



基于非负约束的稀疏化ONLSTM及其工业软测量建模

郭迎宸, 隋璘, 熊伟丽

引用本文:

郭迎宸, 隋璘, 熊伟丽. 基于非负约束的稀疏化ONLSTM及其工业软测量建模[J]. *智能系统学报*, 2025, 20(6): 1366–1378.

GUO Yingchen, SUI Lin, XIONG Weili. Sparse ONLSTM and non-negative constrained industrial soft sensing modeling[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(6): 1366–1378.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202502004>

您可能感兴趣的其他文章

不完备数据中面向特征值更新的增量特征选择方法

Incremental approach for feature selection in incomplete data while updating feature values
智能系统学报. 2021, 16(3): 493–501 <https://dx.doi.org/10.11992/tis.202006045>

一种基于深度学习目标检测的长时目标跟踪算法

A long-term object tracking algorithm based on deep learning and object detection
智能系统学报. 2021, 16(3): 433–441 <https://dx.doi.org/10.11992/tis.201910029>

用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction
智能系统学报. 2021, 16(3): 518–527 <https://dx.doi.org/10.11992/tis.202008036>

深度自编码与自更新稀疏组合的异常事件检测算法

Abnormal event detection method based on deep auto-encoder and self-updating sparse combination
智能系统学报. 2020, 15(6): 1197–1203 <https://dx.doi.org/10.11992/tis.202007003>

记忆神经网络在机器人导航领域的应用与研究进展

Research progress and application of memory neural network in robot navigation
智能系统学报. 2020, 15(5): 835–846 <https://dx.doi.org/10.11992/tis.202002020>

多标记学习自编码网络无监督维数约简

Unsupervised dimensionality reduction of multi-label learning via autoencoder networks
智能系统学报. 2018, 13(5): 808–817 <https://dx.doi.org/10.11992/tis.201804051>

DOI: 10.11992/tis.202502004

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20251009.1946.008>

基于非负绞杀的稀疏化 ONLSTM 及其工业软测量建模

郭迎宸¹, 隋璘¹, 熊伟丽^{1,2}

(1. 江南大学物联网工程学院, 江苏 无锡 214122; 2. 江南大学轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

摘要: 实际工业过程往往具有多变量、非线性和动态性等特点, 建模数据包含过多冗余信息和时序依赖特征, 从而导致建模复杂度增加和模型性能下降。因此, 提出一种基于非负绞杀的稀疏化有序神经元长短期记忆网络 (ordered neurons long short-term memory, ONLSTM) 用于工业软测量建模。将非负绞杀收缩系数嵌入 ONLSTM 输入层权重矩阵, 对其进行收缩绞杀, 剔除冗余输入节点的同时实现变量选择。将非负绞杀收缩系数与 ONLSTM 隐藏层权重矩阵相结合, 根据不同隐藏神经元重要性设计权重分配规则, 剔除网络隐藏层冗余节点及其对应的信息传递通路, 进行网络结构稀疏优化。通过数值仿真验证了所提算法的有效性, 并将其应用于某火电厂烟气脱硫过程排放净烟气 SO₂ 浓度预测。实验结果表明所提算法能有效实现变量选择, 并在保证预测性能的前提下, 使模型结构得到稀疏优化, 展现出比较广阔的应用前景。

关键词: 软测量; 长短期记忆网络; 有序神经元; 非负绞杀; 冗余信息; 变量选择; 稀疏优化; 深度学习

中图分类号: TP274 **文献标志码:** A **文章编号:** 1673-4785(2025)06-1366-13

中文引用格式: 郭迎宸, 隋璘, 熊伟丽. 基于非负绞杀的稀疏化 ONLSTM 及其工业软测量建模 [J]. 智能系统学报, 2025, 20(6): 1366-1378.

英文引用格式: GUO Yingchen, SUI Lin, XIONG Weili. Sparse ONLSTM and non-negative constrained industrial soft sensing modeling[J]. CAAI transactions on intelligent systems, 2025, 20(6): 1366-1378.

Sparse ONLSTM and non-negative constrained industrial soft sensing modeling

GUO Yingchen¹, SUI Lin¹, XIONG Weili^{1,2}

(1. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; 2. Key Laboratory of Advanced Process Control for Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China)

Abstract: Industrial processes often exhibit characteristics such as multivariable interactions, nonlinear behaviors, and time-varying changes. Thus, the resulting modeling data contain excessive redundant information and complex time-dependent patterns, which increase modeling complexity and degrade the model performance. To address these challenges, an ordered neurons long short-term memory (ONLSTM) network integrated with non-negative garrote-based regularization is proposed herein for industrial soft sensor modeling. First, the shrinkage coefficients of the non-negative garrote are embedded into the weight matrix of the ONLSTM input layer to eliminate redundant input nodes and enable variable selection. Second, these coefficients are integrated into the weight matrix of the ONLSTM hidden layer to assign weights based on the importance of hidden neurons. Consequently, redundant nodes and their corresponding information pathways are pruned, achieving sparse optimization of the network structure. The proposed method is validated via numerical simulations and subsequently employed to predict the SO₂ concentration in flue gas emissions from a desulfurization process in a thermal power plant. Experimental results demonstrate that the algorithm effectively selects variables and sparsely optimizes the model structure while maintaining high predictive performance, offering promising prospects for broader industrial applications.

Keywords: soft sensor; long short-term memory networks; ordered neurons; non-negative strangulation; redundant information; variable selection; sparse optimization; deep learning

收稿日期: 2025-02-18. 网络出版日期: 2025-10-10.

基金项目: 国家自然科学基金项目 (61773182); 江南大学“轻工技术与工程”双一流学科与支撑学科协同发展支持计划 (QGJC20230203).

通信作者: 熊伟丽. E-mail: greenpre@163.com.

随着工业智能化的不断推进, 工业生产的复杂程度与信息化水平日益提升。为提高产品质量, 需要对某些难以直接或实时采集但与产品质量密切相关的质量变量进行实时监测^[1-3]。软测

量技术通过构建易于获取的辅助变量与质量变量之间的数学模型来有效预测相关质量变量, 已被广泛应用于工业过程监测和控制领域^[4-6]。

为提升工业过程监测与控制精度, 更高效且鲁棒性更强的软测量建模算法是关键。人工神经网络凭借其强大的非线性映射能力与无需依赖先验知识的学习能力成为软测量领域的研究热点^[7-9]。近年来, 以长短期记忆 (long short-term memory, LSTM) 网络为代表的门控循环神经网络, 通过其独特的门控结构控制信息流动, 更好地捕捉了数据中的长期依赖关系^[10-12]。文献 [13] 通过构建联合 LSTM 网络与卷积网络的特征提取模型, 并利用 LSTM 实现特征融合, 进而利用全连接模块构建预测模型。文献 [14] 通过构建一个结合 LSTM 模块和全连接模块的骨干网络, 并将物理退化属性集成到迁移学习框架中, 提出一种物理引导的 Transfer Learning-LSTM 网络用于锂离子电池的退化轨迹预测。文献 [15] 以逻辑方式指导预训练的 LSTM 网络而获得 LogicLSTM, 进行具有逻辑推理能力的时态数据处理, 以提高分类的预测准确性和可解释性。

然而, LSTM 网络隐藏层神经元相互独立, 未考虑神经元之间的序列层次信息, 导致不同神经元可能学到相同特征信息^[16]。有序神经元长短时记忆网络 (ordered neurons long short-term memory, ONLSTM) 通过对 LSTM 内部隐藏神经元进行重要性排序并引入层级概念, 新增主遗忘门与主输入门以分区更新的方式对特征信息进行选择^[17], 提高了模型对长期依赖关系的捕捉能力, 引起了众多学者的广泛关注。文献 [18] 提出一种带自注意力机制的卷积 ONLSTM 网络时序预测模型, 通过卷积神经网络降低特征维度, 并引入自注意力机制以动态调整特征权重。文献 [19] 提出一种具有注意力引导的 ONLSTM 网络, 通过对齿轮健康特征信息进行层次划分实现了齿轮的剩余寿命预测。文献 [20] 提出一种基于时序迁移的双流加权 ONLSTM 模型, 采用双流结构精确控制信息流传递, 并将模型内嵌于时序迁移框架以解决时间协变量移位问题。

基于深度网络的软测量建模过程中, 建模数据质量及网络结构合理性同样是不可忽视的关键影响因素。然而, 实际工业过程数据中包含过多冗余变量, 导致模型需要更复杂的结构与更多参数学习数据特征, 进而使建模难度增加与性能下降^[21]。因此, 降低模型复杂度并有效提取数据中的关键信息也是软测量建模领域亟待解决的科学

难题^[22]。稀疏模型通过去除冗余信息, 保留模型中最重要的信息流, 可以在有效解决工业数据集数据冗余问题的同时稀疏化模型结构^[23]。非负绞杀 (nonnegative garrote, NNG) 是一种经典的模型稀疏化方法, 常用以对输入变量进行选择压缩^[24]。其基本思想是在回归模型中引入非负约束, 保持模型预测能力的同时降低模型复杂度。文献 [25] 利用 NNG 对人工神经网络的输入权重进行精确收缩, 剔除了冗余输入变量, 与其他先进方法相比, 所提方法在选择变量较少的情况下表现出更好的模型准确性。文献 [26] 结合非平稳干扰提出一种自适应偏最小二乘 NNG 方法, 降低了输入维度, 实现动态过程的高效在线应用, 用于高炉炉膛壁的温度预测和蒸馏塔的杂质成分估计。文献 [27] 提出一种基于注意力机制与非负绞杀的“门控循环单元”, 不仅剔除了无关输入变量, 还在保证模型预测精度的同时对模型结构进行稀疏。

