基于网格模式搜索的支持向量机模型选择

李 兵,姚全珠,罗作民,田 元,王 伟
LI Bing,YAO Quan-zhu,LUO Zuo-min,TIAN Yuan,WANG Wei

西安理工大学 计算机科学与工程学院,西安 710048
School of Computer Science and Engineering,Xi'an University of Technology,Xi'an 710048,China
E-mail:libing@xaut.edu.cn

LI Bing,YAO Quan-zhu,LUO Zuo-min,et al.Gird-pattern method for model selection of support vector machines. Computer Engineering and Applications,2008,44(15):136-138.

Abstract: For fixed functional form of the kernel,model selection amounts to tuning kernel parameters and the slack penalty coefficient *C*.Based on an analysis of the grid algorithm and pattern algorithm,this paper proposes a grid-pattern search algorithm,which combines grid search and pattern search.The main procedure of the proposed method include a fast search in the global domain with grid algorithm,then after obtaining the least interval containing the optimal solution,a pattern algorithm is employed to get the optimal solution in the interval.Experimental results indicate that this method has the advantage of high accuracy and speed when training SVM.

Key words: support vector machine;model selection;grid pattern search

摘 要: 支持向量机的模型选择问题就是对于一个给定的核函数,调节核参数和惩罚因子 *C*。分析了网格搜索算法和模式搜索算法,通过结合上述两种算法的优点提出了网格模式搜索算法。其核心原理是先用网格算法在全局范围内进行快速搜索,找到最优解的最小区间,再在这个最小区间内用模式搜索算法找到最优解。实验证明,网格模式搜索具有学习精度高和速度快的优点。

关键词: 支持向量机;模型选择;网格模式搜索
DOI: 10.3778/j.issn.1002-8331.2008.15.043 **文章编号:** 1002-8331(2008)15-0136-03 **文献标识码:** A **中图分类号:** TP301

1 引言

支持向量机理论最早由 Vapnik[1]提出,是一种基于统计学习理论中 VC 维理论和结构风险最小理论的通用学习方法。它可以解决小样本学习问题,而且对数据的维数、多变性不敏感,可以实现多种传统方法,能够较好地进行模型选择,并具有较好的推广能力。目前已经在许多智能信息获取与处理领域都取得了成功的应用[2]。

支持向量机的成功很大程度上依赖于核函数技术(Kernel tricks)的成功应用,并促进了支持向量机采用该技术处理传统数据的研究。支持向量机的模型选择问题就是给定一个核函数,通过调节核参数和惩罚因子 *C* 来提高精度、同时降低错误率,因此支持向量机的参数选择直接影响着 SVM 的性能[3]。目前,参数的选取缺乏理论指导,大多数核函数中参数的选取还只能凭借先验知识,甚至通过猜的手段来确定。

针对目前的现状,本文分析了通过模式搜索算法和网格搜索算法进行核函数参数选取的方法, 结合这两个算法的优点,改进了核函数参数选取方法,使改进后的算法在性能和效率上比模式搜索算法和网格搜索算法有了很大提高, 从而提高了 SVM 的训练精度并缩短了训练时间。

2 支持向量机原理

支持向量机最初用于数据分类问题的处理。下面针对训练样本集:两类线性、两类非线性两种情况分别加以讨论。

对于两类线性可分问题,已知:训练集包含 l 个样本点:

$T=\{(x_1,y_1),\cdots,(x_l,y_l)\}\in(x\times y)^l$

$\boldsymbol{x}_i\in\chi=R^n,y_i\in Y=\{1,-1\}$

其中 $\boldsymbol{x}_i$ 是输入指标向量,或称输入、模式,其分量称为特征,或属性、或输入指标;Y 是输出指标,或称输出。支持向量就是寻求一个平面 $\omega\cdot x+b=0$,使得训练数据点距离这个分类面尽量的远。这种极大化"间隔"的思想导致求解下列对变量 ω 和 b 的最优问题:

$$\min_{w,b}=\frac{1}{2}\|\omega\|^2 \qquad (1)$$

$$y_i\{(\omega\cdot x_i)+b\}\geqslant 1,i=1,\cdots,n$$

为求解原始问题,根据最优化理论,可以转化为对偶问题来求解:

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_iy_j\alpha_i\alpha_j(x_i\cdot x_j)-\sum_{j=1}^{l}\alpha_j$$

