

基于支持向量机的烟气二氧化硫排放量预测模型^①

薛美盛, 王旭, 冀若阳

(中国科学技术大学 自动化系, 合肥 230026)
通讯作者: 王旭, E-mail: 1362253866@qq.com

摘要: 针对循环流化床锅炉控制系统的烟气 SO_2 对象的非线性特点, 本文建立了一种基于支持向量机的烟气 SO_2 排放量预测模型. 由于直接网格搜索确定支持向量机回归模型参数的方法计算量大、搜索时间长, 本文采用单变量参数搜索结合网格寻优的方法来确定模型参数. 仿真结果表明, 基于支持向量机方法建立的循环流化床锅炉烟气 SO_2 排放量预测模型具有良好的预测效果.

关键词: 二氧化硫; 支持向量机; 网格寻优; 预测模型

引用格式: 薛美盛, 王旭, 冀若阳. 基于支持向量机的烟气二氧化硫排放量预测模型. 计算机系统应用, 2018, 27(2): 186-191. <http://www.c-s-a.org.cn/1003-3254/6188.html>

Prediction Model of SO_2 Emissions in Flue Gas Based on Support Vector Machine

XUE Mei-Sheng, WANG Xu, JI Ruo-Yang

(Department of Automation, University of Science and Technology of China, Hefei 230026, China)

Abstract: In consideration of the nonlinearity of SO_2 in circulating fluidized bed boiler, a prediction model of SO_2 emissions in flue gas based on support vector machine is proposed. It is complex to directly search the parameters of support vector machine regression, so a method combining single variable search and grid search is applied. The simulation shows that the prediction model of SO_2 emissions in circulating fluidized bed boiler based on support vector machine has good prediction performance.

Key words: SO_2 ; support vector machine; grid search; prediction model

1 引言

随着我国经济的高速发展, 电力需求不断增多, 且我国大部分的电能来自燃煤火力发电. 燃煤火电厂在利用燃煤发电的过程中, 会产生各种废气、废水、灰渣, 其中废气中带有 SO_2 是大气的主要污染物之一^[1]. 国家对于排烟的指标要求越来越高, 因此实现脱硫系统的高效稳定运行显得至关重要^[2].

目前, 各热电厂主要采用的烟气检测装备是烟气连续排放检测系统 (CEMS). 因为烟气的检测系统工作情况复杂, 采样的探头、传感器和很多的光学器件会随着时间的积累出现不同程度的腐蚀老化^[3]. 不仅如

此, 安装 CEMS 系统的费用昂贵, 且需要花费大量的人力去维护, 因此建立烟气二氧化硫的预测模型, 得到较为精准的二氧化硫输出预测值很有研究意义, 对锅炉控制系统运行参数的调节也有参考指导的价值.

支持向量机 (Support Vector Machine, SVM) 是一种以统计学理论为基础的新机器学习方法^[4], 在解决小样本、高维度和非线性等问题有特有的优势^[5]. 另外, 该算法通过转化为一个二次规划问题, 可以得到全局最优解, 解决了神经网络训练存在的局部极小值问题^[6], 因此基于支持向量机的系统建模方法非常具有发展前景. 本文基于支持向量机回归方法, 针对循环流化床锅

^① 收稿时间: 2017-04-26; 修改时间: 2017-05-19; 采用时间: 2017-05-25

炉(CFB)控制系统中烟气二氧化硫对象的非线性问题,建立一种多输入单输出的二氧化硫浓度预测模型,仿真结果表明,基于支持向量机的烟气二氧化硫预测模型具有很好的预测效果。

2 支持向量机建模原理

为了把 SVM 推广到回归估计算法中去,我们引入一种损失函数(ε 不敏感函数)来实现^[7],如式(1)。

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & |y - f(x)| \leq \varepsilon \\ |y - f(x)|_\varepsilon - \varepsilon, & \text{其它} \end{cases} \quad (1)$$

对于线性回归,假设所有的训练数据都可以用线性函数在精度 ε 精度精度之外情况,引入 $\xi_i \geq 0$ 和 $\xi_i^* \geq 0$,则回归问题转化成最小化结构风险(SRM)函数的问题^[8],如式(2)。

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

式中惩罚因子 $C > 0$,对经验风险与置信范围这两部分进行折中,指的是对超出误差的样本的惩罚程度。其约束条件可用式(3)表示。

$$\begin{cases} \omega \cdot x_i + b - y_i \leq \varepsilon + \xi_i \\ y_i - \omega \cdot x_i - b \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \quad i = 1, 2, \dots, n \quad (3)$$

