基于 RBF 核函数的支持向量机参数选择

林升梁[1], 刘　志[2]
(1. 浙江工业大学 信息工程学院, 浙江 杭州 310032;
2. 浙江工业大学 软件学院, 浙江 杭州 310032)

摘要: 由于 SVM 在各个领域中得到越来越广泛的应用, 而决定 SVM 性能的因素是核函数的选取. 其中, RBF 核函数是应用最广泛的核函数, 且有两个参数: 惩罚因子 C 和核参数 γ. 因此, 希望能找到最优化参数组(C, γ)使 SVM 具有最好推广性. 首先提出了用 $E=\frac{l_{sv}}{n}$ 代替留一法来评估 SVM 的推广性, 它的优点是速度快、准确性高; 然后, 分析参数 C 和 γ 对 SVM 性能的影响, 由此将问题归结在一个小的"好区"内选取最优参数组(C, γ); 最后, 分别用穷举法和下文所提出的方法进行比较, 得出在"好区"内用 $C\gamma=\tilde{C}$(常数)来确定最优化参数同样能得到很好的推广性, 而且速度上比穷举法快的多. 此方法, 具有一定的实际应用价值.

关键词: 支持向量机; RBF 核参数; 惩罚因子 C; 推广识别率
中图分类号: TP181　　文献标识码: A　　文章编号: 1006-4303(2007)02-0163-05

Parameter selection in SVM with RBF kernel function

LIN Sheng-liang[1], LIU Zhi[2]
(1. College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310032, China;
2. College of Software, Zhejiang University of Technology, Hangzhou, 310032, China)

Abstract: Support vector machines (SVM) is more and more applied in the various fields in recent years. Selection of kernel function is a pivotal factor which decides performance of SVM. The RBF kernel function is most widely used in SVM. There are two parameters in this function: the penalty parameter C and the kernel parameter γ. The optimization parameters (C, γ) will make the SVM have the best performance. Firstly, the $E=\frac{l_{sv}}{n}$ method is proposed to be used to assess the performance of SVM instead of using exhaust algorithm. This algorithm is of high speed and high accuracy. Then through analyzing the influence of the parameters c and γ to the performance of SVM, it is reduced to the problem that the optimization parameters should be selected in a small "good-area"; Finally, through comparing the exhaust algorithm and the method mentioned in this paper separately, we could obtain that using $C\gamma=\tilde{C}$(constant) to define the optimization parameters will get good performance of SVM. This method is of good practical use.

Key words: support vector machine (SVM); parameter of RBF kernel; penalty factor; generalized recognition rate

收稿日期: 2006-09-25
作者简介: 林升梁(1980—), 男, 浙江温州人, 硕士研究生, 从事图像处理、模式识别的研究.

0 引 言

机器学习是继专家系统之后人工智能应用的又一重要研究领域,也是人工智能和神经计算的核心研究课题之一.在模式识别、回归分析和特征提取等方面得到了越来越广泛的应用.目前,在统计学习理论的基础上,发展出来一种新的机器学习方法——支持向量机[1](Support vector machines, SVM).它基于结构风险最小化[1](Structural Rish Minimization, SRM)原则,即是由有限训练样本得到的决策规则对独立的测试集仍能得到小的误差.尽量提高学习机的泛化能力,具有良好的推广性能和较好的分类精确性,能有效的解决过学习问题,现已成为感知器、神经网络的替代性方法.

支持向量机(SVM)是20世纪90年代由Vapnik[3]等人提出的一种新的机器学习方法,与传统的机器学习相比,有较好的推广能力.但是和其他学习算法一样,其性能是依赖参数的选择,到目前还没有一个很好的方法解决这个问题.笔者就是基于SVM推广能力的估计入手,来研究这个问题,并用UCI基准库[7]上的数据来说明.

1 支持向量机的介绍

支持向量机(SVM)理论[1]主要是针对二类模式识别问题提出的.对于二类模式识别问题,设给定的训练集为$\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$,其中$x_i\in R^n$为输入向量,输出向量为$y_i\in\{-1,1\}$,如果该训练集可被一个超平面线性划分,则该超平面为$W\cdot X+b=0$,其中W和b是决定了超平面的位置,$W\cdot X$为两个向量的内积.为了得到最优化的划分,则该问题就转化为求最优化超平面的问题

$$\begin{cases}\min\phi(W,\xi)=\dfrac{1}{2}\|W\|^2+c\sum\limits_{i=1}^{n}\xi_i,\ c\geqslant 0\\ (i=1,2,\cdots,n)\\ s.t\quad y_i[(W\cdot X_1)+b]\geqslant 1-\xi_i,\xi_i\geqslant 0\end{cases}\tag{1}$$

