

# 基于数据驱动算法的再热蒸汽温度动态预测

孙 悅<sup>1</sup>, 姜文娟<sup>2</sup>, 曹玉波<sup>1</sup>

(1. 吉林化工学院 信息与控制工程学院, 吉林 吉林 132022; 2. 东北电力大学 自动化工程学院, 吉林 吉林 132012)

**摘要:** 锅炉再热蒸汽温度具有强非线性和大滞后特性, 为解决其软测量中常规单一模型预测精度不足的问题, 提出一种基于自适应提升算法(Adaptive Boosting, Adaboost)和极限梯度提升(eXtreme Gradient Boosting, XGBoost)的数据驱动建模方法。利用变分模态分解对数据进行深度解析, 通过XGBoost建立预测模型, 将其作为弱学习器, 经过Adaboost算法的不断迭代, 配合误差动态修正(Error Dynamic Correction, EDC)构造出一种再热蒸汽温度动态数据驱动模型。结果表明:模型最终的精度评价指标均方根误差(Root Mean Square Error, RMSE)和平均绝对误差(Mean Absolute Error, MAE)分别为1.733和1.387, 与常规的支持向量回归、随机森林及XGBoost模型相比表现更为优异, 可以实现再热蒸汽温度的快速准确预测, 为后续再热汽温优化控制问题提供有效的参考。

**关键词:** 再热蒸汽温度; 特征选择; 变分模态分解; Adaboost; 误差修正

**中图分类号:** TP181      **文献标识码:** A      **DOI:** 10.16146/j.cnki.rndlgc.2023.06.013

[引用本文格式] 孙 悅, 姜文娟, 曹玉波. 基于数据驱动算法的再热蒸汽温度动态预测[J]. 热能动力工程, 2023, 38(6): 116-121.  
SUN Yue, JIANG Wen-juan, CAO Yu-bo. Dynamic prediction for reheat steam temperature based on data-driven algorithms[J]. Journal of Engineering for Thermal Energy and Power, 2023, 38(6): 116-121.

## Dynamic Prediction for Reheat Steam Temperature based on Data-driven Algorithms

SUN Yue<sup>1</sup>, JIANG Wen-juan<sup>2</sup>, CAO Yu-bo<sup>1</sup>

(1. College of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, China, Post Code: 132022;  
2. School of Automation Engineering, Northeast Electric Power University, Jilin, China, Post Code: 132012)

**Abstract:** The boiler reheat steam temperature has high nonlinearity and great hysteresis characteristics. In order to improve the prediction precision of conventional single model in soft measurement, a data-driven modeling method based on adaptive boosting (Adaboost) and eXtreme Gradient Boosting (XGBoost) was proposed. The variational mode decomposition (VMD) was utilized to analyze the relevant data deeply, and then, XGBoost was used to build a prediction model as weak learner, based on which a dynamic data-driven model of reheat steam temperature was constructed with error dynamic correction (EDC) after continuous iteration of Adaboost algorithm. The results show that the final root mean square error (RMSE) and mean absolute error (MAE) values as the precision indicators of the model are 1.733 and 1.387 respectively, compared with conventional support vector regression (SVR), random forests (RF) and XGBoost models, indicating an improved performance. Thus, the proposed model can realize a rapid and precise prediction of reheat steam temperature, and provide an effective reference for the subsequent control and optimization of reheat steam temperature.

**Key words:** reheat steam temperature, feature selection, variational mode decomposition (VMD), adaptive boosting (Adaboost), error correction

## 引言

电厂锅炉在生产过程中,再热器将汽轮机排出的低温蒸汽再次加热到和过热蒸汽近似温度以实现循环利用。因此,再热器出口蒸汽温度是影响机组工作效率和安全性的重要过程参数<sup>[1]</sup>,异常的蒸汽温度会降低发电效率及损坏机组部件<sup>[2]</sup>,再热蒸汽温度准确预测可对异常状况提前预警,为后续参数优化控制提供基础,具有理论和现实意义。因此,再热蒸汽温度预测方法的研究是当前锅炉燃烧的重要研究方向。

