

最小二乘支持向量机的算法研究

顾燕萍¹, 赵文杰², 吴占松¹

(1. 清华大学 热能工程系, 电力系统国家重点实验室, 北京 100084; 2. 华北电力大学 自动化系, 保定 071003)

摘要: 最小二乘支持向量机(LS-SVM)作为一种新颖的人工智能技术,已越来越广泛地运用于各个学科领域。该文阐述了最小二乘支持向量机的主要思想和基本算法;结合统计学习理论和算例分析了模型参数对模型精度、复杂度和计算量等的影响,为模型参数的确定提供了理论参考;还提出了最小二乘支持向量机的一种改进算法,通过工程实例对比了基于改进算法和原算法的最小二乘支持向量机模型的性能。算例表明该改进算法可以有效地提高模型的整体性能,便于模型在工程上推广使用。

关键词: 最小二乘支持向量机; 支持向量; 精度; 复杂度; 改进算法

中图分类号: TK

文献标志码: A

文章编号: 1000-0054(2010)07-1063-04

Least squares support vector machine algorithm

GU Yanping¹, ZHAO Wenjie², WU Zhansong¹

(1. State Key Laboratory of Power System, Department of Thermal Engineering, Tsinghua University, Beijing 100084, China
2. Department of Automation, North China Electric Power University, Baoding 071003, China)

Abstract The least squares support vector machine (LS-SVM) has been widely used in various fields for artificial intelligence. This paper describes the basic LS-SVM algorithm. The influence of the model parameters on the model precision, complexity and computing time are analyzed using statistical learning theory and numerical examples. The conclusions give a good guide for determining the model parameters. An improved LS-SVM algorithm is developed with the results showing that the improved algorithm having better overall performance and better efficiency for wider application.

Key words: least squares support vector machine (LS-SVM); support vector; accuracy; complexity; improved algorithm

支持向量机是由 Vapnik 最初在统计学习理论的基础上建立起来的一种非常有力的机器学习方法,是一种新颖的人工智能技术,最初用于模式识别,目前在信号处理、系统辨识与建模、先进控制和

软测量等领域都得到了广泛的应用^[1]。

与人工神经网络相比,支持向量机能够很好地克服前者训练时间长、训练结果存在随机性和过学习等不足^[2],因此越来越广泛地被用于复杂非线性系统的建模中。最小二乘支持向量机是支持向量机的一种类型,本文旨在阐明其基本算法,分析模型参数对模型精度、复杂度、计算量和变量存储空间的影响,并提出改进算法以提高模型的整体性能。

1 LS-SVM 的算法

若训练样本集为 (x_i, y_i) , $i = 1, 2, \dots, n$, $x \in \mathbf{R}^d$, $y \in \mathbf{R}$ 。支持向量机建模的主要思想是:首先,用一非线性映射 $\varphi(\cdot)$ 将样本的输入空间 \mathbf{R}^d 映射到特征空间

$$\psi(x) = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n));$$

然后,在这个高维特征空间中构造最优决策函数 $y = w^T \cdot \varphi(x) + b$;最后,以结构风险最小化为原则确定模型参数 w 、 b 。

结构风险的计算式为

$$R = c \cdot R_{\text{emp}} + \frac{1}{2} \|w\|^2. \quad (1)$$

其中: c 为正规化参数; R_{emp} 为损失函数,又称为经验风险。常见的损失函数有一次损失函数、二次损失函数和 Hubber 损失函数,不同的损失函数代表不同的支持向量机模型。最小二乘支持向量机是损失函数为二次损失函数的支持向量机,即 $R_{\text{emp}} = \sum_i \xi_i^2$, ξ_i 为模型对训练样本的预测误差。经验风险最小并不代表模型的期望风险最小,人工神经网络的过学习就是经验风险最小化原则失败的例子^[3]。

根据统计学习理论,期望风险 $R(\alpha)$ 和经验风险

收稿日期: 2009-09-08

作者简介: 顾燕萍(1986—),女(汉),江苏,博士研究生

通讯作者: 吴占松,教授, E-mail: wzs@tsinghua.edu.cn

R_{emp} 之间至少以概率 $1-\eta$ 满足如下关系^[4]：

$$R(\alpha) \leqslant R_{\text{emp}} + \sqrt{\frac{h(\ln(2n/h)+1)-\ln(\eta/4)}{n}}. \tag{2}$$

