

面向空气分离过程的动态集成增量学习软测量算法研究

吴修粮¹, 赵磊², 曹茂永^{1,2}, 孙凯^{2†}

(1. 山东科技大学 电气与自动化工程学院, 山东 青岛 266590; 2. 齐鲁工业大学(山东省科学院) 信息与自动化学院, 山东 济南 250353)

摘要: 针对空气分离过程多模态、多变量、大时延及存在测量离群点等问题, 本文提出了一种基于Lipschitz循环神经网络(LRNN)的动态增量学习软测量算法。首先, 利用易收敛、稳定性好的LRNN处理过程数据中存在的非线性和时延问题。其次, 设计了一种新颖的LRNN损失函数, 该函数结合了Log-Cosh估计对离群点的鲁棒性和L₁正则化的模型稀疏特性, 能够有效处理过程的测量离群点和模型冗余问题。进一步, 对不同鲁棒LRNN基础模型进行动态加权集成, 提高了模型对生产过程模态变化的适应性。最后, 将所提出的算法应用于某钢厂实际空气分离过程中低压塔出口O₂浓度的软测量建模, 并与其他先进算法进行比较, 验证了所提出方法的有效性和优越性。

关键词: 循环神经网络; 空气分离过程; Log-Cosh估计; L₁正则化; 增量学习; 软测量

引用格式: 吴修粮, 赵磊, 曹茂永, 等. 面向空气分离过程的动态集成增量学习软测量算法研究. 控制理论与应用, 2026, 43(4): 843–852

DOI: 10.7641/CTA.2024.40246

Research on dynamic ensemble incremental learning soft sensor algorithms for an air separation process

WU Xiu-liang¹, ZHAO Lei², CAO Mao-yong^{1,2}, SUN Kai^{2†}

(1. College of Electrical Engineering and Automation,

Shandong University of Science and Technology, Qingdao Shandong 266590, China;

2. School of Information and Automation,

Qilu University of Technology (Shandong Academy of Sciences), Jinan Shandong 250353, China)

Abstract: To address the problems of multi-modality, multi-variable, large time delay and existence of measurement outliers in an actual air separation process, a dynamic incremental learning soft sensor algorithm based on a Lipschitz recurrent neural network (LRNN) is proposed. Firstly, the LRNN with easy convergence and excellent stability is used to deal with the nonlinearity and time delay in the process data. Secondly, a novel LRNN loss function is designed by combining the robustness to outliers of the Log-Cosh estimation and the sparsity of the L₁ regularization, which is effective in dealing with the problems of measurement outliers and model redundancy of the process. Further, the adaptability to the modal changes of the production process is improved by the dynamic weighted ensemble of different robust LRNN base models. Finally, the proposed algorithm is applied to the soft sensor modeling of the O₂ concentration at the outlet of the low-pressure tower in an actual air separation process in a steel plant, and the effectiveness and superiority of the proposed algorithm is verified by comparing with other state-of-the-art algorithms.

Key words: recurrent neural networks; air separation process; Log-Cosh estimation; L₁ regularization; incremental learning; soft sensor

Citation: WU Xiuliang, ZHAO Lei, CAO Maoyong, et al. Research on dynamic ensemble incremental learning soft sensor algorithms for an air separation process. *Control Theory & Applications*, 2026, 43(4): 843–852

1 引言

空气分离装置(air separation unit, ASU)是一类可对空气进行液化、分离和精馏的设备, 能够提供化工、能源和冶金等行业所必需的氧气、氮气和其他有用气

体。该系统中, 低压塔出口氧气浓度作为关键性能指标, 能够反映过程分离效率和运行状态, 影响产品质量和生产安全, 对该变量的实时监测和控制具有重要意义^[1]。然而, 现有的在线分析仪存在成本高、易受环

收稿日期: 2024-04-29; 录用日期: 2024-10-04.

[†]通信作者. E-mail: sunkai79@qlu.edu.cn; Tel.: +86 15269190537.

本文责任编辑: 阳春华.

山东省自然科学基金项目(ZR2021MF022), 山东省科技型中小企业创新能力提升工程项目(2023TSGC0399)资助.

Supported by the Shandong Provincial Natural Science Foundation of China (ZR2021MF022) and the Shandong Provincial Innovation Capability Enhancement Engineering Project of Technology-Based Small and Medium-Sized Enterprises (2023TSGC0399).

境影响、需定期维护等问题,而利用生产过程数据构建数学模型的软测量技术具有成本低、响应快、可靠性高等优点,已在流程工业关键参数实时估计方面得到了广泛研究^[2-3]。

ASU生产过程复杂,具有典型的非线性特性,人工神经网络(artificial neural network, ANN)因其在处理未知机理的非线性过程建模上的优势,已在ASU关键参数软测量方面得到了成功的应用^[4]。ANN按其结构是否包含反馈单元,可分为前馈神经网络和循环神经网络(recurrent neural network, RNN)^[5]。前馈神经网络的常见模型有多层感知机(multi-layer perception, MLP)^[6]、径向基函数^[7]和随机配置网络(stochastic configuration network, SCN)^[8]等;RNN则包括长短期记忆(long short-term memory, LSTM)^[9]、门控循环单元(gate recurrent unit, GRU)^[10]等多种结构。前馈神经网络如MLP, SCN缺少记忆功能,当对具有动态时序关系的过程建模时,其模型性能较差。循环神经网络如RNN, LSTM引入了记忆单元,能够捕捉过程的动态依赖关系,更适用于动态过程建模。

ASU系统另一特点是其管道较长,存在较大的惯性和时间延迟,前馈神经网络难以捕捉带未知时滞的过程信息,其模型精度不足。RNN作为一类具有动态特性的神经网络,能够有效地捕捉工业过程的时序信息,因此在时间序列建模方面具有一定优势^[11]。然而,传统RNN对长期依赖关系的建模能力较弱,易出现梯度消失和爆炸问题。其改进模型如LSTM和GRU虽在一定程度上缓解了这些问题,但引入了门控机制和额外的神经元,增加了模型复杂性。

2018年,Chen等^[12]将连续时间动力学引入神经网络,通过常微分方程求解隐藏状态的变化,该网络在处理不规则采样数据和复杂动态系统建模时具有明显优势。在此基础上,Erichson等^[13]将RNN作为具有输入的连续时间动态系统,提出了一种新的Lipschitz递归网络,即Lipschitz RNN(LRNN)。与其他连续时间单元相比,LRNN模型在面对输入和参数扰动时具有更好的稳定性和抗干扰性,从而成功应用于自然语言处理^[14]、图像识别^[15]等领域。作为新型连续时间动态神经网络,LRNN为复杂动态过程建模提供了新的可能,但其理论体系仍不够完善,针对其鲁棒估计、正则化技术的研究较少,当面对非理想数据、多模态等问题时,模型会出现泛化能力不足、训练不稳定以及预测精度下降等问题。