根据建模原理的不同,系统模型主要分为燃烧机理模型和数据模型。张经纬等人<sup>[3]</sup>通过改进粒子群算法对模型进行辨识,建立再热蒸汽的反应过程机理模型。刘森等人<sup>[4]</sup>使用子空间辨识算法得到再热蒸汽系统的多变量模型,为后续优化策略提供重要参考。不同于机理法建模,数据驱动模型是根据历史数据拟合再热蒸汽温度和影响变量间的非线性关系,更加方便易用。唐志炳等人<sup>[5]</sup>使用 RBF 神经网络(Radial Basis Function Neural Network, RBFNN)对再热器出口汽温进行预测建模。冯浩<sup>[6]</sup>将小波神经网络和免疫粒子群法相结合建立二次再热机组再热蒸汽模型。这些方法的应用表明了数据驱动建模的可行性。然而,由于缺乏对数据的深度解析,且并未充分考虑实际建模中参数与预测目标之间的时延性,导致常规单一模型的预测精度并不能完全适应复杂多变的现场环境。因此,在建模前有必要对数据的时延性及深度信息进行分析。

本文通过最大信息系数法(Maximal Information Coefficient, MIC)计算参数历史时间序列和再热蒸汽温度之间的非线性相关程度,推算参数延迟时间。同时,使用变分模态分解<sup>[7]</sup>(Variational Mode Decomposition, VMD)处理建模数据,通过迭代实现频域划分和子分量分离,深度挖掘数据深层信息<sup>[8]</sup>。另外,设计误差动态修正方案,分析历史预测值和真实值之间的误差规律,对初始预测值进行补偿,实现对再热器出口再热蒸汽温度的高性能预测。

## 1 再热蒸汽产生过程

火电机组主要由锅炉、汽轮机和发电机组组成,其

中锅炉系统由热交换设备和循环系统组成。煤粉充分燃烧后,将回水加热产生高压热蒸汽并送至汽轮机,带动机组产生电能,其主要流程如图 1 所示。

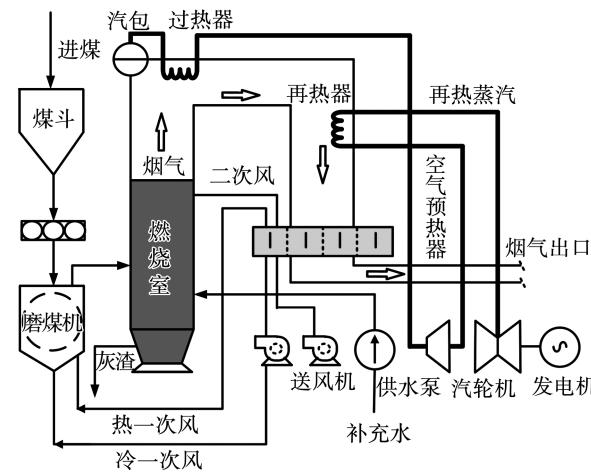


图 1 电厂生产流程图

Fig. 1 Production process chart of power plant

再热蒸汽温度主要是指低温蒸汽经过再热器一次再热后的出口蒸汽温度。蒸汽的再热循环利用,可以有效提升锅炉整体燃烧效率。通过再热蒸汽生产过程结合专家经验,可以分析得出再热蒸汽温度的主要影响因素为锅炉负荷、给煤量、再热蒸汽压力、过热器烟气挡板开度、再热器烟气挡板开度等。从工厂信息管理系统中获取相关数据,作为后续数据分析及建模的基础。

## 2 再热汽温动态预测模型

### 2.1 数据预处理及时延分析

本文所使用的数据来自国内某电厂 300 MW 亚临界一次中间再热、直流式燃烧器四角切圆方式燃煤机组,分析初选了表 1 所示的 19 个影响变量作为初始输入特征变量,以 5 s 为采样间隔从数据管理系统中导出 6 620 组数据。