其中:W是特征空间中分类超平面的系数向量;b是分类面的阈值;ξ_i是考虑分类误差而引入的松弛因子;C是对于错分样本的惩罚因子.这样的话所构造出来的最优化超平面为:$f(x)=W\cdot X+b$

式(1)优化问题可以转化为其对偶问题[2]

$$\begin{cases}\min\limits_{a}\dfrac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}y_iy_ja_ia_j(X_i\cdot X_j)-\\ \qquad\sum\limits_{i=1}^{n}a_i\qquad(i=1,2,\cdots,n)\\ s.t.\quad\sum\limits_{i=1}^{n}y_ia_i=0,0\leqslant a_i\leqslant C\end{cases}\tag{2}$$

对于大多数样本来讲,$a_i=0$,对应$a_i\neq 0$的样本称为支持向量(Support Vector, SV).解出式(2)的最优化函数为

$$f(X)=\mathrm{sign}[\sum_{i=1}^{n}a_iy_i(X_i,X)+b^*],$$

其中$b^*=y_i-\sum\limits_{i=1}^{n}ya_i(X_i\cdot X)$

上式求和实际上由支持向量,即$a_i\neq 0$的样本决定.从这一点可以得出支持向量决定超平面(图1所示)的划分.

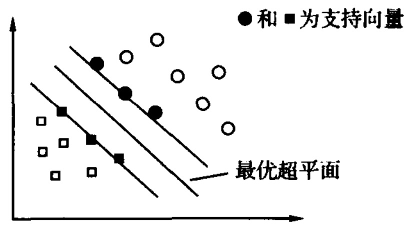

图1 最优化超平面示意图

Fig.1 The sketch of optimization hyperplane

对于非线性可分的情况,可以通过一个映射函数(在SVM称核函数),将低维的输入空间R^n映射到高维的特征空间H使线性可分(图2所示),则低维的线性不可分问题就变成高维空间的线性可分问题.

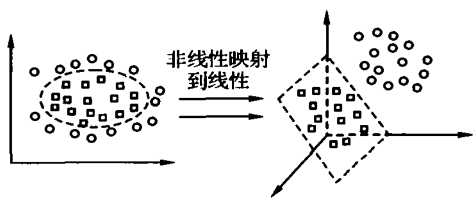

图2 核函数原理图

Fig.2 The elementary diagram of the kernel function

这样问题就可以表述为:输入向量X通过映射$\psi: R^n\longrightarrow H$影射到高维空间$H$中,则核函数$K(X_i,X_j)=\psi(X_i)\cdot\psi(X_j)$,则优化问题转化为

$$\begin{cases}\min\limits_{a}\dfrac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}y_iy_ja_ia_jK(X_i,X_j)-\sum\limits_{i=1}^{n}a_i\\ (i=1,2,\cdots,n)\\ s.t.\quad\sum\limits_{i=1}^{n}y_ia_i=0,0\leqslant a_i\leqslant C\end{cases}\tag{3}$$

解出式(3)的最优化函数为:

$$f(X)=\operatorname{sign}\left[\sum_{i=1}^{n}a_i y_i K(X_i,X_j)+b^*\right]$$

从式(3)的最小化问题可以看出,不需要知道 H 和 ψ,只需要选择合适的核函数 $K(\cdot)$ 和 C 就可以确定 SVM.

现在,应用较常见的核函数有以下四种:

线性核:$K(X_i,X_j)=X_i\cdot X_j$;

多项式核:$K(X_i,X_j)=(X_i\cdot X_j+1)^d$;

径向基(RBF)核:$K(X_i,X_j)=\exp(-\gamma\|X_i-X_j\|)^2$;

Sigmoid 核:$K(X_i,X_j)=\tanh[c_1(X_i,X_j)+c_2]$.

在这四种核函数中,应用最广泛的是 RBF 核,无论是低维、高维、小样本、大样本等情况,RBF 核函数均适用,具有较宽的收敛域,是较为理想的分类依据函数[2].下面就是基于 RBF 核函数的 SVM 来讨论各个参数(C,γ)对其推广能力的影响,寻求优化参数选择的方法.