综上所述, 面向实际工业过程的多变量、非线性和动态性等特点, 提出基于非负绞杀的结构稀疏化 ONLSTM 工业软测量模型。一方面利用 ONLSTM 自动学习层级结构信息对数据进行充分的利用以实现建模; 另一方面通过 NNG 算法在 ONLSTM 模型的输入层中引入非负约束, 在保证模型预测能力的前提下, 剔除冗余输入变量。同时考虑到 ONLSTM 网络结构复杂度相较 LSTM 有所增加, 因此在其隐藏层加入 NNG 算法, 以减少模型计算量并剔除网络模型的冗余节点, 以实现模型的稀疏化。最后基于数值仿真和火电厂脱硫过程的生产实际数据, 建立相对应的软测量模型, 并与不同软测量算法进行性能对比, 验证了算法的有效性。

1 ONLSTM 神经网络

ONLSTM 在 LSTM 的基础上增添了主遗忘门和主输入门对隐藏层神经元进行排序, 进而引入信息层级概念。ONLSTM 通过主遗忘门和主输入门将输入信息分割为 3 部分, 即主遗忘通路、主输入通路和信息重叠通路, 分别计算后合并为输出。其结构如图 1 所示, 图中灰色部分为传统 LSTM 网络结构, 彩色部分为 ONLSTM 新增结构。传统 LSTM 门控单元计算过程为

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf} + b_f) \quad (1)$$

$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi} + b_i) \quad (2)$$

$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho} + b_o) \quad (3)$$

$$\hat{c}_t = \tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \quad (4)$$

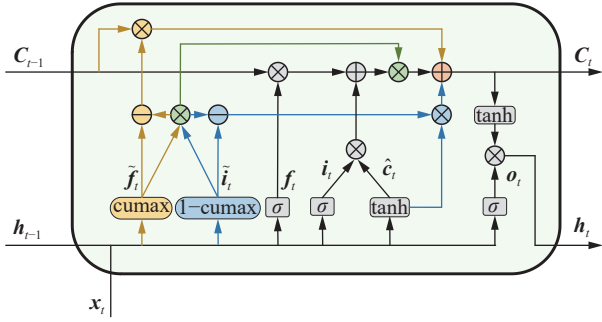


图 1 ONLSTM 网络结构

Fig. 1 ONLSTM network structure

式中: f_t 、 i_t 、 o_t 、 \hat{c}_t 分别为遗忘门、输入门、输出门与候选单元状态的输出; x_t 表示当前时刻的输入; h_{t-1} 表示前一时刻的隐含状态; W_{x*} 和 W_{h*} 分别表示输入权重与隐藏层权重; b_* 为对应的偏置; $\sigma(\cdot)$ 表示 Sigmoid 激活函数; $\tanh(\cdot)$ 表示双曲正切激活函数。LSTM 中隐含状态和单元状态的更新计算过程为

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{c}_t \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

式中: C_t 和 h_t 为更新后的单元状态和隐含状态, \odot 为向量点乘运算。ONLSTM 在此基础上引入了一个新的激活函数, 其新增参数更新计算过程为

$$\text{cumax}(\cdot) = \text{cumsum}(\text{Softmax}(\cdot)) \quad (7)$$

$$\tilde{f}_t = \text{cumax}(x_t W_{x\tilde{f}} + h_{t-1} W_{h\tilde{f}} + b_{\tilde{f}}) \quad (8)$$

$$\tilde{i}_t = 1 - \text{cumax}(x_t W_{x\tilde{i}} + h_{t-1} W_{h\tilde{i}} + b_{\tilde{i}}) \quad (9)$$

$$w_t = \tilde{f}_t \odot \tilde{i}_t \quad (10)$$

$$w_1 = \tilde{f}_t - w_t \quad (11)$$

$$w_2 = \tilde{i}_t - w_t \quad (12)$$

$$C_t = w_1 \odot C_{t-1} + w_t \odot (f_t \odot C_{t-1} + i_t \odot \hat{c}_t) + w_2 \odot \hat{c}_t \quad (13)$$

$$h_t = o_t \odot \tanh(C_t) \quad (14)$$

式中: $\text{cumax}(\cdot)$ 函数由 cumsum 和 Softmax 函数嵌套组成, 是 ONLSTM 新增主遗忘门的激活函数; $1 - \text{cumax}(\cdot)$ 为反向的 cumax 函数, 是 ONLSTM 新增主输入门的激活函数; w_t 表示 \tilde{f}_t 和 \tilde{i}_t 的信息重叠部分, 对应信息重叠通路; w_1 表示 \tilde{f}_t 中高于 w_t 的部分, 对应主遗忘通路; w_2 表示 \tilde{i}_t 中低于 w_t 的部分, 对应主输入通路。

ONLSTM 隐藏神经元顺序被定义为信息层级, 其中高层级神经元存储长期保持的长期信息, 而低层级神经元存储可以快速忘记的短期信息^[17]。式 (10)~(12) 对应的 3 条通路对信息流中的长期信息与短期信息进行划分提取, 式 (13) 表示 C_t 的更新逻辑, 直观地显示不同通路的融合计算过程。ONLSTM 可在不同层级上分别处理信息流向, 进而实现更好的模型预测效果。

2 非负绞杀算法

非负绞杀 (NNG) 算法是一种基于约束函数的嵌入式模型优化算法, 能将信息节点选择与权重系数的优化过程相结合, 最初被用以解决线性优化问题^[24]。相较于过滤式与包裹式的特征选择方法, NNG 算法将特征选择过程与模型训练过程相融合, 同时实现了效率与性能之间的有效权衡^[28]。NNG 算法的基本原理是在最小二乘算法的基础上添加一个约束函数, 构造一个带有限制条件的非线性函数, 然后求得该函数的极值以获取在该限制条件下参数的最优解。以线性模型为例:

$$y = \beta x + \varepsilon \quad (15)$$

式中: $x = [x_1 \ x_2 \ \dots \ x_p]$ 和 y 分别表示输入和输出变量, $\beta = [\beta_1 \ \beta_2 \ \dots \ \beta_p]$ 表示系数, p 为输入维度, ε 是随机误差。令 $X \in \mathbf{R}^{n \times p}$ 为输入矩阵, n 为样本总数; $Y \in \mathbf{R}^n$ 为输出矩阵。令 $\hat{\beta}$ 表示系数估计, 得到带有限制条件的非线性函数:

$$\begin{aligned} \theta^*(S) = \operatorname{argmin}_{\forall (x,y) \in (X,Y)} \{ & \sum \{ (y - (\theta \odot \hat{\beta})x_i)^2 \} \\ \text{s.t. } & \theta_i \geq 0, \sum_{i=1}^p \theta_i \leq S \end{aligned} \quad (16)$$

式中: S 为绞杀超参数; $\theta = [\theta_1 \ \theta_2 \ \dots \ \theta_p]$ 为收缩向量; θ_i 为对应的收缩系数。求解式 (16) 可得收缩向量 $\theta^* = [\theta_1^* \ \theta_2^* \ \dots \ \theta_p^*]$, 将 θ^* 代入式 (15) 即可获得更新后的系数:

$$\tilde{\beta} = \theta^* \odot \hat{\beta} \quad (17)$$

则优化后得到的输出为

$$\tilde{y} = \tilde{\beta} x \quad (18)$$

算法通过绞杀超参数对收缩向量进行约束, 以在约束条件下取得最优解。在寻优过程中, 绞杀超参数越小, 则约束越大, 其收缩系数更趋于 0。当收缩系数趋于 0 时, 则对应系数估计也将趋于 0, 即对应输入变量被删除。

3 DNNG-ONLSTM 算法设计

3.1 模型架构

尽管 ONLSTM 能更为有效地捕捉时序数据中的长期依赖信息。但其性能易受建模数据质量和其固有网络结构的影响。因此, 提出了一种基于非负绞杀的稀疏化 ONLSTM 软测量算法 (double NNG-ONLSTM, DNNG-ONLSTM), 该算法实现了模型输入变量选择和网络结构的同步优化, 解决了输入变量和隐藏神经元节点的冗余问题, 其架构如图 2 所示。

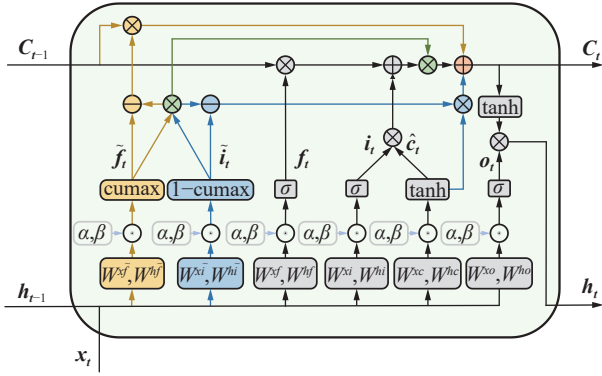


图 2 基于非负绞杀的稀疏化 ONLSTM 网络结构

Fig. 2 Structure of sparse ONLSTM network based on non-negative strangulation

DNNG-ONLSTM 算法的搭建主要通过网络预训练和非负绞杀收缩系数嵌入两个步骤实现。第 1 步通过超参数寻优获得训练结果良好的初始 ONLSTM 网络模型。第 2 步将非负绞杀的收缩系数依次分别嵌入 ONLSTM 输入和隐藏层权重矩阵中, 获得基于非负绞杀的稀疏化 ONLSTM 网络。其内部详细拓扑结构如图 3 所示。

