

# 基于非线性高斯混合回归的燃煤锅炉 $\text{NO}_x$ 排放浓度预测模型

李延臣, 孙智慧, 赵建华, 丁学东

(华能吉林发电有限公司长春热电厂, 吉林 长春 130216)

**摘要:**为了准确测量锅炉出口的  $\text{NO}_x$  排放浓度, 针对燃煤锅炉的复杂非线性, 提出了一种基于非线性高斯混合回归(Nonlinear Gaussian Mixture Regression, NGMR)的  $\text{NO}_x$  排放浓度预测方法。采用滑动时间窗方法, 结合奇异值分解实现稳态判定; 进一步采用互信息(Mutual Information, MI)判断不同变量与  $\text{NO}_x$  排放浓度的相关性, 确定模型输入变量; 利用选定的输入变量, 基于 NGMR 建立  $\text{NO}_x$  排放浓度预测模型; 基于某 660 MW 燃煤机组运行数据, 将提出的 NGMR 模型分别与人工神经网络(Artificial Neural Network, ANN)模型、支持向量回归(Support Vector Machine, SVR)模型以及极限学习机(Extreme Learning Machine, ELM)模型进行对比分析。结果表明: NGMR 模型预测均方根误差(Root Mean Square Error, RMSE)为 4.66 mg/m<sup>3</sup>, 平均绝对误差(Mean Absolute Error, MAE)为 3.98 mg/m<sup>3</sup>; 绝对误差系数( $R^2$ )为 0.9; 十折交叉验证结果也表明 NGMR 模型具有良好的预测精度和泛化能力。

**关键词:**  $\text{NO}_x$  浓度预测; 非线性高斯混合回归; 燃煤机组; SCR

中图分类号: TK221 文献标识码: A DOI: 10.16146/j.cnki.rndlgc.2023.09.013

[引用本文格式] 李延臣, 孙智慧, 赵建华, 等. 基于非线性高斯混合回归的燃煤锅炉  $\text{NO}_x$  排放浓度预测模型[J]. 热能动力工程, 2023, 38(9): 103-112. LI Yan-chen, SUN Zhi-hui, ZHAO Jian-hua, et al. NO<sub>x</sub> emission concentration prediction in Coal-fired boiler based on nonlinear Gaussian mixture regression[J]. Journal of Engineering for Thermal Energy and Power, 2023, 38(9): 103-112.

## NO<sub>x</sub> Emission Concentration Prediction in Coal-fired Boiler based on Nonlinear Gaussian Mixture Regression

LI Yan-chen, SUN Zhi-hui, ZHAO Jian-hua, DING Xue-dong

(Changchun Thermal Power Plant, Huaneng Jilin Power Generation Co., Ltd.,  
Changchun, China, Post Code: 130216)

**Abstract:** In order to accurately measure the concentration of  $\text{NO}_x$  at the boiler outlet, aiming at complex nonlinearity of coal-fired boiler, a  $\text{NO}_x$  concentration prediction model based on nonlinear Gaussian mixture regression (NGMR) was proposed. A time window sliding method, along with singular value decomposition, was used to realize steady judgement; furtherly, mutual information (MI) was adopted to identify the correlation between different variables and  $\text{NO}_x$  emission concentrations, thereby the input variables of model were determined; with the selected input variables, a  $\text{NO}_x$  concentration prediction model was established based on NGMR; based on the operation data of a 660 MW coal-fired power unit, the proposed NGMR model was compared and analyzed with artificial neural network (ANN), support vector machine (SVR) and extreme learning machine (ELM) models. The results show that for NGMR, root mean square error (RMSE) is 4.66 mg/m<sup>3</sup>, average absolute error (MAE) is 3.98 mg/Nm<sup>3</sup>, and absolute error coefficient ( $R^2$ ) is 0.9. And the 10-fold cross-validation results also show that NGMR model has better prediction accuracy and generalization ability.

**Key words:** NO<sub>x</sub> concentration prediction, nonlinear Gaussian mixture regression, coal-fired power units, SCR

## 引言

近年来,我国新能源电力装机容量占总装机容量比例逐年上升,但是燃煤机组装机和发电量仍然占据相当大的比例,且将长时间占据主导地位<sup>[1]</sup>。燃煤机组运行过程中,煤粉燃烧产生的 NO<sub>x</sub> 所引发的大气污染问题日趋严重,NO<sub>x</sub> 排放量的预测和控制得到了研究者们的广泛关注<sup>[2-3]</sup>。虽然随着测量技术的发展,检测仪器逐渐可靠且普及,可以快速得到 NO<sub>x</sub> 的测量数据,但是直接测量的 NO<sub>x</sub> 排放浓度无法反映出锅炉燃烧时不同运行参数对其生成的影响。因此,准确建立锅炉燃烧后 SCR 处理前的 NO<sub>x</sub> 排放量预测模型能够为调整燃烧过程的控制参数和优化 SCR 系统的喷氨量提供模型基础,具有重要的理论和运行指导意义。

建模方法对于 NO<sub>x</sub> 排放浓度的预测精度具有重要的影响。目前,主流的方法包括机理建模和数据驱动建模<sup>[4]</sup>。对于机理建模,NO<sub>x</sub> 的产生受到一次风量和配风方式等多种因素的影响,炉膛内的煤粉燃烧过程也表现出异常复杂的物理化学特性,部分数理方程无法充分反映完整的燃烧过程特性,利用机理模型进行准确的定量预测在某些时候仍存在困难<sup>[5]</sup>。此外,燃烧机理建模往往涉及多个复杂的耦合迭代计算过程,计算量普遍较大、耗时较长,且定量预测的准确性有待提升,因此难以直接用于在线调整锅炉的燃烧参数。随着人工智能和信息技术的快速发展,电厂生产过程积累了大量的运行数据,数据驱动方法为建立 NO<sub>x</sub> 排放浓度预测模型提供了切实可行的新思路。Zhou 等人<sup>[6]</sup>以支持向量机(Support Vector Machine, SVM)为基础,建立了不同运行参数与 NO<sub>x</sub> 排放浓度的映射模型。LYU 等人<sup>[7-8]</sup>和赵征等人<sup>[9]</sup>对 SVM 进行了改进和优化,提升了预测模型的精度。此外,ANN<sup>[10-12]</sup>、ELM<sup>[13]</sup>、径向基<sup>[14]</sup>和反向传播<sup>[15]</sup>神经网络等算法也被用来建立 NO<sub>x</sub> 排放浓度预测模型,并得到了一定的预测效果。但是,在深度调峰和新能源电力消纳等新形

