

# 支持向量机在锅炉汽包水位预测中的应用研究

安 鹏, 赵林涛\*

(北京航天指挥控制中心 技术勤务站, 北京 100094)

**摘要:** 汽包水位预测的准确性对于锅炉火力发电和供暖设备的安全运行,起着重要的作用。针对汽包水位时滞、非线性特性,在统计学习理论(SLT)和结构风险最小化(ERM)准则基础上,建立了基于支持向量机(SVM)理论的汽包水位预测模型。采用本单位供暖锅炉的实际汽包水位数据,按照不同的汽包水位日属性和历史汽包水位数据进行了样本选择,并将该预测结果同实际汽包水位进行了比较;根据供暖规模不同,分别建立了工作日和双休日预测模型。实验结果表明,所提出的预测方法具有较高的精度。

**关键词:** 支持向量机; 汽包水位; 预测; 统计学习理论; 结构风险最小化

中图分类号: TH69; TP181

文献标识码: A

文章编号: 1001-4551(2010)07-0029-04

## One support vector machine method's application in prediction of drum-water level

AN Peng, ZHAO Lin-tao

(Unit 63933, PLA, Beijing 100094, China)

**Abstract:** Water-level measurement system for boiler is one of the key measuring systems in power plant. And it is important to measure and predict the drum water level. As drum water level of boiler is a time-delay, time-varying and nonlinearity system, a new mathematical model according to the support vector machines (SVM) theory was brought out, which was based on the statistical learning theory (SLT) and the empirical risk minimization (ERM) principle. Adopting the actual data from the heating boiler of author affiliation, the samples were chosen according to different attributes of daily water level and historical load data. The forecasted results were compared with actual boiler water level according to the scale of heating. The results show that the presented prediction method is more accurate.

**Key words:** support vector machine(SVM); drum boiler water level; prediction; statistical learning theory(SLT); empirical risk minimization(ERM)

## 0 引 言

工业锅炉中汽包满水和缺水事故是火力发电厂的重大恶性事故之一。维持汽包水位值在一定范围是保证锅炉和汽轮机安全运行的必要条件,也得到了越来越广泛的关注。

工业锅炉是一个多输入因素影响的非线性系统,如汽包压力、汽包温度、汽包内饱和蒸汽密度、凝结水密度等,并且这些变量相互之间又存在交叉影响。传统 PID 控制通过调整测量系统模型,如采用具有一定自我补偿能力的双室平衡容器的水位测量装置<sup>[1]</sup>,或

三冲量带前馈的串级控制调节测量系统<sup>[2]</sup>等,但这些测量装置会给系统带来很大的超调。随着智能控制研究的发展,人工神经网络、专家系统、模糊逻辑以及模糊神经等非线性系统模型<sup>[3]</sup>也不断应用到汽包水位测量与预测中,这些模型有更快的响应和更小的超调,对过程参数有一定的鲁棒性,但仍有一些难以克服的缺陷,如优化过程可能陷入局部极值,神经网络采用的经验风险最小化准则泛化能力不强,无法控制其收敛以及收敛的速度等<sup>[4]</sup>。

支持向量机(Support Vector Machines, SVM)是 Vapnik 等人根据统计学理论提出的机器学习领域的工具方法之一<sup>[5-6]</sup>,能较好地解决小样本、非线性、高维

数和局部极小点等实际问题,并被成功地应用于分类<sup>[7]</sup>、函数逼近和预测等方面<sup>[8]</sup>,而汽包水位的预测问题也可以看作是一种多影响因子的非线性函数关系的逼近问题。

