关于支持向量机参数选择方法分析*

王 睿
(重庆师范大学 数学与计算机科学学院 ,重庆 400047)

摘 要:在分析支持向量机(SVM)原理基础上,分析了 SVM 中核函数、核参数及惩罚参数 C 的影响。介绍了两种 SVM 参数选择方法,作了深入比较,并提出了一种改进的最优化方法。

关键词:支持向量机;参数选择
中图分类号:TP183 **文献标识码**:A **文章编号**:1672-6693(2007)02-0036-03

Method Analye About Support Vector Machine's Parameter

WANG Rui
(College of Mathematics and Computer Science, Chongqing Normal University, Chongqing 400047, China)

Abstract: On the foundation of support vector machine (SVM) principle in the analysis, it analyzes kernel function among the SVM, kernel parameter and the influence of penalty parameter C. It introduces two kinds of SVM's parameter selection methods, and compares them deeply. It provides a method which works as perfect as possible.

keyword: support vector machine; the parameter selected

20 世纪 70 年代,V. Vapnik 等人提出的统计学习理论是专门研究有限样本下机器学习规律的理论。它指出学习机器的期望风险 $R(\alpha)$ 和经验风险 $Remp(\alpha)$ 的关系为 $R(\alpha) \le Remp(\alpha) + \Phi(h/n)$,其中 $\alpha \in \Lambda$ 是学习机器的广义参数,h 是反映学习机器学习能力(复杂度)的参数,称为 VC 维,n 表示训练样本的个数,$\Phi(h/n)$ 被称为置信范围,它随 h 的增加而增加。可见,学习机器的推广能力不但与经验风险有关,而且和学习机器的复杂性有关,基于此,统计学习理论提出了结构风险最小化原则和一种实现它的通用学习算法,即支持向量机(SVM)。SVM 基于统计学习理论的结构风险最小化原则,将最大分界面分类器思想和基于核的方法结合在一起,表现出了很好的泛化能力,和其他学习算法一样,其性能依赖于学习算法的参数。而 SVM 参数选择没有固定的方法,因此,SVM 的参数选择一直是一个研究的热点问题[1]。

1 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,基本思想可用图 1 的两维情况说明。图中,圆点和方点分别代表两类样本,H 为分类线,$H1$、$H2$ 分别为各类中离分类线最近的样本且平行于分类线的直线,它们之间的距离叫做分类间隔。所谓最优分类线就是要求分类线不但能将两类样本正确分开(训练错误率为 0),而且使分类间隔最大[2]。

设有 n 个样本 x_i 及其所属类别 y_i,表示为

$$(x_i, y_i), x_i \in \mathbf{R}^N, y_i \in \{1, -1\} \quad (i = 1, \cdots, n)$$

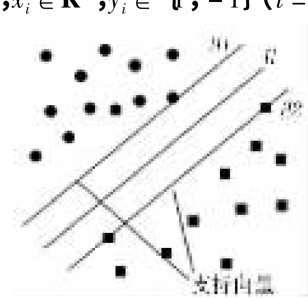

图 1 最优分类面

超平面 $W \cdot X + b = 0$ 方程,能将两类样本正确区分,并使分类间隔最大的优化问题可表示为在

$$y_i[W \cdot X + b] - 1 \ge 0 \quad (i = 1, \cdots, n) \qquad (1)$$

的约束下求

* 收稿日期:2006-12-26
资助项目:国家社科基金(No. 04CJY012)
作者简介:王睿(1978-),男,重庆人,实验师,硕士研究生,研究方向为系统研发、数据挖掘。

$$\varphi(W)=\frac{1}{2}\|w\|^2=\frac{1}{2}(w\cdot w) \quad (2)$$

的最小值。考虑到可能存在一些样本不能被正确分类,为了保证分类的正确性,引入松弛因子 $\xi_i\geq 0(i=1,\cdots,n)$ 此时的约束条件式(1)变为

$$y_i[W\cdot X+b]-1+\xi_i\geq 0(i=1,\cdots,n) \quad (3)$$

寻优目标函数(2)式变为

$$\varphi(W)=\frac{1}{2}(w\cdot w)+C\sum_{i=1}^{n}\xi_i \quad (4)$$

(4)式中的 C 为某个指定的常数,起到对错分样本惩罚程度控制的作用,实现在错分样本的比例和算法复杂程度之间的'折衷'。该问题可转化为其对偶问题。即在

$$\sum_{i=1}^{n}y_i\alpha_i=0 \quad (5)$$

和

$$0\leq\alpha_i\leq C \quad (i=1,2,\cdots,n) \quad (6)$$

的约束下求

$$Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_jy_iy_jk(x_i\cdot x_j) \quad (7)$$

的最大值。其中,若 $\alpha_i=0$,样本 x_i 称为非支持向量;若 $\alpha_i>0$,x_i 称为支持向量;若 $\alpha_i=C$,x_i 称为有界支持向量;若 $C>\alpha_i>0$,x_i 称为非有界支持向量。求解出上述各系数 α、W、b 对应的最优解 α^*、W^*、b^* 后,得到最优分类函数为

$$f(\alpha)=\mathrm{sgn}\left(\sum_{i=1}^{n}\alpha_i^*y_i(x_i\cdot x)+b^*\right) \quad (8)$$