势下,机组燃烧过程表现为复杂多模态运行特性,现有方法多假设运行数据服从单一运行模态,无法对其进行准确描述,模型预测精度难以进一步获得提升。虽然深度学习<sup>[4,16]</sup>方法对复杂燃烧过程具有更强的解释能力,但是基于梯度下降的训练过程计算复杂度大,而且模型的层数和每层神经元的个数难以确定,其实用性有待提升。

针对锅炉燃烧过程的复杂多模态特征,高斯混合模型(Gaussian Mixture Model, GMM)具有更强的数据特征提取能力<sup>[17]</sup>。GMM 假设不同模态各自服从特定高斯分布,通过对各高斯组分赋予概率权重来逼近多模态特性。同时,对于每个运行模态,分别建立局部模型,从而使 GMM 能够处理多模态问题。此外,锅炉燃烧时的高温、高压和高尘特性容易对传感器产生干扰,造成测量数据出现噪声以及分布发生偏离,GMM 从概率角度出发,能够处理过程不确定性<sup>[18]</sup>。Mei 等人<sup>[19]</sup>在 GMM 的基础上,提出了高斯混合回归(Gaussian Mixture Regression, GMR)模型,实现了多模态过程的关键变量预测,取得了良好的预测性能。Zhu 等人<sup>[17]</sup>提出了变分贝叶斯 GMR 模型,进一步克服了 GMR 可能出现的奇异性问题,并成功应用到工业过程中。

现有的大多数基于 GMR 的预测方法只能计算变量间的线性关系,且容易造成过拟合的问题,对于复杂非线性过程的适用性不足。因此,尝试将非线性分析纳入 GMR 建模过程,构建 NGMR 模型,有助于提升模型对锅炉燃烧过程的非线性特征提取能力。除了模型的选择,辅助变量的确定对预测结果的准确性也会产生直接的影响。吕游等人<sup>[20]</sup>采用偏最小二乘(Partial Least Squares, PLS)分析 NO<sub>x</sub> 和辅助变量的相关性来确定输入特征。Hong 等人<sup>[21]</sup>利用奇异值分解剔除辅助变量之间的相关性。上述方法改变了模型输入变量的原始信息,且只能分析变量间的线性关系,不利于提升模型的准确性。MI 通过分析变量间的信息熵衡量一个随机变量包含另一随机变量的信息量<sup>[22]</sup>,合理地使用 MI 分析变量之间的相关性,能有效删除冗余变量,为辅助变量选择提供依据。

本文在充分考虑燃煤锅炉运行特性的基础上,提出了一种基于 NGMR 模型的  $\text{NO}_x$  排放浓度预测方法。首先,在建模前根据锅炉燃烧理论分析选择可能影响  $\text{NO}_x$  生成的辅助参数。其次,采用滑动时间窗方法和奇异值分解实现稳态判定。然后,利用 MI 分析不同辅助变量与  $\text{NO}_x$  的相关性,选择相关性高的变量作为输入特征。最后,利用提出的 NGMR 方法建立预测模型,并对  $\text{NO}_x$  排放浓度进行预测。通过和 ANN 模型、SVR 模型和 ELM 模型对比分析,验证了提出的方法具有更高的精确度和泛化能力。

## 1 研究对象及变量初选

以某 660 MW 超临界燃煤直流锅炉为研究对象,如图 1 所示。该锅炉的炉膛横截面尺寸为  $18.82 \text{ m} \times 18.14 \text{ m}$ ,高度为  $64.65 \text{ m}$ ,采用四角切圆燃烧的形式。锅炉整体按照 II 形布置,配备低  $\text{NO}_x$  燃烧系统,所有的燃烧器均为垂直方向的摆动燃烧器,炉膛四角从下到上依次布置 A ~ F 共 6 层燃烧器,并在主燃烧器上下间隔布置一、二次风喷口提供煤粉燃烧所需的空气,在主燃烧区域顶部分别布置了 5 层分离燃尽风(Separated Over Fire Air, SOFA)和 2 层紧凑燃尽风(Close Coupled Over Fire Air, CCOFA)。试验锅炉采用淮南煤进行燃烧,且保持整个试验过程中煤种不发生改变。

$\text{NO}_x$  的生成形式主要包括热力型、燃料型和快速型,其中快速型占比一般不到 5%。 $\text{NO}_x$  主要通过

热力型和燃料型的途径产生,影响因素主要包括锅炉的结构、煤质特性和运行参数等。本文中锅炉的结构已经固定,且试验过程中煤质保持不变,因此主要分析相关运行参数与  $\text{NO}_x$  浓度之间的关系。此外,飞灰含碳量和 CO 含量等参数是燃烧后的输出量,与  $\text{NO}_x$  排放浓度之间不存在直接关系,作为输入建模难以指导燃烧优化从而实现  $\text{NO}_x$  排放控制,且现场缺乏这些变量的实时测点,故未纳入输入参数。如表 1 所示,初步选择 29 个辅助过程变量,包括机组负荷、总燃料量、总空气量、氧气浓度、6 台给煤机的给煤量(A ~ F)、6 层一次空气量(A ~ F)、6 层二次空气喷嘴的风门开度(AA ~ EF)、5 层 SOFA 风门开度(A ~ E)和 2 层 CCOFA 风门开度(A ~ B)作为输入变量,炉膛出口  $\text{NO}_x$  排放浓度作为模型输出变量。