基于上述考虑,本研究尝试将支持向量机理论引入汽包水位预测之中。

## 1 支持向量机回归基本原理

支持向量机是一种基于统计学习的理论方法,它是通过某种事先选择的非线性映射  $\Psi$  将  $n$  维输入样本向量  $\mathbf{x}$  从原空间  $(x_1, y_1), \dots, (x_l, y_l) \in R^n (y_i \in \{+1, -1\}, i=1, \dots, l)$  映射到一个高维特征空间  $Z$ , 在这个空间中构造最优分类超平面,且基于最优超平面的分类规则如下:

$$f(x, \alpha) = \text{sgn}(\mathbf{w} \cdot \Psi(x) + b) \quad (1)$$

其中,  $b$  是常数(阈值)。将结果推广到估计实函数(回归)中,在上式中引入  $\varepsilon$  不敏感损失函数,它不但使估计具有鲁棒性,而且是稀疏的。这样,在高维特征空间的线性回归便对应于低维输入空间的非线性回归,免去了在高维空间  $\mathbf{w}$  和  $\Psi(x)$  点积的计算。那么回归估计问题就定义为对  $\varepsilon$  不敏感损失函数进行风险最小化的问题:对于输入向量的训练数据,寻找  $\mathbf{w}_l$  和  $b_l$  使得经验风险最小化(Empirical Risk Minimization, ERM)<sup>[9-10]</sup>,且有限样本定义的经验风险为:

$$R_{emp} = \frac{1}{l} \sum_{i=1}^l |y_i - (\mathbf{w} \cdot \mathbf{x}) - b|_{\varepsilon} \quad (2)$$

最小化问题可以等价于下面的问题:寻找  $\mathbf{w}, b$  对,使得引入的松弛变量  $\xi_i, \xi_i^* (i=1, \dots, l)$ , 定义的最小化下述量:

$$F(\xi, \xi^*) = \sum_{i=1}^l \xi_i^* + \sum_{i=1}^l \xi_i \quad (3)$$

约束条件为:

$$\begin{cases} y_i - (\mathbf{w} \cdot x_i) - b \leq \varepsilon + \xi_i^*, i=1, \dots, l \\ (\mathbf{w} \cdot x_i) + b - y_i \leq \varepsilon + \xi_i, i=1, \dots, l \\ \xi_i^* \geq 0, i=1, \dots, l \\ \xi_i \geq 0, i=1, \dots, l \\ (\mathbf{w} \cdot \mathbf{w}) \leq c_n \end{cases} \quad (4)$$

这是一个凸二次优化问题,必须找到下面拉格朗日泛函的鞍点:

$$L(\mathbf{w}, \xi^*, \xi; \alpha^*, \alpha, C^*, \gamma, \gamma^*) = \sum_{i=1}^l (\xi_i^* + \xi_i) - \sum_{i=1}^l \alpha_i [y_i - (\mathbf{w} \cdot x_i) - b + \varepsilon + \xi_i] - \sum_{i=1}^l \alpha_i^* [(\mathbf{w} \cdot x_i) + b - y_i + \varepsilon + \xi_i^*] -$$

$$\frac{C^*}{2}(c_n - (\mathbf{w} \cdot \mathbf{w})) - \sum_{i=1}^l (\gamma_i^* \xi_i^* + \gamma_i \xi_i) \quad (5)$$

其中,常量  $C$  为容量控制,作为对超出误差  $\varepsilon$  样本的惩罚。此泛函的鞍点便是这个不等式约束下凸二次优化问题的解,且其存在全局唯一解。

通常训练数据集中的数据是线性不可分的,对于这类问题,根据泛函的有关理论,笔者引入核函数  $K(x_i, x_j)$ , 利用其使非线性问题变成线性分类<sup>[11]</sup>。则此时式(5)目标函数变为:

$$L = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (6)$$

其中常用的核函数有:①线性核: $K(x_i, x_j) = x_i \cdot x_j$ ;②多项式内核: $K(x_i, x_j) = [(x_i \cdot x_j) + 1]^q$ ;③径向基函数(RBF)内核: $K(x_i, x_j) = \exp\left\{-\frac{|x_i - x_j|^2}{\sigma^2}\right\}$ ;④Sigmoid 内核: $K(x_i, x_j) = \tan(v(x_i, x_j) + c)$  等等。

拉格朗日泛函和核函数的引入,  $f(x)$  最终可表示为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (7)$$

