

基于自适应遗忘因子极限学习机的高炉煤气预测

孙雪莹^{1,2}, 胡静涛^{1,2}, 王卓^{1,2}, 张吉龙¹

(1. 中科院沈阳自动化研究所, 沈阳 110016; 2. 中国科学院大学, 北京 100049)

摘要: 高炉煤气是钢铁企业重要的二次能源, 其产生量和消耗量的实时准确预测对高炉煤气系统的平衡调度具有重要作用; 但由于高炉煤气系统工况多变、产销量数据波动较大, 给高炉煤气产销量的准确预测带来了很大的挑战; 为此, 通过对煤气产销量数据特征的深入分析, 提出了一种基于自适应遗忘因子极限学习机 (AF-ELM) 的在线预测算法; 在序贯极限学习机的基础上, 引入遗忘因子逐步遗忘旧样本, 通过预测误差反馈机制, 自适应的调节遗忘因子, 从而提高预测方法对系统工况的动态变化的适应能力, 提高预测精度; 将该算法应用于钢铁企业的高炉煤气产销量在线预测, 实验结果表明与序贯极限学习机相比, 该预测方法在系统工况变化的情况下能保持较高的预测精度, 更适合于高炉煤气产销量的在线预测。

关键词: 高炉煤气; 在线预测; 极限学习机; 遗忘因子

Online Prediction Method for Generation and Consumption of Blast Furnace Gas Based on Adaptive Forgetting Factor Extreme Learning Machine

Sun Xueying^{1,2}, Hu Jingtao^{1,2}, Wang Zhuo^{1,2}, Zhang Jilong¹

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Blast furnace gas is an important byproduct in iron and steel plants, and prediction of its generation and consumption has a great effect on balance and scheduling of gas system. However, the accurate prediction poses a significant challenge because of the unstable conditions of the blast furnace gas system and the fluctuation of data. To solve this problem, an online prediction method based on adaptive forgetting factor extreme learning machine (AF-ELM) is proposed. Dynamic adaptability of online sequential extreme learning machine is improved by introducing forgetting factor to gradually forget of the old samples. And the forgetting factor is adaptively updated by prediction error, which improves the prediction accuracy. The case study on the online prediction in iron and steel plants shows that compared with online sequential extreme learning machine, the proposed method achieve higher prediction accuracy in changing conditions, and more suitable for online prediction of generation and consumption of blast furnace gas.

Keywords: blast furnace gas; online prediction; extreme learning machine (ELM); forgetting factor

0 引言

钢铁企业中, 副产煤气是一种重要的二次能源, 主要包括焦炉煤气、高炉煤气和转炉煤气, 其中高炉煤气 (Blast furnace gas, BFG) 产量最大, 其有效的回收与利用对钢铁企业的节能减排具有重要的意义。高炉煤气产生量和消耗量预测的准确度直接影响高炉煤气调度的决策方案, 从而影响煤气系统的平衡和稳定。实际生产中, 高炉煤气的产销量受到生产设备工况的影响, 当工况发生改变时, 会发生大幅度的波动, 表现出很强的时变性, 给高炉煤气产销量的准确预测带来很大的

挑战。

目前国内外对钢铁企业副产煤气产销量预测方法的研究取得了一些研究成果^[1-9]。张琦等^[1]通过分析影响煤气发生量的主要因素, 采用关联度较大的因素作为输入, 并建立了神经网络预测模型。肖冬峰等^[2]提出了一种改进 BP 网络高炉煤气发生量预测模型, 采用机理与现场实际数据结合的方法确定煤气发生量的影响因素。由于影响煤气产销量的因素众多, 部分数据难以实时获取, 因此以上方法在实际应用中受到一定的限制。近年来, 基于数据的时间序列建模方法受到很多研究人员的关注。刘颖等^[3]采用改进的回声状态网络建立了煤气产生量预测模型。岳有军等^[4]将小波分析与 ARIMA 和 LSSVM 结合, 建立高炉煤气组合预测模型。张颜颜等^[5]提出了一种基于数据驱动的子空间方法, 用于预测钢铁企业各生产工序的能耗, 并采用粒子群算法对模型参数进行了优化。以上方法在平稳工况时预测精度高, 但由于模型结构复杂, 模型参数在预测过程中难以及时更新, 因此在高炉煤气系统发生状态漂移时, 预测精度会有所降低。