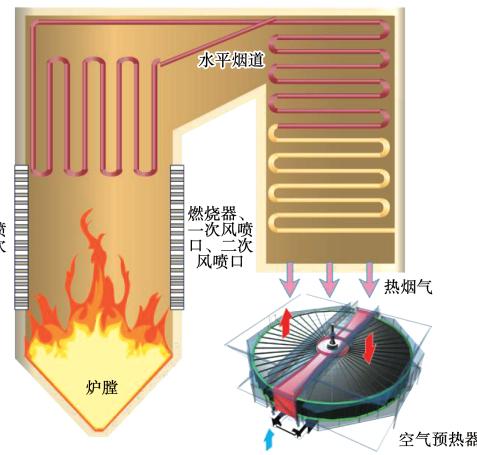


图 1 660 MW 燃煤锅炉示意图

Fig. 1 Schematic diagram of 660 MW coal-fired boiler

表 1 参与建模的过程变量和目标变量

Tab. 1 Process and target variables selected for modeling

参 数	变量数值范围
机组负荷/MW	[318, 661]
总给煤量/t·h <sup>-1</sup>	[142.3, 294.5]
总一次风流量/t·h <sup>-1</sup>	[1311.5, 2243.6]
氧量/%	[1.7, 5.5]
A-F 磨给煤量/t·h <sup>-1</sup>	[43.69, 51.48], [36.58, 54.26], [39.02, 52.86], [30.06, 56.05], [0, 51.21], [0.43, 21]
A-F 层一次风量/t·h <sup>-1</sup>	[95.21, 115.84], [95.14, 110.31], [92.02, 114.86], [99.86, 132.05], [0, 103.82], [0, 113.16]
AA-EF 层二次风挡板开度/%	[21.3, 54.4], [18.8, 46.5], [19.2, 47.3], [14.6, 54.9], [0, 51.3], [0, 98.7]
A-E 层分离燃尽风(SOFA)/%	[52.7, 98.5], [64.2, 99.2], [63.3, 98.8], [51.6, 99.4], [48.7, 94.2]
A-B 层紧凑燃尽风(CCOFA)/%	[3.7, 82.5], [1.1, 87.3]
$\text{NO}_x$ 排放浓度/mg·m <sup>-3</sup>	[256.4, 393.7]

## 2 锅炉出口 $\text{NO}_x$ 排放浓度预测算法

### 2.1 GMR 算法

给定  $m$  维输入变量  $x \in R^m$  和目标变量  $y \in R$ , 其联合分布为高斯分布的混合形式, 即:

$$p(x, y) = \sum_{k=1}^K \pi_k N(x, y | \mu_k, \Sigma_k^{-1}) \quad (1)$$

式中:  $N(\cdot)$ —高斯分布;  $K$ —高斯组分数;  $\mu_k$ —均值;  $\Sigma_k^{-1}$ —协方差矩阵;  $\pi_k$ —高斯组分的权重;  $P$ —概率密度。

基于期望最大化 (Expectation Maximization, EM) 算法, 高斯混合模型能通过迭代方式估计参数  $\{\pi_k, \mu_k, \Sigma_k^{-1}\}_{k=1}^K$ 。进而, GMR 对  $\mu_k$  和  $\Sigma_k$  进行分解得到:

$$\mu_k = \begin{pmatrix} \mu_k^i \\ \mu_k^o \end{pmatrix}, \Xi_k = \sum_k^{-1} = \begin{pmatrix} \Xi_k^i & \Xi_k^{io} \\ \Xi_k^{oi} & \Xi_k^o \end{pmatrix} \quad (2)$$

式中:  $\mu_k^i \in R^m$ ;  $\mu_k^o \in R$ ;  $\Xi_k^i \in R^{m \times m}$ ;  $\Xi_k^{io} = (\Xi_k^{oi})^T \in R^m$ ;  $\Xi_k^o \in R$ 。

给定输入变量  $x$ , 目标变量  $y$  的条件概率密度函数仍然服从高斯混合的形式, 即:

$$p(y | x) = \sum_{k=1}^K \pi_k^x N(y | \mu_k^x, \Xi_k^x) \quad (3)$$

式中:  $\mu_k^x = \mu_k^o + \Xi_k^{oi} (\Xi_k^i)^{-1} (x - \mu_k^i)$ ;  $\pi_k^x = (\pi_k N(x | \mu_k^i, \Xi_k^i)) / (\sum_{j=1}^K \pi_j N(x | \mu_j^i, \Xi_j^i))$ ;  $\Xi_k^x = \Xi_k^o - \Xi_k^{oi} (\Xi_k^i)^{-1} \Xi_k^{io}$ 。

基于式(3), 目标变量预测值  $\hat{y}$  的计算公式如式(4)所示:

$$\hat{y} = \sum_{k=1}^K \pi_k^x (\mu_k^o + \Xi_k^{oi} (\Xi_k^i)^{-1} (x - \mu_k^i)) \quad (4)$$

如式(4)所示, 基于 GMR 实现  $\text{NO}_x$  排放浓度的预测可对不同运行模态的预测结果进行概率加权, 充分考虑了不同的机组负荷的情况, 因此不需要在各负荷段单独建立预测模型。

### 2.2 NGMR 算法

GMR 算法容易产生协方差矩阵奇异等问题, 且该方法只能处理变量间的线性关系, 对于复杂非线性系统的建模效果容易受到限制。因此, 在 GMR 的基础上, 本文结合非线性映射, 提出了 NGMR 算法, 有助于提升模型对锅炉燃烧过程的非线性特征

提取能力。

基于式(1), 给定二元变量  $c = \{c_k\}_{k=1}^K, c_k \in \{0, 1\}$ , 均值  $\mu = \{\mu_k\}_{k=1}^K$  和协方差矩阵  $\Sigma = \{\Sigma_k^{-1}\}_{k=1}^K$ , 存在式(5)所示的联合概率密度函数。

$$p(x | c, \mu, \Sigma) = \prod_{k=1}^K [N(x | \mu_k, \Sigma_k^{-1})]^{c_k} \quad (5)$$

如式(5)所示, 输入变量  $x$  由  $K$  个不同运行工况下的样本汇总而成, 有必要对其进行非线性扩展, 从而获得  $\text{NO}_x$  排放浓度与输入变量之间的非线性关系。结合单隐藏层前馈神经网络, 可定义关于自变量  $x$  的非线性映射关系  $\varphi$ :

$$\varphi = (f_1(x), f_2(x), \dots, f_m(x), 1)^T \quad (6)$$

非线性映射  $f(x)$  有多种函数形式, 如正弦函数、指数函数和径向基函数等, 本文选择径向基函数, 如式(7)所示:

$$f(x) = \frac{1}{1 + \exp(-(\omega_l^T x + b_l))} \quad (7)$$

式中:  $\omega_l$  和  $b_l$ —随机生成参数, 下标  $l = 1, \dots, m$ 。

因此, 非线性映射后的输入变量矩阵可表示为式(8):

$$\Phi = \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) & 1 \\ f_1(x_2) & \dots & f_m(x_2) & 1 \\ \vdots & \ddots & \vdots & 1 \\ f_1(x_n) & \dots & f_m(x_n) & 1 \end{pmatrix} \quad (8)$$

式中:  $n$ —样本数。

基于式(8), 对于第  $k$  个高斯组分, 目标变量  $y$  可以看作是非线性映射量  $\varphi$  的线性关系, 即:

$$y = \varphi^T \alpha_k + \varepsilon_k \quad (9)$$

式中:  $\alpha_k$ —回归系数向量;  $\varepsilon_k \sim N(\varepsilon_k | 0, \lambda_k^{-1})$ —噪声;  $\lambda_k^{-1}$ —噪声方差。