作者简介: 李兵(1972-),男,讲师,主要研究方向:数据挖掘、网络安全;姚全珠(1960-),男,教授,主要研究方向:软件工程、数据挖掘;罗作民(1963-),男,副教授,主要研究方向:算法、数据库技术与系统集成、网络软件技术及应用;田元(1975-),男,硕士研究生,主要研究数据挖掘、网络安全;王伟(1973-),男,讲师,主要研究方向:文本挖掘与机器学习。

收稿日期: 2007-09-06 **修回日期:** 2007-12-05

$$\text{s.t.} \quad \sum_{i=1}^{l} y_i\alpha_i=0 \tag{2}$$

$$\alpha_i\geqslant 0 \quad i=1,\cdots,l$$

解上述问题后得到的分类规则函数是:

$$f(x)=\text{sgn}\{(\omega\cdot x)+b\}=\text{sgn}\{\sum_{i=1}^{l}\alpha_i\cdot y_i(x_i\cdot x)+b\} \tag{3}$$

对于两类线性不可分问题,为第 i 个训练点(x_i,y_i)引入松弛变量(Slack Variable)$\xi_i\geqslant 0$,把约束条件放松到 $y_i\{(\omega\cdot x_i)+b\}+\xi_i\geqslant 1,i=1,\cdots,n$(即"软化"约束条件)。显然 ξ_i 可以描述为训练集错划的程度。现在就有两个目标:希望超平面间隔最大,即$(2/\|\omega\|$ 最大),又希望训练集错划程度 $\sum_{i=1}^{l}\xi_i$ 尽可能小,所以引入惩罚参数 C。在实际使用时,可以选择 C 来修改这两个目标的权重。这样新的目标函数成为:

$$\min_{\omega,b,\xi} \quad \frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}\xi_i$$

$$\text{s.t.} \quad y_i\{(\omega\cdot x_i)+b\}+\xi_i\geqslant 1,i=1,\cdots,l,\xi_i\geqslant 0 \tag{4}$$

$\sum_{i=1}^{l}\xi_i$ 体现了经验风险,而 $\|w\|$ 则体现了表达能力。所以惩罚参数 C 实质上是对经验风险和表达能力匹配的一个裁决。当 $C\to\infty$ 时,线性不可分的原始问题退化为线性可分。

对于非线性分类,通过引入核函数,将原空间样本数据通过非线性变换映射到高维特征空间 H:Φ:$R^d\to H$。在这高维空间中求最优或广义最优分类面。

常用的核函数为:

(1)多项式核函数:

$K(x,x_i)=[((x\cdot x_i)+1)]^d$

(2)RBF(径向基 Radial Basis Function)核函数:

$$K(x,x_i)=\exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right)$$

(3)Sigmoid 核函数:

$K(x,x_i)=\tanh(b(x\cdot x_i)-c)$

式中 b,c 为常数。

这样对于非线性分类问题,最终归结为一个二次规划问题:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_iy_j\alpha_i\alpha_jK(x_i\cdot x_j)-\sum_{j=1}^{l}\alpha_j=-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a}+\frac{1}{2}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}a$$

$$\text{s.t.} \quad \sum_{i=1}^{l}y_i\alpha_i=0$$

$$0\leqslant\alpha_i\leqslant C \quad i=1,\cdots,l$$

3　支持向量机参数的选择方法

目前,国内外已经提出了很多种关于支持向量机参数选择的标准和方法[4],其中有很多种方法是基于梯度的方法。如果一些参数选择标准是不可微函数,那么这类基于梯度的参数选择方法将会失效。因此,本文提出了网格模式搜索算法。首先介绍网格搜索算法和模式搜索算法。

3.1　网格搜索算法

网格搜索算法是将惩罚参数 C 和核参数 γ 分别取 M 个值和 N 个值,对 $M\times N$ 个(C,γ)的组合,分别训练不同的 SVM,再估计其推广识别率,从而在 $M\times N$ 个(C,γ)的组合中得到推广识别率最高的一个组合作为最优参数。

3.2　模式搜索算法

模式搜索算法[5]是由 Hooke 和 Jeeves 提出,因此又称为 Hooke—Jeeves 方法。模式搜索算法由两个过程组成:一是试探过程,也称探测移动。这一过程是确定有利的搜索方向;二是加速过程,也称模式移动,即在有利的方向上加速搜索。

模式搜索算法是一个直接搜索算法,是一个简单而有效的优化技术。它不要求目标函数可导,迭代简单[6]。

3.3　网格搜索算法和模式搜索算法的比较

网格搜索算法与模式搜索算法相比,网格搜索算法搜索范围较大,目标函数收敛速度快。因为每个 SVM 是相互独立的,所以网格搜索算法可以并行训练。网格搜索算法的缺点是:学习精度略低。