构造拉格朗日函数,通过求偏导转化为求解对偶二次规划问题,最终得到回归函数,如式(4)。

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \alpha_i)(x \cdot x_i) + b \quad (4)$$

另外,对于非线性问题,可以通过映射将原始低维样本转化为某个高维空间的线性问题。由于在上面的求解对偶问题中,只涉及到训练样本之间的内积运算($x_i \cdot x_j$),因此我们用核函数 $K(x_i, x_j)$ 替代原来的内积运算来实现高维空间上的非线性函数拟合^[9]。本文采用的核函数是径向基核函数(RBF): $K(x_i, x) = \exp(-\|x - x_i\|^2 / g^2)$ 。

3 烟气二氧化硫预测模型仿真

3.1 烟气对象

石灰石和燃煤分别放在石灰石仓和燃煤仓内,由各自的传送带送入炉膛内,通过床料加热着火,在一次风的作用下呈流化状态燃烧。燃烧产生的烟气经过旋风分离器分离,利用余温加热过热器、空气预热器等,

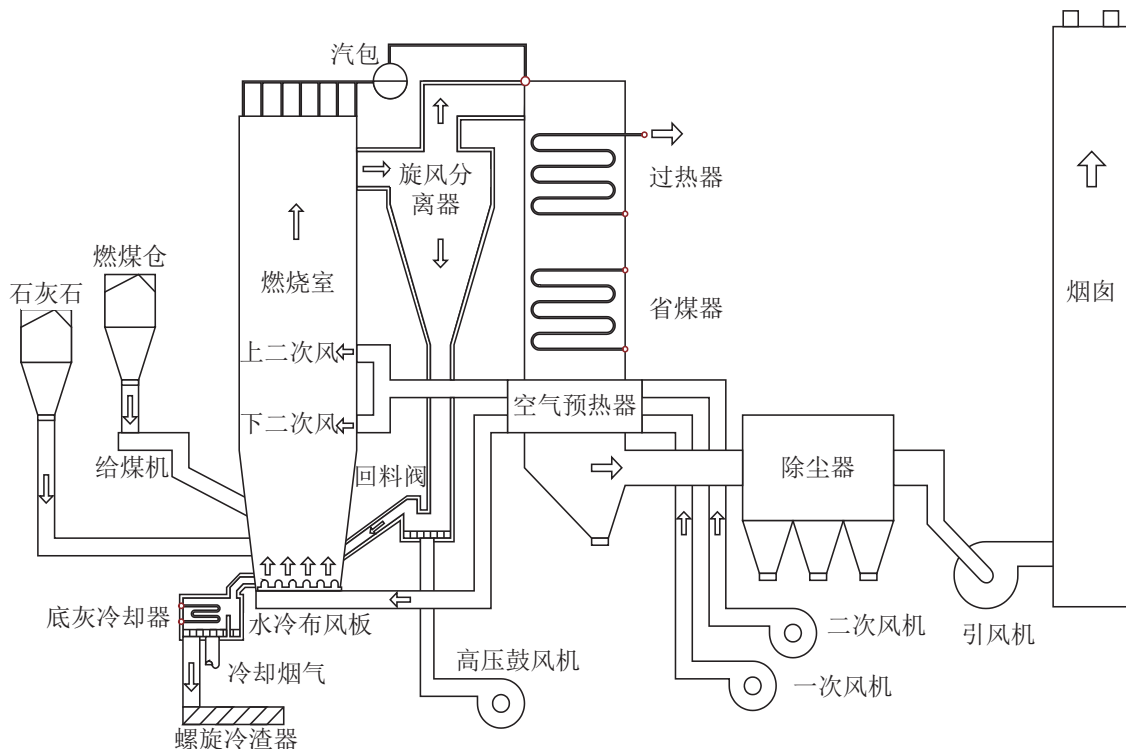


图1 循环流化床锅炉燃烧工作原理图

最后通过除尘器除尘后由烟道排除. 其工作原理如图1所示.

石灰石和煤燃烧生成的 SO_2 在炉膛内反应生成硫化盐, 从而达到脱硫的目的. 根据循环流化床锅炉燃烧的工作原理和添加石灰石的脱硫方法, 可知石灰石量、给煤量影响着烟气二氧化硫的浓度. 一次风、二次风影响着锅炉燃料的燃烧状态及炉内氧含量, 所有对二氧化硫的浓度也有很大影响. 床温不仅影响着煤燃烧的速度还影响石灰石脱硫的效率, 因此也是影响二氧化硫浓度的重要因素.