基于式(9), 可以进一步确定锅炉燃烧过程中不同工况条件下  $\text{NO}_x$  排放浓度的条件概率密度函数, 进而获得相应的预测值。因此, 给定输入变量  $X = \{x_i\}_{i=1}^n$ , 关于目标变量  $Y = \{y_i\}_{i=1}^n$  的条件概率分布形式可以表示为:

$$p(Y | X, C, \alpha, \lambda) = \prod_{i=1}^n \prod_{k=1}^K [N(y_i | \varphi^T \alpha_k, \lambda_k^{-1})]^{c_k} \quad (10)$$

针对式(10)所示的概率预测形式, 需要估计

NGMR 的模型参数  $\{\pi, \mu, \Sigma, \alpha, \beta, \lambda\}$ , 从而实现  $\text{NO}_x$  排放浓度的预测, 本文选择变分贝叶斯方法<sup>[23]</sup> 实现参数估计。为了简化计算过程, 采用共轭先验分布对模型参数进行假设, 即:

$$p(\pi) = \text{Dir}(\pi | \chi_0) \quad (11)$$

$$\begin{aligned} p(\mu, \Sigma) &= \prod_{k=1}^K p(\mu_k, \Sigma_k) \\ &= \prod_{k=1}^K N(\mu_k | d_0, (\gamma_0 \sum_k)^{-1}) \times \\ &\quad W(\sum_k | W_0, v_0) \end{aligned} \quad (12)$$

$$p(\alpha | \beta) = \prod_{k=1}^K p(\alpha_k | \beta_k) = \prod_{k=1}^K N(\alpha_k | 0, \beta_k^{-1} I) \quad (13)$$

$$p(\beta) = \prod_{k=1}^K p(\beta_k) = \prod_{k=1}^K G(\beta_k | \kappa_0, \rho_0) \quad (14)$$

$$p(\lambda) = \prod_{k=1}^K p(\lambda_k) = \prod_{k=1}^K G(\lambda_k | \tau_0, \zeta_0) \quad (15)$$

式中:  $\text{Dir}(\cdot)$ —Dirichlet 分布;  $W(\cdot)$ —Wishart 分布;  $G(\cdot)$ —Gamma 分布; 下标为 0 的变量均为各分布的先验超参数。

基于式(11)~式(15)所示关于不同参数的共轭分布假设, 本文基于后验概率与先验概率的一致性, 通过循环迭代实现参数估计。所提 NGMR 方法充分结合不同模态信息和非线性关系进行建模, 从而能更好地适应  $\text{NO}_x$  排放浓度的预测, 并为锅炉的燃烧优化提供指导。

### 3 $\text{NO}_x$ 排放浓度预测模型构建

为了获得高精度的  $\text{NO}_x$  排放浓度预测模型, 提出了考虑锅炉非线性特性的混合高斯回归算法。该算法包含稳态判定, 辅助变量筛选, 非线性高斯混合回归建模 3 个部分。

#### 3.1 稳态判定

针对试验锅炉, 在 300~660 MW 负荷范围内, 从 DCS (Distributed Control System) 系统采集 10 天, 共 28 800 组实际运行数据进行训练和测试, 采样频率为 30 s, 并采用滑动时间窗方法<sup>[4]</sup> 来筛选稳态数据。给定输入变量  $X = \{x_i\}_{i=1}^n \in R^{n \times m}$ , 以  $s$  为时间窗宽度, 构建扩展输入矩阵  $Z_s$ :

$$Z_s = \begin{bmatrix} x_1^T & x_2^T & \cdots & x_s^T \\ x_2^T & x_{2+s}^T & \cdots & x_{s+1}^T \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-s+1}^T & x_{n-s+2}^T & \cdots & x_n^T \end{bmatrix} \quad (16)$$

定义  $S$  为扩展输入矩阵的协方差矩阵, 其计算表达式如式(17)所示。

$$S = \frac{1}{n-1} Z_s^T Z_s \quad (17)$$

对  $S$  进行奇异值分解, 得到从大到小排列的特征值  $\Lambda = (\zeta_1, \zeta_2, \dots, \zeta_m)^T$ , 对应特征向量矩阵为  $Q = (q_1, q_2, \dots, q_m)$ , 根据累计方差贡献度<sup>[24]</sup> 选择前  $l$  个主成分实现稳态判定。定义  $E = Q_l \Lambda_l = (e_{t-s+1}, \dots, e_t)$  为  $t$  时刻的稳态因子向量, 计算其标准差作为稳态评价指标 SI, 如式(18)所示。

$$SI = \sqrt{\frac{1}{s} \sum_{i=t-s+1}^t (e_i - \bar{e})^2} \leq \eta \quad (18)$$

式中:  $\bar{e}$ — $E$  中各元素的均值;  $\eta$ —稳态判定阈值;  $s$ —时间窗宽度。

对采集的样本进行稳态判定, 部分数据筛选结果如图 2 所示。

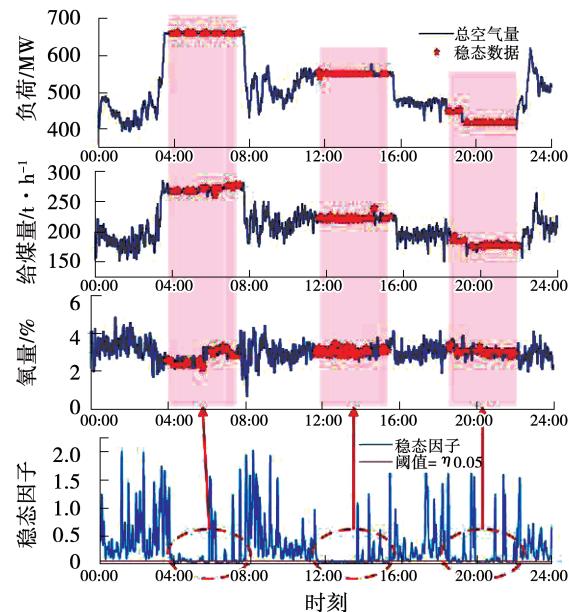


图 2 稳态判定结果示意图

Fig. 2 Schematic diagram of steady-state judgement results

由图可以看出, 选择时间窗宽度为 20, 设定阈值  $\eta = 0.05$ , 稳态因子 SI 能成功实现稳态数据筛

选。执行稳态判定后,根据不同稳态运行时间段的数据量在全部稳态数据集中所占比例进行蒙特卡洛采样<sup>[25]</sup>,在保留样本多样性的同时,删除重复的稳态数据。经过筛选,确定 4 248 组稳态数据用来建立 NO<sub>x</sub> 排放浓度预测模型。