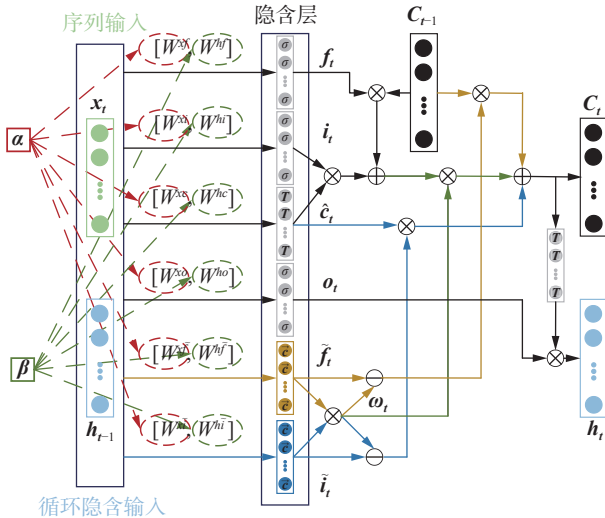


图 3 基于非负绞杀的稀疏化 ONLSTM 网络内部结构

Fig. 3 Internal structure of sparse ONLSTM network based on non-negative strangulation

3.2 基于 NNG 的变量选择

预训练阶段, 采用 Adam 算法训练 ONLSTM 网络, 并选取训练结果良好的模型作为初始模型。此时模型输出可以表示为

$$y_t = g(W_y(o_t \odot \tanh(C_t)) + b_y) \quad (19)$$

式中: $g(\cdot)$ 表示全连接层的激活函数; W_y 和 b_y 分别是全连接输出层的学习权重矩阵和偏差向量。

再训练阶段, 为避免耦合冗余数据影响, 将 NNG 算法收缩向量 α 嵌入 ONLSTM 输入权重矩阵, 实现输入变量选择并削弱变量耦合对模型预测性能的影响。此时算法输出更新为

$$y'_t = g(W_y(o'_t \odot \tanh(C_{t-1} \odot (f'_t \odot \omega'_t + \tilde{f}'_t - \omega'_t) + \hat{c}'_t \odot (i'_t \odot \omega'_t + \tilde{i}'_t - \omega'_t))) + b_y) \quad (20)$$

式中: 各个门的更新过程为

$$f'_t = \sigma \left(\sum_{i=1}^p x'_t (\alpha_i \odot W_{ik}^{xf}) + \sum_{j=1}^q h_j^{t-1} W_{jk}^{hf} + b_j^f \right) \quad (21)$$

$$i'_t = \sigma \left(\sum_{i=1}^p x'_t (\alpha_i \odot W_{ik}^{xi}) + \sum_{j=1}^q h_j^{t-1} W_{jk}^{hi} + b_j^i \right) \quad (22)$$

$$o'_t = \sigma \left(\sum_{i=1}^p x'_t (\alpha_i \odot W_{ik}^{xo}) + \sum_{j=1}^q h_j^{t-1} W_{jk}^{ho} + b_j^o \right) \quad (23)$$

$$\hat{c}'_t = \tanh \left(\sigma \left(\sum_{i=1}^p x'_t (\alpha_i \odot W_{ik}^{xc}) + \sum_{j=1}^q h_j^{t-1} W_{jk}^{hc} + b_j^c \right) \right) \quad (24)$$

$$\tilde{f}'_t = \text{cumax} \left(\sigma \left(\sum_{i=1}^p x'_t (\alpha_i \odot W_{ik}^{xf}) + \sum_{j=1}^q h_j^{t-1} W_{jk}^{hf} + b_j^f \right) \right) \quad (25)$$

$$\tilde{i}'_t = 1 -$$

$$\text{cumax} \left(\sigma \left(\sum_{i=1}^p x'_t (\alpha_i \odot W_{ik}^{xi}) + \sum_{j=1}^q h_j^{t-1} W_{jk}^{hi} + b_j^i \right) \right) \quad (26)$$

$$C'_t = C_{t-1} \odot (f'_t \odot \omega'_t + \tilde{f}'_t - \omega'_t) + \hat{c}'_t \odot (i'_t \odot \omega'_t + \tilde{i}'_t - \omega'_t) \quad (27)$$

式中: $k=1, 2, \dots, q$, p 和 q 分别为输入变量维度与模型隐藏层节点数。收缩系数 α_i 确定了输入变量 x_i 的收缩程度, 通过将其限制在 $[0, 1]$ 范围内增强了模型系数估计的可解释性。具体来说, $\alpha_i = 1$ 表示 x_i 完全保留, 而 $\alpha_i = 0$ 表示 x_i 被完全删除。最佳收缩向量 α 求解式:

$$\alpha = \underset{\forall (x,y) \in \{X,Y\}}{\text{argmin}} \sum [(y - g(W_y(o'_t \odot \tanh(C'_t)) + b_y))^2]$$

$$\text{s.t. } 0 \leq \alpha_i \leq 1, \sum_{i=1}^p \alpha_i \leq S_1 \quad (28)$$

显然, 式 (28) 是一个带有边界约束和线性约束的非线性不等式最优化问题。对于给定的绞杀超参数 S_1 , 可通过 COBYLA (constrained optimization by linear approximation) 算法计算模型在 S_1 条件下的最优解, 其通过构造目标函数和约束的连续线性逼近, 并在每一步的信赖区域中优化这些逼近进行求解。

在 NNG 对 ONLSTM 权重系数进行收缩时, 其通过 L1 约束迫使冗余特征对应权重趋近于 0, 从而剔除无效输入变量。该机制可有效降低输入维度与输入层参数规模, 在实现变量选择的同时显著减少模型计算量。式 (28) 将模型的预测输出嵌入 NNG 算法以扩展其应用范围至非线性最优求解问题, 能有效剔除无关变量, 稀疏模型结构。

将 α 代入式 (21)~(26) 可得到一组新的权重矩阵:

$$\tilde{W}_{i*}^{x*} = \alpha_i \odot W_{i*}^{x*}, \quad i = 1, 2, \dots, p \quad (29)$$

NNG 算法嵌入输入层强制其权重非负, 增强了输入特征可解释性, 有效缓解模型过拟合, 提升了模型泛化能力。

3.3 基于 NNG 的结构优化

除了建模数据, 模型本身的结构同样会影响模型性能。因此本文所提网络在进行变量选择的同时, 对 ONLSTM 结构进行稀疏优化。图 4 给出了模型的结构稀疏化过程。

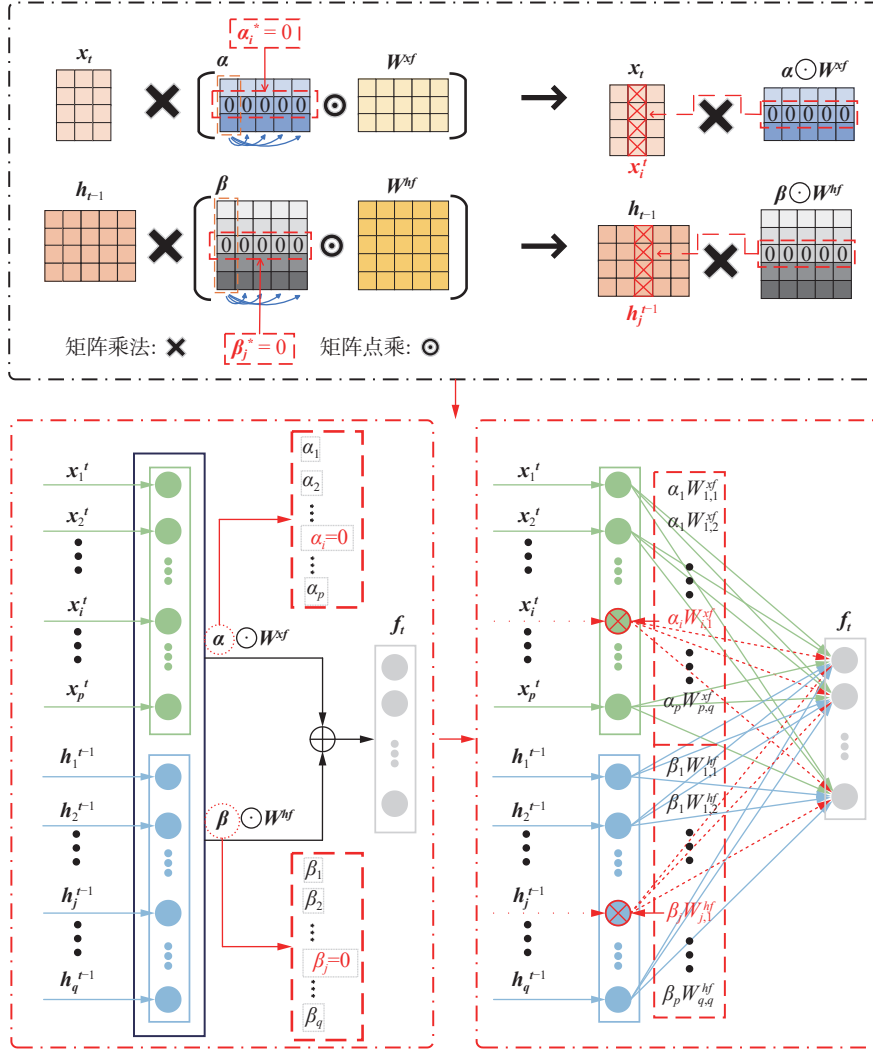


图 4 模型结构稀疏过程

Fig. 4 Model structure sparsity process

如图 4 所示, 算法将优化的收缩系数 α 与模型输入矩阵点乘, 绞杀系数为零的输入节点权重置零, 其所对应输入节点及其信息通路被删除, 实现变量选择。同时, 收缩系数 β 与模型隐藏层输入矩阵点乘, 将部分隐藏层节点的输入权重归零, 实现模型结构稀疏化, 更新后算法的前向计算过程为

$$y_t'' = g(W_y(o_t'' \odot \tanh(C_{t-1} \odot (f_t'' \odot \omega_t'' + \tilde{f}_t'' - \omega_t'') + \tilde{c}_t'' \odot (\tilde{i}_t'' \odot \omega_t'' + \tilde{j}_t'' - \omega_t'')))) + b_y) \quad (30)$$