其中： h 是函数集的 VC 维， n 是样本数。式 (2) 表明：同时减小经验风险和置信范围可以降低期望风险；而降低函数集的 VC 维可以缩小置信范围。又由于函数集的 VC 维与 $\|\boldsymbol{w}\|$ 有关，减小 $\|\boldsymbol{w}\|$ 可以降低 VC 维的大小^[5]，因此将 $\|\boldsymbol{w}\|^2/2$ 作为结构风险最小化原则的第二项。

基于结构风险最小化原则确定决策函数参数 \boldsymbol{w} 、 b ，可等效为求解以下优化问题：

$$\begin{aligned} \min R &= c \cdot \sum_{i=1}^n \xi_i + \frac{1}{2} \cdot \|\boldsymbol{w}\|^2, \\ \text{s. t. } y_i &= \boldsymbol{w}^T \cdot \varphi(\boldsymbol{x}_i) + b, \quad i = 1, 2, \cdots, n. \end{aligned} \tag{3}$$

用 Lagrange 方法求解这个优化问题：

$$\begin{aligned} L(\boldsymbol{w}, b, \xi_i, \alpha) &= c \cdot \sum_{i=1}^n \xi_i + \frac{1}{2} \|\boldsymbol{w}\|^2 - \\ &\sum_{i=1}^n (\alpha_i \cdot (\boldsymbol{w}^T \cdot \varphi(\boldsymbol{x}_i) + b - y_i)). \end{aligned} \tag{4}$$

其中 $\alpha=[\alpha_1, \alpha_2, \cdots, \alpha_n]$ 是 Lagrange 乘子。

根据优化条件

$$\frac{\partial L}{\partial \boldsymbol{w}} = 0, \quad \frac{\partial L}{\partial b} = 0, \quad \frac{\partial L}{\partial \xi_i} = 0, \quad \frac{\partial L}{\partial \alpha_i} = 0, \tag{5}$$

可得

$$\boldsymbol{w} = \sum_{i=1}^n \alpha_i \cdot \varphi(\boldsymbol{x}_i), \tag{6}$$

$$\sum_{i=1}^n \alpha_i = 0, \tag{7}$$

$$2c\xi_i = \alpha_i, \tag{8}$$

$$y_i = \boldsymbol{w}^T \cdot \varphi(\boldsymbol{x}_i) + b + \xi_i. \tag{9}$$

将式 (6) 和式 (8) 代入式 (9)，得

$$y_i = \sum_{j=1}^n (\alpha_j \cdot \langle \varphi(\boldsymbol{x}_j), \varphi(\boldsymbol{x}_i) \rangle) + b + \frac{1}{2c} \alpha_i. \tag{10}$$

若定义核函数 $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \varphi(\boldsymbol{x}_j), \varphi(\boldsymbol{x}_i) \rangle$ ，则

$$y_i = \sum_{j=1}^n (\alpha_j \cdot K(\boldsymbol{x}_i, \boldsymbol{x}_j)) + b + \frac{1}{2c} \alpha_i. \tag{11}$$

将式 (7) 和 (11) 合成线性方程组如下：

$$\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & K(\boldsymbol{x}_1, \boldsymbol{x}_1) + \frac{1}{2c} & K(\boldsymbol{x}_1, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ 1 & K(\boldsymbol{x}_2, \boldsymbol{x}_1) & K(\boldsymbol{x}_2, \boldsymbol{x}_2) + \frac{1}{2c} & \cdots & K(\boldsymbol{x}_2, \boldsymbol{x}_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K(\boldsymbol{x}_n, \boldsymbol{x}_1) & K(\boldsymbol{x}_n, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_n, \boldsymbol{x}_n) + \frac{1}{2c} \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

$$\begin{bmatrix} b \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \tag{12}$$

基于训练样本集 (\boldsymbol{x}_i, y_i) ， $i = 1, 2, \cdots, n$ ，求解线性方程组 (12)，可以得到模型参数 $[b \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$ 。最后所确定的决策函数为

$$f(\boldsymbol{x}) = \sum_{i=1}^n \alpha_i \cdot K(\boldsymbol{x}, \boldsymbol{x}_i) + b. \tag{13}$$