3.2 变量选择及数据预处理

选取某动力站1#炉CFB锅炉的稳定运行数据, 根据该循环流化床锅炉燃烧特性和相关系数法计算, 烟气二氧化硫预测模型的输入变量选择为: 给煤量、石灰石给料传送带转速、一次风A导叶、一次风B导叶、二次风A液偶、二次风B液偶、床温等7个参数.

选取2015.03.22一天的稳定运行数据进行仿真实验. 建模前, 对训练数据和测试数据进行归一化预处理, 采用式(5)进行归一化.

$$x' = \frac{x - \mu}{\sigma} \quad (5)$$

式中, x 为原始数据, x' 为归一化之后的数据, μ 为原始样本数据均值, σ 为原始样本数据标准差.

3.3 仿真实验

在MATLAB仿真平台上进行仿真研究, 操作系统为Windows 7, 本文采用通用的支持向量机工具箱, 其作者是Steve Gunn, 将工具箱添加到“Matlab search path”中, 通过调用工具包的功能函数来实现预测模型的训练和测试. 调用svr函数来训练模型, 见式(6)

$$[nsv, beta, bias] = svr(X_1, Y_1, ker, C, loss, e) \quad (6)$$

其中 X_1 和 Y_1 分别为输入输出数据集, ker 表示核函数的选择; C 为惩罚因子的参数设置, $loss$ 为损失函数的选择设置, e 为设置的不敏感系数值. 函数返回值 nsv 为支持向量个数, $beta$ 和 $bias$ 为返回的模型拉格朗日乘子解.

采用svroutpu函数计算模型的输出, 如式(7), $tstY_1$ 为返回的模型输出值.

$$tstY_1 = svroutput(X_1, X_1, ker, beta, bias) \quad (7)$$

3.4 参数寻优

采用高斯径向基函数RBF作为寻优核函数, 经过

多年的应用实践看出, RBF核函数是一个较为的普适函数^[10], 可以通过参数的选取, 使其适用于任意分布的样本. 这样SVR可调的模型参数有惩罚系数 C 、核函数参数 g 和不敏感系数 ϵ ^[11]. 其中 C 的大小表示对误差的惩罚程度, 若是 C 越小, 惩罚程度越小. g 的变化会引起映射核函数的变化, 导致样本空间分布的复杂程度改变^[12]. ϵ 反映输入变量所含噪声对模型的影响程度, 影响着模型的拟合精度. 网格搜索法进行参数寻优的基本原理是使 $\log_2\epsilon$ 、 \log_2C 、 \log_2g 在一定范围内变化, 使用交叉验证来评估参数对模型性能的影响, 以此来选择出最佳的参数值. 一般 $\log_2\epsilon$ 、 \log_2C 、 \log_2g 的取值范围分别为 $[-10, 2]$ 、 $[-5, 15]$ 、 $[-15, 5]$. 但网格搜索方法的缺点是计算量大, 搜索时间长.

为减少参数寻优的计算量, 需要确定一个更为紧凑的参数搜索范围. 为此, 首先分析单个参数变化与模型均方根误差和支持向量个数之间的关系, 在此基础上确定参数精选搜索范围. 最后在参数精选范围内, 利用网格搜索法确定模型的参数. 具体步骤如下:

(1) 确定 $\log_2\epsilon$ 、 \log_2C 、 \log_2g 的粗选搜索范围, 为 $[-10, 2]$ 、 $[-5, 15]$ 、 $[-15, 5]$;

(2) 测试 ϵ 时, C 、 g 设为默认值. 使 $\log_2\epsilon$ 在 $[-10, 2]$ 范围内变化, 将训练样本用于交叉验证, 然后得到模型对应的均方误差和支持向量所占百分比;

(3) 画出 $\log_2\epsilon$ 与均方误差以及支持向量所占百分比之间的关系曲线;

(4) 根据曲线的形状, 得到 $\log_2\epsilon$ 的精选搜索范围;

(5) 采用同样的步骤对其它两个参数进行实验, 得到3个参数的精选搜索范围, 参数 ϵ 通常对于模型的训练影响较小, 一般取一个合理值即可;

(6) 在精选搜索范围内, 使用网格搜索确定参数 C 、 g 的最优设定值.