### 3.2 输入变量筛选

经过机理分析初步确定了表 1 所示的 29 个输入变量,当部分变量与 NO<sub>x</sub> 排放浓度的相关性较弱时,会出现信息冗余,且变量过多容易导致模型计算复杂度增加,降低模型计算效率,因此需要剔除与 NO<sub>x</sub> 排放浓度相关性小的变量。以熵和条件熵为基础,MI 是衡量变量之间相互依赖程度的度量,表示为  $I(X;Y)$ 。定义熵是对变量的不确定性的量度,如式(19)所示:

$$H(X) = - \sum_x p_X(x) \log p_X(x) \quad (19)$$

式中: $p_X(x)$ —关于  $X$  的边缘概率分布。

给定目标变量  $Y$ ,定义输入变量  $X$  和  $Y$  之间的条件熵  $H(X|Y)$  为变量  $X$  的条件不确定性,即:

$$H(X|Y) = - \sum_{y \in Y} p_Y(y) \left[ - \sum_x p(x|y) \log p(x|y) \right] \quad (20)$$

式中: $p(x|y)$ — $x$  关于  $y$  的条件概率; $p_Y(y)$ —关于  $Y$  的边缘概率分布。

在熵和条件熵的基础上, $x$  和  $y$  的互信息表示为式(21):

$$I(x;Y) = H(x) - H(x|Y) \quad (21)$$

对于任意两个变量  $x$  和  $y$ ,有  $I(x;Y) \geq 0$ ,当且仅当  $x$  和  $y$  相互独立时等号成立。可选择最小冗余最大相关性 (Minimal Redundancy Maximal Relevance, mRMR) 为变量筛选评价标准,mRMR 定义如式(22)所示:

$$mRMR = \max \left\{ I(x_i;Y) - \frac{1}{m-1} \sum_{x_j \in S_x} I(x_i; x_j) \right\} \quad (22)$$

式中: $S_x$ —除  $x_i$  已外的其他输入变量; $m$ —变量的个数;下标  $i, j = 1, 2, \dots, m$ 。

当  $I(x_i;Y) \geq mRMR$ ,则表明输入变量  $x_i$  与目标变量  $Y$  之间存在显著的相关关系,保留该变量作为预测模型输入。图 3 给出了初选的 29 个输入变量

与 NO<sub>x</sub> 排放浓度之间互信息分布情况,根据计算得到的  $mRMR = 0.804$ ,最终确定  $x_1 - x_4, x_6 - x_8, x_{12} - x_{14}, x_{18}, x_{20} - x_{21}, x_{23} - x_{26}, x_{28} - x_{29}$  共 19 个变量建立 NO<sub>x</sub> 排放浓度预测模型。在变量筛选过程中,主要剔除了 A 和 F 两台磨煤机的相关变量。因为试验过程中,低负荷段时 F 磨煤机基本处于停运状态,只有在高负荷运行时,F 磨煤机才启动运行,因此锅炉负荷与 F 磨煤机的状态有强相关性,模型的输入中保留了负荷,已经间接考虑了 F 磨煤机的影响。此外,试验过程中,A 磨煤机的给煤量较大,且保持相对稳定的状态,因此 A 磨煤机的相关参数与 NO<sub>x</sub> 排放浓度的相关性较低。综上,采用 MI 方法筛选出了与 NO<sub>x</sub> 排放浓度相关性较高的变量进行建模,有助于提升了模型的计算效率。

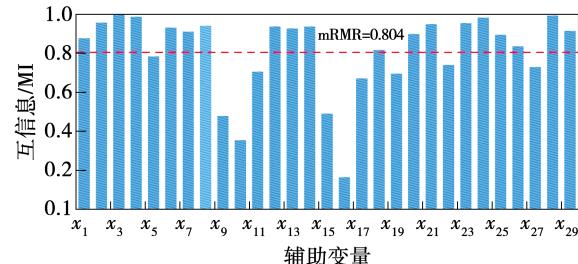


图 3 基于 MI 的变量筛选结果

Fig. 3 The result of selected variables based on MI

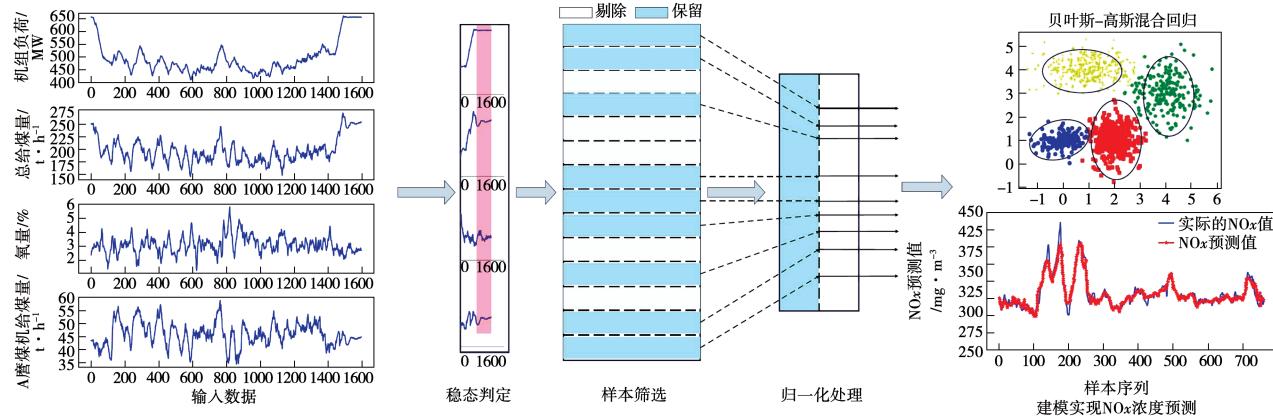
### 3.3 非线性高斯混合回归建模

在训练预测模型之前,需要首先对样本进行归一化处理,采用 Z-score 标准归一化方法,如式(23)所示:

$$x_{\text{scale}} = \frac{x - x_{\text{mean}}}{x_{\text{std}}} \quad (23)$$

式中: $x_{\text{std}}$ —样本的标准差; $x_{\text{mean}}$ —样本均值; $x_{\text{scale}}$ —归一化后的数据。