对于非线性问题,可通过一个变换 $\phi:R^N\rightarrow F$ 映射到一个高维特征空间 F,在新空间中求最优分类面。在上面的对偶问题中,不论是寻优目标函数(7)式还是分类函数(8)式都只涉及训练样本之间的内积运算 $(x_i\cdot x_j)$。而这种内积运算可以用原空间中的核函数实现。根据泛函的有关理论,只要核函数 $K(x_i,x_j)$ 满足 Mercer 条件,它就对应某一变换空间中的内积。因此,在最优分类面中用适当的内积核函数 $K(x_i,x_j)$ 就可以实现从低维空间向高维空间的映射,从而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。此时的寻优目标函数(7)式变为

$$Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_jy_iy_jk(x_i,x_j) \quad (9)$$

而相应的分类函数式(8)也变为

$$f(\alpha)=\mathrm{sgn}\left(\sum_{i=1}^{n}\alpha_i^*y_i(x_i,x)+b^*\right) \quad (10)$$

2 核函数作用及核参数的影响

核函数,映射函数以及特征空间是一一对应的,确定了核函数 $K(x,y)$,就隐含地确定了映射函数和特征空间 F。核参数的改变实际上是隐含地改变映射函数,从而改变样本数据子空间分布的复杂程度(维数)[3]。数据子空间的维数决定了能在此空间构造的线性分类面的最大 VC 维,也就决定了线性分类面能达到的最小经验误差。同时,每一个数据子空间对应唯一的推广能力最好的分类超平面,如果数据子空间维数很高,则得到的最优分类面就可能比较复杂,经验风险小但置信范围大;反之亦然。所以,只有首先选择合适的核函数,将数据投影到合适的特征空间,才可能得到推广能力良好的 SVM 分类器[4]。

常用的核函数有

1)多项式核函数

$$K(x,x')=(x\cdot x'+c)^d \quad (d=1,2,\cdots) \quad (11)$$

2)径向基核函数

$$K(x,x')=\exp(-\gamma\|x-x'\|^2) \quad (12)$$

3)Sigmoid 核函数

$$K(x,x')=\tanh(k(x\cdot x')+v) \quad (13)$$

3 支持向量机中参数 C 的影响

(4)式中误差惩罚参数 C 的作用是在确定的数据子空间中调节学习机器置信范围和经验风险的比例以使学习机器的推广能力最好。在确定的数据子空间中,C 的取值小表示对经验误差的惩罚小,学习机器的复杂度小而经验风险值较大;反之亦然。前者称为'欠学习'现象,而后者则为'过学习'。每个数据子空间至少存在一个合适的 C 使得 SVM 推广能力最好。当 C 超过一定值时,SVM 的复杂度达到了数据子空间允许的最大值,此时经验风险和推广能力几乎不再变化。然而,目前还没有一个统一的方法来决定 C 的最佳取值[5]。

4 参数选择方法分析

4.1 试凑法(穷举法)

该方法是在模型(SVM、核函数)选择以后,首先为常数 C 和核函数固有的参数赋初始值,然后开始实验测试,根据测试精度重复调整参数值,直至得到满意的测试精度为止。实验表明,随着 C 的增加,测试精度首先增高,超过一定值以后,精度开始下降。同时,随着 C 的增加,支持向量的个数严格减少,处于边界值的支持向量的个数迅速减少,直到为 0。通过实验比较认为,C 参数的值对训练结果

有很大影响,但它的最佳取值与具体问题有很大关系,一般来说,用于训练的数据量越大,训练结果对 C 的变化越不敏感;如果训练数据很少,C 的较大取值很容易使模型过拟合训练数据。于是建议在各类样本数目不平衡的情况下,对于样本较少的类别施加较大的错分惩罚系数,惩罚系数的大小应该与各类样本数成反比。

试凑法是目前比较常用且非常行之有效的方法。但基本是凭经验调整,缺乏足够的理论依据,对不同的核函数,不同的样本其调整方法可能不同,因此在参数调整过程中带有一定的盲目性,且当需要调整幅度较大时调整次数较多,实验比较复杂。

4.2　最优化方法

从前面的分析知道,训练一个 SVM 就是求解使(9)式最大化的解 α 和 b,选择常数 C 和核函数固有参数(核参数)就是最小化推广能力的估计值。所以参数选择可以归结为最小最大化问题:最大化(9)式并在解的基础上最小化推广能力的估计值,由此可以得到选择 SVM 参数的最优化方法[6-7]。

1)为常数 C 和核函数固有参数赋初值;

2)最大化(3)式,得到 α 和 b;

3)更新常数 C 和核参数,最小化推广能力的估计值;

