

# 基于改进 Joachims 上界的 SVM 泛化性能评价方法

宋小杉, 蒋晓瑜, 汪 熙, 姚 军

(装甲兵工程学院控制工程系, 北京 100072)

**摘 要:** 留一法(Leave One Out, LOO)错误率是评价支持向量机(Support Vector Machine, SVM)性能最准确方法, LOO 错误率越小, SVM 泛化性能越好. 但 LOO 实现起来较为费时. 因此人们提出了 LOO 错误率的各种上界作为近似, 最有代表性的是 Joachims 上界和 Jaakkola-Haussler 上界. 基于 LOO 上界的 SVM 泛化性能评价方法不但较为准确, 而且耗时大大减小. 本文首先证明了在径向基函数(Radial Basis Function, RBF)为核函数的 SVM 中, Joachims 上界和 Jaakkola-Haussler 上界是等价的; 其次对 Joachims 上界进行理论分析, 指出了其不足之处, 并予以改进, 得到了改进的 Joachims 上界; 最后通过实验对 LOO 错误率、Jaakkola-Haussler 上界、Joachims 上界和改进的 Joachims 上界进行了比较, 结果显示改进的 Joachims 上界比 Jaakkola-Haussler 上界和 Joachims 上界更加接近 LOO 错误率, 是一种更加准确的 SVM 泛化性能评价方法.

**关键词:** 支持向量机; 高斯核; 泛化性能评价; 改进 Joachims 上界

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 0372-2112 (2011) 06-1379-05

## Evaluating the Generalization Performance of SVMs Based on the Advanced Joachims Bound

SONG Xiao-shan, JIANG Xiao-yu, WANG Xi, YAO Jun

(Department of Control Engineering, Academy of Armored Force Engineering, Beijing 100072, China)

**Abstract:** LOO (Leave One Out) is commonly used to evaluate the generalization performance of an SVM (Support Vector Machine), the disadvantage of which is time consuming. In order to decrease the time cost, several LOO bounds are proposed. The most famous bounds are Joachims bound and Jaakkola-Haussler bound. Based on the two bounds, the generalization performance of an SVM can be evaluated properly with decreased time cost. This paper gives the proof of the equivalence of the two bounds in an SVM with RBF (Radial Basis Function) kernel, analyzes the two bounds theoretically, proposes an advanced Joachims bound, compares the LOO error, Joachims bound, Jaakkola-Haussler bound and the advanced Joachims bound by simulated experiments. Results show that the advanced Joachims bound is closer to the LOO error, and is a better method to evaluate the generalization performance of an SVM.

**Key words:** support vector machine; radial basis function kernel; evaluation of the generalization performance; advanced Joachims bound

## 1 引言

支持向量机(Support Vector Machine, SVM)由 Vapnik 于 1995 年提出<sup>[1]</sup>, 是一种智能学习机器, 相对于人工神经网络, 具有 3 个优点: (1) SVM 以统计学习理论为基础, 从结构风险最小理论和 VC (Vapnik-Chernovenkis) 维理论<sup>[1]</sup>发展而来, 具有完善的理论基础; (2) SVM 同时考虑了经验风险最小和学习机器的复杂度, 并进行了恰当的折中, 使其具有了很好的泛化能力; (3) SVM 算法是一个凸二次规划问题, 因此具有全局最优解. 正是由于这些优点, 十几年来, 人们对 SVM 的研究一直非常活跃,

从理论到应用被人们不断完善和扩展.

为解决非线性分类问题, SVM 引进了核函数, 从而使得: (1) 将样本空间中线性不可分的样本映射到更高维的特征空间中, 使其线性可分; (2) 通过核函数用原空间中的运算实现更高维的特征空间中的运算, 避免了“维数灾难”问题.