本文提出的基于 NGMR 的 NO<sub>x</sub> 排放浓度预测算法流程如图 4 所示。详细步骤如下:步骤 1,基于机理分析初步确定输入变量,并从 DCS 数据库采集历史运行数据;步骤 2,采用滑动时间窗方法,结合奇异值分解对采集的数据进行稳态判定;步骤 3,基于 MI 理论实现输入变量筛选,确定最终参与建模的变量;步骤 4,根据 2.2 节提出的非线性高斯混合回归算法构建 NO<sub>x</sub> 排放浓度预测模型。

图 4 锅炉 NO<sub>x</sub> 排放浓度预测流程示意图Fig. 4 Schematic diagram of NO<sub>x</sub> emissions concentration prediction process in a coal-fired boiler

#### 4 模型验证与分析

为了验证模型的可靠性,采用十折交叉验证,分 10 次随机选择稳态样本的 80% (3 398 组)作为训练集,剩余的 850 组数据作为测试集。分别建立 SVR 模型<sup>[6]</sup>、ANN 模型<sup>[10]</sup>、ELM 模型<sup>[13]</sup>,并与本文构建的基于 NGMR 的 NO<sub>x</sub> 预测模型进行预测性能对比。

##### 4.1 模型性能指标

采用均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均绝对误差系数 ( $R^2$ ) 作为指标评价模型的预测性能。具体的计算表达式为:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (24)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (26)$$

式中: $\bar{y}$ —NO<sub>x</sub> 排放浓度实测均值; $y_i$ —NO<sub>x</sub> 排放浓度真实值; $\hat{y}_i$ —NO<sub>x</sub> 排放浓度预测值; $n$ —样本数。

上述性能指标中, RMSE 和 MAE 的值越小, 则模型预测 NO<sub>x</sub> 排放浓度的精确度越高,  $R^2$  越接近于 1, 则预测模型的拟合效果越好。

##### 4.2 模型测试分析

基于 NGMR 的 10 组交叉验证测试数据的预测

模型性能指标如图 5 所示。

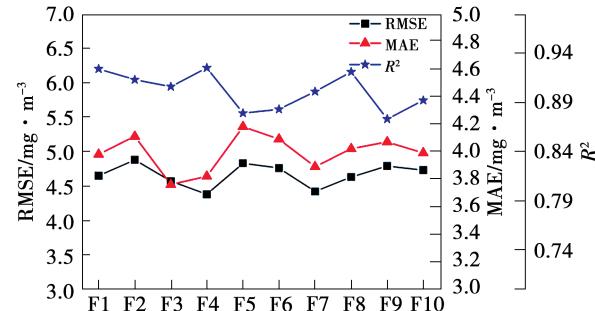


图 5 NGMR 预测模型 10 组交叉验证结果

Fig. 5 Ten-fold cross-validation results of the NGMR prediction model

由图可以看出,各组测试集的 RMSE 均在 4.6 mg/m<sup>3</sup> 附近, MAE 均在 4 mg/m<sup>3</sup> 上下波动,  $R^2$  则稳定在 0.9 左右,说明提出的模型具有较高的预测精度,且对不同的数据集具有泛化能力。基于 SVR、ANN、ELM 和 NGMR 的 NO<sub>x</sub> 排放浓度预测模型经过交叉验证后,不同性能指标的均值如表 2 所示。可以看出, ANN 模型的预测性能最差,测试集的各项性能指标分别为 RMSE = 9.03 mg/m<sup>3</sup>, MAE = 7.45 mg/m<sup>3</sup>,  $R^2$  = 0.75,而训练集的各项性能指标则显著好于测试结果。这是因为 ANN 模型训练容易产生过拟合现象,模型泛化能力不足,且模型的层数和每层神经元的个数难以确定。相较而言,SVR 和 ELM 算法的训练结果与测试结果相接近,但是 RMSE 和 MAE 的数值较大,且  $R^2$  在 0.8 左右,模型的预测精度较低。

提出的 NGMR 预测模型有  $RMSE = 4.66 \text{ mg/m}^3$ ,  $MAE = 3.98 \text{ mg/m}^3$ ,  $R^2 = 0.90$ , 且训练集对应的

$RMSE = 4.03 \text{ mg/m}^3$ ,  $MAE = 3.46 \text{ mg/m}^3$ ,  $R^2 = 0.92$ , 均表现出较高的预测精度。

表 2 不同建模方法预测结果对比

Tab. 2 Comparison of predicted results by different modeling methods

模型	RMSE		MAE		$R^2$	
	训练	测试	训练	测试	训练	测试
ANN	6.39	9.03	4.93	7.45	0.86	0.75
SVR	7.35	7.93	5.92	6.64	0.86	0.82
ELM	7.92	8.42	6.08	6.71	0.81	0.78
NGMR	4.03	4.66	3.46	3.98	0.92	0.90

图 6 给出了不同模型  $\text{NO}_x$  预测排放浓度的分布情况。可以看出, ANN 的预测值与理想情况存在较大的偏离度, 仅有少量结果与理想情况契合。 ELM 模型的预测结果中, 多数预测值位于理想线的