模式搜索算法相对于网格搜索算法具有学习精度较高,但存在迭代次数较高,计算量大等缺点。同时模式搜索存在容易使目标函数陷入局部最小点而不是全局最小点的可能。

本文结合两种方法的优点,提出一种能够快速、精确选择支持向量机模型的方法–网格模式搜索算法。

4　网格模式搜索算法

为了使该算法更具通用性,本文首先对网格搜索算法进行了改造,将 γ 变为 n 维向量($\gamma\in R^n$)。使其能应用于多参数模型,因而具有通用性。在实际应用中根据不同的核函数设置 C 和 γ 的变化范围以及 γ 的维数,例如:C 的变化范围为$\lfloor 2^{-4},2^{-2},\cdots,2^8,2^{10}\rfloor$。改造后,网格算法在最坏情况下的时间复杂度为 $O((N+1)^{N+1})$。

从上面的分析来看,网格搜索比较耗时。针对上述情况,提出了网格模式搜索算法。网格模式搜索算法的核心原理是:先用网格算法在全局范围内进行快速的搜索,也就是说通过网格算法在全局范围内进行粗粒度的搜索,从而找到最优解的最小区间。因为这个阶段是在全局范围内粗粒度的搜索,因此搜索的步长应该设置得较大。当这一步骤完成后,在这个最小区间内用模式搜索算法,找到最优解。

下面是网格模式搜索算法的具体步骤。

步骤 1　初始化参数的选择范围。

步骤 2　先用一个步长为 2^a 的(C,γ)的组合求得一个最高学习精度,然后缩小步长,以 $2^{a/2}$ 为步长进行一次更精细的网格搜索。按照这种方法,在这一步不断循环,直到在 C 和含有 n 维向量的 γ 中得到推广识别率最高的一个组合、并作为最优参数。其中,a 为步长参数,可以控制学习步长。

步骤 3　将步骤 2 得到的最优参数 $t_0\in(C,\gamma)$设为中心,同时给定初始点同时给定 n 个坐标方向:$e_i=[0,\cdots,0,\overset{i}{1},0,\cdots,0]$ $(i=1,2,\cdots,n)$。初始步长 δ,加速因子 $\alpha\geqslant 1$,缩减因子 $\beta\in(0,1)$,允许误差 $\varepsilon>0$,设置 $y_0=t_0,k=1,i=1$。

步骤 4　从 t_i 出发,沿第 i 个坐标轴方向 $e_i=[0,\cdots,0,\overset{i}{1},0,\cdots,0]$,以步长 δ 进行探索,取得使目标函数下降的点:

$$y_{i+1}\begin{cases}y_i+\delta e_i & \text{if } f(y_i+\delta e_i)<f(y_i)\\ y_i-\delta e_i & \text{if } f(y_i-\delta e_i)<f(y_i)\\ y_i & \text{if } f(y_i)\leqslant\min[f(y_i-\delta e_i),f(y_i+\delta e_i)]\end{cases}$$

步骤 5　如果 $j<n$,则 $j=j+1$,转到步骤 4;否则进行步骤 6。

步骤 6　如果 $f(y_{n+1})<f(t_k)$,转到步骤 7;否则进行步骤 8。

步骤 7　令 $t_{k+1}=y_{n+1},y_1=t_{k+1}+\alpha(t_{k+1}-t_k)$,同时置 $k=k+1,j=1$,转到步骤 4。

步骤 8 如果$\delta \leqslant \varepsilon$,则停止迭代,得到最优参数$(C,\gamma)=t_k$;否则,令$\delta=\beta\delta$,$y_1=t_k$,$t_{k+1}=t_k$,转到步骤 4。

5 实验结果及分析

为了验证网格模式搜索算法的有效性,从 UCI 机器学习知识库[7]中选取了五组分类数据集(Iris,Glass Identification,wine,Cleveland Heart,Wisconsin Breast Cancer)进行了实验。

实际应用中核函数的种类有很多,研究表明,当缺少过程的先验知识时,选择高斯核函数比选择其他核函数好[8]。因此,多数应用研究都采用高斯核函数。考虑到这个原因,本次实验选取了径向基核函数作为第一组实验。实验结果见表 1、表 2。

表 1 基于径向基核函数的三种算法的检测精度 %

数据集	网格搜索算法	模式搜索算法	网格模式搜索算法
Iris	96.578 4	96.775 3	98.135 7
Glass	71.853 4	72.632 7	75.662 1
wine	98.987 2	99.013 5	99.771 1
Cleveland Heart	56.174 2	57.188 6	60.015 3
Breast Cancer	96.765 3	97.623 1	98.537 9