在实际生产中,设备和环境的波动会导致过程数据分布特性的变化,从而导致软测量模型性能退化^[16]。例如,压缩机是ASU的关键设备,可将环境气体压缩至所需的高压水平,然而长时间运行和高压操作会使其内部活塞、气缸和密封件磨损,从而引起轻

度泄漏及工作条件变化。同时这些部件位于压缩机内部,无法直接观察磨损情况。为应对此类工况变化,需通过新样本对已建立模型进行更新调整,以提高模型的适应性。已有的软测量模型更新方法主要包括滑动窗口更新、递归更新和增量学习(incremental learning, IL)等^[17-19]。文献[20]提出了一种改进的增量式深度学习的方法,用于催化裂化装置吸收稳定系统的在线质量预测。文献[21]提出了对不同工况具有识别能力的半监督IL模型,用于解决发电机转子削刮过程刀片破损监测。然而,上述文献主要研究前馈神经网络的模型IL策略,当处理多工况时间序列数据驱动的循环神经网络模型连续学习问题时,上述方法无法解决。

此外,ASU的过程自动化系统提供了大量数据,为软测量建模提供了便利,但也带来了大量的冗余信息,使模型复杂度增加、性能降低。 L_1 正则化通过向损失函数引入绝对值惩罚来简化复杂模型,可缓解ANN的过拟合问题^[22]。另一方面,由于设备故障、传感器漂移等异常干扰,测量数据会存在一些离群点,也会影响软测量模型的性能。为解决此问题,可利用各种鲁棒损失函数如Huber M-Loss^[23]、最大相关性准则^[24]和Log-Cosh^[25],提高模型的鲁棒性。其中,Log-Cosh损失函数是可微和平滑的,对带离群点的回归问题具有良好的性能。

本研究针对ASU过程软测量所面临的多模态、多变量、大时延及存在测量离群点等难题,提出了一种基于鲁棒LRNN的动态加权集成软测量算法,具体创新点如下:

- 1) 针对过程数据中存在多变量、时滞和测量离群点的问题,以LRNN为基础模型,融合Log-Cosh损失函数和 L_1 正则化技术,提高了模型的鲁棒性和泛化能力;
- 2) 针对过程的多模态特性,设计了一种新的动态加权集成IL策略,对不同的鲁棒LRNN模型进行动态加权,克服了过程模态变化时造成的模型性能下降问题,提高了模型的稳定性;
- 3) 将所提出的软测量算法应用于某钢厂实际ASU生产过程,预测关键性能参数,为过程监测、先进控制策略实施提供了坚实的模型基础。

2 研究对象

2.1 ASU工艺描述

ASU过程采用先进的物理分离技术,从空气中分离 O_2 , N_2 和Ar等高纯度产品,主要由蒸馏塔、主热交换器(main heat exchanger, MHE)、冷凝器和辅助设备组成,如图1所示。

过程开始时,压缩空气首先进入MHE调节温度和压力,然后进入加压塔。在加压塔中,从塔顶部抽出压

缩气体中的 N_2 , 并将其分成两部分: 一部分经过主冷凝器和次冷凝器的液化处理, 形成液态 N_2 产品; 而另一部分则通过MHE, 保持气态成为气态 N_2 产品. 加压塔底部的气体经过次冷凝器, 在粗氩塔2和纯氩塔中进行冷却. 随后, 冷却后的气体被送至低压塔, 从中抽出剩余的 N_2 , 并在底部析出液态 O_2 , 经泵送至MHE加

压, 分别得到 32 kg/cm^2 的高压 O_2 产品和 12 kg/cm^2 的中压 O_2 产品. 气态Ar经粗氩塔1, 2和纯氩塔提纯并压缩至液态, 以方便储存和运输. 最后, 从Ar罐中提取的液态Ar, 经MHE汽化降温, 通过控制液态Ar的汽化速率可以稳定系统的热量传递过程, 降低系统的能耗和温度波动, 从而提高压缩空气的效率.

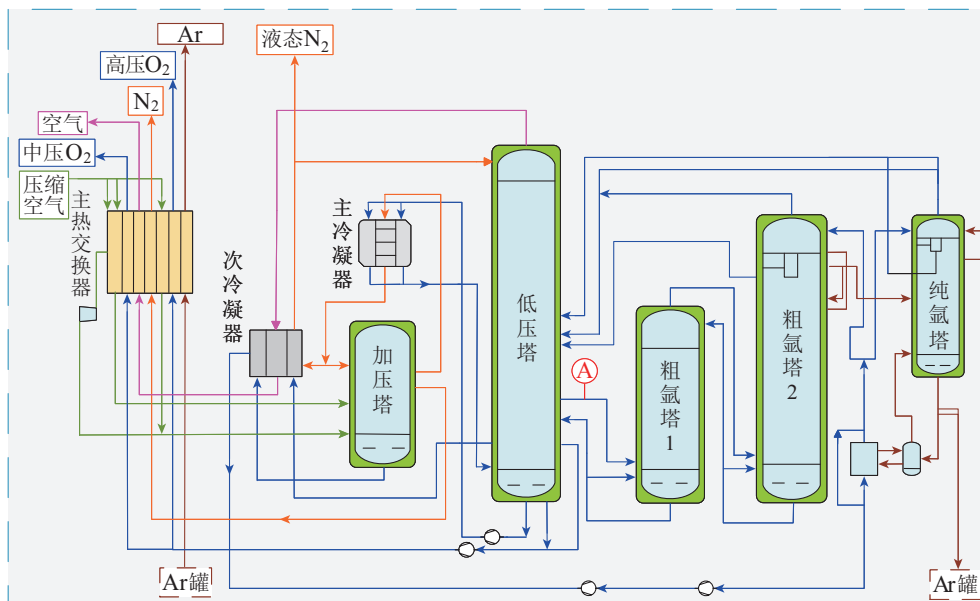


图1 空气分离过程示意图

Fig. 1 Schematic diagram of air separation process

实际生产中, 当 O_2 浓度过低时, 粗氩塔1的入口处将聚集大量 N_2 , 导致粗氩塔2顶部压强升高, 容易发生倒塔; O_2 浓度过高会影响Ar的有效分离和提纯, 导致产量减少. 因此, 实时监控 O_2 浓度并将其控制在合理的范围内对于生产至关重要. 图1中A为安装在低压塔出口处的某款在线分析仪, 可对 O_2 浓度进行监测. 然而, 该分析仪对环境敏感、稳定性较差, 有时会出现故障, 因此亟需建立有效的软测量模型以验证分析仪测量结果并进行校准, 从而确保该参数测量的准确性和稳定性.

2.2 建模数据分析

本研究数据来自某钢集团实际ASU生产过程, 以每隔5 min一次的频率收集了6300个连续观测样本. 根据工艺机理分析和现场测量条件, 选取了30个输入变量, 名称及单位如表1所示, 低压塔的出口 O_2 浓度作为目标变量.