## 2 基于 SVM 的汽包水位预测

由于引起汽包水位变化的因素较多,如给水流量、蒸汽流量、燃烧工况等,以及锅炉汽包特有的虚假水位现象,因而常规的方法难以满足对汽包水位预测及控制的较高要求。

本研究在 SVM 基本原理的基础上进行汽包水位预测,首先要确定影响汽包水位的主要因素,其次选择由历史数据构成的样本数据集,然后采用 SVM 回归方法得到学习训练汽包水位数据,最后根据训练后获得的参数进行预测。

### 2.1 构造样本数据集

本研究所用的单位供暖锅炉是建立在双室平衡容器差压式锅炉模型上,模型如图 1 所示。

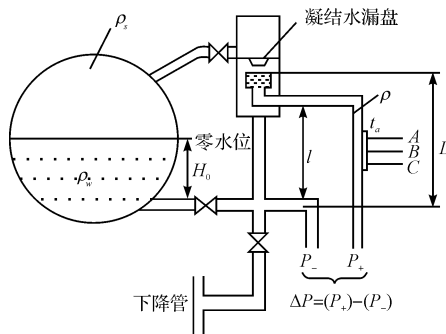


图 1 双室平衡容器的水位测量示意图

其中,图中  $\Delta P$  为差压值,是汽包水位测量的主要影响值。

根据汽包水位数据和环境因素可知,影响汽包水位测量的因子主要有汽包内压力  $P_b$ 、汽包内饱和水密度  $\rho_w$ 、汽包内饱和蒸汽密度  $\rho_s$ ,而这些因子也受汽包内温度  $t$  的影响。并且利用工业汽包水位变送器测量得到的汽包水位数据,在处理中还要考虑正负压管输出的差压值  $\Delta P$  对汽包水位的影响。因此本研究选用 5 输入 ( $P_b$ 、 $\rho_w$ 、 $\rho_s$ 、 $t$ 、 $\Delta P$ ) 和 1 输出汽包水位高度构成样本数据集。

## 2.2 训练算法及核函数构造

SVM 经典算法的实验可归纳为解决一个线性约束的二次规划问题,该算法复杂,收敛速度慢。1998 年,Platt 提出 SMO (Sequential Minmial Optimization) 作为 SVM 的训练算法。该算法将大型二次规划问题分解为一系列小规模二次规划子问题,避免了在内循环中使用数值方法进行最优化,可获得较好的性能。本研究即采用 SMO 方法作为 SVM 回归的训练算法。

核函数  $K(x_i, x_j)$  的选择对汽包水位预测的精度影响很大。本研究试算表明 RBF 核函数在锅炉汽包水位预测方面较其他核函数有更高的精度。此外, SVM 回归模型算法还主要由损失函数  $\varepsilon$  和容量控制  $C$  来确定。而目前还没有一个有效的设置参数的方法。在本研究实验中遵循以下两个原则<sup>[12-13]</sup>:

(1) 根据训练样本集选择参数  $C$ ;

(2) 根据训练集中已知的或者预测的误差和训练样本的数量选择  $\varepsilon$ 。

## 2.3 汽包水位预测的具体步骤

用 SVM 算法进行锅炉汽包水位预测,其具体的步骤如下:

(1) 对获取的历史数据进行平滑预处理;

(2) 建立预测样本,包括历史汽包水位数据、汽包内压力、汽包内饱和水密度、汽包内饱和蒸汽密度、汽包内温度、正负压管输出的差压值,然后建立系统模型;

(3) 利用 SMO 训练算法的 SVM 计算  $\alpha^*$ ,  $\alpha$ ;

(4) 将  $\alpha^*$ ,  $\alpha$  代入式(7),利用预测样本对未来某一时刻的锅炉汽包水位进行预测。

## 3 SVM 回归模型仿真与算法分析

本研究将本单位锅炉作为研究对象,收集其历史数据应用到 SVM 算法模型中。在每个工作日中,样本选取自锅炉开炉到正常运行随着汽包温度升高而测得

的 100 组数据,以及锅炉运行稳定后的 120 组数据。

核函数选定 RBF 核函数:  $K(x_i, x_j) = \exp \left\{ \frac{|x_i - x_j|^2}{\sigma^2} \right\}$ 。