此时, 各个门的更新过程为

$$f_t'' = \sigma \left(\sum_{i=1}^p x_i^t \tilde{W}_{ik}^{xf} + \sum_{j=1}^q h_j^{t-1} (\beta_j \odot W_{jk}^{hf}) + b_j^f \right) \quad (31)$$

$$\tilde{i}_t'' = \sigma \left(\sum_{i=1}^p x_i^t \tilde{W}_{ik}^{xi} + \sum_{j=1}^q h_j^{t-1} (\beta_j \odot W_{jk}^{hi}) + b_j^i \right) \quad (32)$$

$$o_t'' = \sigma \left(\sum_{i=1}^p x_i^t \tilde{W}_{ik}^{xo} + \sum_{j=1}^q h_j^{t-1} (\beta_j \odot W_{jk}^{ho}) + b_j^o \right) \quad (33)$$

$$\tilde{c}_t'' = \tanh \left(\sigma \left(\sum_{i=1}^p x_i^t \tilde{W}_{ik}^{xc} + \sum_{j=1}^q h_j^{t-1} (\beta_j \odot W_{jk}^{hc}) + b_j^c \right) \right) \quad (34)$$

$$\tilde{f}_t'' = \text{cumax} \left(\sigma \left(\sum_{i=1}^p x_i^t \tilde{W}_{ik}^{xf} + \sum_{j=1}^q h_j^{t-1} (\beta_j \odot W_{jk}^{hf}) + b_j^f \right) \right) \quad (35)$$

$$\tilde{j}_t'' = 1 - \text{cumax} \left(\sigma \left(\sum_{i=1}^p x_i^t \tilde{W}_{ik}^{xi} + \sum_{j=1}^q h_j^{t-1} (\beta_j \odot W_{jk}^{hi}) + b_j^i \right) \right) \quad (36)$$

$$\begin{aligned} C_i'' &= C_{i-1} \odot (f_i'' \odot w_i'' + \tilde{f}_i'' - w_i'') + \\ &\quad \tilde{c}_i'' \odot (\tilde{i}_i'' \odot w_i'' + \tilde{i}_i'' - w_i'') \end{aligned} \quad (37)$$

可获得最佳收缩向量 β :

$$\begin{aligned} \beta &= \operatorname{argmin}_{V(x,y) \in (X,Y)} \sum [(y - g(W_y(o_i'' \odot \tanh(C_i'') + b_y)))^2] \\ \text{s.t. } \beta_j &\geq 0, \sum_{j=1}^q \beta_j \leq S_2 \end{aligned} \quad (38)$$

式 (38) 将 NNG 算法嵌入模型的前向计算过程中, 其中 S_* 值的大小决定了算法约束程度。具体而言, 当 S_* 不小于对应节点数时, 算法的绞杀能力无效, 即所有待绞杀节点均可保留; 当 S_* 逐渐减小时, 算法的绞杀能力随之增强, 更多节点的系数将被压缩至 0, 实现网络的结构稀疏化处理。因此, 可以通过调整 S_* 值来改变算法的绞杀强度, 从而控制删除的冗余输入变量与冗余隐藏神经元的个数。通过网格搜索法选择最优绞杀超参数, 并采用 COBYLA 算法求解式 (38)。将 β 代入方程 (31)~(36) 可得到一组新的权重矩阵:

$$\tilde{W}_{j*}^{h*} = \beta_j \odot W_{j*}^{h*}, \quad j = 1, 2, \dots, q \quad (39)$$

$\beta_i = 0$ 表示从模型中删除递归隐藏神经元 h_j^{t-1} 和候选状态。NNG 算法嵌入隐藏层权重矩阵, 使神经元间时序传递方向明确化。通过压缩冗余循环连接, 剔除无效隐藏神经元, 降低时序迭代计算成本。同时, 强化关键时序神经元作用, 结合 ONLSTM 的有序神经元结构, 更精准捕捉工业过程多时间尺度滞后特性, 提升模型对时变特性的适配性。

3.4 算法流程

DNNG-ONLSTM 网络的建立过程分为两步。首先通过网格搜索法进行超参数寻优, 同时使用 Adam 算法^[29] 进行模型训练以获得预训练结果良好的 ONLSTM 网络, 利用其引入的层级概念对隐藏层神经元进行分级排序。其次引入两组 NNG 算法, 分别设置关于输入和隐藏神经元输出的约束函数, 将求解 NNG 算法得到的收缩向量 α 和 β 分别嵌入 ONLSTM 输入权重矩阵与隐藏层权重矩阵, 得到基于非负绞杀的稀疏化 ONLSTM 网络。NNG 算法求解得到两组收缩向量 α 和 β , 分别实现变量选择与结构优化功能, 从而实现模型的结构稀疏化。

DNNG-ONLSTM 预测模型算法流程描述如下。

1) 数据预处理: 对数据进行归一化、扩维等处理, 使其适应算法的输入要求; 将数据的前 80% 划分为训练集, 用以进行模型超参数调优和模型初始化训练; 后 20% 划分为测试集, 用以评估模型的预测性能。

2) 设置一系列的网络超参数组合, 在相同的实验环境下分别进行网络训练并重复多次, 以实现超参数寻优并生成初始 ONLSTM 网络;

3) 根据输入维度与模型隐藏节点数确定绞杀超参数的理论上限值, 在实际取值中往往取上限的 1/4~1/2, 根据网格搜索法首先以较大的网格间距初步筛选, 在初筛出的网格区间内进行更细致的筛选, 最终找出最优绞杀超参数 S_* ;

4) 根据寻得的最优绞杀超参数 S_* , 设置合适的收缩系数初始值, 分别求解非线性不等式约束优化表达式 (28) 和式 (38) 得出优化收缩向量;

5) 将求得的优化收缩向量嵌入预训练完成的 ONLSTM 网络权重矩阵, 得到基于非负绞杀的稀疏化 ONLSTM 网络;

6) 将测试集数据输入模型中得到预测结果, 并进行结果评估并与其他先进模型进行比较。

4 数值仿真验证与分析

4.1 数值仿真

为验证算法的有效性, 本节模拟复杂工业过程非线性、动态性及变量冗余等特点, 构建了一个基于弗里德曼数据集构建的数值示例^[30]。该数据集每个输入变量都是分布在 [0, 1] 的随机数。

$$\begin{aligned} X_1^t &= 0.4x_1^{t-1} + 0.6(x_2^{t-1})^2 + e_1^t, \\ X_2^t &= 0.8x_1^{t-1} - 0.3x_2^{t-2} + 0.5x_3^{t-3} + e_2^t, \\ X_3^t &= 0.2(x_3^t)^2 + 0.6x_4^{t-3} + e_3^t, \\ X_4^t &= 0.6x_4^{t-1} + 0.2x_3^{t-1} + 0.2x_3^{t-2} + e_4^t, \\ X_5^t &= 0.3x_5^t + 0.5x_5^{t-4} + 0.2(x_5^{t-9})^2 + e_5^t, \\ y &= 10 \sin(\pi X_1^t X_2^t) + 20(X_3^t - 0.5)^2 + 10X_4^t + 5X_5^t + e_y^t \end{aligned} \quad (40)$$

式中: x_* 和 y 分别为输入变量和输出变量, e_*^t 均为值为零、标准差为 0.01 的高斯噪声。仿真实验通过式 (40) 生成数据, 数据中共有 30 个输入变量与一个目标变量, 输入变量中仅有 5 个变量与目标变量相关, 共生成 2000 组样本数据。

4.2 模型性能评价指标

将仿真数据集以 4:1 的比例划分为训练集和测试集, 通过超参数调优搭建软测量模型, 并通过测试集验证模型性能。采用以下参数作为算法性能的评价指标。

1) 模型输入规模 (model input size, MIS): 模型有效输入变量的个数。其计算公式为

$$M_{IS} = \sum_{i=1}^n V_i, \quad V_i = \begin{cases} 1, & \lambda_i \geq 0.05 \\ 0, & \lambda_i < 0.05 \end{cases} \quad (41)$$