收稿日期:2017-01-13; 修回日期:2017-02-21。

基金项目:中国科学院重点部署项目(KGZD-EW-302);中国科学院科技服务网络计划(KTJ-SW-STS-159);辽宁省科学技术计划项目(2015020140)。

作者简介:孙雪莹(1988-),女,山东德州人,博士研究生,主要从事能源预测与调度,人工智能算法方向的研究。

极限学习机 (Extreme learning machine, ELM) 是一种新型的神经网络^[10], 与 BP 网络和 SVM 相比, 具有结构简单、参数少、训练速度快等优点。目前, ELM 已经应用于许多在线回归和预测问题^[11-13]。序贯极限学习机^[14] (Online sequential extreme learning machine, OS-ELM) 是极限学习机的递推模式, 随着样品的到达在线更新输出权值。相比于极限学习机, OS-ELM 能更好地适应系统的动态变化, 减少样本更新时的模型训练时间。因此, 可以考虑将 OS-ELM 引入到煤气预测领域。但是由于 OS-ELM 只是将新样本加入到原有的模型中, 没有考虑新旧样本对模型的不同影响。对于高炉煤气系统而言, 工况变化时系统状态的改变会导致旧样本失效, 从而影响预测精度。

针对目前高炉煤气产销量在线预测方法所存在的问题, 结合基于数据的时间序列预测方法, 本文提出一种基于自适应遗忘因子极限学习机 (Adaptive forgetting factor extreme learning machine, AF-ELM) 的高炉煤气产销量在线预测方法。在 OS-ELM 的基础上, 引入了遗忘因子方法对样本进行加权, 通过误差反馈机制自适应调节遗忘因子, 使预测方法能够适应高炉煤气系统工况的动态变化。本文利用现场采集的高炉煤气产生量和热轧工序高炉煤气消耗量数据对提出的预测方法进行了实验验证, 实验结果表明了该方法的有效性。

1 序贯极限学习机

1.1 极限学习机

极限学习机具有典型的三层神经网络结构, 即输入层、输出层和隐含层, 如图 1 所示。给定具有 T 个不同样本的训练集 $(X_t, y_t), t=1, \dots, T, X_t \in R^m, y_t \in R$ 。假设 ELM 模型具有 L 个隐含层节点, 其激活函数为 $g(x)$, 则其数学表达式为:

$$\sum_{i=1}^L \beta_i g(a_i, b_i, X_t) = y_t, t = 1, \dots, T \quad (1)$$

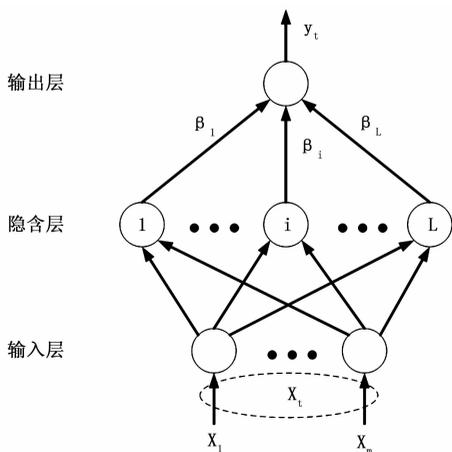


图 1 极限学习机基本原理示意图

其中: $a_i = [a_{i1}, a_{i2}, \dots, a_{im}]^T$ 表示连接输入节点和第 i 个隐含层节点的权值向量, b_i 表示第 i 个隐含层节点偏置值, β_i 表示连接第 i 个隐含层节点和输出层节点之间的权值, $i=1, \dots, L$ 。其中 a_i 和 b_i 随机生成, 且在整个训练过程中保持不变, 仅有输出权值 β_i 是需要确定的参数。由上述分析, (1) 式可以简写为:

$$Y = HB \quad (2)$$

其中:

$$H = \begin{bmatrix} g(a_1, b_1, X_1) & \dots & g(a_L, b_L, X_1) \\ \dots & \dots & \dots \\ g(a_1, b_1, X_T) & \dots & g(a_L, b_L, X_T) \end{bmatrix},$$

$$\beta = [\beta_1, \dots, \beta_L]$$

$$Y = [y_1, \dots, y_T]$$

根据 Moore-Penrose 广义逆理论可以得到 β 最小二乘意义下的最优估计值:

$$\beta = (H^T H)^{-1} H^T Y \quad (3)$$

由 (3) 式可以看出当样本出现共线性的时候, 由于难以保证 $H^T H$ 的非奇异性, 造成模型的泛化性能降低。为了增强模型的泛化性能, 文献^[15]综合考虑系统的经验风险和结构风险, 将 (2) 式转化为如下优化问题:

$$Y = \min_{\beta} \{ \|HB - T\|^2 + \lambda \|\beta\|^2 \} \quad (4)$$

由此得到:

$$\beta = (H^T H + \lambda I)^{-1} H^T Y \quad (5)$$

1.2 序贯极限学习机

序贯极限学习机是极限学习机的递推形式。该算法分为离线和在线两个阶段。

离线训练阶段: 随机给定模型的输入权值 a_i 和偏置 $b_i, i=1, \dots, L$, 计算初始隐含层节点输出矩阵 H_0 。由 (5) 式可得到:

$$\beta_0 = (H_0^T H_0 + \lambda I)^{-1} H_0^T Y_0 \quad (6)$$

令 $P_0 = (H_0^T H_0 + \lambda I)^{-1}$, 则:

$$\beta_0 = P_0 H_0^T Y_0 \quad (7)$$

在线预测阶段: 在 $k+1$ 时刻, 根据新增样本, 计算隐含层输出矩阵 $h_{k+1} = [g(a_1, b_1, x_{k+1}) \dots g(a_L, b_L, x_{k+1})]$ 。递推计算:

$$P_{k+1} = P_k - \frac{P_k h_{k+1}^T h_{k+1} P_k}{1 + h_{k+1}^T P_k h_{k+1}} \quad (8)$$

$$\beta_{k+1} = \beta_k + P_{k+1} h_{k+1}^T (y_{k+1} - h_{k+1} \beta_k) \quad (9)$$

2 基于 AF-ELM 的高炉煤气预测方法

2.1 自适应遗忘因子极限学习机

在线预测过程中, 随着新样本的不断加入, 旧样本的时效性不断降低。然而, OS-ELM 赋予新旧样本以同样的权值, 不能充分的利用新样本包含的信息, 从而难以适应系统的动态变化。因此为了改善 OS-ELM 对系统动态变化的适应性, 本文在序贯极限学习机的基础上提出了自适应遗忘因子极限学习机, 并给出了一种自适应遗忘因子计算方法。

在初始阶段, 假设初始样本集为:

$$\mathfrak{S}_0 = \{(X_t, y_t)\}_{t=1}^{N_0},$$

根据 (6) 式, 得到 $\beta_0 = (H_0^T H_0 + \lambda I)^{-1} H_0^T y_0$ 。令 $F_0 = H_0^T H_0 + \lambda I$, 则 $\beta_0 = F_0^{-1} H_0^T y_0$ 。

当新样本 $\mathfrak{S}_1 = \{(X_t, y_t)\}_{t=N_0+1}^{N_0+N_1}$ 到达时, 计算输出权值:

$$\beta_1 = F_1^{-1} \begin{bmatrix} H_0 \\ h_1 \end{bmatrix}^T \begin{bmatrix} Y_0 \\ y_1 \end{bmatrix} \quad (10)$$

$$F_1 = \begin{bmatrix} H_0 \\ h_1 \end{bmatrix}^T \begin{bmatrix} H_0 \\ h_1 \end{bmatrix} + \lambda I \quad (11)$$

对历史数据加入遗忘因子 α ,

$$F_1 = \begin{bmatrix} \alpha H_0 \\ h_1 \end{bmatrix}^T \begin{bmatrix} \alpha H_0 \\ h_1 \end{bmatrix} + \lambda I = \alpha^2 H_0^T H_0 + h_1^T h_1 + \lambda I = \alpha^2 F_0 + (1 - \alpha^2)\lambda I + h_1^T h_1 \quad (12)$$

$$\beta_1 = F_1^{-1} \begin{bmatrix} \alpha H_0 \\ h_1 \end{bmatrix}^T \begin{bmatrix} \alpha Y_0 \\ y_1 \end{bmatrix} = F_1^{-1} (\alpha^2 H_0^T y_0 + h_1^T y_1) \quad (13)$$

由 $F_0 \beta_0 = H_0^T y_0$ 得到

$$\beta_1 = \beta_0 + F_1^{-1} (h_1^T (y_1 - h_1 \beta_0) - (1 - \alpha^2)\lambda \beta_0) \quad (14)$$