通过试凑法对多组  $\varepsilon$ 、 $C$  参数进行实验,最终选定  $\varepsilon = 0.1$ ,  $C = \infty$ 。

利用训练好的模型预测该锅炉 2009 年 1 月 5 日至 1 月 9 日及 2009 年 1 月 12 日至 1 月 16 日共计 10 个工作日的 220 点汽包水位数据。将预测所得的结果与预测日的实际锅炉汽包水位做比较。定义预测日各点的预测符合数值为  $k_{pi}$ , 实际汽包水位为  $k_i$  ( $i = 1, 2, \dots, 220$ ), 则相对误差为  $e_i = \frac{|k_{pi} - k_i|}{k_i} \times 100\%$ 。本研究以 4% 为评判标准,若某点的  $e_i > 4\%$ , 则判定该点的预测结果不合格。定义  $r_i$  为该工作日汽包水位预测结果的不合格率,则:

$$r_i = \frac{\text{num}(e_i > 4\%)}{220} \times 100\% \quad (8)$$

式中  $\text{num}(e_i > 4\%)$ —某工作日预测结果不合格率超过 4% 的点数。

本研究用 Matlab 实现了 SVM 锅炉汽包水位预测程序,并建立了完备的历史数据库,较好地实现了所提出方法的通用性及实用性验证。预测结果如表 1 所示,其中 num 为预测结果不合格 ( $e_i > 4\%$ ) 点数,  $r_i$  为当日预测结果的不合格率。

表 1 一般工作日预测结果比较

日期	num	$r_i$ (%)
1-5	12	5.455
1-6	15	6.818
1-7	14	6.363
1-8	11	5.000
1-9	12	5.455
1-12	10	4.545
1-13	13	5.909
1-14	14	6.363
1-15	10	4.545
1-16	15	6.818

从表 1 所示的预测结果计算可知,用 SVM 方法预测的平均不合格率为 5.727%,因而 SVM 方法在锅炉汽包水位预测方面,其结果具有可实现的精度。

对于双休日,由于锅炉供暖规模变化,其相关数据与正常工作日差别很大。以 2009 年 1 月 17 日为例,预测当日 100 组数据,预测曲线及实际曲线的比较如图 2 所示,预测结果的相对误差如图 3 所示。

