第 34 卷 第 3 期 计 算 机 工 程 2008 年 2 月
Vol.34 No.3 Computer Engineering February 2008
·博士论文· 文章编号：1000—3428(2008)03—0027—02 文献标识码：A 中图分类号：TP391

医学图像决策支持系统中的 SVM 算法

孙 蕾
(西安电子科技大学经济管理学院，西安 710071)

摘 要：支持向量机(SVM)方法是利用最优分类面(线)将两类样本在特征空间或输入空间中无错误地分开，而且要使两类的分类空隙最大。因此标准的 SVM 方法需要求解二次规划问题，计算量很大。该文以一个医学决策支持系统为应用背景，介绍一种解决该问题的新方法。在 UCI 数据集和所开发的决策支持系统上的应用表明，该算法简便可行，具有更高的精度和更快的速度。

关键词：支持向量机；分类算法；决策支持

Algorithm of Support Vector Machine for Medical Image Decision Support System

SUN Lei
(School of Economic and Management, Xidian University, Xi'an 710071)

【Abstract】Support Vector Machine (SVM) is to correctly classify samples into two parallel planes in input or feature space by optimal planes (lines). And the margin between the two classes is made to be the largest. The standard SVM requires to solve quadratic program that needs considerable computational time. Based on a concrete decision support system of medical images, a novel algorithm is introduced to solve the problem. Experimental results on UCI and a developed decision support system demonstrate that the presented algorithm is simple, feasible, and faster with better precision.

【Key words】Support Vector Machine(SVM); classification algorithm; decision support

1 概述

决策支持系统是针对某一类型的半结构化决策问题，集数据库技术、人工智能理论和建模理论等技术为一体的综合体系，为管理者做出正确决策提供帮助的人机交互系统。面对复杂的决策问题，试图完全用数学模型进行精确刻画是不现实的，即使对某些问题可行，但求解与分析也是非常困难的。因此，从 20 世纪 90 年代初开始，人们就借助新发展的信息技术来处理或支持处理复杂的决策问题。

基于数据的机器学习是现代智能信息技术中十分重要的一个方面，主要研究如何从一些观测数据中得出目前尚不能通过学习原理分析得到的规律，利用这些规律去分析客观对象，对未来数据或无法观测的数据进行预测。而预测是决策的基础，正确的预测是正确决策的前提和依据。在机器学习方面，主要采用如神经元网络、基于案例的推理等传统的方法。但传统的机器学习算法，如神经元网络，存在着过学习和推广性不好、网络结构的确定必须依靠经验、容易陷入局部极小点问题等缺点。支持向量机(Support Vector Machine, SVM)作为一种新的机器学习方法，是在统计学习理论的基础上发展起来的[1]。它基于结构风险最小化原则，能有效地解决过学习问题，具有良好的推广性能和较好的分类精确性。当前，SVM 已被用于人脸识别、医疗诊断、数据挖掘等方面[2]。标准的支持向量机因具有数学形式简单、几何解释直观、全局最优、学习速度快、泛化能力优良、适合处理高维数据等特点而被成功地用于许多分类和回归的问题中[1]。其线性分类器是基于数据的输入空间实现分类，而非线性分类器则是基于高维特征空间，用核函数 $K(x,x')$ 代替非线性映射 $\phi(x)$ 来实现分类，为非线性预测提供可能。总之，SVM 的训练分类问题实质上是一个二次规划(QP)问题，本文所提算法，能够在很大程度上减少其计算量，从而提高速度。

2 线性 SVM 二类分类器

Vapnik给出的标准线性SVM(其核函数为线性核)[3] 如下：

$$\min \nu e'y+\frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\omega} \qquad \nu \text{ 是大于 0 的参数}$$

$$\text{s.t.} \quad \boldsymbol{D}(\boldsymbol{A}\boldsymbol{\omega}-e\gamma)+y\geq \boldsymbol{e} \quad \text{and} \quad y\geq 0 \tag{1}$$