由此可以递推得到

$$\beta_{k+1} = \beta_k + F_{k+1}^{-1} (h_{k+1}^T (y_{k+1} - h_{k+1} \beta_k) - (1 - \alpha^2)\lambda \beta_k) \quad (15)$$

$$F_{k+1} = \alpha^2 F_k + (1 - \alpha^2)\lambda I + h_{k+1}^T h_{k+1} \quad (16)$$

进一步, 考虑到高炉煤气系统在不同的时间段工况的变化情况不同, 为了加强模型对系统变化的适应性, 我们采用变遗忘因子代替固定的遗忘因子。根据系统的动态变化, 自适应地调整遗忘因子。预测误差可以反映当前系统状态的变化, 当误差较小时, 说明当前系统比较平稳, 此时应该增大 α , 增加旧样本的权值; 当误差较大时, 说明当前系统不平稳, 此时应该减小 α , 减少旧样本的权值。据此, 本文根据预测误差, 给出了遗忘因子的计算方法

$$\alpha_k = e^{-\eta MSE_{k-1}} \quad (17)$$

其中, MSE_{k-1} 表示前一时刻的均方误差, η 是遗忘速率的调节参数, 对不同的系统可以设置不同的遗忘速率。公式 (17) 可以保证 $\alpha \in (0, 1]$, 使 α 的选取符合之前的假设, 即新的样本比旧样本包含更多的当前系统的信息。同时, 在预测误差增大时, α 减小, 即当系统状态发生变化时, 减小旧样本的权值。

2.2 高炉煤气在线预测模型

本文采用基于数据的时间序列建模方法建立高炉煤气产消量的预测模型。时间序列预测的基本思想是通过系统本身的发展状态去构造系统的模型, 基本方法是采用相空间重构的方法。高炉煤气产消量的时间序列可以表示为 $\{x(1), x(2), \dots, x(t)\}$ 。根据相空间重构理论, 假设其延迟时间为 τ , 嵌入维数为 m , 可以得到训练样本集 $S: \{(X_1, y_1), (X_2, y_2), \dots, (X_t, y_t)\}$, 其中 $X_t = [x(t), x(t-\tau), \dots, x(t-(m-1)\tau)]^T, y_t = x(t+\tau)$ 。 X_t 为预测模型的输入, y_t 为模型的输出。

结合上文提出的在线预测算法, 给出基于 AF-ELM 的高炉煤气产生量和消耗量在线预测的具体的步骤如下:

- 1) 构造样本集: 读入高炉煤气产生量 (或消耗量) 的历史数据, 采用时间序列预测的思想, 利用相空间重构理论, 构造样本集, 将样本集分为训练样本集 D^{train} 和测试样本集 D^{test} ;
- 2) 参数初始化: 给定 ELM 模型的隐含层的权值以及遗忘速率 η ;
- 3) 利用训练样本集 D^{train} 得到初始模型 f^0 , 并令 $t = m + 1$;
- 4) 在线预测: While $t \leq T$ do:
 - a) 利用当前样本计算预测误差 MSE ;
 - b) 更新遗忘因子 α 以及各子模型的输出权值 β ;
 - c) 根据更新后的模型和当前的样本得到输出预测值 \hat{y} ;
 - d) 令 $t = t + 1$;
 End while

3 实验结果与分析

3.1 实验数据及预处理

为了验证文中提出的方法的有效性, 本文采用来自某大型钢铁企业高炉煤气系统 2015 年 5 月的现场数据作为实验数据。高炉煤气的产生单元主要是高炉炼铁, 高炉煤气的消耗单元主要有焦化、烧结、冷轧、热轧、棒线等生产工序。文中选取了一座 5 800 m³ 高炉的煤气产生量作为高炉煤气发生量预测的实验数据, 选取热轧工序的煤气消耗量作为高炉煤气消耗量预测的实验数据。煤气产消量的数据采样周期为 5 min, 选取 144 个 (即 12 小时) 连续的样本点作为训练数据集, 预测之后 96 个点 (即 8 小时) 的高炉煤气的产消量。根据自相关分析法 [17], 选择延迟时间为 1, 嵌入维数为 6, 即样本的特征值为 $x(t), x(t-1), \dots, x(t-6)$ 。由于在实际煤气管理系统中, 煤气的调度周期为 15min, 因此本文进行超前三步预测, 即令输出值为 $y = x(t+3)$ 。