式中: V_i 代表输入变量 x_i 是否为模型有效输入,

n 为输入变量的个数, λ_i 为输入变量所对应的系数。

2) 正选率 (correct selection ratio, CSR): 模型有效输入变量中相关变量所占比例。其计算公式为

$$C_{SR} = \frac{n_r}{M_{IS}} \times 100\% \quad (42)$$

式中: n_r 为模型有效输入变量中相关变量的个数。

3) 决定系数 R^2 : 用以衡量模型输出与真实值的拟合程度, 表示模型对数据的拟合程度, 值越接近 1 表示模型拟合得越好。其计算公式为

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (43)$$

式中: n 为测试集样本总数, y_i 是测试集样本输出的实际值, \hat{y}_i 是模型的输出预测值, \bar{y} 是 y_i 的平均值。

4) 均方误差 (mean square error, MSE): 反映估计量与被估计量之间的差异程度, 是预测值与真实值之差的平方和的平均值。MSE 可以将模型预测的准确性进行量化, 从而判断模型的拟合程度。MSE 值越小, 表示模型的预测能力越强, 反之则说明模型存在较大的预测误差。其计算公式为

$$E_{MS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (44)$$

5) 平均绝对误差 (mean absolute error, MAE): 衡量预测值与真实值之间的差异, 是预测值与实际值之间偏差的绝对值之和的平均值。MAE 的值越小, 说明模型的预测结果越接近真实值, 预测性能越好。其计算公式为

$$E_{MA} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (45)$$

4.3 仿真实验设置与结果分析

为确保实验设计的严谨性, 本文在相同的数据集和仿真环境下, 对所提出的算法以及各对比算法进行多轮次重复运行。为更好地验证所提算法的优势, 将其与 LSTM、ONLSTM、STA-LSTM (spatiotemporal attention-based lstm)^[31]、SA-CNN-ONLSTM (self-attention-convolutional neural networks-onlstm)^[18] 算法的性能进行对比。为消除单次实验带来的偶然性影响, 结合多次实验的结果获取更为准确和可靠的评估, 选择统计结果中的最优值作为衡量模型性能的指标。表 1 是不同算法在多次实验后所得到的最优仿真结果汇总。

表 1 数值仿真实验结果

Table 1 Numerical simulation results

模型	MIS	CSR/%	MSE	MAE	R^2
LSTM	30	16.67	4.6843	1.6490	0.7071
ONLSTM	30	16.67	4.3624	1.6020	0.7282
STA-LSTM	30	16.67	4.3901	1.6280	0.7255
SA-CNN-ONLSTM	30	16.67	4.3731	1.6075	0.7266
NNG-ONLSTM	13	38.46	4.2312	1.6161	0.7354
DNNG-ONLSTM	13	38.46	3.8432	1.5116	0.7651

由表 1 结果可知, 在进行对比分析的 6 种算法中, 所提算法展示明显优势。具体而言, 该算法有着最小的模型输入规模 MIS, 这正体现了其在处理复杂系统时的精简性; 最高的正选率 CSR, 误差评价指标 MSE、MAE 与决定系数 R^2 优于其他算法, 充分验证了算法的有效性和优越性。

为更直观地评估 DNNG-ONLSTM 网络的性能, 图 5 和图 6 分别给出了不同算法数值仿真的预测结果散点图和预测误差箱线图。从图中可以看出, DNNG-ONLSTM 网络预测结果最优, 其误差波动最小, 预测精度最高。

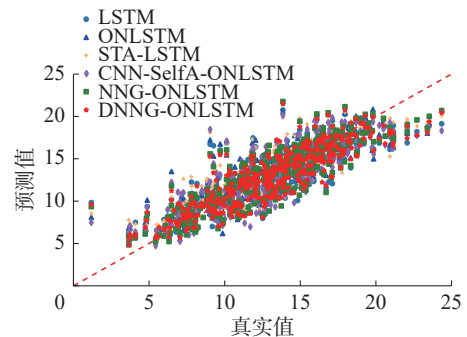


图 5 仿真预测结果散点图

Fig. 5 Scatter plot of simulation prediction results

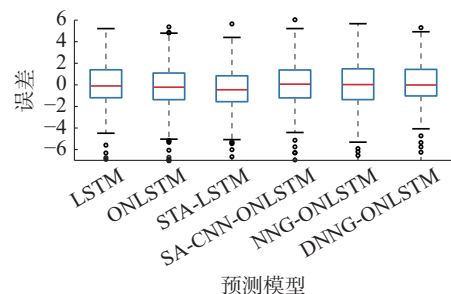


图 6 仿真预测结果误差箱线图

Fig. 6 Box plot of simulation prediction result errors

此外, 为量化分析非负熵杀与稀疏化对模型性能的优化提升, 图 7 将 DNNG-ONLSTM 算法的候选输入变量收缩系数进行统计并绘图, 其中前 5 个变量为相关变量。

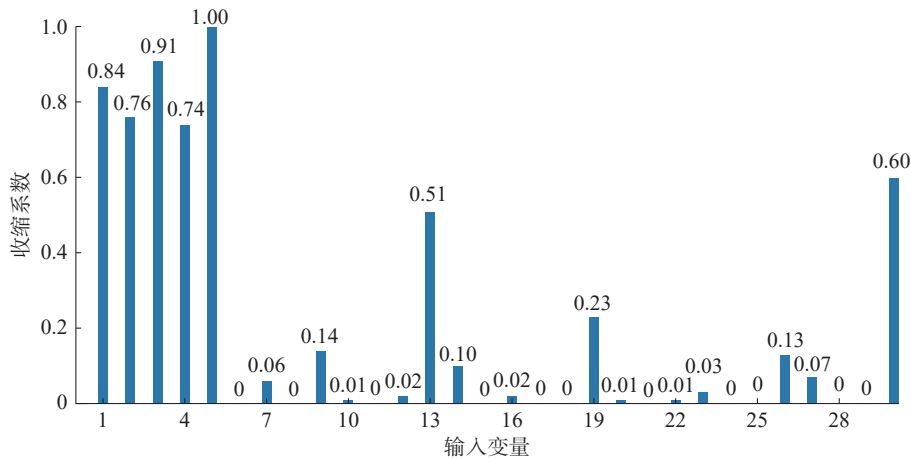


图 7 数值仿真候选输入变量收缩系数统计

Fig. 7 Numerical simulation candidate input variable contraction coefficient statistics

5 工业实例验证与分析

本节以某燃煤电厂烟气脱硫工艺过程的实际生产数据为研究对象, 将所提算法应用于脱硫过程排放的 SO₂ 浓度的软测量建模预测, 并与其他算法进行对比, 验证了所提算法的优越性。

5.1 研究对象

该燃煤电厂在烟气脱硫工艺过程中, 选用双

吸收塔结构和石灰石-石膏湿法技术, 这一技术以其反应迅速、脱硫高效以及成本低廉等特点而受到广泛采用, 是大规模火电厂烟气脱硫的主流工艺。其基本工作原理是基于石灰石浆液中的 CaCO₃ 与烟气中的 SO₂ 发生化学反应, 生成石膏, 以实现烟气脱硫的目标^[32]。完整的脱硫工艺系统由多个子系统组成, 如图 8 所示。其中, 烟气吸收系统是整个流程的核心, 其决定了脱硫效率的高低。

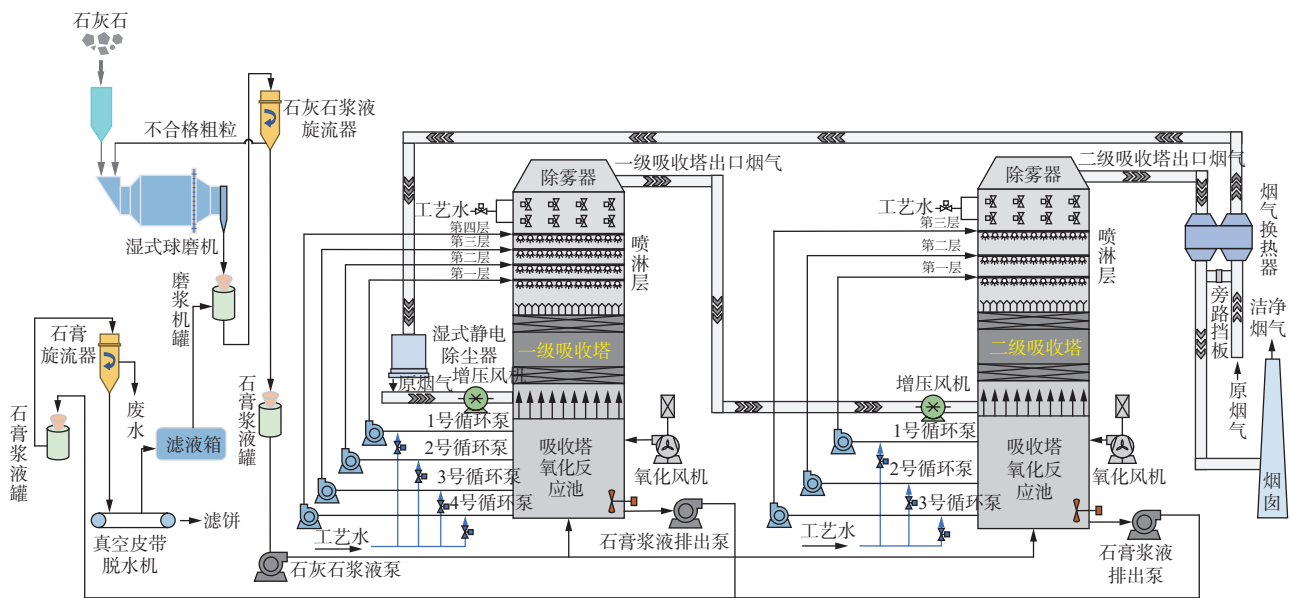


图 8 烟气脱硫系统工艺流程图

Fig. 8 Process flow diagram of flue gas desulfurization system

5.2 实验设置

通过对该燃煤电厂在烟气脱硫工艺过程的研究以及初步的数据筛选分析, 确定了由 30 个过程变量和 1 个质量变量组成的数据集。数据集包含 2 676 个样本, 包括炉原烟气 SO₂ 浓度、一级吸收塔出口烟气 SO₂ 浓度、出口烟气温度、机组脱硫效率中间值等 30 个过程变量以及质量变量排放的 SO₂ 浓度。