支持向量系数 α 中不等于零的元素 α_i 所对应的样本 (\boldsymbol{x}_i, y_i) 称为支持向量^[4]。从决策函数式 (13) 可以看出，对象的特性主要是由支持向量决定的。

算法中的核函数 $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 是高维特征空间的内积，根据泛函的有关理论，只要满足 Mercer 条件的函数都可以作为核函数^[4]。不同的核函数构造不同的支持向量机，常见的核函数形式有：

- 1) 线性核函数： $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \boldsymbol{x}_j$ ，
- 2) 多项式核函数： $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \boldsymbol{x}_j + 1)^d$ ，
- 3) 径向基函数 (RBF)： $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left[-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right]$ 。

2 LS-SVM 参数的确定方法

本文以径向基函数为核函数的 LS-SVM 为例，分析模型参数对模型性能的影响，为模型参数的确定提供理论参考。

以 RBF 为核函数的 LS-SVM 模型结构为

$$f(\boldsymbol{x}) = \sum_{i=1}^n \alpha_i \cdot K(\boldsymbol{x}, \boldsymbol{x}_i) + b.$$

其中： $K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left[-\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|^2}{2\sigma^2}\right]$ ，模型参数 $[b \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$ 通过求解线性方程组 (12) 得到，而结构风险计算式中的正规化参数 c 和核参数 σ 则需要事先假设。

正规化参数 c 和核参数 σ 对模型的预测精度和复杂度都有很大的影响，寻找最佳正规化参数和核参数属于最佳模型选择问题，目前文献中还没有具体的参数选取方法。本文以非线性函数 $z = x^2 + y^2$ 为例，结合统计学习理论，分析正规化参数 c 和核参数 σ 对模型性能的影响。

算例中共准备 49 个样本，其中 40 个作为训练样本，9 个为测试样本。

2.1 核参数 σ

保持正规化参数 $c=50$ 不变, 分析核参数对模型性能的影响。

图 1 为核参数 $\sigma=0.1$ 和 $\sigma=0.5$ 的核函数示意图。结合图 1, 分析核函数的表达式可知, 核参数 σ 越大, 式(12)中系数矩阵的稀疏性就越弱, 模型训练时的计算量就越大, 训练时间也就越长。

图 2 显示了核参数 $\sigma=0.1$ 和 $\sigma=0.5$ 时, 由式(12)计算得到的参数向量 α 的值, 从图中可以看出核参数越小, 向量 α 中接近 0 的就越多, 模型支持向量的个数就越少, 模型复杂度也越低。

图 3 显示了模型的预测相对误差 e_r 与核参数 σ

的关系, 从图中可以看出随着核参数 σ 的增大, 模型对训练样本的预测误差大致呈增大趋势, 逼近能力变差; 模型对测试样本的预测误差是先减小后增大的, 在核参数取某一值时模型具有最佳的泛化能力。

2.2 正规化参数 c

保持核参数 $\sigma=0.2$ 不变, 分析正规化参数对模型性能的影响。

图 4 显示了正规化参数 $c=1$ 和 $c=100$ 时由式(12)计算得到的参数向量 α 的值, 从图中可以看出 c 越小, 参数 α 中接近 0 的就越多, 说明模型的支持向量个数就越少, 模型复杂度就越低。

图 5 显示了模型预测的相对误差与正规化参数 c 的关系。从图中可以看出, 随着 c 的增大, 模型对训练样本和测试样本的预测误差均下降, 但下降的幅度逐渐减小。也就是说, 模型的逼近能力和泛化能力都会随着正规化参数 c 的增大而变好, 但当 c 增大到一定值后, 模型精度提高的幅度不大。