常用的核函数有以下几种:

(1) 径向基函数(Radial Basis Function, RBF):  $K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$

(2)  $q$  阶多项式函数:  $K(\mathbf{x}, \mathbf{x}') = (\gamma \langle \mathbf{x}, \mathbf{x}' \rangle + b)^q$

(3) 感知器核函数:  $K(\mathbf{x}, \mathbf{x}') = \tanh(\gamma \langle \mathbf{x}, \mathbf{x}' \rangle + b)$

(4) B 样条函数:  $K(\mathbf{x}, \mathbf{x}') = B_{2n+1}(\mathbf{x} - \mathbf{x}')$

SVM 的应用研究中,模型选择至关重要,好的模型训练得到的 SVM 判别函数会具有好的泛化能力,反之,则得到泛化能力低下的 SVM 判别函数<sup>[2]</sup>.那么 SVM 泛化能力的优劣如果判断呢?由 SVM 原理知,期望风险最小的 SVM 判别函数会具有最好的泛化能力,但期望风险的值往往不易求得.文献[3]中详细证明了 LOO 误差与期望风险的大小具有无偏一致性,所以可以用 LOO 误差来衡量 SVM 判别函数的泛化能力.但 LOO 误差计算耗时,所以人们对 LOO 误差进行了详细研究,并得到了多个上界,其中 Joachims 上界<sup>[3]</sup>和 Jaakkola-Haussler 上界<sup>[4]</sup>最为著名.文献[3]和[4]分别用不同的方法推出了 LOO 误差的两个上界,并分别用实验证明这两个上界对于 SVM 性能评价的有效性.2005 年, Ayat<sup>[5]</sup>提出了一种基于 SVM 分类错误概率估计的 SVM 性能评价准则,并与交叉验证和 VC 维准则进行了比较,显示了该方法选择参数的有效性;文献[6]和[7]将经验误差进行改进,用经验误差的近似梯度来评价 SVM 泛化性能,该方法大大减少了评价所用的时间;文献[8]通过比较特征空间中 5 种不同的距离定义式,指出了类中心之间的 2 范数距离最能有效估计 SVM 判别函数的泛化性,并基于类中心距选择核函数,但该文对类中心距的应用缺乏理论推导和深入分析.本文对 Joachims 上界和 Jaakkola-Haussler 上界进行了详细研究,证明了 RBF 为核函数的 SVM 中,Joachims 上界和 Jaakkola-Haussler 上界是等价的;其次对 Joachims 上界进行理论分析,指出了其不足之处,并予以改进,得到了改进的 Joachims 上界;最后通过实验对 LOO 错误率、Jaakkola-Haussler 上界、Joachims 上界和改进的 Joachims 上界进行了比较,结果显示改进的 Joachims 上界比 Jaakkola-Haussler 上界和 Joachims 上界更加接近 LOO 错误率,是一种更加准确的 SVM 性能评价方法.

## 2 SVM 原理

给定训练样本集:  $\{(\mathbf{x}_i, y_i), i = 1, 2, \dots, l\}$ , 其中  $\mathbf{x}_i \in R^n, y_i \in \{1, -1\}$  代表类别,  $l$  为样本集中样本总数, 则 SVM 的原始优化问题为:

$$\min_{\mathbf{w}, w_0} \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^l \xi_i \quad (1)$$

$$\text{s.t. } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, l$$

其中,  $\mathbf{w}$  是 SVM 决策函数的法向量,  $\xi_i$  称为松弛变量, 是 SVM 分类间隔软化的标志, 意味着可以存在被超平面错分的样本,  $C > 0$  是一个自定义的惩罚因子, 控制着对错分样本的惩罚程度.

引入拉格朗日乘子法, 得其对偶优化问题为:

$$\max \left[ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right] \quad (2)$$

$$\text{s.t. } \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l$$

其中  $\alpha_i$  是每个样本向量对应的拉格朗日系数,  $\alpha_i > 0$  对应的样本向量叫做支持向量(Support Vector, SV),  $0 < \alpha_i < C$  对应的支持向量是界内支持向量,  $\alpha_i = C$  对应的支持向量是界上支持向量<sup>[3]</sup>.

对式(2)求解得到 SVM 判决函数:

$$f(\mathbf{x}) = \text{sgn} \left\{ \sum_{\mathbf{x}_i \in S} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right\} \quad (3)$$

其中,  $S$  表示所有支持向量组成的集合.