表 2 基于径向基核函数的三种算法的学习时间表 min

数据集	网格搜索算法	模式搜索算法	网格模式搜索算法
Iris	18.16	20.21	17.80
Glass	23.35	30.05	20.56
wine	25.27	26.83	22.58
Cleveland Heart	30.17	35.62	27.37
Breast Cancer	35.63	38.96	31.19

为了验证该算法的通用性,选择多项式核函数作为第二组实验。实验结果见表 3、表 4。

表 3 基于多项式核函数的三种算法的检测精度 %

数据集	网格搜索算法	模式搜索算法	网格模式搜索算法
Iris	86.112 6	86.965 1	90.102 8
Glass	63.254 7	65.597 2	77.995 3
wine	93.157 2	95.315 6	97.689 1
Cleveland Heart	43.352 8	48.496 9	51.612 3
Breast Cancer	87.056 3	89.667 2	93.553 2

每组实验,分别用网格搜索算法、模式搜索算法、网格模式搜索算法进行测试,三种算法均采用标准 C++实现,使用 Microsoft Visual C++6.0,缺省编译器优化选项进行编译。系统平台为 2.66 GHZ Pentium 4 处理器,Windows 2000 标准版,256 MB RAM。

表 4 基于多项式核函数的三种算法的学习时间表 min

数据集	网格搜索算法	模式搜索算法	网格模式搜索算法
Iris	17.28	20.56	16.02
Glass	18.65	28.22	17.36
wine	22.18	24.71	20.37
Cleveland Heart	27.92	30.40	25.61
Breast Cancer	32.06	36.64	29.94

通过两组实验数据表明,虽然网格搜索算法的训练时间短于模式搜索算法,但是网格搜索算法的学习精度却不如模式搜索算法。而本文提出的网格模式搜索算法正是综合网格搜索算法和模式搜索算法的优点,因而具有较高的性能。从表 1、表 3 的几组数据可以看出,网格模式搜索算法的学习精度均高于网格搜索算法和模式搜索算法。从表 2、表 4 可以看出网格模式搜索算法的速度也高于后面两种算法。

6 结论

本文首先介绍了网格搜索算法和模式搜索算法,并分析了这两种算法的优缺点。通过综合这两种算法的优点,本文提出了网格模式搜索算法。实验表明,这种方法可以明显提高 SVM 的模型的学习精度,缩短学习时间。该算法简单有效,没有繁琐复杂的运算。从本文的实验可以看出,在实际应用中该算法可以不受核函数类型的限制,具有良好的通用性。

参考文献:

[1] Vapnik V N.The nature of statistical learning theory[M].New York:Springer,1995.

[2] Sanchez A D.Advanced support vector machines and kernel methods[J].Neurocomputing,2003,55(1):5-20.

[3] Müller K R,Mika S,Ratsch G,et al.An introduction to kernel-based learning algorithms[J].IEEE Transactions on Neural Networks,2001,12(2):181-202.

[4] Chapelle O,Vapnik V.Choosing multiple parameters for support vector machines[J].Machine Learning,2002,46:131-59.

[5] Hooke R,Jeeves T A.Direct search solution of numerical and statistical problems[J].Ass Comput Mach,1961,8:212-229.

[6] 陈宝林.最优化理论与算法[M].2 版.北京:清华大学出版社,2005.

[7] Murphy P M, Irvine A D W.UCI repository of machine learning databases.[EB/OL].(1994).http://www.ics.uci.edu/~mlearn/MLRepository.html.

[8] Smola A J.Learning with kernels[D].Berlin:Technical University of Berlin,1998.

(上接 93 页)

参考文献:

[1] Shamir A.How to share a secret[J].Communications of the ACM,1979,22(11):612-613.

[2] Zhou Lidong,Hass Z J.Securing Ad Hoc networks[J].IEEE Networks,1999,13(6):24-30.

[3] 熊焰,苗付友,张伟超,等.移动自组网中基于多跳步加密签名函数签名的分布式认证[J].电子学报,2003,31(2):161-165.

[4] Yao Jun,Zeng Gui-hua.A distributed authentication algorithm based on GQ signature for Mobile Ad Hoc networks[J].Journal of Shanghai Jiaotong University:Science,2006,11(3):346-350.

[5] 李方伟.移动通信系统认证协议与密码技术[M].北京:人民邮电出版社,2007.