为提升建模效率,先对所有数据进行归一化处理,利用式(1)将数据全部映射到[0,1]之间。

$$Y'_{i} = \frac{Y_i - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (1)$$

式中: $Y_i$ — $i$ 时刻样本值; $Y'_{i}$ —规范化后的样本值; $Y_{\max}$ 和 $Y_{\min}$ —样本最大值和最小值。

后续为进行特征选择和重构建模数据集,提出采用 TD-MIC 时延算法计算各相关变量的时间延迟,再重构建模数据集,消除时延特性对模型精度的影响。

最大信息系数法是一种相关性分析方法,可衡量变量之间的非线性关系,其计算公式为:

$$\text{MIC}(x, y) = \max_{xy < B(n)} \frac{I(x, y)}{\log_2(\min(x, y))} \quad (2)$$

$$I(x, y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (3)$$

式中: $I(x, y)$ —变量  $x$  和  $y$  之间互信息值; $P(x)$  和  $P(y)$ — $x$  和  $y$  的概率密度函数; $P(x, y)$ —联合概率密度函数; $B(n)$ —样本总量  $n$  的平方根。

令数据集的输入变量为  $X = [x_1(t), x_2(t), \dots, x_m(t)]$ ,输出变量为  $Y = y(t)$ ,输入变量的最大延迟为  $k$ ,时间延迟计算如图 2 所示。

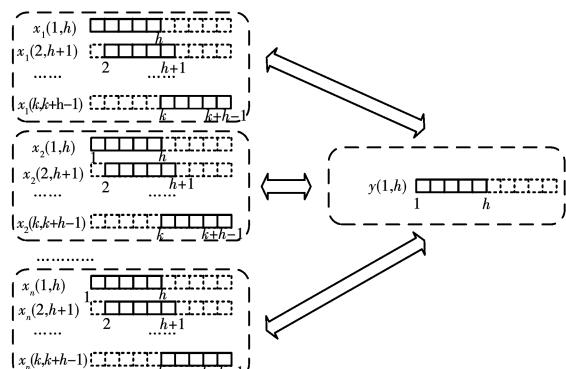


图 2 时间延迟计算示意图

Fig. 2 Schematic diagram for calculating  
the time delay

通过 TD-MIC 算法对  $m$  个特征依次求取最佳延迟时间,最后实现数据重构。根据现场专家经验,锅炉内部燃烧过程持续时间大约为 5~10 min,由于数据采样间隔为 5 s,因此  $k$  设置为 60,最终时间延迟计算结果如表 1 所示。

## 2.2 特征选择及数据分解

针对参数冗余的问题,在 Spearman 分析的基础上结合随机森林(Random Forests, RF)<sup>[9]</sup>和 XGBoost 实现加权特征选择。

表 1 主要变量信息及延迟时间

Tab. 1 Key variables information and delay time

序号	变量名称	范围
1	负荷/MW	[200.559, 277.215]
2	给煤量/t·h <sup>-1</sup>	[6.237, 13.369]
3	主蒸汽流量/t·h <sup>-1</sup>	[707.269, 1 022.609]
4	主蒸汽压力/MPa	[15.178, 17.473]
5	主蒸汽温度/℃	[521.061, 545.289]
6	再热蒸汽压力/MPa	[0.020, 6.607]
7	汽包压力/MPa	[16.779, 18.636]
8	给水压力/MPa	[17.401, 19.261]
9	给水温度/℃	[254.991, 276.225]
10	给水流量/t·h <sup>-1</sup>	[28 015.511, 118 534.406]
11	炉膛负压/MPa	[-265.083, 11.511]
12	再热器烟气挡板开度/%	[67.017, 99.854]
13	再热器入口温度/℃	[282.995, 321.942]
14	再热器喷水流量(左)/t·h <sup>-1</sup>	[0, 21]
15	再热器喷水流量(右)/t·h <sup>-1</sup>	[0, 18]
16	过热器烟气挡板开度(左)/%	[52.219, 100.758]
17	过热器烟气挡板开度(右)/%	[51.824, 100.599]
18	再热器一级减温水流量/t·h <sup>-1</sup>	[0, 23.073]
19	再热器二级减温水流量/t·h <sup>-1</sup>	[0, 23.064]