松弛变量  $\xi_i$  可由下式求得:

$$\xi_i = 1 - y_i \left( \sum_{\mathbf{x}_j \in S} \alpha_j^* y_j K(\mathbf{x}_j, \mathbf{x}_i) + b^* \right) \quad (4)$$

## 3 Joachims 上界和 Jaakkola-Haussler 上界

**定理 1** (Joachims 上界) 若  $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)$  和  $(\mathbf{w}^*, b^*, \xi^*)$  分别是式(1)和对偶问题的解, 则 SVM 的 LOO 误差满足

$$R_{\text{LOO}}(T) \leq \frac{d}{l} \quad (5)$$

其中  $d$  是集合  $\{i: 2\alpha_i^* R^2 + \xi_i^* \geq 1\}$  中元素的个数,  $R^2$  为核函数的一个上界:  $R^2 = \max \{K(\mathbf{x}_i, \mathbf{x}_j) | i, j = 1, \dots, l\}$ .

**证明** 见文献[3].

**定理 2** (Jaakkola-Haussler 上界) 若  $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)$  是式(1)的解, 则 SVM 的 LOO 误差满足

$$R_{\text{LOO}}(T) \leq \frac{1}{l} \sum_{i=1}^l \text{step}(-y_i f(\mathbf{x}_i) + 2\alpha_i^* K(\mathbf{x}_i, \mathbf{x}_i)) \quad (6)$$

**证明** 见文献[4].

**定理 3** 当 SVM 核函数为 RBF 时, Joachims 上界和 Jaakkola-Haussler 上界是等价的.

**证明** 由式(3)和式(4)知  $\xi_i^* = 1 - y_i f(\mathbf{x}_i)$ , 所以式(6)可变为

$$R_{\text{LOO}}(T) \leq \frac{1}{l} \sum_{i=1}^l \text{step}(\xi_i^* + 2\alpha_i^* K(\mathbf{x}_i, \mathbf{x}_i) - 1) \quad (7)$$

又因为非支持向量对应的  $\xi_i^* = 0, \alpha_i^* = 0$ , 所以式(7)等价于

$$R_{\text{LOO}}(T) \leq \frac{1}{l} \sum_{\mathbf{x}_i \in S} \text{step}(\xi_i^* + 2\alpha_i^* K(\mathbf{x}_i, \mathbf{x}_i) - 1) \quad (8)$$

当核函数为 RBF 时,  $\max \{K(\mathbf{x}_i, \mathbf{x}_j) | i, j = 1, \dots, l\} = K(\mathbf{x}_i, \mathbf{x}_i) = R^2$

所以式(8)可变为

$$R_{\text{LOO}}(T) \leq \frac{1}{l} \sum_{\mathbf{x}_i \in S} \text{step}(\xi_i^* + 2\alpha_i^* R^2 - 1) \quad (9)$$

$\text{step}(\cdot)$  是阶跃函数, 即

$$\text{step}(\eta) = \begin{cases} 1, & \eta \geq 0 \\ 0, & \eta < 0 \end{cases}$$

所以  $\sum_{x_i \in S} \text{step}(\xi_i^* + 2\alpha_i^* R^2 - 1)$  实质表示集合  $\{t: 2\alpha_i^* R^2 + \xi_i^* \geq 1, x_i \in S\}$  中元素的个数, 与式(5)中的  $d$  相等, 所以当 SVM 核函数为 RBF 时, Joachims 上界和 Jaakkola-Haussler 上界等价. 证毕.