最大信息系数(maximal information coefficient, MIC)是一种检测变量之间非线性关系, 且对离群点和噪声鲁棒性较强的数据分析方法^[26]. ASU过程不同变量间的MIC热力图如图2所示, 其中第31列表示目标变量. 显然, 输入变量之间存在显著的互相关性, 即输入数据中存在明显的冗余, 这将降低模型性能, 因此需引入正则化技术以降低模型的冗余.

为识别不同的生产模态, 本研究采用高斯聚类的方法, 通过拟合多个高斯分布来描述数据分布情况^[27]. 由图2可知, 共有10个输入变量与目标变量相关系数大于0.5, 存在显著相关. 为便于模态分类的可视化, 将这10个变量中的每一对进行高斯聚类. 在尝试了所有变量组合后, 变量17和18组合的数据分布反映生产过程存在明显不同的模态, 如图3所示. 在图3中, 不同颜色表示不同的数据分布模式, 可以看出, 样本的数据分布与采样顺序有明显的相关性, 反映了过程模态的时变特性. 另外, 由专家经验可知, 高压氧气产品流量和中压氧气产品流量对ASU的运行和氧气生产具有关键作用, 通过调节它们可以控制氧气产量、调节分离效率和平衡操作参数, 因此这两个变量组合能够反映不同的生产模态.

进一步利用异常检测工具包^[23]对历史数据进行离群点检测. 该方法使用滑动窗口遍历历史数据, 利用当前数据与窗口中统计值的残差生成新序列. 若该残差超出新序列箱线图的边界, 则该数据为异常. 检测结果如图4所示, 可以看出数据中存在一些显著的离群点.

3 基于DIL-RLRNN的软测量建模

本节首先介绍所提出算法的LRNN基础模型和

Log-Cosh损失函数的基本原理, 然后详细介绍动态集成增量学习的鲁棒 Lipschitz 循环神经网络 (dynamic incremental learning robust Lipschitz recurrent neural network, DIL-RLRNN)算法的具体原理和算法流程.

表 1 ASU 的输入变量
Table 1 ASU input variables

序号	变量名称	单位
1	N ₂ 产品流量设定值	Nm ³ /h
2	冷冻水流量	m ³ /h
3	压缩空气流量	Nm ³ /h
4	加压塔液位	%
5	加压塔O ₂ 浓度	%
6	液N ₂ 产品流量	Nm ³ /h
7	经过低压塔的液态空气流量	Nm ³ /h
8	回流至低压塔的液N ₂ 流量	Nm ³ /h
9	低压塔压力	mmWC
10	经过涡轮机的流量	Nm ³ /h
11	第1阶段O ₂ 压缩机排气压强	kg/cm ²
12	主热交换器的液O ₂ 流量	Nm ³ /h
13	第2阶段O ₂ 压缩机排气压强	kg/cm ²
14	再生气体流量	Nm ³ /h
15	N ₂ 产品流量	Nm ³ /h
16	粗Ar塔流量	Nm ³ /h
17	中压O ₂ 产品流量	Nm ³ /h
18	高压O ₂ 产品流量	Nm ³ /h
19	粗氩塔1液位	%
20	粗氩塔1压差	mmWC
21	粗氩塔2液位	%
22	粗氩冷凝器吹扫	%
23	纯氩冷凝器压强	kg/cm ²
24	纯氩塔压强	kg/cm ²
25	低压塔出口温度	°C
26	回流至低压塔的液N ₂ 流量	Nm ³ /h
27	中间变量1	—
28	低压塔压差	mmWC
29	低压塔液位	%
30	中间变量2	—

3.1 Lipschitz RNN基本原理

LRNN将标准RNN视为连续时间动力系统, 提出了一种描述隐藏状态随时间演化的递归单元, 该单元采用了对称-斜对称分解方案来构建隐藏层之间的矩阵, 解决了梯度消失和梯度爆炸的问题. 模型由一个已知的线性分量和一个Lipschitz非线性分量组成, 这种特殊的结构保证了信息长期传递的稳定性. 假设 N 和 p 分别代表输入数据集的大小和维度, 输入数据集 $X = [x_1 \ x_2 \ \cdots \ x_N]^T$, 其中 $x_i = [x_{i1} \ x_{i2} \ \cdots \ x_{ip}]$, $i = 1, 2, \dots, N$. LRNN的前向计算公式表示如下:

$$\dot{h} = A_{\beta_A, \gamma_A} h + \tanh(W_{\beta_W, \gamma_W} h + Ux^T + b), \quad (1)$$

$$f(x) = D^T h, \quad (2)$$

其中: 隐藏层的矩阵 $A_{\beta_A, \gamma_A} \in \mathbb{R}^{d \times d}$ 和 $W_{\beta_W, \gamma_W} \in \mathbb{R}^{d \times d}$ 表示为

$$A_{\beta_A, \gamma_A} = (1 - \beta_A)(M_A + M_A^T) + \beta_A(M_A - M_A^T) - \gamma_A I, \quad (3)$$

$$W_{\beta_W, \gamma_W} = (1 - \beta_W)(M_W + M_W^T) + \beta_W(M_W - M_W^T) - \gamma_W I, \quad (4)$$

式中: $\beta_A, \beta_W \in [0, 1]$; $\gamma_A, \gamma_W > 0$; $M_A, M_W \in \mathbb{R}^{d \times d}$ 分别为可调与可训练参数; 内部隐藏状态 $h = h(t) \in \mathbb{R}^d$, $\dot{h} = \frac{\partial h(t)}{\partial t}$ 为其时间导数; $U \in \mathbb{R}^{d \times p}$ 和 $b \in \mathbb{R}^d$ 分别表示输入权重矩阵和偏置向量; $D \in \mathbb{R}^d$ 表示最终回归系数向量. 具体结构如图5所示.

基于其循环单元的特殊结构, 通过稳定性分析和 Hessian 矩阵评估, 发现 Lipschitz 循环单元对于输入和参数扰动具有鲁棒性, 与其他连续时间单元相比表现出更好的稳定性和抗干扰性, 使 LRNN 适用于处理复杂工业过程中的长期和短期时序依赖关系.

3.2 Log-Cosh损失函数基本原理

Log-Cosh函数作为一种鲁棒性回归的损失函数, 具有基于阈值分段 Huber 函数的所有优点, 但 Log-Cosh 以平滑连续的方式组合了不同幂次的函数, 其函数表达式为

$$L(x) = \log(\cosh x), \quad (5)$$

其中: $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ 表示双曲余弦函数, $\log(\cdot)$ 表示对数函数, x 表示真实值与预测值之差. 式 (5) 中双曲余弦函数对较大残差处的损失较小, 近似等于 $|x - \log(2)|$, 能够降低 Log-Cosh 损失对离群点的敏感度. 当残差较小时, 可使模型更快地收敛, 与均方误差 (mean square error, MSE) 具有相似效果, 但具有更稳定的梯度计算, 避免了计算复杂度指数级增长问题, 因此降低了梯度爆炸的风险. 此外, Log-Cosh 函数具有可微性和凸性, 基于该函数的神经网络模型训练可获得全局最优解.