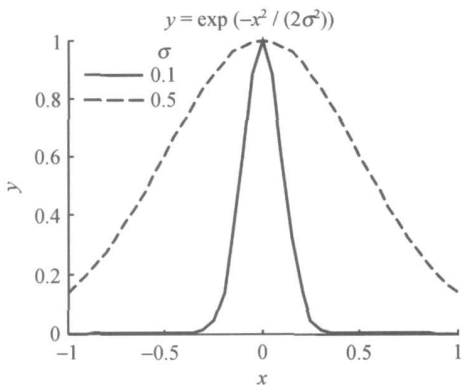


图 1 不同核参数的核函数图

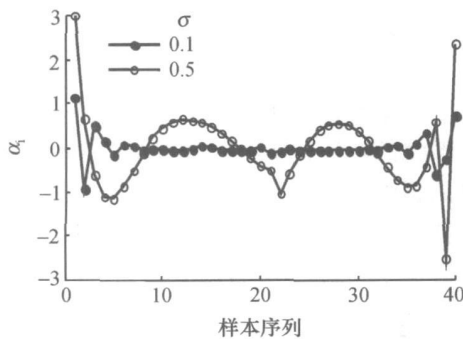


图 2 核参数对参数 α 的影响

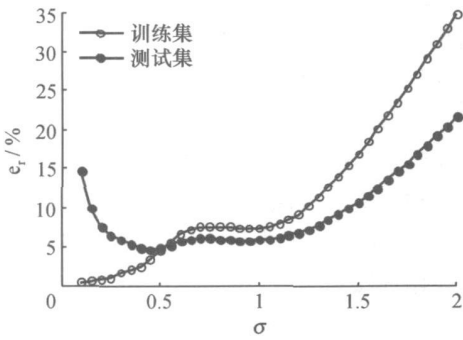


图 3 模型预测相对误差与核参数的关系

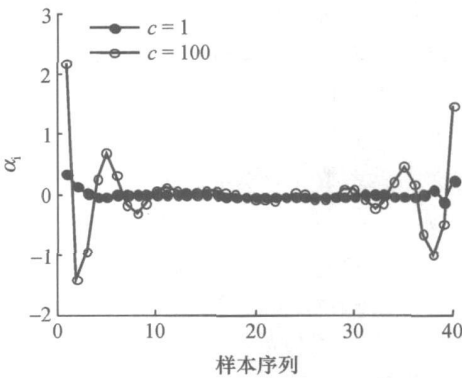


图 4 正规化参数对参数 α 的影响

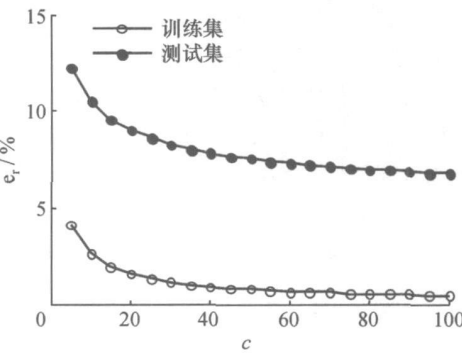


图 5 模型预测相对误差与正规化参数的关系

以上结论还可以从结构风险计算式中得到进一步验证。分析结构风险计算式(1)可知, 正规化参数 c 越小, 表明经验风险在基于结构风险最小化参数优化原则中所占的比重越弱, 式中第二项更重要, 也

即结构风险侧重通过减少 VC 维来降低模型的实际风险,因此模型的复杂度会相应的下降,而精度会变差;反之,正规化参数 c 越大,结构风险则更侧重经验风险项,因此模型的精度会提高,但会更复杂。

以上分析了核参数和正规化参数对支持向量个数、模型逼近能力、泛化能力以及训练时计算量的影响。通过分析可知,模型精度和复杂度对参数的要求是相互矛盾的,实际运用时应根据需要进行最佳的模型参数。

3 LS-SVM 的改进算法

从决策函数式(13)可以看出,实际运用时,为实现 LS-SVM 模型的预测,不仅需要保存模型参数 $[b \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$,还需要保存支持向量集,因此需要大量的变量存储空间,而且计算量也很大,特别是对于输入维数较多、对象又较复杂的情况。

但在研究的过程中发现,通过求解线性方程组(12)得到的系数 $\alpha_i, i=1, 2, \cdots, n$ 中,绝对值较大的只有少数,很多 α_i 都在 0 附近,这说明对应样本在反映对象特性中所起的作用很小,称之为边界样本。为减少模型所需的变量存储空间,降低计算量,本文提出最小二乘支持向量机的改进算法。改进算法的描述如下:

定义临界系数 α_c ,规定通过式(12)计算得到的模型参数 $[b \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$ 中,若 $|\alpha_i| < \alpha_c$,则令 $\alpha_i = 0$,临界系数 α_c 可以视工程要求具体规定,如果对象对模型预测精度要求较高, α_c 取较小的值;反之可加大 α_c 的取值以降低模型的复杂度。

本文借用文[6]中的锅炉热态实验数据,分别建立基于原算法和改进算法的电站锅炉 NO_x 排放量的最小二乘支持向量机模型,并对比两模型的性能。

样本数据共 105 组,其中 90 组作为训练样本,15 组作为测试样本。根据前面的分析,确定模型正规化参数 $c=70$ 、核参数 $\sigma=0.4$ 。基于原算法和改进算法建立的支持向量机模型性能对比如表 1 所示,模型对 NO_x 排放量的预测如图 6 所示,图 6 中改进算法设置临界系数 $c=0.05$ 。

表 1 改进算法与原算法的对比

算法	预测相对误差/ %		支持向量个数
	训练集	测试集	
原算法	0.11	4.62	90
改进算法	$\alpha_c=0.02$	4.60	76
	$\alpha_c=0.05$	4.34	68
	$\alpha_c=0.08$	4.80	60

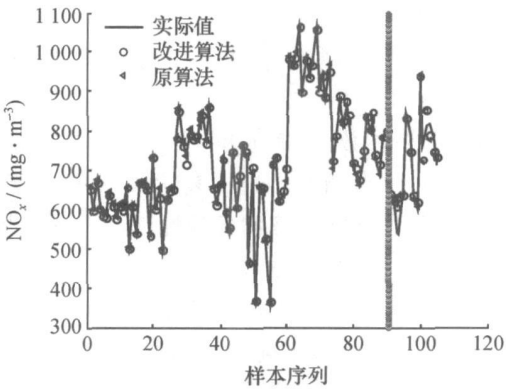


图 6 模型对 NO_x 排放量(标准状态下)的预测

图 6 表明,基于原算法和改进算法建立的电站锅炉 NO_x 排放量模型具有较高的精度,都可以实现 NO_x 排放的预测。

从表 1 中可以看出:改进算法与原算法相比,模型精度略有下降,但只要适当选取临界系数 α_c ,模型精度仍可以控制在允许的范围内;与原算法相比,改进算法可以很大程度地减少支持向量的数目;临界系数 α_c 在一定范围内选取的越大,支持向量的个数越少,模型对训练集的预测误差越大,但对测试集的预测误差影响不大。也就是说:模型复杂度会随着临界系数的增大而降低,临界系数对模型的泛化能力影响不大,但模型的逼近能力对临界系数比较敏感,会随着其增大而变弱。因此应用时应结合实际情况,合理选取临界系数。

理论上,改进算法只取消边界样本的作用,仍然保留决定对象特征的主体样本,因此模型仍然能够反映出系统的特性;算例结果表明:这种改进仍可使模型的预测精度满足要求,但却可以很大程度地减少支持向量的数目,降低模型的复杂度,减少变量所需的存储空间,减少运算量,缩短计算时间。

4 结 论

最小二乘支持向量机的模型参数对模型精度、复杂度、变量存储空间和计算量等都有很大的影响。模型精度和复杂度对参数的要求是互相矛盾的,选择模型参数时应根据实际需要,综合考虑模型的整体性能。

本文所提出的最小二乘支持向量机改进算法,可以有效地降低支持向量数目,减少变量存储空间和计算量,而对模型精度没有太大的影响,因此,提高了模型的整体性能,有利于该模型在工程上推广使用。

(下转第 1071 页)

器内敷设耐火材料, 整体热流密度小, 停滞产生的破坏后果就远没有发生在煤粉锅炉中的严重。炉膛侧墙水冷壁 No. 9 停滞校核安全系数较高, 表明在炉膛主体内发生停滞的概率较小。

使用这种结构的炉排-循环床垃圾焚烧炉已经运行 5 a 以上, 分离器的中心筒出现过变形故障, 分离器入口也因磨损减薄, 但没有出现因受热不均匀、局部停滞而导致管壁温度升高诱发受热面爆管事故。可见, 尽管从计算上表明分离器内可能出现停滞, 但事实上并没有因此而出现事故。原因一方面可能是由于计算方法比较保守; 另一方面, 鉴于停滞发生在整体受热弱的管内, 这种停滞可以认为是安全的。