2　以 RBF 核为核函数的支持向量机

RBF 核为:$K(X_i,X_j)=\exp(-\gamma\|X_i-X_j\|)^2$,则对应式(3)的最优化问题就转化就下面最小化问题

$$\min_{a_i}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}y_i y_j a_i a_j\exp(-\gamma\|X_i-X_j\|)^2-\sum_{i=1}^{n}a_i,$$

其中 $s.\ t.\ \sum_{i=1}^{n}y_i a_i=0,0\leqslant a_i\leqslant C$　(4)

这样求式(4)的最小值就取决于参数(C,γ)的选择.这样选择最佳的参数就可以使 SVM 分类器性能最好,即推广能力最强.

2.1　SVM 推广能力的估计

一个 SVM 分类器的好坏,主要看它的推广性和学习机器的复杂性,即对未知数据进行测试时的准确性.而估计 SVM 推广能力的方法很多,都是基于 SRM 原则,在每一组参数组合上均能求得对实际风险的估计,通过比较不同的参数组合就可以找到最好的 SVM,此时的推广能力也最好.

留一法(Leave One Out)[4],首先从训练集中去掉一个样本,再在其他样本上训练判决准则,并利用该判决准则对去掉的样本进行分类,如果分类错误则产生了一个留一法错误.f^i 令表示去掉第个样本后在剩余样本上得到的分类准则,$f^i(X_i)$ 表示使用该规则对样本 X_i 进行分类,$P(f^i(X_i),Y_i)$ 表示分类结果,用这个结果和去掉的第样本进行比较,如果分类正确取 0,反之取 1.则最后得到的推广能力估计为

$$\mathrm{LOO}=\frac{1}{n}\sum_{i=1}^{n}P(f^i(X_i),Y_i)$$

从上式可以看出 LOO 越小,则推广能力越强,即识别错误率越低.而且已经证明用留一法对测试错误率的估计是无偏估计的[6],作为估计性能的评估标准.但该算法估计效率却很低,对 n 个样本需要进行 n 次学习和分类判决,其复杂度为 $o(n^2)$,所以随着样本的增加,估计所需要的运算量也急剧增加,显然不适合大样本的推广能力估计.

支持向量记数法[5],从留一法的原理得知,对于非支持向量($a_i=0$ 的样本),在留一法测试时不会产生错误.所以 $\mathrm{LOO}\leqslant\frac{l_{sv}}{n}$,其中 l_{sv} 是支持向量的个数,n 是总样本数.该式 $E=\frac{l_{sv}}{n}$ 计算方便,只要训练好一个 SVM,就可以马上得到支持向量(SV)的个数,特别是对于大样本 SVM 准确度比较高,从而可以作为推广能力的估计.

下面通过选取 UCI 基准库的 breast-cancer-wisconsin(BCW),Wisconsin Diagnostic Breast Cancer(WDBC),pima-indians-diabetes(PID),ionosphere,tie-tae-toe 样本数据比较对同一参数用两种方法得到的识别错误率(表 1)以及 E 和 LOO 变化趋势(图 3)

表 1　分别用 LOO 和支持向量记数法得到的错误率

Table 1　The error rate based on LOO and SVM separately

样本名称	BCW	WDBC	PID	ionosphere	tie-tae-toe
样本数	699	569	768	351	958
LOO	0.032 9	0.040 4	0.021 8	0.091	0.054 3
E	0.087 2	0.094 9	0.061 5	0.131	0.090 8

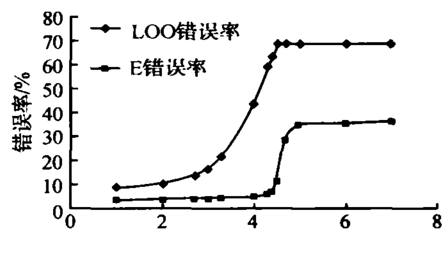

图 3　LOO 错误率和 E 错误率的变化趋势

Fig. 3　The change trend of LOO Error rate and E error rate

从以上表 1 和图 3 实验结果得知,用支持向量记数法得到的错误率和留一法得到的错误很接近,也比较准确的反映了 SVM 的推广能力;并且它的

变化趋势与留一法的变化趋势很相似，当 LOO 很小时，E 也很小. 由于该方法计算简单，实现方便，运行速度快，所以本文用 E 来估计 SVM 的推广能力，并最终用 LOO 检验.

2.2 参数(C, γ)对 SVM 的影响

从式(3)中可以看出，C 的作用是在确定的数据子空间中调节学习机器置信区间范围，不同数据子空间中最优化的 C 不同. 而核参数 γ 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，即线性分类面的最大 VC 维[1]，也就决定了线性分类达到最小误差. Vapnike 等人的研究表明了，核参数 γ 和误差惩罚因子 C 是影响 SVM 性能的关键因素[1]. 下面通过实验，分别表示核参数 γ(图 4(a)所示)和惩罚因子 C(图 4(b)所示)对 SVM 的影响.