矩阵$\boldsymbol{A}(m\times n)$表示给定$n$维空间$R^n$中的 m个样本，矩阵$\boldsymbol{D}$是一个$m\times m$的对角阵，其对角线上的元素是对应$\boldsymbol{A}$中元素的所属类别，即 1 或-1，其他元素为 0。常量X是超平面距原点的距离，$\boldsymbol{e}$是元素全部为 1 的m维列向量，$\boldsymbol{\omega}$ 是分类超平面的法向量：

$$x'\boldsymbol{\omega}=\gamma+1 \qquad x'\boldsymbol{\omega}=\gamma-1 \tag{2}$$

常量 γ 决定了分类面相对原点的距离。而当两类样本严格的线性可分时，即式(1)中的错分误差变量y=0，式(2)可将两类样本A^+和A^-正确无误地分类，表示如下：

$$\boldsymbol{A}\boldsymbol{\omega}\begin{cases}\geq \gamma+1 & \text{for} \quad D=+1\\ \leq \gamma-1 & \text{for} \quad D=-1\end{cases} \tag{3}$$

其最优分类超平面为

$$x'\boldsymbol{\omega}=\gamma \tag{4}$$

而当两类样本线性不可分时，式(3)则变为

作者简介：孙 蕾(1968－)，女，副教授、博士，主研方向：数据挖掘，决策分析

收稿日期：2007-02-10 **E-mail：**leisun68@yahoo.com

$$\begin{cases} \boldsymbol{A\omega}+y\geq\gamma+1 & \text{for} \quad D=+1 \\ \boldsymbol{A\omega}-y\leq\gamma-1 & \text{for} \quad D=-1 \end{cases} \tag{5}$$

此时两个分类面之间存在一个“ 软间隔 ”, 这是因为存在一个非负的分类误差 y，在式(1)中权值 v 使得误差变量 y 最小，从而得到一个似分类面式(4)。由此似分类面得到如下的线性分类器：

$$x'\boldsymbol{\omega}-\gamma\begin{cases} >0 & \text{then} \quad x\in A^{+} \\ =0 & \text{then} \quad x\in A+ \ \text{ or } \ A^{-} \\ <0 & \text{then} \quad x\in A^{-} \end{cases} \tag{6}$$

现在将式(1)做进一步的变化：将y的 1 范数改为 2 范数的平方，这样可以取消条件$y\geq 0$。再将X^2加到 $\boldsymbol{\omega}'\boldsymbol{\omega}$ 后，这样通过优化 $\boldsymbol{\omega}$ 和X就可以有效地扩大两个平行分类平面的距离，从而增强其泛化能力(文献[2, 4]表明此方法比标准的式(1)更有效)。

$$\min\frac{\nu}{2}\|y\|^2+\frac{1}{2}(\boldsymbol{\omega}'\boldsymbol{\omega}+\gamma^2)$$
$$\text{s.t.} \quad \boldsymbol{D}(\boldsymbol{A\omega}-e\gamma)+y\geq\boldsymbol{e} \tag{7}$$

用条件等式代替条件不等式，对式(7)做一步的简化：

$$\min\frac{\nu}{2}\|y\|^2+\frac{1}{2}(\boldsymbol{\omega}'\boldsymbol{\omega}+\gamma^2)$$
$$\text{s.t.} \quad \boldsymbol{D}(\boldsymbol{A\omega}-\boldsymbol{e}\gamma)+y=\boldsymbol{e} \tag{8}$$

式(8)在几何学上可以这样解释：分类面 $x'\boldsymbol{\omega}-\gamma=\pm 1$不再是最优分类超平面，但是可以认为它是一个似最优分类超平面。因为每个类中的点都聚集在各分类面周围，而且被($\boldsymbol{\omega}'\boldsymbol{\omega}+\gamma^2$)尽可能地分开。而在目标函数中的($\boldsymbol{\omega}'\boldsymbol{\omega}+\gamma^2$)就是在R^{n+1}的($\boldsymbol{\omega}$, γ)空间的两个分类面之间的距离的平方的倒数。