4 结 论

水循环计算表明, 炉排-循环床复合垃圾焚烧炉的水冷分离器受热面内发生停滞的可能性较大, 但由于其表面敷设有绝热材料, 降低了传热热流, 因而不会出现因停滞导致的传热恶化、管壁温度升高, 引起锅炉爆管事故。5 a 以上的运行时间内也未出现过分离器内爆管事故。这表明: 布置紧凑的炉排-循环床复合垃圾焚烧炉, 虽然在结构上牺牲了对水循环安全性的保障, 但综合考虑了炉内传热与水循环的匹配, 事实上是安全的。

致谢 在焚烧炉的水循环计算结构分区、方法选择方面, 杨瑞昌教授给予了宝贵的指导和建议。

参考文献 (References)

[1] 李清海, 张衍国, 陈昌和, 等. 水分对垃圾焚烧燃烧影响的实验研究 [J]. 中国电机工程学报, 2008, 28(8): 58-64.
LI Qinghai, ZHANG Yanguo, CHEN Changhe, et al. Experimental study of moisture impact on municipal solid waste incineration [J]. *Proceedings of the CSEE*, 2008, 28(8): 58-64. (in Chinese)

[2] 李清海, 张衍国, 党文达, 等. 炉排-循环床复合垃圾焚烧炉中垃圾模拟干燥实验 [J]. 清华大学学报 (自然科学版), 2008, 48(5): 824-827.
LI Qinghai, ZHANG Yanguo, DANG Wenda, et al. Experimental investigation of simulated waste drying in a grate-CFB incinerator [J]. *J Tsinghua Univ (Sci & Tech)*, 2008, 48(5): 824-827. (in Chinese)

[3] 陈昌和, 李清海, 张衍国, 等. 炉排-循环床复合垃圾焚烧炉燃烧过程模型 [J]. 清华大学学报 (自然科学版), 2008, 48(5): 832-835.
CHEN Changhe, LI Qinghai, ZHANG Yanguo, et al. Model for waste combustion in a grate-CFB [J]. *J Tsinghua Univ (Sci & Tech)*, 2008, 48(5): 832-835. (in Chinese)

[4] LI Qinghai, ZHANG Yanguo, CHEN Meiqian, et al. Study on drying and combustion process in grate-CFB incinerator [J]. *Science in China Series E: Technological Sciences*, 2009, 52(5): 1153-1160

[5] 冯俊凯, 沈幼庭, 杨瑞昌. 锅炉原理与计算 [M]. 北京: 科学出版社, 2003.
FENG Junkai, SHEN Youting, YANG Ruichang. *Boiler Principle and Calculation* [M]. Beijing: Science Press, 2003. (in Chinese)

[6] 冯俊凯, 杨瑞昌. 锅炉水自然循环原理计算及实验方法 [M]. 北京: 清华大学出版社, 1994.
FENG Junkai, YANG Ruichang. *Boiler Natural Circulation Principle and Experiment* [M]. Beijing: Tsinghua University Press, 1994. (in Chinese)

(上接第 1066 页)

参考文献 (References)

[1] Smola A J, Scholkope B. A tutorial on support vector regression [J]. *Statistics and Computing*, 2004, 14: 199-222.

[2] Zheng L G, Zhou H, Wang C L, et al. Combing support vector regression and ant colony optimization to reduce NO_x emissions in coal-fired utility boilers [J]. *Energy and Fuels*, 2008, 22: 1034-1040.

[3] 张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报, 2000, 26(1): 32-41.

ZHANG Xuegong. Introduction to statistical learning theory and support vector machines [J]. *Acta automatica sinica*, 2000, 26(1): 32-41. (in Chinese)

[4] Vapnik V N. *Statistical Learning Theory* [M]. New York: J Wiley, 1998.

[5] Vapnik V N. *The Nature of Statistical Learning Theory* [M]. New York: Springer Verlag, 1999.

[6] 张毅. 电站锅炉燃烧优化控制理论及应用研究 [D]. 北京: 清华大学, 2006.
ZHANG Yi. *Research on Theory and Application of Combustion Optimization Control of Utility Boilers* [D]. Beijing: Tsinghua University, 2006. (in Chinese)