classification[A]. Proceedings of the Second SIAM International Conference on Data Mining[C]. SIAM, 2002. 247 - 260.

[21] Yang M H, Ahuja N. A Geometric Approach to Train Support Vector Machines[A]. Proceeding of CVPR 2000[C]. Hilton Head Island, 2000. 430 - 437.

[22] 张文生, 丁 辉, 王 珏. 基于邻域原理计算海量数据支持向量的研究[J]. 软件学报, 2001, 12(5):711 - 720.

[23] Burges C J C, Schölkopf B. Improving The Accuracy and Speed of Support Vector Learning Machines[A]. Advances in Neural Information Processing Systems[C]. Cambridge, MA: MIT Press, 1997. 375 - 381.

[24] Bennett K P, Blue J A. A Support Vector Machine Approach to Decision Trees[A]. Proceedings of IJCNN98[C]. Alaska: Anchorage, 1997. 2396 - 2401.

[25] Weston J, Watkins C. Multi - class Support Vector Machines[R]. Royal Holloway College, TR CSD - TR - 98 - 04, 1998.

[26] Platt J. Large Margin DAGs for Multiclass Classification[A]. Advances in Neural Information Processing Systems, Cambridge [C]. MA: MIT Press, 2000. 547 - 553.

[27] Schölkopf B, Simard P Y, Smola A J, et al. Prior Knowledge in Support Vector Kernels[A]. Advances in Neural Information Processings Systems[C]. Cambridge, MA: MIT Press, 1998. 640 - 646.

[28] Pontil M, Verri A. Support Vector Machines for 3 - d Object Recognition[J]. IEEE Trans. PAMI, 1998, 20(5): 637 - 646.

[29] 李晓黎,刘继敏,史忠值.基于支持向量机与无监督聚类相结合的中文网页分类器[J].计算机学报,2001,24(1):62 - 68.

[30] Gish H, Schimdt M. Text - Indepentent Speaker Identification[J]. IEEE Trans. Signal Processing Magazine, 1994, 42(1): 18 - 32.

[31] Kumar V, Poggio T. Learning - Based Approach to Real Time Tracking and Analysis of Faces[A]. Proceedings of the Fourth International Conference on Face and Gesture Recognition[C]. France: Grenoble, 2000. 96 - 101.

[32] Huang J. Face Pose Discrimination Using Support Vector Machines[A]. Proceedings of 14TH International Conference on Pattern Recognition[C]. 1998. 154 - 156.

[33] Walavalkar L. Support Vector Learning for Gender Classification Using Audio and Visual Cues[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2003, 17(3): 417 - 439.

[34] Mukherjee S. Classifying Microarray Data Using Support Vector Machines[A]. A Practical Approach to Microarray Data Analysis[C]. Boston: Kcuwer Academic, 2003. 166 - 185.

[35] Zhao Q, Principe J C. SVM for SAR ATR[J]. IEEE. Trans. AES, 2001, 37(2): 643 - 653.

[36] Brown M, Lewis H G, Gunn S R. Linear Spectral Mixture Models and Support Vector Machines for Remote Sensing[J]. IEEE Trans. Geoscience and Remote Sensing, 1998, 38(5): 2346 - 2360.

[37] Vapnik V N, Golowich S, Smola A. Support Vector Method for Function Approximation, Regression, Estimation, and Signal Processing[A]. Advances in Neural Information Processing Systems[C]. Cambridge, MA: MIT Press, 1997. 281 - 287.

[38] Weston J. Support Vector Density Estimation[A]. Advances in Kernel Methods: Support Vector Learning[C]. Cambridge, MA: MIT Press. 1999. 293 - 306.

[39] Müller K R, Smola A J, Rtsch G, et al. Predicting Time Series With Support Vector Machines[A]. Proceedings of ICANN'97 [C]. Springer, 1997. 999 - 1005.

[40] Haiqin Yang, Laiwan Chan, Irwin King. Support Vector Machine Regression for Volatile Stock Market Prediction[A]. Third International Conference on Intelligent Data Engineering and Automated Learning(IDEAL'02)[C]. Springer, 2002. 391 - 396.

[41] Mike Fugate, James R, Gattikers. Computer Intrusion Detection with Classification and Anomaly Detection Using SVMs[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2003, 17(3): 441 - 458.

(编辑:田新华)

Research and Application of Support Vector Machine

WANG Xiao - dan, WANG Ji - qin

(The Missile Institute, Air Force Engineering University, Sanyuan, Shaanxi 713800, China)

Abstract: A support vector machine is a new machine learning technique developed on the basis of statistical learning theory, and it is the most successful realization of statistical learning theory. For machine learning tasks involving pattern classification, regression estimation, and function approximation, the support vector machine has increasingly become a popular tool. In this paper a survey of the recent new development on the research and application of the support vector machines is made. Some important issues to be investigated in application and the direction of research of the support vector machine have been pointed out simultaneously.

Key words: support vector machine; optimal algorithm; training; classification