由预测结果可以计算出平均预测误差为 4.053%,误差在可容许范围内,并且预测结果精度高于一般工作日的预测精度。通过调节参数  $C$ ,可以使误差尽可

能小,也可以使回归函数尽可能平滑,从而提高了泛化(预测)能力。实验中预测时间共耗 4.368 7 s,远远快于同类预测模型。

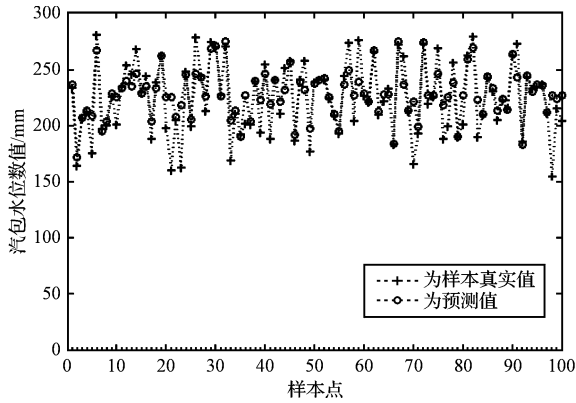


图2 锅炉汽包水位预测结果

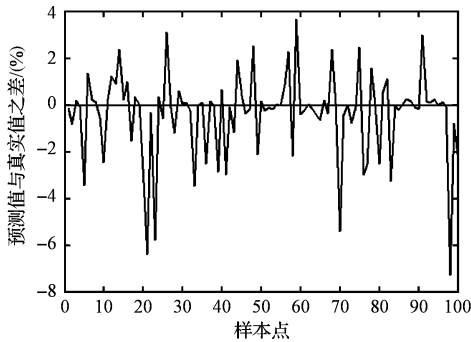


图3 测试数据与预测结果的误差

从上述仿真结果可见,在预测精度方面,利用 SVM 预测锅炉汽包水位的方法,充分考虑了影响汽包水位的主要因素,如汽包内压力、汽包内饱和水密度、汽包内饱和蒸汽密度、汽包内温度、正负压管输出的差压值等,对历史数据进行聚类分析,找出与预测点同一类的数据进行训练,通过算法内部高维投影避免了计算复杂多因素控制的过程。无论是对于一般工作日还是节假日都有较高的预测精度。

## 4 结束语

支持向量机回归建模将低维非线性的输入映射到高维线性的输出,模型简单,具有良好的应用前景,对于小样本系统具有广泛的适用性。对于锅炉汽包水位变化影响因素复杂的情况,这类模型能得到比较好的模拟和预测结果。通过与神经网络等学习方法对比可以看出,SVM 方法能充分利用训练样本的分布特性,根据部分训练样本构建判别函数,不需要过多的先验信息和使用技巧,最终能得到一个全局最优解;并且 SVM 算法最终转化为二次寻优问题,从理论上可以证

明,能得到唯一的全局最优解,避免了局部极值的问题;同时通过引入非线性变换和核函数,巧妙地解决了高维数的问题,使得算法复杂度与样本维数无关,加快了训练学习速度。本研究只是用 SVM 回归方法对有限样本进行了汽包水位的训练和预测,而对于大规模的预测问题仍有待进一步细致地研究,适宜核函数的选用也是需要进一步研究的方向之一。

## 参考文献 (References):

- [1] 吴业飞,时敏. 双室平衡容器汽包水位测量及其补偿系统的引用[J]. 自动化仪表,2004,25(7):33-37.
- [2] 孙伟,谭德健,许世范. 三冲量汽包水位自适应神经元控制算法[J]. 中国矿业大学学报,1996,25(3):43-47.
- [3] 彭道刚,杨平,杨艳华,等. 具有对负荷变化前馈补偿的锅炉汽包水位系统的神经网络内模控制[J]. 中国电机工程学报,2005,25(5):141-145.
- [4] SUN Zhong-hai, SUN You-xian. Fuzzy support vector machine for regression estimation[C]//IEEE International Conference on Systems Man and Cybernetics,2003:3336-3341.
- [5] VAPNIK V. The Nature of Statistical Learning Theory[M]. 2nd ed. Berlin: Springer,1999.
- [6] 张海军,王学军. 基于 RBF 核函支持向量机分类器的多导脑电信号分类识别研究[J]. 机电工程技术,2008,37(8):73-75.
- [7] VAPNIK V, GOLOWICH S E, SMOLA A J. A Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing[M]. Neural Information Processing Systems. MIT Press,1997.
- [8] 王亮,张宏伟,牛志广. 支持向量机在城市用水量短期预测中的应用[J]. 天津大学学报,2005,38(11):1021-1024.
- [9] FRANCIS E H, CAO L J. Application of support vector machines in financial time series forecasting[J]. Omega, 2001,29(4):309-317.
- [10] 张浩然,汪晓东,张长江. 一种鲁棒回归支持向量机及其学习算法[J]. 南京理工大学学报,2006,30(3):311-314.
- [11] MERCIER G, LENNON M. Support vector machines for hyperspectral image classification with spectral-based kernels[C]//Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International, 2003:288-290.
- [12] THORSTEN J. Making Large-scale SVM Learning Practical, Advanced in Kernel Methods-support Vector Learning[M]. MIT Press,1999.
- [13] CHERKASSKY V, MA Yun-qian. Practical selection of SVM parameters and noise estimation for SVM regression[J]. Neural Networks,2004,17(1):113-126.