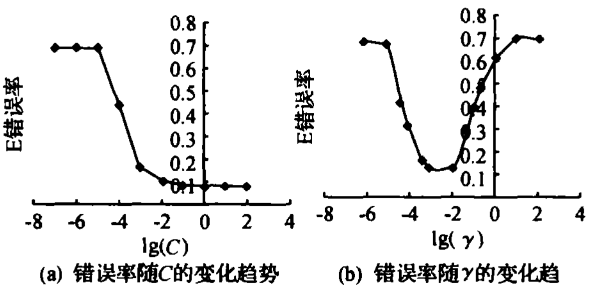

图 4 错误率随 C 和 γ 的变化趋势

Fig. 4 The change trend of error rate with parameter C and γ

从以上两个图 4(a)可以得知，当 C 较小时推广错误率的估计值比较高；当 C 增加时急剧降低，即性能得到迅速的提高；当继续增大时性能的变化就不明显了，而且增加到一定值后，错误率不再变化了，即此时推广能力对 C 的变化不敏感. 也就是说此时的 C 变化几乎不影响 SVM 的推广能力，所以在这个区域中，就通过核参数的变化来近一步得到 SVM 的最优化值. 从图 4(b)中还可以看出 γ 的变化，错误率有大到小，然后再从小到大的过程，也就是说取一定的 γ 就可以得到 SVM 的最优值. 所以，希望通过对参数(C, γ)的变化，选取最优参数组合来得到最终 SVM 的最优值，即此时的错误率最低.

2.3 参数(C, γ)最优化选择

对于一个基于 RBF 核函数的 SVM，其性能是由参数(C, γ)决定，选取不同的 C 和 γ 就会得到不同的 SVM. 我们的目的是为了寻找最佳的参数组合使该 SVM 的性能最好，即推广错误率最低.

最简单的方法是分别选取不同的参数组合，得出不同的错误率；分别比较这些错误率选取其中错误率最小的参数组合作为最优化选择，这种方法也叫做“穷举法”. 参数 C 和 γ 分别取 N 个值和 M 个值，对 $N\times M$ 个(C, γ)的组合分别训练不同的 SVM，再估计其推广错误率，从而在 $N\times M$ 个组合中得到错误率最低的一个组合作为最优参数，如图 5 所示参数(1 000, 0.000 01)为最优化参数，此时的错误率最低. 虽然用这种方法最终能找出最优化参数，但是其复杂度为 $o(N^2)$，显然运算量非常大，花费时间很大，特别对大样本数据来讲是不切实际的.

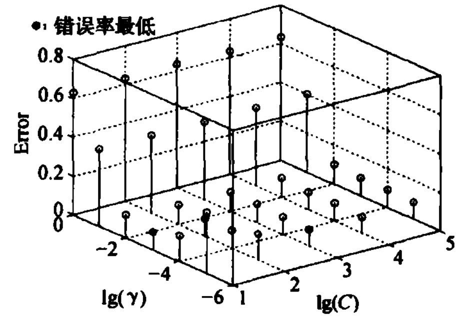

图 5 “穷举法”得到不同参数(C, γ)不同的错误率

Fig. 5 The error rate with different parameters (C, γ) based on the exhaust algorithm

从以上图 5 分析知道，对于在一定区域内的(C, γ)组合得到错误率都非常低，即 SVM 推广识别能力都非常高. 对这个区域叫“好区”[8]，如图 6(a)所示. 为了确定“好区”内最优化参数组合，该文的思想如图 6(b)所示，通过对曲线 $C\gamma=C$ 上的点(C, γ)来估计最优化参数，则用该曲线得到的最优化参数为(C_0, γ_0)，作为 SVM 的最优化参数组合，其中 $\tilde{C}$ 是常数，用该思想得到复杂度为 $o(N)$. 而且，从 3.2 节中已经得知当 C 取得一定值时，使 SVM 最优化，此时用线性 SVM 来得到最优化参数，把该参数作为常数 $\tilde{C}$，以此来确定曲线 $C\gamma=\tilde{C}$，使错误率最低的参数组合集中出现在“好区”中该曲线的附近，并通过以下实验证明.