#### 4 改进的 Joachims 上界

当 SVM 核函数为 RBF 时,  $R^2 = \max\{K(x_i, x_i) | i, j = 1, \dots, l\} = 1$ , 这时 Joachims 上界可写为  $R_{\text{LOO}}(T) \leq \frac{d}{l}$ , 其中  $d$  是集合  $\{i: 2\alpha_i^* + \xi_i^* \geq 1\}$  中元素的个数. 由文献[1]知, 随着惩罚因子  $C$  的增大,  $\alpha_i^*$  ( $i = 1, \dots, l$ ) 单调增大,  $\xi_i^*$  ( $i = 1, \dots, l$ ) 单调减小, 当  $C > 0.5$  时,  $\{\alpha_i^* | i = 1, \dots, l\}$  中将有一部分值大于 0.5, 从而导致  $d$  值的过分增大. 也就是说, 当  $C > 0.5$  时, Joachims 上界将过大地估计 LOO 误差, 从而过低地估计 SVM 的泛化性能, 使得以 Joachims 上界为准则不能准确选择出 SVM 的最好模型. 为此我们在文献[3]的基础上对其进行了改进, 得到了改进的 Joachims 上界.

**定理 4**(改进的 Joachims 上界) 若  $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)$  和  $(w^*, b^*, \xi^*)$  分别是式(1)和对偶问题的解, 则 SVM 的 LOO 误差满足

$$R_{\text{LOO}}(T) \leq \frac{d}{l} \quad (10)$$

其中  $d$  是集合  $\{i: 2\alpha_i^* (R^2 - M^2) + \xi_i^* \geq 1\}$  中元素的个数,  $R^2$  为核函数值的一个上界:  $R^2 = \max\{K(x_i, x_j) | i, j = 1, \dots, l\}$ ,  $M^2$  为核函数值的一个下界:  $M^2 = \min\{K(x_i, x_j) | i, j = 1, \dots, l\}$ .

**证明** 设 LOO 法中的测试样本为  $(x_t, y_t)$ , 那么基于训练样本  $\{(x_i, y_i) | i = 1, 2, \dots, t-1, t+1, \dots, l\}$  的 SVM 对偶问题表示为

$$W_t(\alpha^t) = \max_{0 \leq \alpha_i^t \leq C} \left\{ \mathbf{1}^T \alpha^t - \frac{1}{2} \alpha^t H \alpha^t \right\} \quad (11)$$

$$y^t \alpha^t = 0$$

其中  $\alpha^t$ ,  $y^t$  和  $H$  均表示把第  $t$  个样本去掉后的向量. 在这种情况下, 如果表达式  $y_t \left( \sum_{i \neq t} \alpha_i^t y_i K(x_i, x_i) + b^t \right) > 0$  为真, 则说明 LOO 法对测试样本  $(x_t, y_t)$  判断正确, 反之, 则判断不正确.

基于所有样本的 SVM 对偶问题表示为

$$W(\alpha) = \max_{0 \leq \alpha_i \leq C} \left\{ \mathbf{1}^T \alpha - \frac{1}{2} \alpha^T H \alpha \right\} \quad (12)$$

$$y^T \alpha = 0$$

在上面两个式子的基础上, 文献[3]讨论了 3 种情况:

**情况 1**  $\alpha_t = 0$ . 这种情况下  $y_t \left( \sum_{i \neq t} \alpha_i^t y_i K(x_i, x_i) + b^t \right) > 0$  永为真.

**情况 2**  $0 < \alpha_t < C$ . 这种情况下首先基于式(11)构造一个拉格朗日系数向量  $\beta$ , 使得

$$\beta_i = \begin{cases} \alpha_i^t, & \text{if } \alpha_i^t = 0 \vee \alpha_i^t = C \\ \alpha_i^t - y_i y_t \nu_i, & \text{if } i \in S^t \\ \alpha_i, & \text{if } i = t \end{cases} \quad (13)$$