图2、图3分别为单变量参数 ϵ 变化对模型均方根误差、支持向量占比的影响, 这里参数 $g=3$, $C=2$ 设为默认值. 当 $\log_2\epsilon \leq -4$ 时, ϵ 对模型的均方误差和支持向量占比影响较小; 从 $\log_2\epsilon \geq -1$ 时模型误差开始增大, 而当 $\log_2\epsilon \geq 1$ 时, 支持向量个数小于总训练样本的20%, 此时出现了欠拟合, 模型过于简单, 均方误差有增大的趋势. 由此可以看出, $\log_2\epsilon$ 的最优值位于 $[-4, 1]$. 其中不敏感系数 ϵ 作为一个控制精度的参数, 通常情况对模型的训练学习影响较小, 一般取一个合理值即可, 这里取 $\epsilon=0.5$.

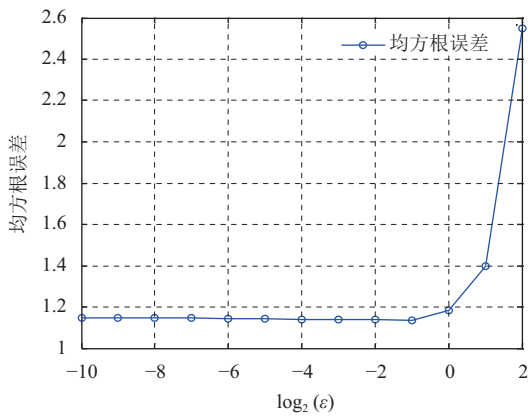


图2 参数 ϵ 对模型均方根误差的影响

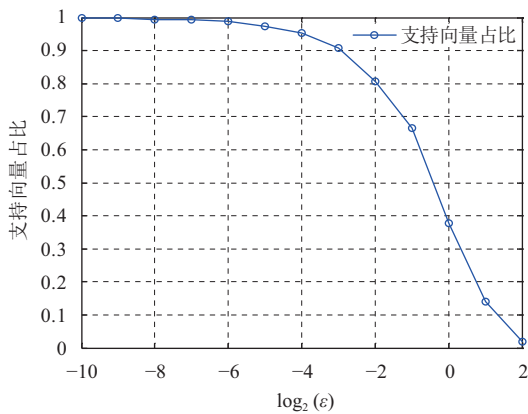


图3 参数 ϵ 对模型支持向量占比的影响

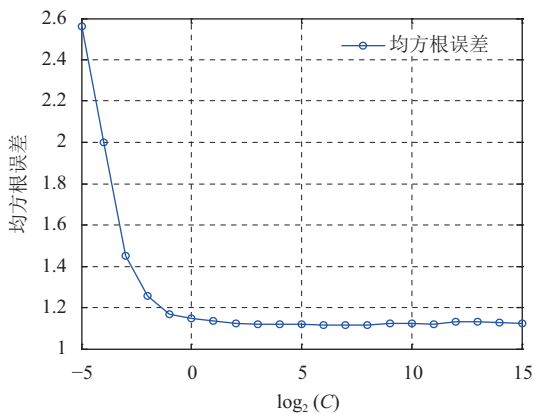


图4 参数 C 对模型均方根误差的影响

图4、图5分别为参数 C 对模型均方根误差、支持向量占比的影响, 另两个参数设为默认值 $\epsilon=0.5, g=3$. 开始模型的误差随着 $\log_2 C$ 的增大而减小, 支持向量占比也不断降低; 但在 $\log_2 C \geq 5$ 后, 模型在训练样本上的误差和支持向量个数都基本不变. 而 C 越大, 对模型误差的惩罚程度越大, 模型的泛化能力较低. 因此, 可以

确定最优 $\log_2 C$ 的值位于 $[-3, 5]$.

图6、图7分别为参数 g 对模型均方根误差、支持向量占比的影响, 另两个参数设为默认值 $\epsilon=0.5, C=2$. 首先随着 $\log_2 g$ 不断增大, 模型误差先减小后增大, 支持向量的个数也是先减少, 然后增加, 在 $\log_2 g = -1$ 附近模型的误差达到最小值. 由此可以确定参数 $\log_2 g$ 的搜索范围为 $[-4, 4]$.