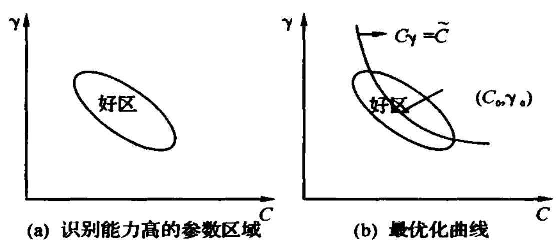

图 6 识别能力高的参数区域内的最优化曲线

Fig. 6 The optimization curve of the parameter district with the high identification ability

基于该思想得出本文优化参数(C, γ)的算法步骤:

(1) 用线性 SVM 求解最优化参数 $\tilde{C}$, 使之该参数的 SVM 推广识别错误率最低;

(2) 对 RBF 的 SVM, 固定 $\tilde{C}$, 取满足 $C\gamma = \tilde{C}$ 的(C, γ), 训练 SVM, 根据对推广识别错误率的估计, 取错误率最低的参数(C_0, γ_0), 把该参数作为 SVM 的最优化参数. 其中错误率用第 2.1 节中的表示, 最后用 LOO 检验.

2.4 仿真实验

本文实验所有数据来源于 UCI 基准库[7], 用 Matlab7.0 进行编程, 分别用"穷举法"和本文方法得到以下实验结果如表 2 和表 3 所示. 其中用"穷举法"实验时, lg(C)为[-10, 10] 而 lg(γ)为[-5, 5] .

表 2　用"穷举法"得到最优化参数

Table 2　The optimization parameters based on the exhaust algorithm

样本名称	lg(C)	lg(γ)	E	LOO
BCW	4	-5	0.082 975 7	0.030 904 1
WDBC	4	-5	0.094 903 3	0.040 421 8
PID	7	-3	0.061 527 6	0.021 875 6
ionosphere	7	-5	0.093 568 0	0.054 138 7
tic tac toe	6	-3	0.080 871 5	0.031 573 5

表 3　用本文方法得到的最优化参数

Table 3　The optimization parameter based on the method mentioned in this paper

样本名称	$\tilde{C}$	lg(C)	lg(γ)	E	LOO
BCW	1	9	-9	0.064 377 7	0.025 641 2
WDBC	10	8	-7	0.066 783 8	0.036 897 1
PID	10 000	7	-3	0.061 527 6	0.021 875 6
ionosphere	10 000	6	-2	0.094 817 4	0.045 369 7
tic tac toe	100	6	-4	0.090 814 2	0.044 108 9

从以上实验可以得知, 用本文方法大大减少了训练量的情况下, 同样可以达到与"穷举法"的推广识别率, 甚至比"穷举法"得到的更低.

另外, 在本文算法的基础上, 可以引入一个修正量 δ 使曲线 $C\gamma = \tilde{C}$ 成为 $C\gamma = \delta\tilde{C}$, 其中 $0 < \delta < 1$, 通过这样的修正量可使曲线更加接近"好区"中最优化参数组合.

3 总　结

笔者重点对以 RBF 核为核函数的 SVM 性质分析, 用实验证明了用支持向量记数法来代替留一法在本文中的可用性, 并分别得出参数 C 和 γ 对 SVM 的推广能力的影响. 基于此提出了本文优化参数的算法. 实验证明, 该算法比"穷举法"复杂度要低, 并且同样达到"穷举法"得出的 SVM 推广识别准确率, 使用该文方法优化参数来构造 SVM 分类有一定意义. 对于修正量 δ的选取可以更深入的研究

参考文献:

[1] 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.

[2] 李盼池, 许少华. 支持向量在模式识别中的核函数特性分析[J] . 计算机工程与设计, 2005, 26(2): 302-304.

[3] VAPINK V N. The nature of statistical learning theory[M] . New York: Springer Verlag, 2000.

[4] LUNTS A, BRAILOVSKIY V. Evaluation of attributes obtained in Statistical Decision Rules[J] . Engineering Cybemetics, 1967(3): 98-109.

[5] CHAPELLE O, VAPINK V N. Choosing multiple parameters for support vector machines[J] . Machine Learning, 2002, 46: 131-159.

[6] LUNTS A, BRAILOVSKIY V. Evaluation of attribtes obtained in statistical decision rules[J] . Engineering Cybernetics, 1967, 3(1): 982-1009.

[7] MURPHY P M, AHA IRVINE D W . CA: University of califomia, department of information and computer science[EB/OL] . http://www.ics.uci.edu/~mlearn/MLRepository.html, 1994.

[8] KEERTHI S, CHIH J. Asymptotic behavior of support vector machines with gaussian kernel[J]. Neural Computation, 2003, 15: 1667-1689.

(责任编辑: 翁爱湘)