训练集和测试集数据分别取全部样本的前 80% 和后 20%。选取模型输入规模 MIS、均方误差 MSE、平均绝对误差 MAE 和决定系数 R² 作为模型性能评价指标。同时, 为进一步验证 DNNG-ONLSTM 网络的预测性能优势, 本文将其与 LSTM、ONLSTM、STA-LSTM 及 SA-CNN-ONLSTM 神经网络在相同的实验环境下的预测结果进行对比分析。

5.3 实验结果与分析

6种网络模型对应的4种仿真指标结果如表2所示。各网络的基本结构及超参数设置如表3所示。如表2所示,与LSTM和ONLSTM算法相比,引入注意力机制的STA-LSTM和SA-CNN-ONLSTM算法与引入变量选择机制的NNG-ONLSTM和DNNG-ONLSTM算法的MSE和MAE相对较小,决定系数 R^2 更接近1,表明对不同的输入变量

予以正确区分能有效提高模型预测精度。NNG-ONLSTM与DNNG-ONLSTM通过对输入变量进行合理地取舍,不仅有效地降低了无关变量对预测结果的影响,而且显著提升了网络的预测性能。其中, DNNG-ONLSTM算法所选择的关键生产参数最少,同时其性能评价指标更优,具体而言, MSE和MAE最低,决定系数也明显高于其他算法,这一结果充分验证了DNNG-ONLSTM算法的有效性。

表 2 SO₂ 浓度预测结果
Table 2 Prediction results of SO₂ concentration

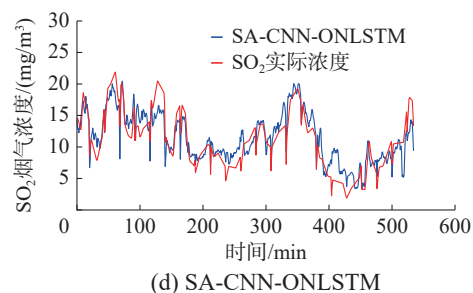
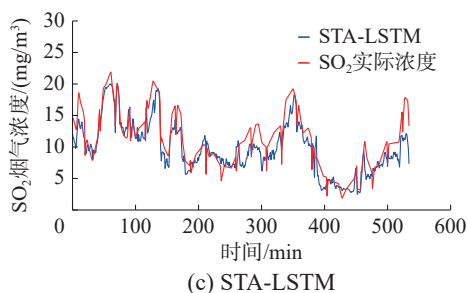
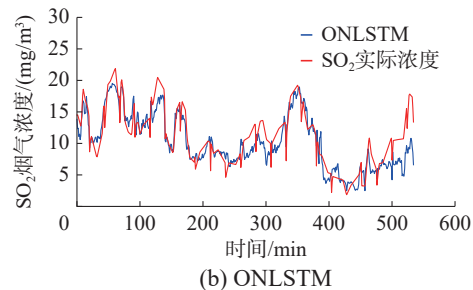
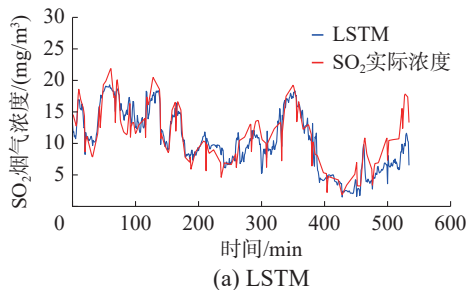
模型	MIS	MSE	MAE	R ²
LSTM	30	4.9280	1.7296	0.7591
ONLSTM	30	4.8116	1.6426	0.7648
STA-LSTM	30	4.8294	1.7529	0.7639
SA-CNN-ONLSTM	30	4.6217	1.7490	0.7741
NNG-ONLSTM	12	3.2953	1.4651	0.8389
DNNG-ONLSTM	12	3.2086	1.4600	0.8431

表 3 不同网络基本结构及超参数设置
Table 3 Different network basic structure and hyperparameter settings

参数名称	LSTM	STA-LSTM	SA-CNN-ONLSTM	ONLSTM	NNG-ONLSTM	DNNG-ONLSTM
输入维度	30	30	30	30	30	30
隐藏层数	1	1	1	1	1	1
隐藏节点数	36	36	36	36	36	36
初始学习率	0.01	0.001	0.005	0.005	0.01	0.01
时间步长	2	2	2	2	2	2
输出维度	1	1	1	1	1	1
最大迭代数	1700	1000	1700	1000	1700	1700

不同算法的预测性能有所差异,为更好地凸显这种差异,本文给出了不同算法的脱硫工艺预测结果曲线和预测误差频率分布直方图,分别如图9和图10所示。

由图9可以看出,相比于其他算法, DNNG-ONLSTM算法在数据拟合方面展现出最优性能,能更精确地捕捉SO₂浓度的细微波动和变化趋势。



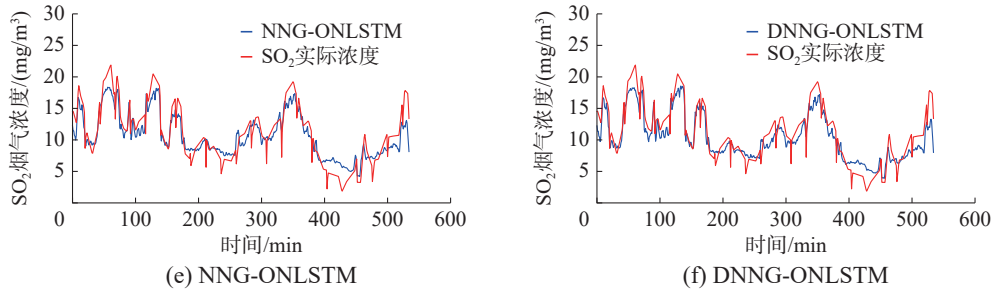


图 9 不同模型下脱硫工艺预测结果曲线

Fig. 9 Prediction curves of desulfurization process under different models

由图 10 可知, DNNG-ONLSTM 算法在预测误差分布方面展现出了卓越的稳定性, 其标准差 σ 值与其他算法相比最小, 表明该算法的预测结

果更加可靠和准确, 具有高预测精度和强稳定性, 有助于更好地实现对实际生产过程中脱硫工艺输出的实时监测与控制。

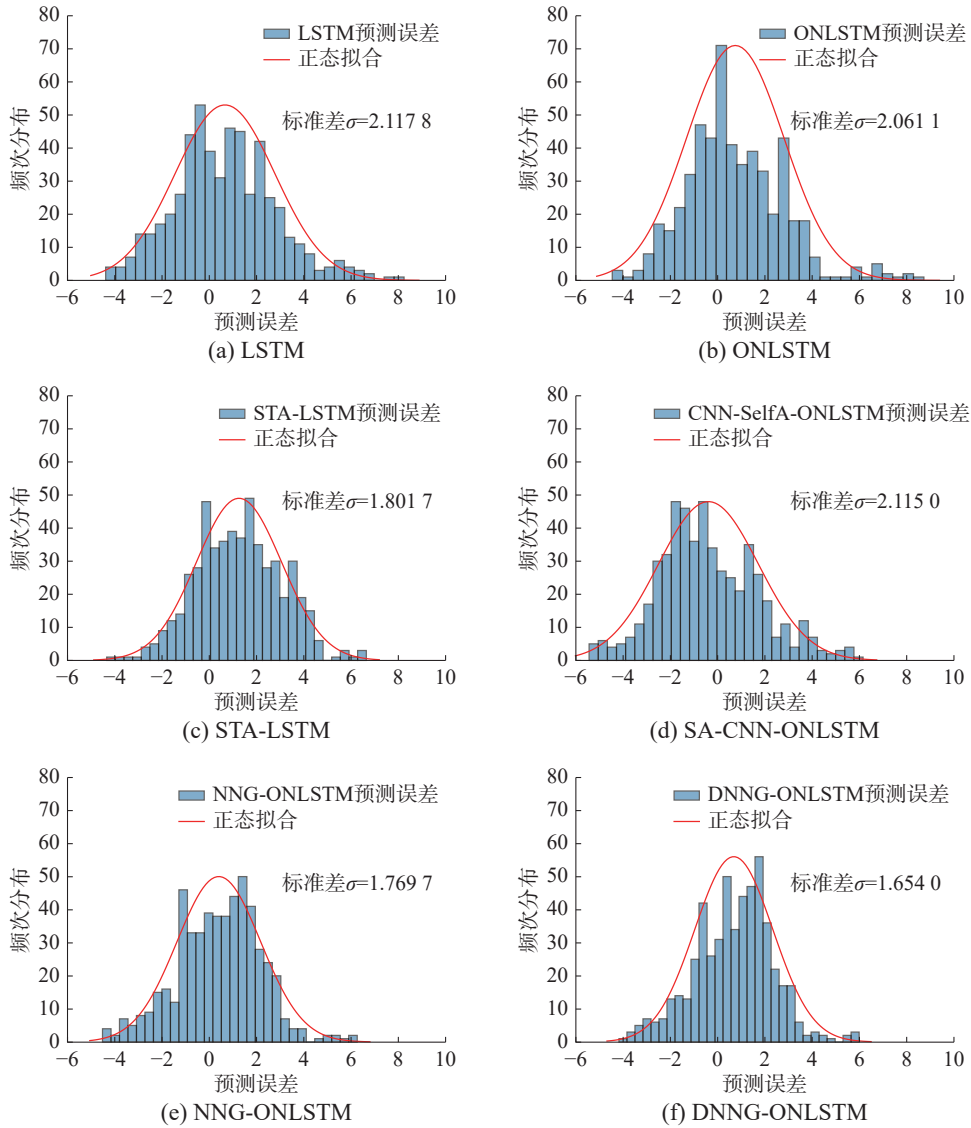


图 10 不同模型下脱硫工艺预测误差分布直方图

Fig. 10 Histogram of prediction error distribution of desulfurization process under different models