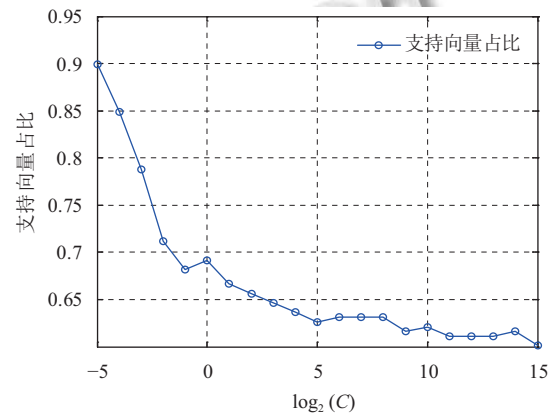


图5 参数 C 对模型支持向量占比的影响

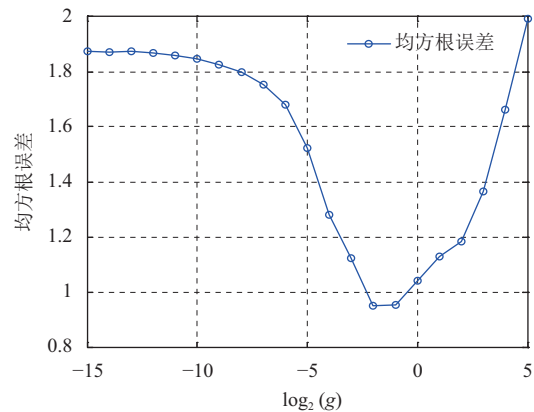


图6 参数 g 对模型均方根误差的影响

以上基于单变量的参数搜索方法得到了参数精选范围, $\log_2 C$ 在区间 $[-3, 5]$ 之间, $\log_2 g$ 在区间 $[-4, 4]$ 之间, 参数 $\epsilon=0.5$. 在精选的搜索范围内, 利用基于交叉验证的网格搜索方法确定参数最优值. 如图8、图9所示的是在精选范围内进行参数网格寻优的3D视图和等高线图, 为保证模型更高的泛化能力, 在最小的模型误差参数组中选择较小的惩罚因子 C , 由此得到最佳的参数组合 $C=4, g=0.5, \epsilon=0.5$.

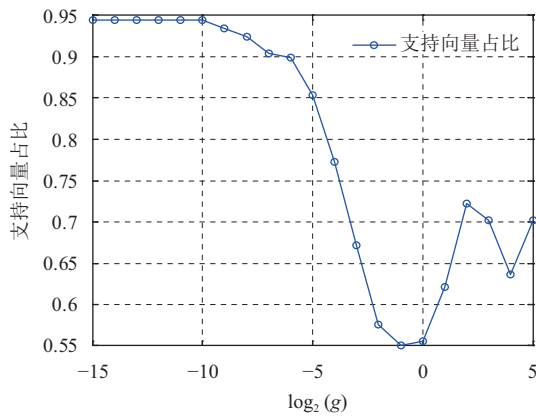


图7 参数g对模型支持向量占比的影响

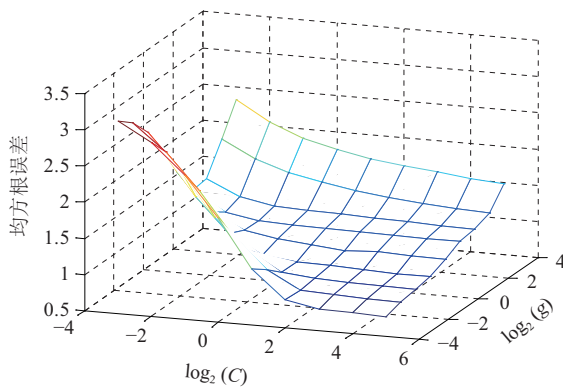


图8 网格搜索结果3D图

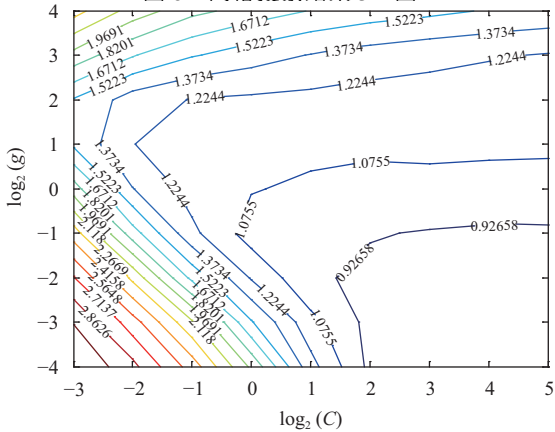


图9 网格搜索结果等高线图

3.5 仿真建模结果

由于循环流化床锅炉燃烧工况随着时间和条件的变化在不断改变,利用离线数据建立的固定模型不能真实地反映当前的实时工况,因此建立一个滑动更新的二氧化硫预测模型来验证二氧化硫浓度的预测效果.这里,选取2015.03.22这一天的某段稳定运行数据,采样时间为10s.建立一个随时间滑动的建模数据区间,