3.3 鲁棒 LRNN 算法设计

LRNN 可有效地捕捉动态过程时序信息, 但其在模型训练过程中, 若样本存在离群点, 会增加过拟合和陷入局部最优的风险. 为增强模型对离群点的抵抗力, 采用 Log-Cosh 函数取代原有 LRNN 的 MSE 损失函数. L_1 正则化具有特征选择的能力, 可以将不重要的特征系数压缩到零, 从而得到稀疏模型, 减少计算成本和存储需求, 并能提高模型的泛化能力^[22]. 因此, 算法引入 L_1 正则化以减少冗余特征对模型的影响. 综上, 所提出的鲁棒 LRNN (robust LRNN, RLRNN) 损失函数如下所示:

$$L_\varepsilon(y, f(x)) = \frac{1}{N} \sum_{i=1}^N \log(\cosh(y - f(x))) + \alpha |M, U, D|_1, \quad (6)$$

$$|M, U, D|_1 = |M_A|_1 + |M_W|_1 + |U|_1 + |D|_1, \quad (7)$$

其中 α 为控制正则化强度的非负可调参数, 随着 α 的增加, RLRNN权重系数 M, U, D 不断缩小, 直到其中一些系数压缩为零, 表示从模型中消除了这些冗余单元. 最佳的 α 值由交叉验证的方法来确定.

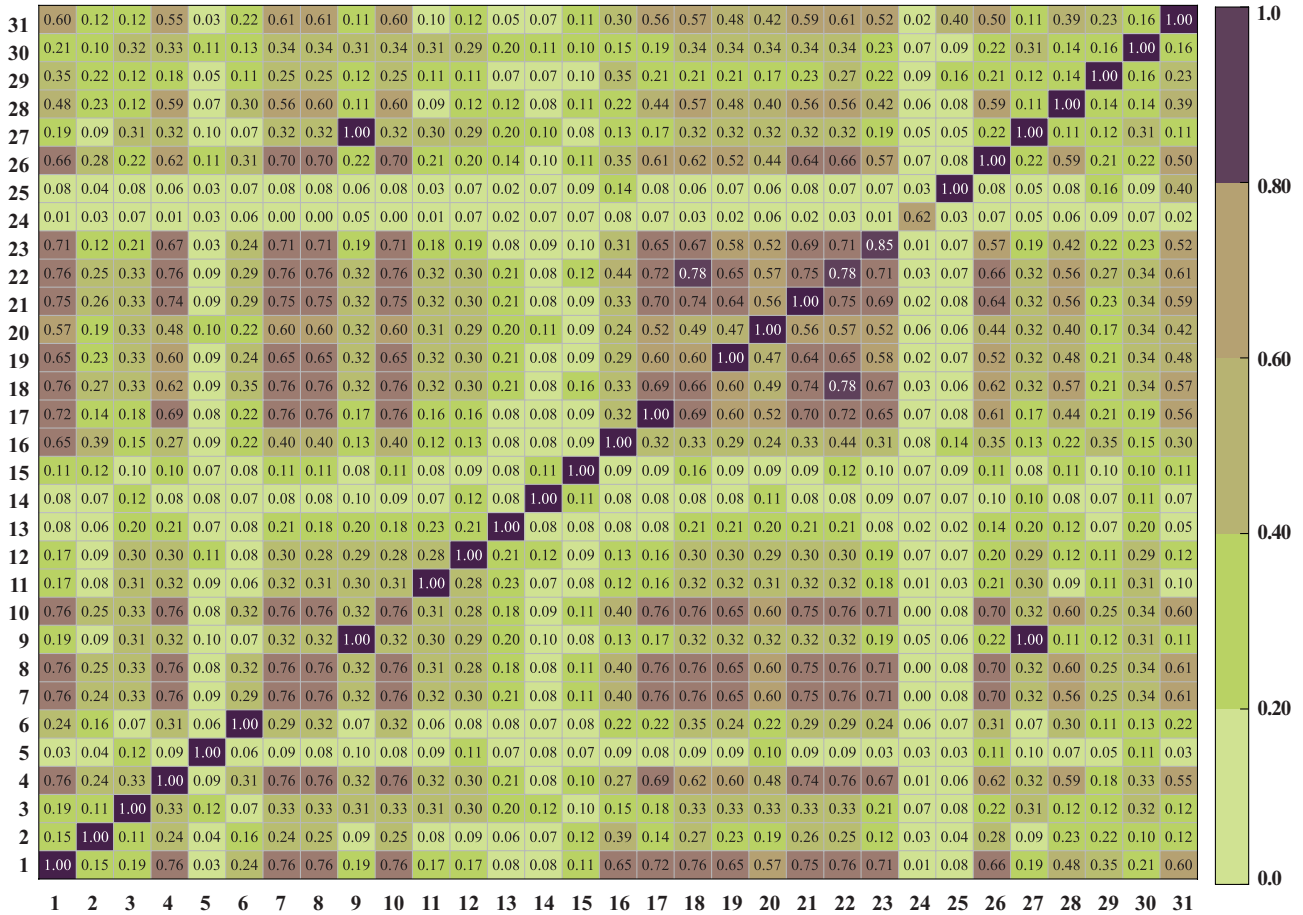


图 2 各变量MIC热点图

Fig. 2 MIC heatmap for all variables

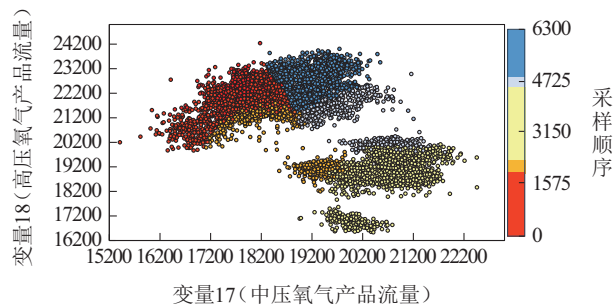


图 3 变量17和18的高斯聚类结果

Fig. 3 Gaussian clustering results for variables 17 and 18

采用自适应矩估计 (adaptive moment estimation, Adam)对式(6)–(7)进行优化, 该方法是一种高效的梯度下降算法, 已在深度学习中得到广泛应用^[28]. Adam算法利用损失函数的一阶导数计算梯度, 从而保证了较低的计算量. 从式(6)中可看出, 所提出的RLRNN仅修改了LRNN的损失函数, 并未显著增加模型训练的计算复杂度. 另外, 当数据存在离群点或呈

现复杂分布时, Log-Cosh函数的平滑梯度性、鲁棒性提高了训练的收敛性和稳定性.