其中  $S^t$  是  $W_t(\alpha^t)$  的解中的非边界支持向量组成的集合,  $\nu_i$  是非负实数, 在  $i \notin S^t$  时等于 0, 在  $i \in S^t$  时满足  $\sum_{i \in S^t} \nu_i = \alpha_t$ . 这样, 将  $W_t(\alpha^t)$  的解加上  $\alpha_t$  构造了向量  $\beta$ , 并满足  $y^T \beta = 0 \wedge 0 \leq \beta_i \leq C$ . 所以  $\beta$  可以看成是  $W(\cdot)$  的一个解. 将式(13)代入式(12)得

$$\begin{aligned} & \alpha_t y_t \left( \sum_{i \in S^t} \alpha_i^t y_i K(x_i, x_i) + b^t \right) \\ &= -W(\beta) + W_t(\alpha^t) - \frac{1}{2} \alpha_t^2 K(x_t, x_t) + \alpha_t \\ & \quad - \frac{1}{2} \sum_{i \in S^t} \sum_{j \in S^t} \nu_i \nu_j K(x_i, x_j) + \alpha_t \sum_{i \in S^t} \nu_i K(x_i, x_t) \end{aligned} \quad (14)$$

同样, 基于  $W(\cdot)$  构造系数向量  $\chi$ , 使其满足  $W_t(\cdot)$ ,

$$\chi_i = \begin{cases} \alpha_i, & \text{if } \alpha_i = 0 \vee \alpha_i = C \\ \alpha_i + y_i y_t \mu_i, & \text{if } i \in S \setminus \{t\} \end{cases} \quad (15)$$

$S$  是  $W(\alpha)$  的解中的非边界支持向量组成的集合,  $S \setminus \{t\}$  指从  $S$  中去掉第  $t$  个支持向量,  $\mu_i$  是非负实数, 在  $i \in S \setminus \{t\}$  时满足  $\sum_{i \in S \setminus \{t\}} \mu_i = \alpha_t$ . 这样, 将  $W(\alpha)$  的解减去  $\alpha_t$  构造了向量  $\chi$ , 并满足  $y^T \chi = 0 \wedge 0 \leq \chi_i \leq C$ . 所以  $\chi$  可以看成是  $W_t(\cdot)$  的一个解. 由于  $W(\alpha) \geq W(\beta)$  且  $W_t(\alpha^t) \geq W_t(\chi)$ , 所以由式(14)和式(15)有

$$\begin{aligned} & \alpha_t y_t \left( \sum_{i \in S^t} \alpha_i^t y_i K(x_i, x_i) + b^t \right) \geq \alpha_t y_t \left( \sum_{i \in S \setminus \{t\}} \alpha_i y_i K(x_i, x_i) + b \right) \\ & \quad - \frac{1}{2} \sum_{i \in S^t} \sum_{j \in S^t} \nu_i \nu_j K(x_i, x_j) + \alpha_t \sum_{i \in S^t} \nu_i K(x_i, x_t) \\ & \quad - \frac{1}{2} \sum_{i \in S \setminus \{t\}} \sum_{j \in S \setminus \{t\}} \mu_i \mu_j K(x_i, x_j) + \alpha_t \sum_{i \in S \setminus \{t\}} \mu_i K(x_i, x_t) \end{aligned} \quad (16)$$

由于  $\frac{1}{2} \sum_{i \in S^t} \sum_{j \in S^t} \nu_i \nu_j K(x_i, x_j) \leq \frac{1}{2} \alpha_t^2 R^2$ ,  $\frac{1}{2} \sum_{i \in S \setminus \{t\}} \sum_{j \in S \setminus \{t\}} \mu_i \mu_j K(x_i, x_j) \leq \frac{1}{2} \alpha_t^2 R^2$ ,  $\alpha_t \sum_{i \in S^t} \nu_i K(x_i, x_t) \geq \alpha_t^2 M^2$ ,

$$\begin{aligned} & \alpha_t \sum_{i \in S \setminus \{t\}} \mu_i K(x_i, x_t) \geq \alpha_t^2 M^2, \text{ 所以式(16)可以变为} \\ & y_t \left( \sum_{i \in S^t} \alpha_i^t y_i K(x_i, x_i) + b^t \right) \\ & \geq y_t \left( \sum_{i \in S \setminus \{t\}} \alpha_i y_i K(x_i, x_i) + b \right) - \alpha_t R^2 + 2\alpha_t M^2 \end{aligned} \quad (17)$$

从式(17)可以看出, 当  $y_t \left( \sum_{i \in S \setminus \{t\}} \alpha_i y_i K(x_i, x_i) + b \right) - \alpha_t R^2$

+2 $\alpha_i M^2 \leq 0$  时 LOO 误差将产生, 两边分别加上  $y_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_i)$ , 由于  $(\mathbf{x}_i, y_i)$  是非边界支持向量, 所以有  $y_i (\sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_i) + b) = 1$ , 从而得到  $2\alpha_i R^2 - 2\alpha_i M^2 \geq 1$ .