区间长度为100步并保持不变.模型更新时间 T 为30步,即每间隔30步,滑动一次建模数据区间,更新模型.设置参数 $C=4, g=0.5, \epsilon=0.5$.训练模型,得到如图10所示的二氧化硫浓度预测值与实际值对比的效果图.

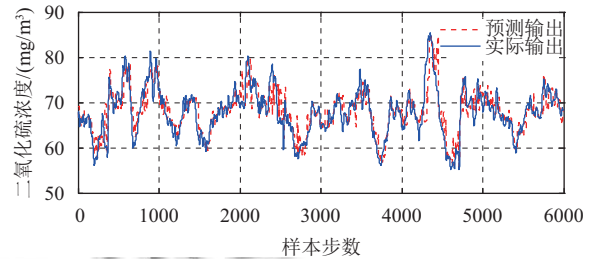


图10 二氧化硫浓度模型预测输出

由图10可见,基于SVM的二氧化硫模型预测输出曲线很好拟合实际值曲线,77.54%的样本预测输出相对误差小于4.39%,90.58%的样本预测输出相对误差小于6.88%.

表1为基于SVM的二氧化硫模型预测精度,可见预测精度较高.其相关系数为0.84,相关性很强,平均误差为1.96 mg/m³,均方根误差为2.85 mg/m³.

表1 二氧化硫浓度模型预测精度

相关系数	平均误差	平均相对误差	均方根误差
0.84	1.96 mg/m ³	2.87%	2.85 mg/m ³

4 总结

CFB锅炉运行工况复杂,烟气的二氧化硫对象具有很强的非线性,传统的建模方法很难获得较为精确的模型,为解决此问题,本文建立了一种基于支持向量机烟气二氧化硫排放量预测模型.同时,为了减少支持向量机回归模型参数寻优的计算量,本文采用单变量参数搜索结合网格寻优的方法来确定模型参数.通过将模型预测值与实际测量值对比,实验结果表明基于SVM二氧化硫浓度预测模型有很好的预测效果.

参考文献

- 1 郑海明, 杨志. 燃煤电厂二氧化硫排放质量浓度的软测量技术. 动力工程学报, 2013, 33(2): 130-134.
- 2 吴晓蔚. 火电厂执行大气污染物排放新标准的达标研究. 环境监测管理与技术, 2015, 27(6): 57-60.
- 3 刘林杰. 基于过程监控的烟气排放软测量预测研究[硕士学位论文]. 北京: 华北电力大学, 2015.
- 4 王松, 王东风, 董宇. 基于支持向量机的CFB锅炉烟气含

- 氧量建模和预测. 电力科学与工程, 2013, 29(6): 57-60, 72.
- 5 王定成, 方延健, 高理富, 等. 支持向量机回归在线建模及应用. 决策与控制, 2003, 18(1): 89-91, 95.
- 6 Kaneko H, Funatsu K. Application of online support vector regression for soft sensors. AICHE Journal, 2014, 60(2): 600-612. [doi: 10.1002/aic.14299]
- 7 李先知. 污水处理过程远程监控系统设计与支持向量机技术应用研究[硕士学位论文]. 合肥: 中国科学技术大学, 2016.
- 8 Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): Article No. 27.
- 9 Basak D, Pal S, Patranabis DC. Support vector regression. Neural Information Processing-Letters and Reviews, 2007, 11(10): 203-224.
- 10 荣海娜, 张葛祥, 金炜东. 系统辨识中支持向量机核函数及其参数的研究. 系统仿真学报, 2006, 18(11): 3204-3208, 3226. [doi: 10.3969/j.issn.1004-731X.2006.11.050]
- 11 管晓晨. 基于支持向量机的循环流化床锅炉燃烧系统建模与控制[硕士学位论文]. 南京: 东南大学, 2015.
- 12 薄翠梅, 张滢, 王执铨, 等. 基于滑动时间窗的支持向量机软测量建模研究. 自动化仪表, 2006, 27(1): 45-48.