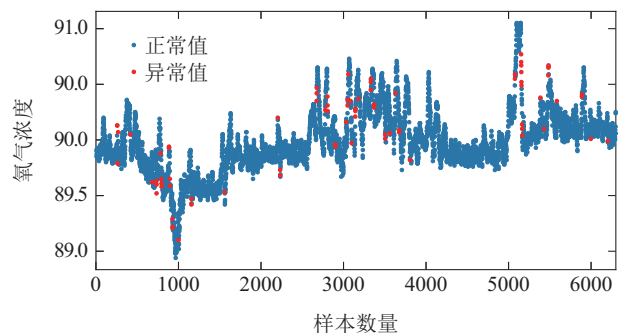


图 4 目标变量测量值的数据分布情况

Fig. 4 Data distribution of measurement values of the target variable

3.4 DIL-RLRNN软测量算法流程

本研究以RLRNN为基础学习器, 采用动态加权集成IL策略, 提出了一种称为DIL-RLRNN的软测量算

法, 算法结构如图6所示.

算法流程详述如下:

步骤 1 对初始时间序列数据进行预处理, 构造多步时间序列数据集 D , 并将其划分为 T 个子集 D^t , $t = 1, 2, \dots, T$.

步骤 2 将 Log-Cosh 估计和 L_1 正则化改进的 RLRNN 作为基础学习器.

步骤 3 对当前 D^t 中的每个样本, 进行权重初始化.

$$\omega_i^t = \frac{1}{\delta^t}, \quad (8)$$

其中: $\omega^t = [\omega_1^t \ \omega_2^t \ \dots \ \omega_{\delta^t}^t]$ 表示 D^t 中样本的权重, δ^t 表示 D^t 中样本的数量.

步骤 4 基于 D^t 和 ω^t 训练基础模型 q_t .

步骤 5 计算每个基模型 $q_k, k = 1, 2, \dots, t$. 在 D^t 的预测值, 并与真实值比较获得残差加权和 ε_k^t .

$$e = \operatorname{argmax}(|q_k(x_i) - y_i|), \quad (9)$$

$$\varepsilon_k^t = \sum_{i=1}^{\delta^t} \omega_i^t \left(\frac{|q_k(x_i) - y_i|}{e} \right)^2, \quad (10)$$

如果 $\varepsilon_{k=t}^t > \frac{1}{2}$, 产生一个新的 q_k , 如果 $\varepsilon_{k<t}^t > \frac{1}{2}$, 令 $\varepsilon_{k<t}^t = \frac{1}{2}$.

步骤 6 根据每个基模型在 D^t 的预测结果计算权重 W_k^t , 并对其归一化处理.

$$w_k^t = \frac{1}{1 + e^{-(t-k)}}, \quad w_k^t = \frac{w_k^t}{\sum_{j=0}^{t-k} w_k^{t-j}}, \quad (11)$$

$$\beta_k^t = \sum_{j=0}^{t-k} w_k^{t-j} \left(\frac{\varepsilon_k^{t-j}}{1 - \varepsilon_k^{t-j}} \right), \quad (12)$$

$$W_k^t = \log\left(\frac{1}{\beta_k^t}\right), \quad W_k^t = \frac{|W_k^t|}{\sum_{k=1}^t |W_k^t|}. \quad (13)$$

步骤 7 利用加权平均法获得最新的集成模型.

$$Q^t = \sum_{k=1}^t W_k^t q_k. \quad (14)$$

步骤 8 令 $t = t + 1$, 在下一个数据集 D^t 中, 首先赋予每个样本相同的权重 ξ_i^t , 然后评估 Q^{t-1} 在 D^t 的预测结果, 更新权重.

$$\xi_i^t = \frac{1}{\delta^t}, \quad i = 1, 2, \dots, \delta^t, \quad (15)$$

$$e = \operatorname{argmax}(|Q^{t-1}(x_i) - y_i|), \quad (16)$$

$$E^t = \sum_{i=1}^{\delta^t} \xi_i^t \left(\frac{|Q^{t-1}(x_i) - y_i|}{e} \right)^2, \quad (17)$$

$$\omega_i^t = \frac{1}{\delta^t} \times (E^t)^{1 - \frac{|Q^{t-1}(x_i) - y_i|}{e}}, \quad (18)$$

$$\omega^t = \frac{\omega^t}{\sum_{i=1}^{\delta^t} \omega_i^t}. \quad (19)$$

步骤 9 循环迭代, 重复步骤4到8, 输出所更新的模型.

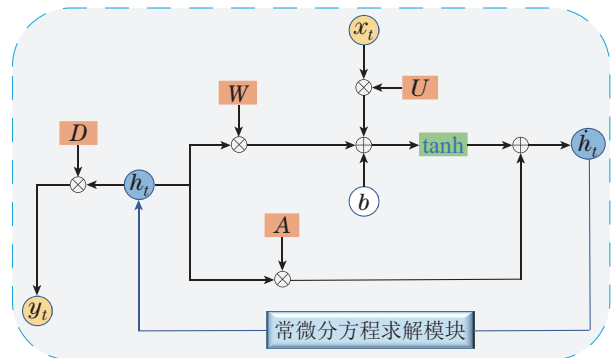


图5 LRNN 结构图

Fig. 5 Structure of the LRNN

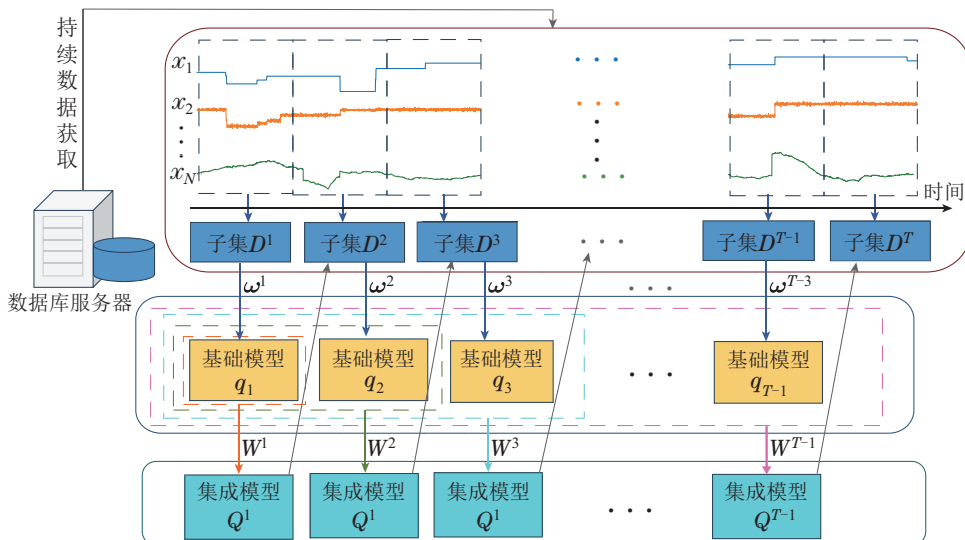


图6 DIL-RLRNN软测量算法结构图

Fig. 6 Structure of the DIL-RLRNN soft sensor algorithm

在本算法中, 步骤6的基础模型权重更新和步骤8的子集样本权重更新是关键步骤. 由步骤6中式(12)–(13)可知, 不同基础模型根据 ϵ 的大小分配不同权重, 模型对当前子集预测精度越高, 其权重越大, 这样所建立的集成模型能够更适应过程最新的模态. 在步骤8中, 由式(16)–(18)可知, 若集成模型对下一子集中的样本预测残差较大, 则其相应样本权重减小. 这样通过动态机制调整样本权重, 可使训练过程在处理模态变化的数据时更加稳健.