上方, 模型的整体偏差较大。 SVR 和 NGMR 模型的结果均匀分布于理想线两侧, 但是基于 SVR 的预测结果与理想线的距离大于 NGMR 模型的预测结果。

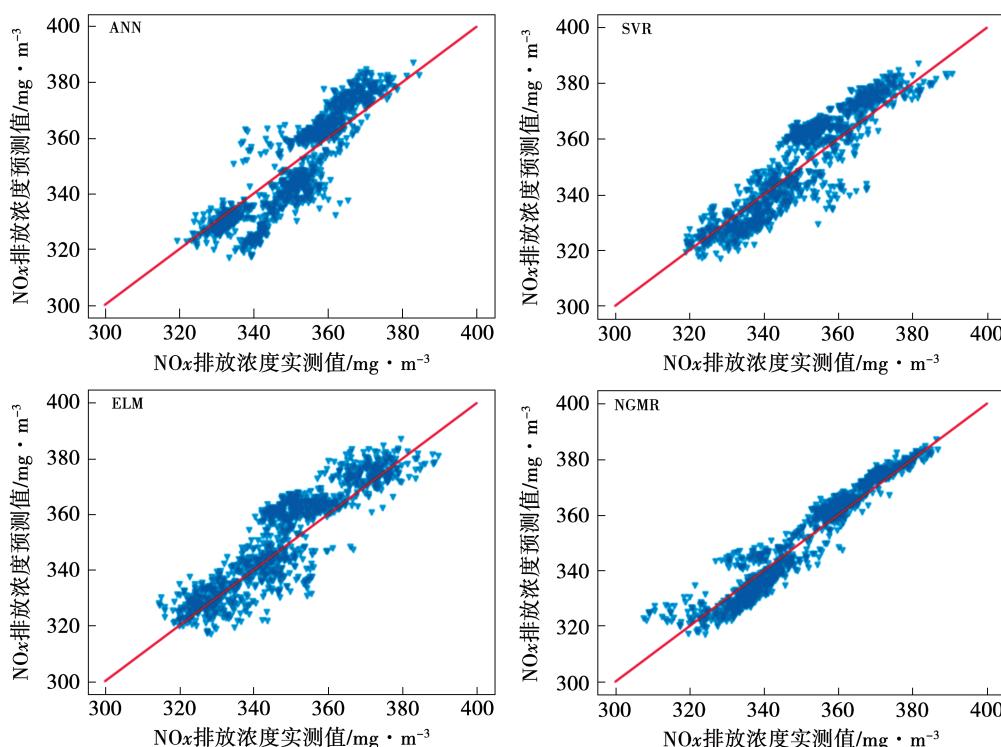


图 6 不同模型预测值分布

Fig. 6 Distribution of predicted values for different models

图 7 为不同模型绝对预测误差的频率分布直方图。可以看出, 4 个模型预测绝对误差中位数分别为  $ANN = 6.53$ ,  $SVR = 6.47$ ,  $ELM = 5.58$ ,  $NGMR = 3.96$ 。基于 NGMR 模型的  $\text{NO}_x$  预测排放浓度, 大多数绝对预测误差均小于  $10 \text{ mg/m}^3$ , 而另外 3 种模型

对应的绝对预测误差则分布在较宽的范围内, 表明预测精度低于 NGMR 模型。综上所述, 证明了基于 NGMR 的  $\text{NO}_x$  排放浓度预测模型具有更高的准确性和更好的泛化能力。

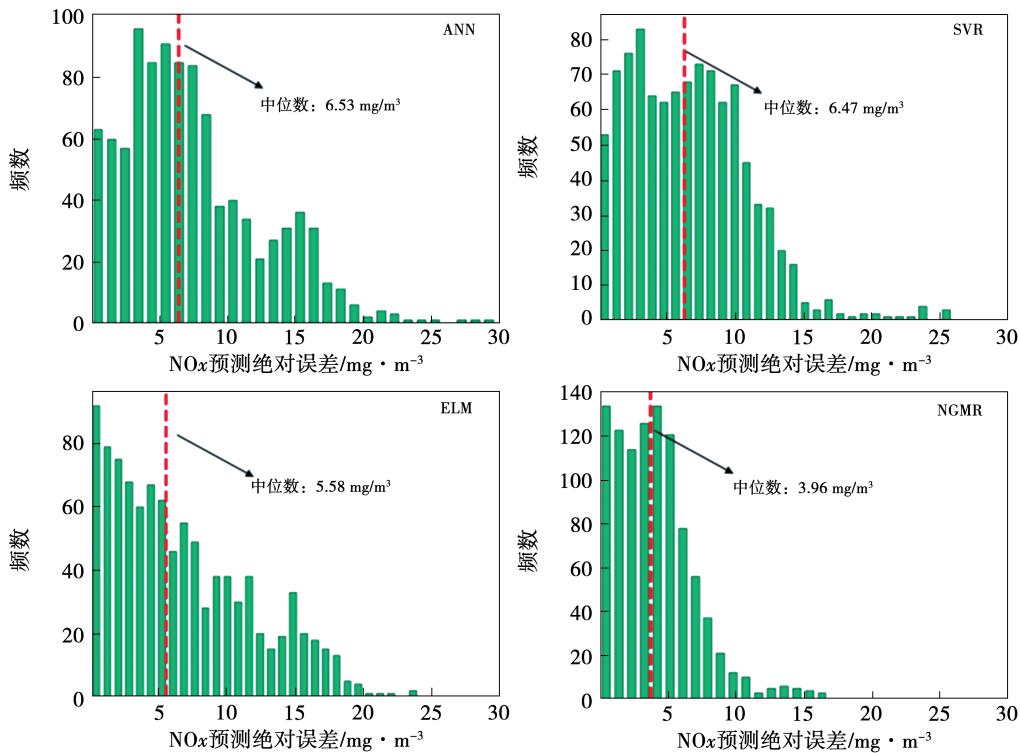


图 7 不同模型绝对预测误差频率直方图

Fig. 7 Absolute prediction error frequency histogram of different models

## 5 结 论

本文提出了基于非线性高斯混合回归方法的燃煤锅炉  $\text{NO}_x$  排放浓度预测模型,并结合实际锅炉运行数据对提出的模型进行了验证,得到了以下结论:

(1) 试验结果表明,经过稳态判定、变量筛选和非线性高斯混合回归建模,获得了锅炉燃烧相关的过程变量与  $\text{NO}_x$  排放浓度之间的复杂非线性关系,提升了锅炉  $\text{NO}_x$  排放浓度预测的精度和泛化能力。

(2) 建立的 NGMR 模型能够很好反应运行参数间的相关关系,与 ANN 模型、SVR 模型、ELM 模型相比,具有更好的预测效果。

(3) 基于 NGMR 的燃煤锅炉出口  $\text{NO}_x$  预测模型具有显著的自适应能力,能为锅炉燃烧参数调整和 SCR 系统优化提供模型基础。

## 参考文献:

[1] 郑明光,刘海滨. 以低碳能源为基础的创新体系建设研究[J]. 上海质量,2021(8):23-25.

ZHENG Ming-guang, LIU Hai-bin. Research on the construction of innovation system based on low-carbon energy[J]. Shanghai Quali-

ty, 2021(8):23-25.