情况 3  $\alpha_i = C$ . 这时

$y_i (\sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_i) + b) = 1 - \xi_i$ , 所以对边界支持向量来说, 结论应该为  $2\alpha_i (R^2 - M^2) + \xi_i \geq 1$ .

证毕.

可以看出, 在改进的 Joachims 上界中, 系数  $(R^2 - M^2)$  对由于惩罚因子  $C$  过大时  $\alpha_i$  的快速增长有所抑制, 从而降低了 Joachims 上界对 LOO 误差的过高估计, 使其更加接近 LOO 误差.

### 5 模拟实验

由于 RBF 核函数的普适性<sup>[9]</sup>, 实验中 SVM 核函数采用 RBF, 惩罚因子  $C$  取值范围为  $\{2^{-5}, 2^{-4}, \dots, 2^9\}$ . 基于 UCI (University of California, Irvine) 公开实测特征数据库<sup>[10]</sup> Australian, German, Vehicle 和 Satimage 进行模拟实验, 分别得出 4 个数据库的 LOO 误差曲线、Joachims 上界曲线、Jaakkola-Haussler 上界曲线和改进的 Joachims 上界曲线, 并进行详细的比较和讨论.

核参数  $\gamma = 2^{-10}$  时, 实验基于上述 4 个数据库得到 4 幅曲线图如图 1 所示. 每一幅图中有 4 条曲线, 分别是 LOO 错误率曲线、Joachims 上界曲线、Jaakkola-Haussler 上界曲线和改进的 Joachims 上界曲线, 然而图中只显示出 3 条曲线, 是因为每一幅图中的第 2 和第 3 组曲线完全重合, 说明当核函数是 RBF 时, Joachims 上界和