4 实验验证

4.1 仿真实验设置

本节将所提出算法应用于某钢铁企业ASU生产过程中低压塔出口O₂浓度预测, 解释变量如第2.2节中的表1所示. 所有模型训练及预测都在以下实验环境中实现: AMD Ryzen 7 5800H CPU @ 3.20 GHz处理器、16 GB 内存、64位Windows11操作系统、Python

3.6版本PyTorch深度学习框架. 采用决定系数(coefficient of determination, R²)、均方根误差(root MSE, RMSE)、平均绝对误差(mean absolute error, MAE)来评估模型性能.

4.2 仿真结果及分析

4.2.1 基础模型性能对比

本小节将数据集 D 的前 80% 作为训练集, 另外 20% 作为测试集. 用于对比 RLRNN 与 SCN, MLP, RNN 和 LSTM 等 4 种基础模型的算法性能. 仿真结果如表 2 所示. 可以看出, 基于前馈结构的 MLP 和 SCN 模型性能明显较差, 这是由于 ASU 较长的管道结构和复杂的反应机制, 使得输入输出关系呈现时延特性, 前馈网络在对这类具有时延特性的过程建模时性能不足. 另一方面, 在上述基础模型中 RLRNN 性能最好, 但其 R² 仍低于 0.7, 表明通过离线训练所建立的基础模型, 难以适应生产过程的模态变化.

表 2 各基础模型性能比较

Table 2 Performance comparison of different base models

模型	R ²			RMSE			MAE		
	Mean	Std	Best	Mean	Std	Best	Mean	Std	Best
SCN	0.357	0.030	0.383	0.180	0.004	0.176	0.126	0.008	0.134
MLP	0.454	0.034	0.463	0.166	0.005	0.164	0.113	0.002	0.112
RNN	0.586	0.032	0.617	0.144	0.005	0.139	0.112	0.009	0.107
LSTM	0.642	0.021	0.652	0.134	0.003	0.132	0.095	0.008	0.089
RLRNN	0.678	0.013	0.693	0.127	0.002	0.124	0.081	0.004	0.076

4.2.2 增量学习模型性能对比

本节将数据集 D 划分为 T 个子集, 第 1 个子集用于离线模型训练, 其余子集用作增量学习和测试. 首先, 对 DIL-RLRNN 算法中数据子集数目 T 进行了分析. 该参数是影响模型性能的关键超参数, 如果取值过大, 则每个数据子集所包含的信息太少, 会导致基础模型的估计出现偏差; 若取值过小, 则两次模型更新的间隔时间较长, 不利于模型及时响应过程模态的变化. 图 7 展示了 DIL-RLRNN 算法在不同 T 值下的平均预测 RMSE, 可看出当 T 为 7 时, 所提算法 RMSE 值最小.

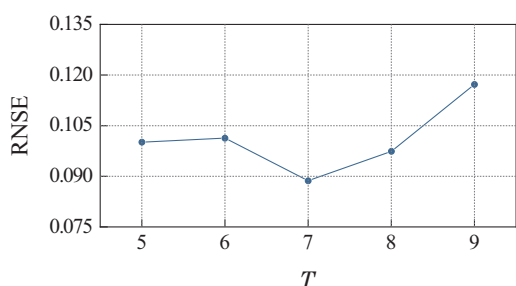


图 7 不同 T 值下 DIL-RLRNN 的 RMSE

Fig. 7 RMSE of DIL-RLRNN at different values of T

其次, 将其他先进的时间序列过程软测量算法如 RNN, STA-LSTM^[29], L₁-HM-LSTM^[23] 与动态集成增量学习结合, 设计了对比动态集成增量学习的循环神经网络(dynamic incremental learning RNN, DIL-RNN)算法、动态集成增量学习的时空注意力长短期记忆网络(DIL spatio-temporal attention LSTM, DIL-STA-LSTM)算法、动态集成增量学习的正则化及 Huber-M 损失长短期记忆网络(DIL regularization and Huber M-loss LSTM, DIL-HM-LSTM)算法, 与所提算法进行性能对比.

各算法经 10 次运行, 其平均和最佳预测结果如表 3 所示. 首先, 各增量学习模型性能均明显优于表 2 中的离线模型, 验证了所提动态增量学习策略的有效性. 此外, DIL-RLRNN 在所有指标上均优于其他算法, 结果表明在对具有离群点的时间序列过程建模时, 所设计的 RLRNN 模型具有显著的优越性. 这是因为所提算法结合了 DIL、稳健估计和 LRNN 的优点. 首先, 鲁棒 LRNN 强大的模型简化能力和抗干扰性提高了基础模型的性能. 其次, 将不同鲁棒 LRNN 基础模型进行动态加权集成, 提高了模型对过程模态变化的适应性.

从表3中可以看出,所提算法比其他算法的训练时间成本略高,这主要是因为DIL-RLRNN利用 L_1 正则化约简模型并求解常微分方程,因此导致了额外的时间成本.尽管如此,该算法的模型训练可以在8 s内完

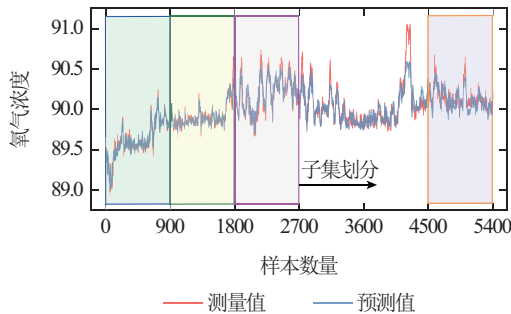
成,为现场操作提供了足够的快速性.此外,DIL-RLRNN的平均测试时间小于0.3 s.在实际应用中,由于目标变量相邻两次测量值的时间间隔为5 min,因此所提方法的训练时间和预测时间完全满足实时性需求.