4)如果估计值满足要求结束运算;否则重复 2)。

步骤 3)中的推广能力是指学习机对未知数据进行测试时的分类性能。该方法是一个标准的最小最大化优化问题,可以使用已有的优化工具来实现,但问题的难点是如何更新常数 C 和核参数。因为无法得到推广能力的估计值与这些参数的显式表达式,而且变化不连续,所以不能使用常用的最陡梯度法、牛顿法等方法。

5　改进的最优化方法

该方法主要针对最优化方法步骤 3)进行改进。从(4)式可知,惩罚因子 C 控制的是训练错误率与模型复杂度间的折衷。从(7)式可知,惩罚因子 C 并没有出现在(4)式的 Wolfe 对偶式中,而是改变了 Lagrange 系数的取值范围,因此,对于一个 SVM,如果无限制地增大惩罚因子 C,当 SVM 中没有边界支持向量时,C 的改变不会再影响分类性能。

由图 1 和(4)式可知 SVM 是最大化间隔算法,以径向基核为例,根据(9)式和(10)式经适当变换后可得

$$\| w \|^2 = 2\sum_{i=1}^{n} - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \exp(-\gamma \| x_i , x_j \|^2) \quad (14)$$

从(12)式可以看出,核参数 γ 相当于对样本间欧式距离的归一化,判定了特征空间中向量间的距离,另一方面还可以得到

$$\| w \|^2 = \sum_{i=1}^{l}\alpha_i \le NC \quad (15)$$

因此,选用支持向量与样本数据比例估计推广能力并根据(14)式调整 γ 是合理的。(14)式对 γ 求导有

$$\frac{\mathrm{d}\| w \|^2}{\mathrm{d}\gamma} = \lambda\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \| x_i - x_j \|^2 \exp(-\gamma) \| x_i , x_j \|^2 \quad (16)$$

由此可得 γ 的调整规则为

$$\gamma_{i+1} = \gamma_i - \rho\frac{\mathrm{d}\| w \|^2}{\mathrm{d}\gamma} \quad (17)$$

根据以上分析可以得到下面的规则。

1)以边界支持向量的个数或在支持向量集合中的比例调整惩罚因子 C;

2)根据(17)式调整核参数;

3)同时考虑以支持向量个数不再变化为结束条件。

通过实验表明,使用该方法,支持向量与总样本数的比值与实际测试结果随参数的变化虽然在数值上有较大差别。但趋势非常相似,取得最小值点的位置也很接近,能够指导人们选择最优参数。当惩罚因子 C 较小时,无论是推广能力错误率的估计值还是实际测试的错误率都比较高,当 C 增加时,性能迅速提高;而继续增加 C,性能的变化并不明显,当 C 增加到一定值后,性能不再随 C 的变化而变化。即在较大的范围内,推广能力对 C 的变化不敏感。

这个方法与其它方法的不同之处在于:

1)不再寻求一个推广能力随参数连续、且可微变化的直接解析表达式,而是通过一个与之有关的量来调整核参数,从而简化了运算;

2)不再尝试将惩罚因子与核参数合并到同一个表达式中等同对待。从而能够分别考虑推广能力对不同性质参数不同的敏感性,提高了参数更新的效率;

3)通过分析,简化了多项式核和 Sigmoid 核 SVM 的参数,只需考虑最简单的标准形式。从而能够在不影响性能的情况下,提高参数选择的效率;

(下转第 42 页)

(上接38页)

4)参数选择使用典型的最小最大优化方法,从而可以使用许多现有的优化工具。使得参数选择易于实现。

6 结 论

SVM分类器推广能力的好坏直接受核参数及误差惩罚参数 C 的影响,然而,目前还没有一个统一的方法来决定其最佳取值,本文介绍了两种参数选择方法,并提出了一种改进的优化方法。其中,穷举法比较行之有效,但比较复杂;最优化方法尽管能指导人们选择参数,但仍然比较复杂。改进的最优化方法虽然有所进步,但是复杂程度也没有显著降低。因此,寻找一个行之有效且计算简单的参数选择方法依然是需要SVM研究工作者继续深入研究的一个难点课题。

参考文献:

[1] VAPNIK V N. 计算学习理论的本质[M]. 张学工译. 北京:清华大学出版社,2000.

[2] 姬水旺,姬旺田. 支持向量机训练算法的实验比较[J]. 计算机应用研究,2004(11):18-19.

[3] 董春曦,饶鲜,杨绍全,等. 支持向量机参数选择方法研究[J]. 系统工程与电子技术,2004,8(26):1118-1119.

[4] 侯风雷,王炳锡. 基于说话人聚类和支持向量机的说话人确认研究[J]. 计算机应用,2002,22(10):33-35.

[5] 王鹏,朱小燕. 基于RBF核的SVM的模型选择及其应用[J]. 计算机工程与应用,2003,24:72-73.

[6] 邓乃扬,田英杰. 数据挖掘中的新方法:支持向量机[M]. 北京:科学出版社,2004.

[7] 董景荣. 汇率预报的非线性组合建模与预测方法研究[J]. 重庆师范学院学报(自然科学版),2003,20(3):1-4,18.

(责任编辑 游中胜)