通过对拉格朗日函数式(9)中的 ω,γ,μ,y 分别求梯度，并令其为 0 而得到 KKT 条件：

$$L(\boldsymbol{\omega},\gamma,y,\mu)=\frac{\nu}{2}\|y\|^2+\frac{1}{2}\left\|\begin{bmatrix}\boldsymbol{\omega}\\ \gamma\end{bmatrix}\right\|^2-\mu'(\boldsymbol{D(A\omega}-\boldsymbol{e}\gamma)+y-\boldsymbol{e}) \tag{9}$$

KKT 条件:

$$\boldsymbol{\omega}-\boldsymbol{A}'\boldsymbol{D}\mu=0\,,\ \gamma+\boldsymbol{e}'\boldsymbol{D}\mu=0\,,\ \ \nu y-\mu=0\,,$$
$$\boldsymbol{D(A\omega}-\boldsymbol{e}\gamma)+y-\boldsymbol{e}=0 \tag{10}$$

用式(10)的前 3 个等式可以计算出 $\boldsymbol{\omega},\gamma,y$:

$$\boldsymbol{\omega}=\boldsymbol{A}'\boldsymbol{D}\mu\,,\ \ \gamma=-\boldsymbol{e}'\boldsymbol{D}\mu\,,\ \ y=\frac{\mu}{\nu} \tag{11}$$

再将以上公式代入式(10)中的最后一个等式中，得到如下μ的计算公式：

$$\mu=(\frac{I}{\nu}+\boldsymbol{D(AA'}+\boldsymbol{ee')D})^{-1}\boldsymbol{e}=(\frac{I}{\nu}+\boldsymbol{HH}')^{-1}\boldsymbol{e} \tag{12}$$

其中， $\boldsymbol{H}=\boldsymbol{D}[\boldsymbol{A}\quad -\boldsymbol{e}]$ (13)

利用 Sherman-Morrison-Woodbury 公式将式(12)转化为

$$\mu=\nu(I-\boldsymbol{H}(\frac{I}{\nu}+\boldsymbol{H'H})^{-1}\boldsymbol{H'})\boldsymbol{e} \tag{14}$$

这样，式(14)就将式(12)中的 $m\times m$ 维矩阵的逆降为(n+1)×(n+1)维，从而提高计算速度。线性支持向量机的算法如下：

(1)用式(13)定义矩阵 $\boldsymbol{H}$。

(2)用式(12)计算μ。

(3)将(2)中得出的μ代入式(11)的 3 个等式中，得到 $\boldsymbol{\omega},\gamma,y$。

(4)用式(6)对新的样本进行分类。

3 非线性 SVM 二类分类器

为了得到非线性分类器，将 $\boldsymbol{\omega}=\boldsymbol{A}'\boldsymbol{D}\mu$ 代入式(8)的限制条件中：

$$\min\frac{\nu}{2}\|y\|^2+\frac{1}{2}(\mu'\mu+\gamma^2)$$
$$\text{s.t.} \quad \boldsymbol{D}(\boldsymbol{AA}'\boldsymbol{D}\mu-\boldsymbol{e}\gamma)+y=\boldsymbol{e} \tag{15}$$

在这里用非线性核函数 $K(\boldsymbol{A},\boldsymbol{A}')$ 代替 $\boldsymbol{AA}'$ ，可得

$$\min\frac{\nu}{2}\|y\|^2+\frac{1}{2}(\mu'\mu+\gamma^2)$$
$$\text{s.t.} \ \boldsymbol{D}(K(\boldsymbol{A},\boldsymbol{A}')\boldsymbol{D}\mu-\boldsymbol{e}\gamma)+y=\boldsymbol{e} \tag{16}$$

以下用 K 代替 $K(\boldsymbol{A},\boldsymbol{A}')$ ，为此可以得到拉格朗日函数如下：

$$L(\mu,\gamma,\upsilon,y)=\frac{\nu}{2}\|y\|^2+\frac{1}{2}\left\|\begin{bmatrix}\mu\\ \gamma\end{bmatrix}\right\|^2-\upsilon'(\boldsymbol{D}(K\boldsymbol{D}\mu-\boldsymbol{e}\gamma)+y-\boldsymbol{e}) \tag{17}$$