表3 各增量学习算法性能比较

Table 3 Performance comparison of various incremental learning algorithms

模型	R^2		RMSE		MAE		训练 时间/s	测试 时间/s
	Mean	Std	Mean	Std	Mean	Std		
DIL-RNN	0.743	0.024	0.127	0.009	0.092	0.008	2.635	0.227
DIL-STA-LSTM	0.809	0.018	0.111	0.006	0.080	0.004	64.481	0.371
DIL- L_1 -HM-LSTM	0.832	0.012	0.098	0.003	0.069	0.002	5.395	0.241
DIL-RLRNN	0.865	0.009	0.092	0.002	0.063	0.001	7.370	0.218

为更直观地体现DIL-RLRNN在不同子集下的预测性能,本文绘制了其预测拟合曲线,如图8所示.可以明显看出,随着增量学习的进行,DIL-RLRNN通过学习最新子集信息并调节基础模型的权重实现模型更新,在不同子集阶段均可准确地跟踪 O_2 浓度的动态变化,说明所研究算法对过程模态变化具有良好的适应性.然而,由于过程工况及物化反应的复杂性,生产过程可能发生数据漂移或产生异常,导致模型在异常子集的预测效果上性能下降.

图8 DIL-RLRNN算法预测 O_2 浓度结果Fig. 8 O_2 concentration prediction results with DIL-RLRNN algorithm

此外,计算了不同模型对 O_2 浓度预测误差的正态分布参数和95%置信区间,如表4所示, $MPE = ((1/N_{test}) \sum_{i=1}^{N_{test}} (y_i - \hat{y}_i))$, $SD = ((1/N_{test}) \sum_{i=1}^{N_{test}} [(y_i - \hat{y}_i) - (y_i - \hat{y}_i)]^2)^{1/2}$ 分别表示平均预测误差和标准差.可以看出,DIL-RLRNN算法与其他算法相比,MPE更小,SD更低.此外,所提算法的置信区间最小且最接近于零.因此,DIL-RLRNN在 O_2 浓度预测中表现出更高的准确性和稳定性.

为分析所提出的动态增量学习模型对过程模态变化的适应性,对最优集成模型中各基础模型的权重系数 W_k^t ($t = 6$)进行对比,如图9所示.其中,基础模型 q_6 的权重系数占主导地位,因为它包含最新的数据和信

息,更适配过程最新的模态.基础模型 q_1 系数仅次于 q_6 ,在图3中数据子集 D^1 属于红色区域,与蓝色区域子集 D^6 相邻,它们的数据分布具有较高的相似性,所以其模型系数相对接近.另外,由图3可以看出基础模型 $q_2 - q_5$ 数据分布与 q_6 存在较大的差异,所以在图9中它们对应的基础模型权重系数相对较小.所设计的动态增量学习策略旨在确保最新的数据和与其模态相近的数据对集成模型具有更大的贡献,这有助于集成模型更好地适应过程数据的模态变化.

表4 不同算法预测误差的置信区间

Table 4 Confidence intervals for prediction errors of different algorithms

模型	MPE	SD	95%置信区间
DIL-RNN	0.015	0.131	[0.011, 0.018]
DIL-STA-LSTM	-0.051	0.101	[-0.053, -0.048]
DIL- L_1 -HM-LSTM	-0.013	0.093	[-0.015, -0.011]
DIL-RLRNN	-0.007	0.077	[-0.009, -0.005]

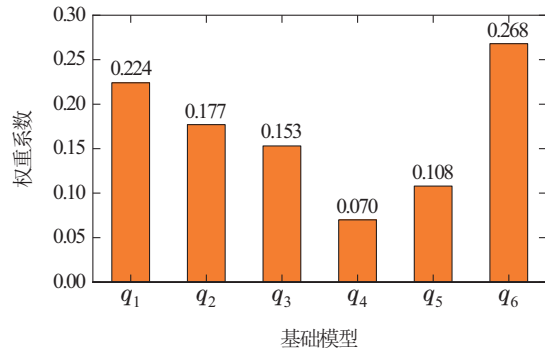


图9 最优集成模型中各基础模型权重系数

Fig. 9 Weighting coefficients of different base models in the optimal ensemble model

4.2.3 消融实验

本节设计了消融实验分析 L_1 正则化和Log-Cosh估计对模型的积极影响,对比结果如表5所示,其中

w/o表示移除某模块。显然, 从该算法中移除任一模块均会导致模型性能明显下降。通过对比, 移除 L_1 正则化对性能的影响比Log-Cosh估计更显著, 这表明ASU采集的过程变量中存在显著互相关性, 导致模型内部权重冗余, 而数据集中离群值分布较少, 因此 L_1 正则化的模型简化能力对提高模型性能贡献更大。

表5 消融实验结果

Table 5 Results of ablation experiments

模型	R^2	RMSE	MAE
w/o L_1 & Log-Cosh	0.791	0.111	0.075
w/o L_1	0.811	0.105	0.071
w/o Log-Cosh	0.835	0.098	0.066
DIL-RLRNN	0.879	0.066	0.061

5 结论

本文研究了一种面向ASU过程关键性能参数建模的动态加权集成IL软测量算法。首先对某钢铁企业ASU过程的运行机理和生产数据进行了统计分析, 该过程呈现明显的时间序列和多模态特性, 且目标变量测量数据中存在较多离群点。为解决上述问题, 采用具有良好稳定性和抗干扰性的LRNN作为基础模型, 并利用Log-Cosh估计和 L_1 正则化改进其损失函数, 以提高模型对离群点的鲁棒性和降低模型复杂度; 通过动态更新由多个基础模型组成的集成模型, 提高模型对生产模态变化的适应能力。实验结果表明所提出的算法可以准确预测目标变量, 精度满足生产现场对于该参数测量与控制的需求。另外, 相对于DIL-RNN, DIL-LSTA-LSTM和DIL- L_1 -HM-LSTM算法, 所提出的DIL-RLRNN算法在泛化能力上具有显著的优势。所提出的动态加权集成策略及鲁棒LRNN模型解决了复杂过程建模所面临的多模态、多变量、大时延及存在测量离群点等难题, 并丰富了常微分方程网络建模的相关理论方法, 具有重要的理论意义和应用价值。

然而, 所提算法是在单一任务场景上进行的, 未来可以探索如何将算法扩展到跨任务学习的场景, 以便于模型在不同任务之间进行迁移学习和知识共享。

参考文献:

[1] QING M L, JIN B, MA J C, et al. Thermodynamic and economic performance of oxy-combustion power plants integrating chemical looping air separation. *Energy*, 2020, 206: 118136.

[2] QIAO Junfei, SUN Zijian, TANG Jian. Overview of concept drift detection for industrial process soft sensor modeling. *Control Theory & Applications*, 2021, 38(8): 1159 – 1174.
(乔俊飞, 孙子健, 汤健. 面向工业过程软测量建模的概念漂移检测综述. 控制理论与应用, 2021, 38(8): 1159 – 1174.)