- [2] 唐振浩,柴向颖,曹生现,等. 考虑时延特征的燃煤锅炉  $\text{NO}_x$  排放深度学习建模[J]. 中国电机工程学报, 2020, 40(20): 6633-6643.  
TANG Zhen-hao, CHAI Xiang-ying, CAO Sheng-xian, et al. Deep learning modeling for the  $\text{NO}_x$  emissions of coal-fired boiler considering time-delay characteristics [J]. Proceedings of the CSEE, 2020, 40(20): 6633-6643.
- [3] TAN Peng, HE Biao, ZHANG Cheng, et al. Dynamic modeling of  $\text{NO}_x$  emission in a 660 MW coal-fired boiler with long short-term memory[J]. Energy, 2019, 176: 429-436.
- [4] FAN Wei, SI Feng-qi, REN Shao-jun, et al. Integration of continuous restricted Boltzmann machine and SVR in  $\text{NO}_x$  emissions prediction of a tangential firing boiler[J]. Chemometrics and Intelligent Laboratory Systems, 2019, 195: 103870.
- [5] 谢磊,毛国明,金晓明,等. 循环流化床锅炉燃烧过程预测控制与经济性能优化[J]. 化工学报, 2016, 67(3): 695-700.  
XIE Lei, MAO Guo-ming, JIN Xiao-ming, et al. Predictive control and economic performance optimization of CFBB combustion process[J]. CIESC Journal, 2016, 67(3): 695-700.
- [6] ZHOU Hao, ZHAO Jia-pei, ZHENG Li-gang, et al. Modeling  $\text{NO}_x$  emissions from coal-fired utility boilers using support vector regression with ant colony optimization[J]. Engineering Applications of Artificial Intelligence, 2012, 25(1): 147-158.

- [7] LYU You, LIU Ji-zhen, YANG Ting-ting, et al. A novel least squares support vector machine ensemble model for  $\text{NO}_x$  emission prediction of a coal-fired boiler [J]. Energy, 2013, 55: 319 – 329.
- [8] LYU You, YANG Ting-ting, LIU Ji-zhen. An adaptive least squares support vector machine model with a novel update for  $\text{NO}_x$  emission prediction [J]. Chemometrics and Intelligent Laboratory Systems, 2015, 145: 103 – 113.
- [9] 赵征,袁洪,宋选锋. 基于 MI-LSSVM 的  $\text{NO}_x$  生成量动态软测量模型[J]. 自动化仪表,2018,39(12):31 – 34,38.  
ZHAO Zheng, YUAN Hong, SONG Xuan-feng. Dynamic soft-sensing model for amount of  $\text{NO}_x$  production based on MI-LSSVM [J]. Process Automation Instrumentation, 2018, 39(12) :31 – 34,38.
- [10] ZHOU Hao, CEN Ke-fa, FAN Jian-ren. Modeling and optimization of the  $\text{NO}_x$  emission characteristics of a tangentially fired boiler with artificial neural networks [J]. Energy, 2004, 29(1) :167 – 183.
- [11] LLAMATHI P, SELLADURAI V, BALAMURUGAN K, et al. ANN-GA approach for predictive modeling and optimization of  $\text{NO}_x$  emission in a tangentially fired boiler [J]. Clean Technologies and Environmental Policy, 2013, 15(1) :125 – 131.
- [12] SHAKIL M. Soft sensor for  $\text{NO}_x$  and  $\text{O}_2$  using dynamic neural networks [J]. Computers & Electrical Engineering, 2009, 35(4) : 578 – 586.
- [13] TAN Peng, XIA Ji, ZHANG Cheng, et al. Modeling and reduction of  $\text{NO}_x$  emissions for a 700 MW coal-fired boiler with the advanced machine learning method [J]. Energy, 2016, 94: 672 – 679.
- [14] ZHENG Li-gang, ZHOU Hao, WANG Chun-lin, et al. Combining support vector regression and ant colony optimization to reduce  $\text{NO}_x$  emissions in coal-fired utility boilers [J]. Energy & Fuels, 2008, 22(2) :1034 – 1040.
- [15] 许昌,吕剑虹. 基于生成机理的燃煤电站锅炉  $\text{NO}_x$  排放量神经网络模型 [J]. 中国电机工程学报, 2004, 24 (10) : 233 – 239.  
XU Chang, LYU Jian-hong. Artificial neural networks model for  $\text{NO}_x$  emissions in a pulverized-coal power station boiler on generation mechanism [J]. Proceeding of the CSEE, 2004, 24 (10) : 233 – 239.
- [16] WANG Fang, MA Su-xia, WANG He, et al. Prediction of  $\text{NO}_x$  emission for coal-fired boilers based on deep belief network [J]. Control Engineering Practice, 2018, 80: 26 – 35.
- [17] ZHU Jin-lin, GE Zhi-qiang, SONG Zhi-huan. Variational Bayesian Gaussian mixture regression for soft sensing key variables in non-Gaussian industrial processes [J]. IEEE Transactions on Control Systems Technology, 2016, 25(3) :1092 – 1099.
- [18] ZHOU Le, CHEN Jung-hui, SONG Zhi-huan, et al. Probabilistic latent variable regression model for process-quality monitoring [J]. Chemical Engineering Science, 2014, 116(3) :296 – 305.
- [19] MEI Cong-li, SU Yong, LIU Guo-hai, et al. Dynamic soft sensor development based on Gaussian mixture regression for fermentation processes [J]. Chinese Journal of Chemical Engineering, 2017, 25(1) :116 – 122.
- [20] 吕游,刘吉臻,杨婷婷,等. 基于 PLS 特征提取和 LS-SVM 结合的  $\text{NO}_x$  排放特性建模 [J]. 仪器仪表学报, 2013, 34(11) : 2418 – 2424.  
LYU You, LIU Ji-zhen, YANG Ting-ting, et al.  $\text{NO}_x$  emission characteristic modeling based on feature extraction using PLS and LS-SVM [J]. Chinese Journal of Scientific Instrument, 2013, 34(11) :2418 – 2424.
- [21] HONG D, BALZANO L, FESSLER J A. Towards a theoretical analysis of PCA for heteroscedastic data [C]//2016 54th Annual Allerton Conference on Communication, Control, and Computing. Allerton: IEEE, 2016: 496 – 503.
- [22] ZHENG Xiu-juan, FANG Hua-jing. An integrated unscented Kalman filter and relevance vector regression approach for lithiumion battery remaining useful life and short-term capacity prediction [J]. Reliability Engineering & System Safety, 2015, 144: 74 – 82.
- [23] BISHOP C M, NASRABADI N M. Pattern recognition and machine learning [M]. New York: Springer, 2006.
- [24] QIN J S. Statistical process monitoring: basics and beyond [J]. Journal of Chemometrics, 2003, 17(8/9) :480 – 502.
- [25] SHAPIRO A. Monte Carlo sampling methods [J]. Handbooks in Operations Research and Management Science, 2003, 10: 353 – 425.

(姜雪梅 编辑)