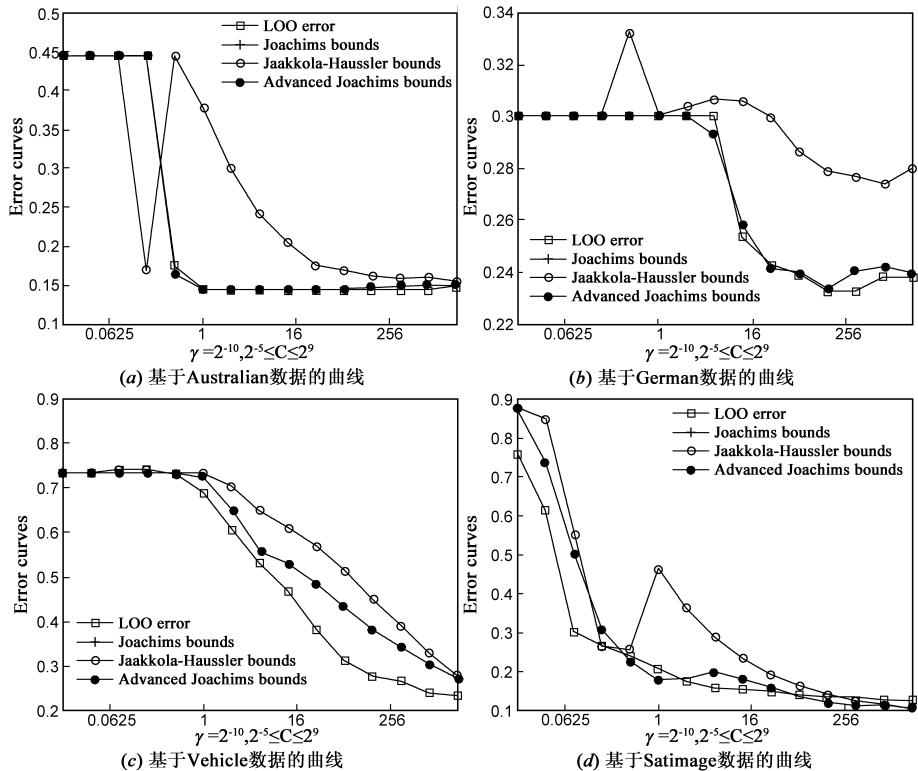


图 1 核参数  $\gamma=2^{-10}$  时, 基于 4 个数据库的 SVM 的 LOO 错误率曲线、Joachims 上界曲线、Jaakkola-Haussler 上界曲线和改进的 Joachims 上界曲线

Jaakkola-Haussler 上界完全等价, 验证了定理 3. 另从图 1 可以看出: (1) Joachims 上界和 Jaakkola-Haussler 上界在  $C > 0.5$  时值突然增大, 在一定范围内其值大于 LOO 误差值, 说明上述的两个上界在  $C > 0.5$  时过高估计了 LOO 误差; (2) 在 4 幅曲线图中, 改进的 Joachims 上界值均比 Joachims 上界更接近 LOO 错误率, 尤其是 Australian 实验曲线, 改进的 Joachims 上界值与 LOO 错误率几乎完全重合, 说明了改进 Joachims 上界作为 SVM 性能评价的显著有效性; (3) Australian 和 German 是二类数据库, 而 Vehicle 和 Satimage 中包含多个类, 说明改进的 Joachims 上界对 SVM 泛化性能的评价不仅适合二类数据库, 也适合多类数据库.

表 1 Australian 数据库在核参数  $\gamma$  分别为  $2^{-10}, 2^{-4}$  和 1 时, 核函数最大值  $R^2$ 、核函数最小值  $M^2$ 、Joachims 上界值以及改进 Joachims 上界值

		$C = 2^{-5}$	$C = 2^{-3}$	$C = 2^{-1}$	$C = 2^0$	$C = 2^3$	$C = 2^5$	$C = 2^7$	$C = 2^9$
$\gamma = 2^{-10}$	$R^2$	1	1	1	1	1	1	1	1
	$M^2$	0.99031	0.99030	0.99030	0.99047	0.99056	0.99095	0.99110	0.99112
	Joachims bound	0.44412	0.44412	0.44412	0.37881	0.20464	0.16836	0.15965	0.15384
	Advanced Joachims bound	0.44412	0.44412	0.16400	0.14368	0.14368	0.14223	0.14804	0.14804
$\gamma = 2^{-4}$	$R^2$	1	1	1	1	1	1	1	1
	$M^2$	0.54825	0.54873	0.56080	0.56864	0.57595	0.57532	0.57589	0.57563
	Joachims bound	0.14368	0.14368	0.17126	0.15239	0.16400	0.16690	0.15529	0.15529
	Advanced Joachims bound	0.14368	0.14368	0.14223	0.14223	0.15674	0.16110	0.15384	0.15529
$\gamma = 1$	$R^2$	1	1	1	1	1	1	1	1
	$M^2$	0.01803	0.01567	0.01479	0.01523	0.01665	0.01677	0.01721	0.01737
	Joachims bound	0.40058	0.09869	0.16400	0.16981	0.19593	0.19593	0.19158	0.18432
	Advanced Joachims bound	0.40058	0.10014	0.15529	0.16836	0.19593	0.19593	0.19158	0.18432