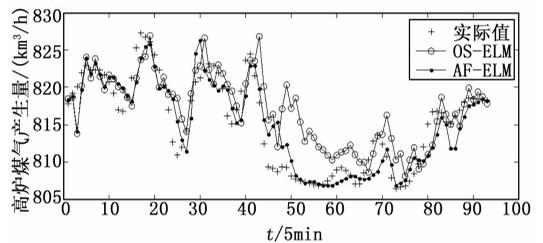


图 2 高炉煤气产生量预测结果

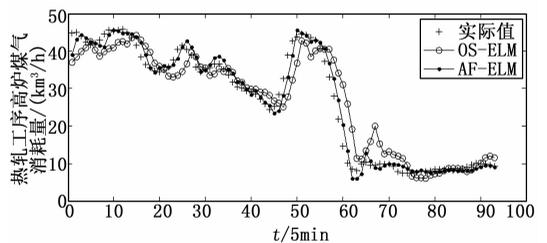


图 3 热轧工序高炉煤气消耗量预测结果

由于文中的实验数据由工业现场采集, 数据中存在异常值和噪声, 因此需要对数据进行预处理。首先, 剔除异常点, 采用插值法进行数据补齐。其次, 由于数据中的噪声会对预测结果产生影响, 因此本文采用小波阈值去噪方法 [16] 进行滤波处理, 在消除噪声的同时较好的保留了数据的有效信息。最后, 对处理后的数据进行归一化。

表 1 不同模型对多组高炉煤气产生量预测性能比较

测试集	AF-ELM			OS-ELM		
	MAPE/%	RMSE	耗时/s	MAPE/%	RMSE	耗时/s
1	0.2644	2.6991	0.0736	0.3639	3.6713	0.0165
2	0.2212	2.2404	0.0709	0.3548	3.6041	0.0178
3	0.2421	2.3943	0.0718	0.3178	3.4319	0.0179
4	0.2448	2.4353	0.0713	0.3494	3.4732	0.0169
5	0.2844	2.8618	0.0767	0.4156	4.3924	0.0165
6	0.2827	2.8875	0.0704	0.3737	3.8629	0.0169
7	0.2686	2.7891	0.0716	0.4051	4.1732	0.0172
8	0.2369	2.3919	0.0696	0.3748	3.9632	0.0173
平均值	0.2556	2.5874	0.0720	0.3694	3.8215	0.0171

表 2 不同模型对多组热轧工序高炉煤气消耗量预测性能比较

测试集	AF-ELM			OS-ELM		
	MAPE/%	RMSE	耗时/s	MAPE/%	RMSE	耗时/s
1	4.1441	1.6868	0.0762	7.7743	3.125	0.0171
2	2.5718	1.0022	0.0891	4.0011	1.5835	0.0167
3	3.6828	1.7961	0.0804	6.9541	3.5257	0.0173
4	6.3311	2.3889	0.0704	13.925	5.1467	0.0164
5	4.3762	1.7404	0.0782	9.0180	3.352	0.0167
6	2.9758	1.2517	0.0702	6.0759	2.5015	0.0185
7	3.6896	1.3137	0.0731	7.0672	2.6393	0.0169
8	3.6745	1.6271	0.0687	6.4702	3.1129	0.0164
平均值	3.9307	1.6008	0.0757	7.6607	3.1233	0.0170

3.2 实验结果比较与分析

为了评价模型的预测性能,采用平均绝对百分误差(MAPE)和均方误差(RMSE)作为评价指标。MAPE指标的定义如式(18)所示,该指标可以对预测误差的平均值进行评价。

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (18)$$

RMSE指标的定义如式(19)所示,该指标可以用来评价预测误差的分散程度。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (19)$$

本实验采用交叉验证的方法选择极限学习机的隐含层节点数为30,激活函数为Sigmoid函数,正则化参数 $\lambda=10^{-4}$ 。为了验证所建立模型的有效性,设计了如下实验,并与OS-ELM方法进行比较,OS-ELM的模型参数与本文模型相同。以5800 m³高炉煤气产生量和热轧工序煤气消耗量为研究对象分别选取9组样本,以其中一组为训练集,其余8组为测试集,验证所提出方法的在线预测性能。实验结果如表1和表2所示。本文采用MAPE和RMSE作为预测精度的评价指标,并比较了两种算法的时间开销。从实验结果可以看出,在所有算例中,AF-ELM的预测精度均要高于OS-ELM。本文提出的AF-ELM算法在高炉煤气产生量预测中,与OS-ELM相比,MAPE平均性能提升了0.1%,RMSE平均精度提升1.3;在热轧工序的高炉煤气消耗量预测中,本文算法与OS-ELM相比,MAPE平均性能提升了3.7%,RMSE平均精度提升1.5。由于加入了自适应遗忘因子的计算AF-ELM方法时间开销要略大于OS-ELM方法。虽然如此,由于AF-ELM方法每次在线预测的时间要远小于1s,因此仍然能够满足工业现场对在线预测时间的要求。