其中，υ为拉格朗日乘子。

对此函数的μ, γ, υ, y 分别求梯度并令其等于 0，得到的 KKT 条件如下：

$$\mu-\boldsymbol{D}K'\boldsymbol{D}\upsilon=0\,,\ \gamma+\boldsymbol{e}'\boldsymbol{D}\upsilon=0\,,\ \nu y-\upsilon=0$$
$$\boldsymbol{D}(K\boldsymbol{D}\mu-\boldsymbol{e}\gamma)+y=\boldsymbol{e} \tag{18}$$

由此得出μ,γ和 y:

$$\mu=\boldsymbol{D}K'\boldsymbol{D}\upsilon\,,\ \ \gamma=-\boldsymbol{e}'\boldsymbol{D}\upsilon\,,\ \ y=\frac{\upsilon}{\nu} \tag{19}$$

再将式(19)代入式(18)中的最后一个等式中，即可求出拉格朗日乘子υ:

$$\upsilon=(\frac{I}{\nu}+\boldsymbol{D}(KK'+\boldsymbol{ee}')\boldsymbol{D}^{-1})^{-1}\boldsymbol{e}=(\frac{I}{\nu}+\boldsymbol{GG}')^{-1}\boldsymbol{e} \tag{20}$$

$$\boldsymbol{G}=\boldsymbol{D}[K\quad -\boldsymbol{e}] \tag{21}$$

非线性 SVM 分类面可以从线性分类面的公式中推导出来。把式 (11) 中的 $\boldsymbol{\omega}=\boldsymbol{A}'\boldsymbol{D}\mu$ 代入到式 (6) 而得到：$x'\boldsymbol{A}'\boldsymbol{D}\mu-\gamma=0$。

用 $K(x',\boldsymbol{A}')$ 代替上式中的内积 $x'\boldsymbol{A}'$，并将式(19)中的 μ 和 γ 代入上式，则得到如下非线性 SVM 分类超平面：

$$\begin{aligned} K(x',\boldsymbol{A}')\boldsymbol{D}\mu-\gamma=K(x',\boldsymbol{A}')\boldsymbol{DD}K(\boldsymbol{A},\boldsymbol{A}')'\boldsymbol{D}\upsilon+\boldsymbol{e}'\boldsymbol{D}\upsilon= \\ (K(x',\boldsymbol{A}')K(\boldsymbol{A},\boldsymbol{A}')'+\boldsymbol{e}')\boldsymbol{D}\upsilon=0 \end{aligned} \tag{22}$$

相应的非线性 SVM 分类器如下：

$$(K(x',\boldsymbol{A}')K(\boldsymbol{A},\boldsymbol{A})'+\boldsymbol{e}')\boldsymbol{D}\upsilon\begin{cases} >0 & x\in A^{+} \\ <0 & x\in A^{-} \\ =0 & x\in A^{+}\text{ or }A^{-} \end{cases} \tag{23}$$

现在给出非线性 SVM 分类器算法步骤：

(1)选择一个核函数 $K(\boldsymbol{A},\boldsymbol{A}')$ 。

(2)用式(21)定义 $\boldsymbol{G}$。

(3)用式(20)计算拉格朗日乘子υ。

(4)应用式(22)得到非线性 SVM 分类超平面。

(5)应用式(23)对新的样本进行分类。

4 在 UCI 数据集上的实验

在实验中选取UCI数据库中的Iris, Mushroom和Galaxy Dim数据集，在训练和测试的精度以及训练时间上与SVM$^{\text{Light}}$算法进行了比较，如表 1 所示，本文所提出算法在精度上和速度上都高于SVM$^{\text{Light}}$算法。

表 1 本算法与SVM$^{\text{Light}}$在UCI数据集上的比较

数据集 $m\times n$	本文算法			SVM$^{\text{Light}}$		
	训练精度/(%)	测试精度/(%)	训练时间/s	训练精度/(%)	测试精度/(%)	训练时间/s
Ionosphere 351×34	92.3	90.6	0.56	91.1	88.0	0.71
Mushroom 8124×22	90.1	85.2	5.70	81.5	81.5	145.59
Galaxy Dim 4192×14	95.1	94.3	1.30	94.2	94.1	28.33

（下转第 55 页）