表 1 给出了 Australian 数据库在核参数  $\gamma$  分别为  $2^{-10}$ 、 $2^{-4}$  和 1 时,核函数最大值  $R^2$ 、核函数最小值  $M^2$ 、Joachims 上界值以及改进 Joachims 上界值.由表 1 可以看出:(1)所有情况下,核函数最大值  $R^2 = 1$ ,这是由 RBF 核函数的定义决定的;(2)核函数最小值  $M^2$  随着  $C$  值的增大变化不大,而随着核参数  $\gamma$  的增大单调减小,这是因为核函数最小值  $M^2$  由核参数  $\gamma$  和特征向量之间的最大距离决定的,当后者值不变时, $M^2$  则主要由  $\gamma$  决定;(3)核参数  $\gamma$  较小时,Joachims 上界值与改进 Joachims 上界值相差较大,而随着  $\gamma$  增大,两组上界值大小逐渐接近,这是因为  $\gamma$  较小时系数( $R^2 - M^2$ )较小,对  $\alpha$  抑制程度较强,使改进 Joachims 上界值远离 Joachims 上界值而靠近 LOO 误差值,但当  $\gamma$  较大时系数( $R^2 - M^2$ )较大,对  $\alpha$  抑制程度变弱,使得改进 Joachims 上界值与 Joachims 上界值慢慢靠近;(4)改进的 Joachims 上界在核参数较小时能准确地评价 SVM 性能,而当核参数较大时其效果与 Joachims 上界相当,是一种准确度介于 LOO 误差率和 Joachims 上界之间的 SVM 性能评价标准.

## 6 小结

SVM 泛化能力的评价是 SVM 研究的基础性内容,其评价方法的准确程度影响着 SVM 模型选择结果的优劣,最终影响着 SVM 判决函数的识别准确率和预测准确率.评价 SVM 泛化能力最基本、最准确的方法是 LOO 误差法,但是 LOO 实现起来太过费时,对于较大规模数据就不再适用.因此人们提出了 LOO 错误率的各种上界作为 LOO 错误率的近似,最有代表性的两个为 Joachims 上界和 Jaakkola-Haussler 上界.本文对上述两种上界进行了详细研究,发现了其中的不足并加以改进,得到了改进的 Joachims 上界.文中分别从理论和实验两个方面分析和验证了改进的 Joachims 上界比 Joachims 上界和 Jaakkola-Haussler 上界更接近于 LOO 错误率,是一种更为准确和实用的 SVM 泛化性能评价方法.

## 参考文献

- [1] Vapnik V. The Nature of Statistics Learning Theory [M]. New York: Springer Verlag, 1995.
- [2] 周伟达,张莉,焦李成.支撑矢量机推广能力分析[J].电子学报,2001,29(5):590-594.  
ZHOU W D, ZHANG L, JIAO L C. An analysis of SVMs gen-

eralization performance [J]. Acta Electronica Sinica, 2001, 29(5):590-594. (in Chinese)

- [3] Joachims T. Estimating the Generalization Performance of a SVM Efficiently [R]. Dortmund: University Dortmund, 2000.
- [4] Jaakkola T S, Haussler D. Exploiting generative models in discriminative classifiers [A]. Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II [C]. Massachusetts: MIT Press, 1998. 487-493.
- [5] Ayat N E, Cheriet M, Suen C Y. Automatic model selection for the optimization of SVM kernels [J]. Pattern Recognition, 2005, 38(10): 1733-1745.
- [6] Adankon M M, Cheriet M. Optimizing resources in model selection for support vector machine [J]. Pattern Recognition, 2007, 40(3): 953-963.
- [7] Adankon M M, Cheriet M. Model selection for the LS-SVM application to handwriting recognition [J]. Pattern Recognition, 2009, 42(12): 3264-3270.
- [8] WU K P, WANG S D. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space [J]. Pattern Recognition, 2009, 42(5): 710-717.
- [9] Hsu C W, Chang C C, Lin C J. A Practical Guide to Support Vector Classification [OL]. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, 2003.
- [10] Blake C L, Merz C J. UCI Repository of Machine Learning Databases [EB/OL]. <http://www.ics.uci.edu/~mlearn/MLRepository.html>, 2003.

## 作者简介



宋小杉 男,1980年10月1日出生于陕西西安,现为装甲兵工程学院控制工程系博士生.  
E-mail: sxsh029@yahoo.com.cn



蒋晓瑜 男,1967年9月生于湖南安仁,现为装甲兵工程学院控制工程系教授、博士生导师.  
E-mail: jiangxiaoyu2007@gmail.com