[3] SUN Kai, SUI Lin, ZHANG Fangfang, et al. Dynamic soft sensor algorithm based on nonnegative garrote and long short-term memory neural network. *Control Theory & Applications*, 2023, 40(1): 83 – 93.

(孙凯, 隋琳, 张芳芳, 等. 基于非负绞杀与长短期记忆神经网络的动态软测量算法. 控制理论与应用, 2023, 40(1): 83 – 93.)

- [4] TSAY C, KUMAR A, FLORES-CERRILLO J, et al. Optimal demand response scheduling of an industrial air separation unit using data-driven dynamic models. *Computers & Chemical Engineering*, 2019, 126: 22 – 34.
- [5] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 1997, 45(11): 2673 – 2681.
- [6] SUN K, SUI L, WANG H X, et al. Design of an adaptive nonnegative garrote algorithm for multi-layer perceptron-based soft sensor. *IEEE Sensors Journal*, 2021, 21(19): 21808 – 21816.
- [7] HU L W, WANG W Y, XIANG Y, et al. Flow field reconstructions with gans based on radial basis functions. *IEEE Transactions on Aerospace and Electronic Systems*, 2022, 58(4): 3460 – 3476.
- [8] WANG D H, LI M. Stochastic configuration networks: Fundamentals and algorithms. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3466 – 3479.
- [9] SAK H, SENIOR A, BEAUFAYS F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *The 15th Annual Conference of the International Speech Communication Association*. Singapore: International Speech Communication Association, 2014: 338 – 342.
- [10] CHUNG J Y, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv Preprint*, 2014, arXiv: 1412.3555.
- [11] YUAN X F, LI L, WANG Y L. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Transactions on Industrial Informatics*, 2019, 16(5): 3168 – 3176.
- [12] CHEN R T Q, RUBANOVA Y, BETTENCOURT J, et al. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 2018, 31: 1 – 13.
- [13] ERICHSON N B, AZENCOT O, QUEIRUGA A, et al. Lipschitz recurrent neural networks. *ArXiv Preprint*, 2020, arXiv: 2006.12070.
- [14] HASANI R, LECHNER M, AMINI A, et al. Closed-form continuous-time neural networks. *Nature Machine Intelligence*, 2022, 4(11): 992 – 1003.
- [15] ZHANG J H, MAEDA K, OGAWA T, et al. Defense against image captioning attacks via a robust and stable recurrent neural network. *IEEE 10th Global Conference on Consumer Electronics*. Kyoto, Japan: Institute of Electrical and Electronics Engineers, 2021: 785 – 786.
- [16] CAO Yue, YU Chong, LI Zhi, et al. Industrial data driven transition state detection with multi-mode switching of a hydrocracking unit. *Journal of Chemical Industry and Engineering*, 2023, 74(9): 3841 – 3854.
(曹跃, 余冲, 李智, 等. 工业数据驱动的加氢裂化装置多工况切换过渡状态检测. 化工学报, 2023, 74(9): 3841 – 3854.)
- [17] CAO Jie, GU Binjie, PAN Feng, et al. Accurate incremental ϵ -twin support vector regression. *Control Theory & Applications*, 2022, 39(5): 1020 – 1032.
(曹杰, 顾斌杰, 潘丰, 等. 精确增量式 ϵ 型孪生支持向量回归机. 控制理论与应用, 2022, 39(5): 1020 – 1032.)
- [18] ALVARENGA W J, CAMPOS F V, HANRIOT V M, et al. Online learning of neural networks using random projections and sliding window: A case study of a real industrial process. *Engineering Applications of Artificial Intelligence*, 2021, 100: 104181.
- [19] ERGEN T, KOZAT S S. Efficient online learning algorithms based on LSTM neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 29(8): 3772 – 3783.
- [20] ZHANG X, ZOU Y Y, LI S Y. Enhancing incremental deep learning for FCCU end-point quality prediction. *Information Sciences*, 2020, 530: 95 – 107.

- [21] SUN Shixu, HU Xiaofeng, XIA Mingyuan. Semi-supervised incremental-learning method for tool breakage detection under variable operation conditions. *Computer Integrated Manufacturing Systems*, 2022, 28(9): 2805 – 2814.
(孙世旭, 胡小锋, 夏铭远. 变工况刀具破损监测的半监督增量学习方法. *计算机集成制造系统*, 2022, 28(9): 2805 – 2814.)
- [22] WANG Qianjin, DAI Wei, LU Qun, et al. A sparse learning method for SCN soft measurement model. *Control and Decision*, 2022, 37(12): 3171 – 3182.
(王前进, 代伟, 陆群, 等. 一种随机配置网络软测量模型的稀疏学习方法. *控制与决策*, 2022, 37(12): 3171 – 3182.)
- [23] ZHAO L, GUO J M, SUN K. A new regularized and robust LSTM learning algorithm for industrial soft sensors. *Journal of Intelligent & Fuzzy Systems*, 2023, 45(1): 323 – 343.
- [24] LU W C, DUAN J D, WANG P, et al. Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and maximum mixture correntropy long short-term memory neural network. *International Journal of Electrical Power & Energy Systems*, 2023, 144: 108552.
- [25] SU W, ZHANG H F, ZHOU Q, et al. Monocular depth estimation using information exchange network. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(6): 3491 – 3503.
- [26] ZHAO Ling, GONG Jiaying, HUANG Darong, et al. Fault feature selection method of gearbox based on Fisher score and maximum information coefficient. *Control and Decision*, 2021, 36(9): 2234 – 2240.
(赵玲, 龚加兴, 黄大荣, 等. 基于Fisher Score与最大信息系数的齿轮箱故障特征选择方法. *控制与决策*, 2021, 36(9): 2234 – 2240.)
- [27] LAI X, DENG C, TANG X P, et al. Soft clustering of retired lithium-ion batteries for the secondary utilization using Gaussian mixture model based on electrochemical impedance spectroscopy. *Journal of Cleaner Production*, 2022, 339: 130786.
- [28] KINGMA D P. Adam: A method for stochastic optimization. *ArXiv Preprint*, 2014, arXiv: 1412.6980.
- [29] YUAN X F, LI L, SHARDT Y A W, et al. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics*, 2020, 68(5): 4404 – 4414.

作者简介:

吴修粮 博士研究生, 目前研究方向为机器学习、软传感器建模,

E-mail: qlwuxiuliang@sina.com;

赵磊 硕士研究生, 目前研究方向为神经网络与数据驱动建模,

E-mail: layzhao98@163.com;

曹茂永 教授, 博士生导师, 目前研究方向为机器视觉、检测技术、监控系统及自动化应用, E-mail: cmy@qlu.edu.cn;

孙凯 教授, 博士生导师, 目前研究方向为神经网络与机器学习、复杂系统建模与优化控制, E-mail: sunkai79@qlu.edu.cn.