如表 4 所示, 模型训练时间与复杂度公式呈现一致性趋势, 尽管 NNG 算法增加了模型在训练阶段的计算量, 但由于模型训练为离线进行,

因此不会影响模型的实时检测性能。此外, 所提算法通过对模型结构进行稀疏优化, 提高了模型预测阶段的响应效率。从表中可以看出, DNNG-

ONLSTM 算法预测时间为 5.984 ms, 优于其他对比模型, 满足工业场景毫秒级响应需求。

表 4 模型复杂度与训练预测时间

Table 4 Model complexity and training prediction time

模型	模型训练复杂度	训练时间/s	预测时间/ms
LSTM	$O(4nh+h+4h^2)$	35.158	8.310
ONLSTM	$O(6nh+h+6h^2)$	29.264	9.983
STA-LSTM	$O(2n^2+4nh+2h+7h^2)$	39.645	13.631
SA-CNN-ONLSTM	$O(3n^2+r_1r_2+6nh+h+6h^2)$	50.438	23.554
NNG-ONLSTM	$O((6+k_1)nh+h+6h^2)$	63.513	6.489
DNNG-ONLSTM	$O((6+k_1)nh+h+(6+k_2)h^2)$	94.216	5.984

绞杀超参数决定了所提算法的绞杀强度, 为确定不同绞杀超参数对模型性能的影响。图 11 分别给出了不同绞杀超参数在变量选择与结构优化过程中对模型预测性能的影响。

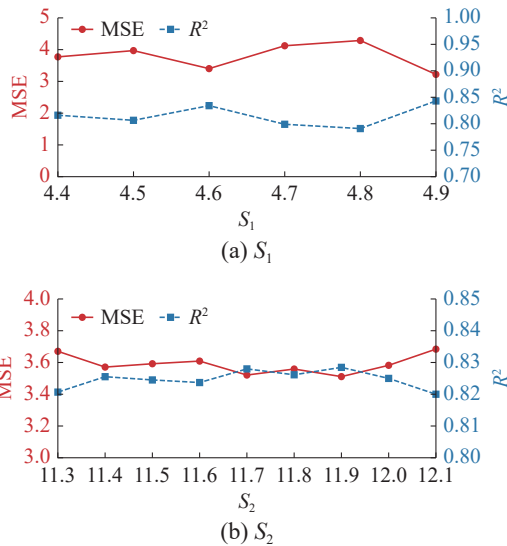


图 11 不同绞杀超参数下的模型性能

Fig. 11 Model performance under different strangulation hyper parameters

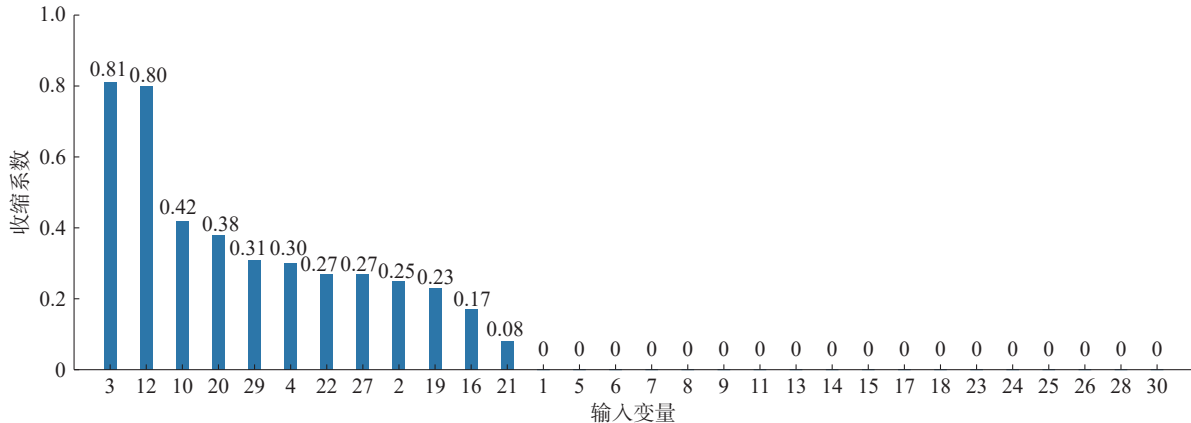


图 12 候选输入变量收缩系数统计

Fig. 12 Statistics of shrinkage coefficients of candidate input variables

图 11 中 S_1 和 S_2 分别对应变量的选择与结构优化的绞杀超参数。在变量选择过程中, 模型性能随绞杀超参数呈非单调波动, 当 S 过小时, 绞杀强度过大, 导致部分相关信息被删除; 当 S 过大时, 则会导致冗余信息保留, 影响模型性能。如图 11 所示, 在 $S_1=4.8$ 时, 模型性能最佳。在结构优化过程中, 当 $S_2=11.9$ 时, 相邻区间内性能波动幅度较小, R^2 稳定于 0.82~0.83, 体现模型的鲁棒性。

此外, DNNG-ONLSTM 算法所选择的输入变量与实际生产过程中的工艺机理和专家经验基本一致。本文根据不同候选输入变量的收缩系数大小, 对其进行了重要性排序(结果保留 2 位小数), 如图 12 所示。其中, θ 值大于 0.05 的变量有 12 个, 变量 3 和变量 12 分别为一级吸收塔出口处烟气 SO_2 浓度和总煤量, 两者所对应 θ 值均大于 0.8, 表明其对目标变量至关重要。这一结果与文献 [33-34] 中所分析的影响工艺输出的关键生产参数相一致。变量 2 为炉原烟气 SO_2 浓度, 变量 20 为机组脱硫效率中间值, 显然其直接与净烟气 SO_2 浓度相关。变量 4 为一级吸收塔石膏浆液 pH 值, 侧面反映了浆液对 SO_2 的吸收程度。变量 10 为吸收塔入口烟气温度, 变量 19 为炉原烟气 O_2 浓度, 变量 21 为二级吸收塔出口温度 1, 变量 22 为二级吸收塔出口压力, 均影响了脱硫过程中发生的化学反应, 与最终排放烟气中的 SO_2 浓度有所关联。

图 13 给出了算法各隐藏层神经元节点所对应的收缩系数的数值统计(保留 2 位小数), 结合图中数据与表 3 的算法预测结果可以得出, 所提算法可在保证模型预测精度的前提下对网络结构进行优化。

综上所述, 所提算法可以有效降低建模所需数据维度、提取关键输入变量, 并在保证预测精度的前提下对模型隐藏神经元进行稀疏优化。

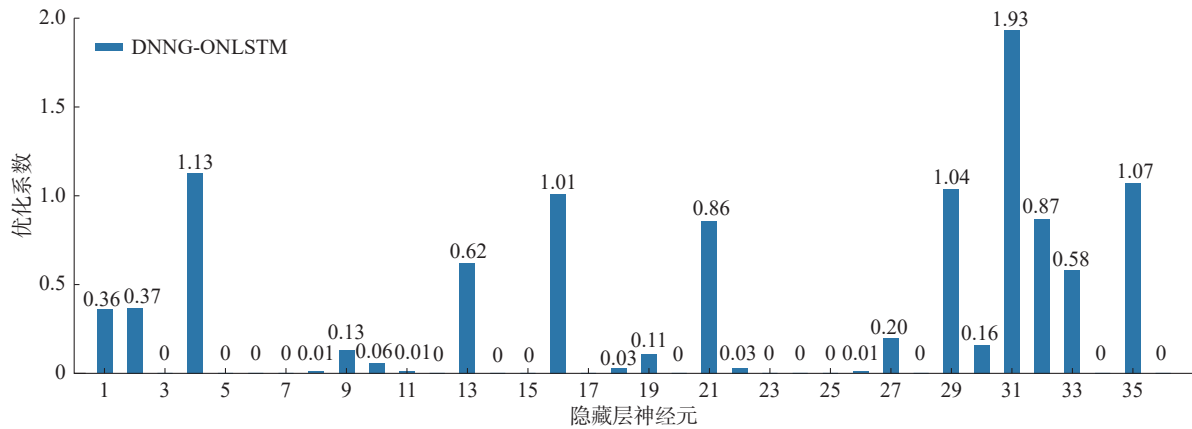


图 13 结构优化系数

Fig. 13 Structural optimization factor

6 结束语

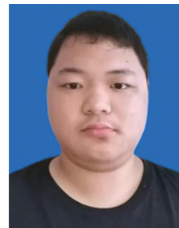
本文提出一种基于非负绞杀的稀疏化 ONLSTM 工业软测量算法, 以解决实际工业软测量建模数据冗余及模型复杂度增加的问题。所提算法通过剔除冗余输入变量与隐藏神经元, 以实现变量选择并降低模型复杂度, 在保证模型预测精度的同时实现其结构稀疏化, 使模型能够更高效地捕捉关键信息。通过数值仿真和某火电厂烟气脱硫过程实际生产数据的模型预测结果表明, 相较于其他算法, 所提算法能有效提高模型预测精度, 降低模型输入维度, 剔除模型中的冗余隐藏神经元。此外, 算法所选择的烟气脱硫过程中的相关生产参数与专家经验相符, 验证了所提算法在真实场景下的可行性, 表明算法可以为烟气脱硫过程的控制与监测提供有力的支持。