在其中分别随机选取一组测试集,预测结果如图2和图3所示。由图2和图3可以看出,在工况稳定的情况下,两种算法精度差异不大;当工况发生变化时,如图2中45~80点和图3中50~70点,煤气的产销量波动比较大,OS-ELM的预测精度下降,AF-ELM由于采用自适应遗忘因子提高了算法的动态适应能力,因而其预测精度高于OS-ELM。

4 结论

1) 本文提出了一种基于自适应遗忘因子极限学习机的高炉煤气产销量在线预测模型。针对OS-ELM在应用于在线预测时精度不高的问题,引入遗忘因子方法,并通过预测误差反

馈机制自适应调节遗忘因子,使其更加适应系统的动态变化。

2) 本文选取了来自实际生产现场的高炉煤气发生量以及热轧工序的煤气消耗量作为实验数据,与OS-ELM方法进行对比实验。实验表明与OS-ELM方法相比,本文提出的预测方法能够跟踪系统的动态变化,预测精度更高。

参考文献:

- [1] 张琦,谷延良,提威,等. 钢铁企业高炉煤气供需预测模型及应用[J]. 东北大学学报:自然科学版,2010,31(12):1737-1740.
- [2] 肖冬峰,杨春节,宋执环. 基于改进BP网络的高炉煤气发生量预测模型[J]. 浙江大学学报(工学版),2012,46(11):2103-2108.
- [3] Liu Y, Zhao J, Wang W. A time series based prediction method for a coke oven gas system in steel industry[J]. ICIC Express Letters, 2010,4(4):1373-1378.
- [4] 刘颖,赵珺,王伟,等. 基于数据的改进回声状态网络在高炉煤气发生量预测中的应用[J]. 自动化学报,2009,35(6):731-738.
- [5] 岳有军,户彦飞,赵辉,等. 基于小波分析的ARIMA与LSSVM组合的高炉煤气预测[J]. 计算机测量与控制,2015,23(6):2128-2131.
- [6] 张颜颜,唐立新. 改进的数据驱动子空间算法求解钢铁企业能源预测问题[J]. 控制理论与应用,2012,29(12):1616-1622.
- [7] 张晓平,汤振兴,赵璿. 钢铁企业高炉煤气发生量的在线预测建模[J]. 信息与控制,2010,39(6):774-782.
- [8] 李红娟,王建军,王华,等. 钢铁企业高炉煤气发生量预测建模及应用[J]. 系统仿真实报,2014,26(6):1308-1314.
- [9] 李红娟,王建军,王华,等. 建立PNN-HP-ENN-LSSVM模型预测钢铁企业高炉煤气发生量[J]. 过程工程学报,2013,13(3):451-457.
- [10] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006,70(1-3):489-501.
- [11] Zhang H, Yin Y, Zhang S. An improved ELM algorithm for the measurement of hot metal temperature in blast furnace[J]. Neurocomputing, 2016,174(Part A):232-237.
- [12] Liu H, Tian H, Li Y. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms[J]. Energy Conversion and Management, 2015,100:16-22.
- [13] Junior J J M S, Backes A R. ELM based signature for texture classification[J]. Pattern Recognition, 2016,51:395-401.
- [14] Liang N Y, Huang G B, Saratchandran P, et al. A fast and accurate online sequential learning algorithm for feedforward networks[J]. Neural Networks, IEEE Transactions on, 2006,17(6):1411-1423.
- [15] Huang G B, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012,42(2):513-529.
- [16] 王控中,朱玉田. 改进小波阈值去噪法的对比性仿真实验与分析[J]. 噪声与振动控制,2012,32(1):128-132.
- [17] 张淑清,贾健,高敏,等. 混沌时间序列重构相空间参数选取研究[J]. 物理学报,2010,59(3):1576-1582.