分别用 3 种方法对各特征进行重要性计算,再按照 Spearman: RF: XGBoost = 2:4:4 的权重比进行加权融合,确定各变量最终的重要性结果。

图 3 为加权融合后的重要性排序结果。将重要程度大于 0.1 的变量作为模型的输入变量,同时考虑到再热器减温水流量是影响出口蒸汽温度的重要因素,保留减温水作为输入变量。即最终选取变量 12、变量 8、变量 5、变量 9、变量 15、变量 10、变量 18 和变量 19 作为输入变量。

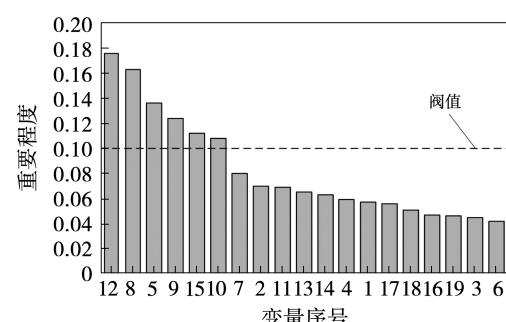


图 3 加权选择算法的特征重要程度排序

Fig. 3 Sorting of importance by a weighted  
feature selection algorithm

特征选择后,采用 VMD<sup>[7]</sup>来进一步提炼数据的深度时频信息。首先给定初始待分解信号并设置模态个数  $n$ ;随后 VMD 分解得到  $n$  个模态分量  $IMF_1, IMF_2, \dots, IMF_n$  (Intrinsic Mode Functions, IMF)。计算模态分量  $IMF_n$  和初始信号的相关系数  $r$ ,若  $r$  大于设定阈值,则继续分解,直到  $r$  小于设定阈值,表明当前的模态分量已不能表征初始信号,予以剔除,同时停止迭代。

## 2.3 再热汽温预测模型

### 2.3.1 XGB-Adaboost 模型基本原理

XGB-Adaboost 模型基本思路是基于 Adaboost 算法<sup>[10]</sup>的循环迭代,将 XGBoost 模型<sup>[11]</sup>作为弱学习器,再将这些弱学习器加权组合得到强学习器。同时,建模过程中使用网格搜索结合 5 折交叉验证对强弱学习器参数进行调优。

### 2.3.2 误差动态修正

为更进一步提升模型预测精度,设计了误差动态修正 (Error Dynamic Correction, EDC) 方案,使模型能够适应复杂的变工况环境。以模型预测误差作为目标对象,建立误差动态修正模型来预测  $t$  时刻的模型误差  $e(t)$ ,再将预测误差  $e(t)$  和  $t$  时刻的预测值  $y(t)$  相叠加,得到最终的再热汽温预测值  $Y(t)$ 。

选取 XGB-Adaboost 模型作为误差动态修正模型,以前 4 个时刻的模型预测误差  $e(t-1), e(t-2), e(t-3), e(t-4)$  及原有输入共同作为误差动态修正模型的输入,以  $t$  时刻的误差  $e(t)$  作为预测目标。

### 2.3.3 数据驱动混合预测模型

将 VMD-XGB-Adaboost 模型与误差动态修正模型 (EDC) 相结合,组成数据驱动混合预测模型 (Data Driven Hyrid Model, DDHM),用于再热器出口再热汽温的预测,如图 4 所示。并基于 Pycharm 集成开发环境 Python 语言开发模型实现程序用于再热器出口再热汽温的预测,分析时延特性和 VMD 分解对于模型的影响,将该模型和多种单一模型预测效果进行对比,评估不同模型预测结果差异。具体研究步骤为:

步骤 1:从分布式信息管理系统 (Distributed Control System, DCS) 中获取现场运行数据并作预处理。

步骤 2:计算各特征变量时延参数,重构数据集。

步骤 3:确定模型的最佳输入特征变量。

步骤 4:对变量进行 VMD 分解,与步骤 3 选取出来的特征变量共同组成模型输入特征  $x(t)$ 。

步骤 5:使用模型预测  $t$  时刻的再热汽温  $y(t)$ 。

步骤 6:以  $e(t-1), e(t-2), e(t-3), e(t-4)$  及  $x(t)$  作为误差动态修正模型的输入,得到补偿误差  $e(t)$ 。

步骤 7:计算最终混合预测模型输出  $Y(t) = y(t) + e(t)$ 。

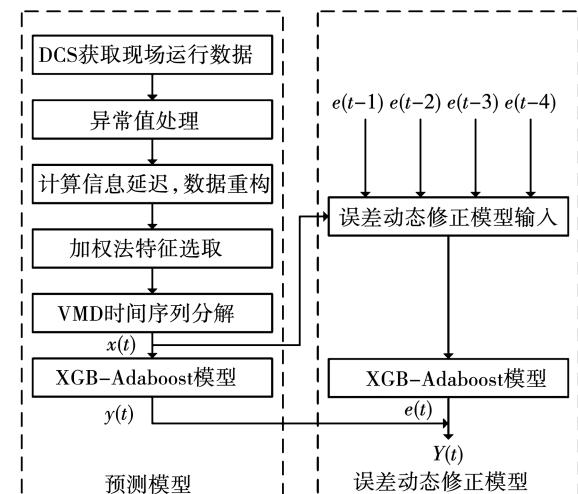


图 4 再热汽温 DDHM 模型框架图

Fig. 4 Frame diagram of reheat steam temperature DDHM model

## 3 模型预测结果与分析

对处理之后的数据集进行划分,选取前 5 620 组数据作为训练集,后 1 000 组数据作为测试集。训练并评估了支持向量回归 (Support Vector Regression, SVR)、随机森林、XGBoost 模型,与 DDHM 模型进行对比,分析不同模型之间的差异。

采用均方根误差 (Root Mean Square Error, RMSE)、平均绝对误差 (Mean Absolute Error, MAE) 和  $R$  方 ( $R$ -Square,  $R^2$ ) 作为预测结果的评价指标。

### 3.1 参数时延特性对预测结果的影响

为验证时延特性分析的有效性,对时延特性分析前后两组不同数据集进行建模分析。为方便区分,令考虑时延的数据集为  $D_y$ ,不考虑时延的数据集为  $D_n$ 。输入特征变量均为加权特征选择算法选出的 8 个变量,且两组数据集下模型参数设置不变。

图 5 为模型预测值和真实测量值的散点图。从图 5 中可以明显看出,时延特性分析前后 DDHM 模型预测效果得到有效改善。

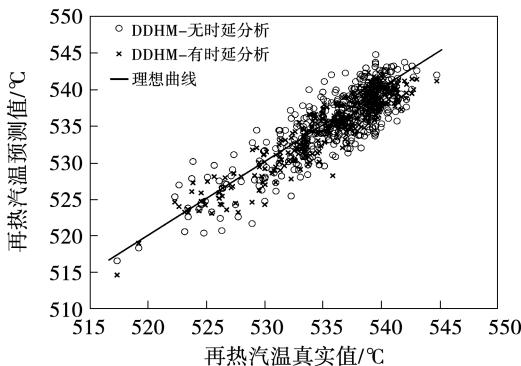


图 5 再热汽温测试集预测结果

Fig. 5 Prediction results of reheat steam temperature in test dataset

为详细分析延迟时间参数对模型精度的影响,分别计算两组数据集下的误差评价指标,结果如表 2 所示。以 RMSE 为例,DDHM 模型降低 7.77%,RF 模型降低 7.86%,SVR 模型降低 1.31%,XGBoost 模型降低 8.23%。

表 2 不同模型的再热温度预测误差对比

Tab. 2 Comparisons of reheat steam temperature predictions errors by differernt models