参考文献:

- [1] DAI Wei, ZHOU Xinyu, LI Depeng, et al. Hybrid parallel stochastic configuration networks for industrial data analytics[J]. *IEEE transactions on industrial informatics*, 2022, 18(4): 2331–2341.
- [2] YUAN Xiaofeng, WANG Yalin, YANG Chunhua, et al. Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes[J]. *IEEE transactions on industrial electronics*, 2018, 65(2): 1508–1517.
- [3] 杜康萍, 隋璘, 熊伟丽. 基于自适应稀疏宽度学习系统的软测量建模[J]. *系统仿真学报*, 2025, 37(6): 1449–1461.
- [4] DU Kangping, SUI Lin, XIONG Weili. Soft sensor modeling based on adaptive sparse broad learning system[J]. *Journal of system simulation*, 2025, 37(6): 1449–1461.
- [5] CHEN Hongtian, HUANG Biao. Fault-tolerant soft sensors for dynamic systems[J]. *IEEE transactions on control systems technology*, 2023, 31(6): 2805–2818.
- [6] YAO Le, GE Zhiqiang. Industrial big data modeling and monitoring framework for plant-wide processes[J]. *IEEE transactions on industrial informatics*, 2020, 17(9): 6399–6408.
- [7] YUAN Xiaofeng, OU Chen, WANG Yalin, et al. A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process[J]. *IEEE transactions on neural networks and learning systems*, 2019, 32(8): 3296–3305.
- [8] YAN Aijun, GUO Jingcheng, WANG Dianhui. Robust stochastic configuration networks for industrial data modelling with Student's-t mixture distribution[J]. *Information sciences*, 2022, 607: 493–505.
- [9] GAO Yunlong, LIN Tingting, PAN Jinyan, et al. Fuzzy sparse deviation regularized robust principal component analysis[J]. *IEEE transactions on image processing*, 2022, 31: 5645–5660.
- [10] GENG Zhiqiang, CHEN Zhiwei, MENG Qingchao, et al. Novel Transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes[J]. *IEEE transactions on industrial informatics*, 2022, 18(3): 1521–1529.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural computation*, 1997, 9(8): 1735–1780.
- [12] YUAN Xiaofeng, LI Lin, WANG Yalin. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network[J]. *IEEE transactions on industrial informatics*, 2019, 16(5): 3168–3176.
- [13] SUI Lin, SUN Wenxin, LIU Wentao, et al. A sparse regularized soft sensor based on GRU and self-interpretation double nonnegative garrote: from variable selection to structure optimization[J]. *Control engineering practice*, 2024, 153: 106074.
- [14] 张凯, 王亚礼, 张晓雯, 等. 基于特征融合的粗轧出口温度建模方法与应用[J/OL]. *控制理论与应用*, 2024, DOI: 10.7641/CTA.2024.40252.
- [15] ZHANG Kai, WANG Yali, ZHANG Xiaowen, et al. A feature fusion-based rough rolling exit temperature modeling method and its applications[J/OL]. *Control theory & applications*, 2024, DOI: 10.7641/CTA.2024.40252.
- [16] LIU Qingqiang, SHANG Zhiqing, LU Shixiang, et al. Physics-guided TL-LSTM network for early-stage degradation trajectory prediction of lithium-ion batteries[J]. *Journal of energy storage*, 2025, 106: 114736.
- [17] HOGEA E, ONCHIŞ D M, YAN Ruqiang, et al. LogicLSTM: logically-driven long short-term memory mod-

- el for fault diagnosis in gearboxes[J]. *Journal of manufacturing systems*, 2024, 77: 892–902.
- [16] YAN Haoran, QIN Yi, XIANG Sheng, et al. Long-term gear life prediction based on ordered neurons LSTM neural networks[J]. *Measurement*, 2020, 165: 108205.
- [17] SHEN Yikang, TAN Shawn, SORDONI A, et al. Ordered neurons: integrating tree structures into recurrent neural networks[C]//Proceedings of the 7th International Conference for Learning Representations. New Orleans: ICLR, 2019.
- [18] 李祥宇, 隋璘, 熊伟丽. 基于自注意力机制与卷积 ONLSTM 网络的软测量算法[J]. *智能系统学报*, 2023, 18(5): 957–965.
LI Xiangyu, SUI Lin, XIONG Weili. Soft sensor algorithm based on self-attention mechanism and convolutional ONLSTM network[J]. *CAAI transactions on intelligent systems*, 2023, 18(5): 957–965.
- [19] XIANG Sheng, QIN Yi, ZHU Caichao, et al. LSTM networks based on attention ordered neurons for gear remaining life prediction[J]. *ISA transactions*, 2020, 106: 343–354.
- [20] 李祥宇, 隋璘, 马君霞, 等. 基于时序迁移与双流加权的 ONLSTM 软测量建模[J]. *化工学报*, 2023, 74(11): 4622–4633.
LI Xiangyu, SUI Lin, MA Junxia, et al. ONLSTM soft sensor modeling based on time series transfer and dual stream weighting[J]. *CIESC journal*, 2023, 74(11): 4622–4633.
- [21] ZHANG Haoran, ZHAO Chunhui. Stable transfer learning-based control: an off-dynamics adaptive approach for unknown nonlinear systems[J]. *Neurocomputing*, 2025, 616: 128951.
- [22] CHIPLUNKAR R, HUANG Biao. Siamese neural network-based supervised slow feature extraction for soft sensor application[J]. *IEEE transactions on industrial electronics*, 2020, 68(9): 8953–8962.
- [23] 刘建伟, 崔立鹏, 刘泽宇, 等. 正则化稀疏模型[J]. *计算机学报*, 2015, 38(7): 1307–1325.
LIU Jianwei, CUI Lipeng, LIU Zeyu, et al. Survey on the regularized sparse models[J]. *Chinese journal of computers*, 2015, 38(7): 1307–1325.
- [24] YUAN Ming, LIN Yi. On the non-negative garrote estimator[J]. *Journal of the royal statistical society series B: statistical methodology*, 2007, 69(2): 143–161.
- [25] SUN Kai, LIU Jialin, KANG Jialin, et al. Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote[J]. *Journal of process control*, 2014, 24(7): 1068–1075.
- [26] WANG Jianguo, JANG Shishang, WONG D S H, et al. Soft-sensor development with adaptive variable selection using nonnegative garrote[J]. *Control engineering practice*, 2013, 21(9): 1157–1164.
- [27] 隋璘, 马君霞, 熊伟丽. 基于注意力绞杀的门控循环单元网络及其工业软测量应用[J/OL]. *控制理论与应用*, 2024, DOI: [10.7641/CTA.2024.305.65](https://doi.org/10.7641/CTA.2024.305.65).
SUI Lin, MA Junxia, XIONG Weili. Gated recurrent unit network based on attention garrote and its application for industrial soft sensors[J/OL]. *Control theory & applications*, 2024, DOI: [10.7641/CTA.2024.305.65](https://doi.org/10.7641/CTA.2024.305.65).
- [28] SUI Lin, SUN Kai, MA Junxia, et al. Input variable selection and structure optimization for LSTM-based soft sensor with a dual nonnegative garrote approach[J]. *IEEE transactions on instrumentation and measurement*, 2023, 72: 1–11.
- [29] KINGMA D P, BA J. Adam: a method for stochastic optimization[C]//Proceedings of the 3rd International Conference for Learning Representations. San Diego: ICLR, 2015.
- [30] FRIEDMAN J H. Multivariate adaptive regression splines [J]. *The annals of statistics*, 1991, 19(1): 1–67.
- [31] YUAN Xiaofeng, LI Lin, SHARDT Y A W, et al. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development[J]. *IEEE transactions on industrial electronics*, 2020, 68(5): 4404–4414.
- [32] SUN Kai, WU Xiuliang, XUE Jingyu, et al. Development of a new multi-layer perceptron based soft sensor for SO₂ emissions in power plant[J]. *Journal of process control*, 2019, 84: 182–191.
- [33] 周祖飞, 金新荣. 影响湿法烟气脱硫效率的因素分析[J]. *浙江电力*, 2001, 20(3): 42–45.
ZHOU Zufei, JIN Xinrong. Analysis of factors affecting the efficiency of wet flue gas desulfurization[J]. *Zhejiang electric power*, 2001, 20(3): 42–45.
- [34] 钟毅, 高翔, 骆仲泐, 等. 湿法烟气脱硫系统脱硫效率的影响因素[J]. *浙江大学学报: 工学版*, 2008, 42(5): 890–894.
ZHONG Yi, GAO Xiang, LUO Zhongyang, et al. Factors influencing desulfurization efficiency of wet flue gas desulfurization system[J]. *Journal of Zhejiang University(engineering science)*, 2008, 42(5): 890–894.

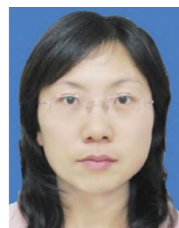
作者简介:



郭迎宸, 硕士研究生, 主要研究方向为机器学习、软测量建模。E-mail: guoyingchen25@163.com。



隋璘, 博士研究生, 主要研究方向为机器学习、软测量建模。E-mail: suilin359@163.com。



熊伟丽, 教授, 博士生导师, 主要研究方向为智能软测量技术、过程监测。主持国家自然科学基金面上项目、国家自然科学基金青年项目、江苏省产学研等省部级以上项目 15 项; 获得江苏省科学技术奖二等奖 1 项; 发表学术论文近百篇, 获发明专利授权 26 项, 其中国际专利 3 项。E-mail: greenpre@163.com。