评价指标	DDHM		SVR		RF		XGBoost	
	$D_y$	$D_n$	$D_y$	$D_n$	$D_y$	$D_n$	$D_y$	$D_n$
RMSE	2.019	2.189	2.219	2.252	2.087	2.265	2.085	2.272
MAE	1.618	1.765	1.703	1.790	1.661	1.810	1.661	1.817
$R^2$	0.801	0.765	0.778	0.752	0.787	0.749	0.788	0.757

### 3.2 参数 VMD 分解对预测结果的影响

为验证 VMD 分解对数据建模的影响,将分解前后的数据组成两组数据集,模型参数做相同设置,结果如图 6 所示。可以看出,各模型在经过 VMD 分解的数据集中均有更好的预测效果。主要原因可能是VMD 分解消除了时间序列的噪声,提取了深度时频信息,进一步增强了模型的非线性拟合能力。

### 3.3 DDHM 模型预测结果及对比分析

表 3 为误差修正前后 DDHM 模型的评价指标。

可以看出,在经过误差修正后预测精度有了明显提升。为进一步验证所搭建混合模型的性能,使用经过了 TD-MIC 时延分析和特征选择后的特征变量作为输入,对比分析 RF,SVR,XGBoost,DDHM 等模型预测结果。图 7 为测试集局部的 40 组数据拟合曲线。图 8 为各模型在测试集下的绝对误差直方图。对比其余模型可以明显看出,DDHM 模型对变化曲线的拟合程度最高,泛化性能最强,由此可得,混合模型可以精准地实现对再热蒸汽温度的实时预测。

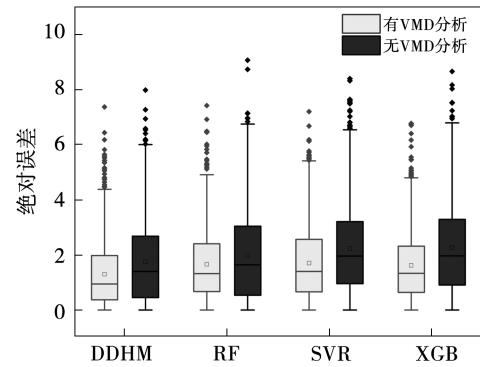


图 6 VMD 分解前后误差箱型图

Fig. 6 Boxplot of prediction absolute errors of model before and after VMD

表 3 误差修正前后评价指标对比

Tab. 3 Comparison of evaluation indicators before and after error correction

评价指标	VMD-XGB-Adaboost	DDHM
RMSE	1.981	1.733
MAE	1.582	1.387
$R^2$	0.803	0.829

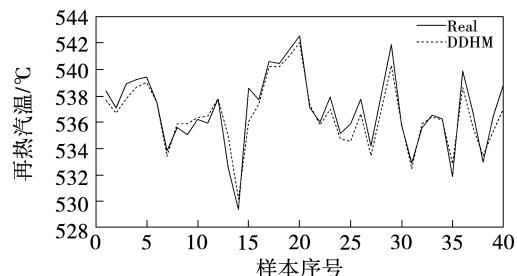


图 7 再热汽温测试集预测结果

Fig. 7 Prediction results of reheat steam temperature in test dataset

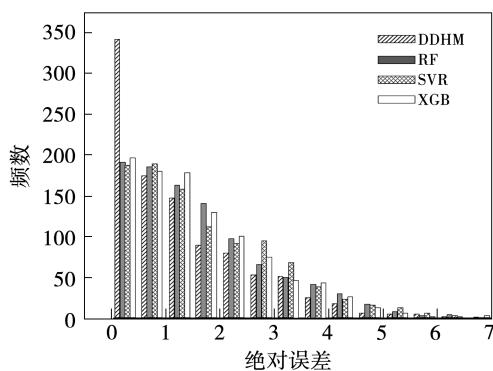


图 8 测量值与预测值绝对误差

Fig. 8 Absolute error between measured and predicted values

## 4 结 论

针对再热系统蒸汽温度测量中存在的时间延迟、时频信息不显著及精度差等问题,提出了一种再热蒸汽温度动态数据驱动模型。基于历史数据进行试验,得出以下结论:

(1) 基于最大互信息法分析参数时延特性,重构数据,以 DDHM 模型为例,重构数据后的评价指标 RMSE, MSE 及  $R^2$  分别为 2.019, 1.618 及 0.801, 与重构前相比, RMSE 和 MSE 分别降低 7.8% 和 8.3%,  $R^2$  提升 4.7%。消除了时延特性对建模精度的影响,有效提升模型预测精度。

(2) 采用 VMD 分解提取数据深度时频信息,去除噪声。数据分解后,模型精度获得了提升。

(3) 设计误差动态补偿方案,补偿后的模型评价指标 RMSE, MSE 及  $R^2$  分别为 1.733, 1.387 及 0.829, 与补偿前相比, RMSE 和 MSE 分别降低 12.5% 和 12.3%,  $R^2$  提升 3.2%, 进一步提升了模型性能。

## 参考文献:

- [1] 刘红波,李少远,柴天佑.一种基于模糊切换的模糊复合控制器及其应用[J].控制与决策,2003,18(5):615–618.  
LIU Hong-bo, LI Shao-yuan, CHAI Tian-you. Fuzzy hybrid controller based on fuzzy switching and its application [J]. Control and Decision, 2003, 18(5): 615–618.
- [2] ZHANG Yong, DANG Jin-geng. Application of PSO-BP neural net-

work in main steam temperature control [C]// 2016 Chinese Control and Decision Conference (CCDC): IEEE, 2016.

- [3] 张经纬,归一数,康英伟,等.基于改进粒子群算法的锅炉再热蒸汽温度模型辨识[J].热力发电,2017,46(7):72–78.  
ZHANG Jing-wei, GUI Yi-shu, KANG Ying-wei, et al. Identification of boiler reheat steam temperature model based on improved particle swarm optimization algorithm [J]. Thermal Power Generation, 2017, 46(7): 72–78.
- [4] 刘森,韩璞,张婷.超超临界二次再热机组再热汽温系统模型辨识[J].计算机仿真,2018,35(5):91–99.  
LIU Miao, HAN Pu, ZHANG Ting. Reheat steam temperature system model identification for ultra-supercritical double reheat units [J]. Computer Simulation, 2018, 35(5): 91–99.
- [5] 唐志炳,王明春,陶成飞,等.人工智能技术在再热汽温建模中的应用[J].发电设备,2015,29(4):252–255.  
TANG Zhi-bing, WANG Ming-chun, TAO Cheng-fei, et al. Application of artificial intelligence technology in modeling of reheat steam temperature [J]. Power Equipment, 2015, 29 (4): 252–255.
- [6] 冯浩.基于小波神经网络的二次再热汽温建模[J].仪器仪表用户,2019,26(12):83–86.  
FENG Hao. Modeling of secondary reheat steam temperature based on wavelet neural network [J]. Instrumentation, 2019, 26 (12) : 83 – 86.
- [7] DRAGOMIRETSKIY K, ZOSO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3):531–544.
- [8] 武小梅,林翔,谢旭泉,等.基于 VMD-PE 和优化相关向量机的短期风电功率预测[J].太阳能学报,2018,39(11):3277–3285.  
WU Xiao-mei, LIN Xiang, XIE Xu-quan, et al. Short-term wind power forecasting based on variational mode decomposition permutation entropy and optimized relevance vector machine [J]. Acta Energiae Solaris Sinica, 2018, 39(11): 3277 – 3285.
- [9] LEO B. Random forests [J]. Machine Learning, 2001, 45 (1): 5–32.
- [10] TAKERKHANI A, COSMA G, McGINNITY T M. AdaBoost-CNN: an adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning [J]. Neurocomputing, 2020, 404:351–366.
- [11] CHEN T, GUESTRIN C. XGBoost:a scalable tree boosting system [C]//The 22nd ACM SIGKDD International Conference: ACM, 2016.

(姜雪梅 编辑)