

控制与决策

Control and Decision

基于双阶段注意力的双流记忆调节GRU软测量建模

廖开继, 隋璘, 熊伟丽

引用本文:

廖开继, 隋璘, 熊伟丽. 基于双阶段注意力的双流记忆调节GRU软测量建模[J]. *控制与决策*, 2025, 40(9): 2848–2858.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.1326>

您可能感兴趣的其他文章

Articles you may be interested in

基于双分支特征融合的场景文本检测方法

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

基于图卷积网络的行为识别方法综述

A survey of action recognition methods based on graph convolutional network

控制与决策. 2021, 36(7): 1537–1546 <https://doi.org/10.13195/j.kzyjc.2020.0514>

一种基于深度学习的时间序列预测方法

A time series prediction method based on deep learning

控制与决策. 2021, 36(3): 645–652 <https://doi.org/10.13195/j.kzyjc.2019.0809>

基于协同聚类和权重注意力稀疏自编码网络的变化检测方法

Change detection approach based on cooperative clustering and weighted attention sparse autoencoder

控制与决策. 2021, 36(10): 2442–2450 <https://doi.org/10.13195/j.kzyjc.2019.1633>

结合注意力机制的循环神经网络复述识别模型

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

基于双阶段注意力的双流记忆调节 GRU 软测量建模

廖开继, 隋 璘, 熊伟丽[†]

(江南大学 物联网工程学院, 江苏 无锡 214122)

摘要: 针对非线性动态工业过程建模中易产生信息冗余和时序信息衰减问题, 提出一种基于双流记忆调节门控循环单元并内嵌双阶段注意力机制的动态软测量算法. 首先, 设计时间相关和动态因果相关的双流信息提取结构, 在门控循环单元中分别引入时间门和因果门, 提取信息中的时序关系与动态因果关系, 从而形成互补信息流, 提高模型的预测性能; 然后, 在特征提取和预测输出阶段分别引入特征注意力和时序注意力机制, 以动态挖掘输入特征与目标特征间的潜在相关性, 捕捉关键特征, 并评估不同历史时间点对于待预测时刻的重要程度, 从而选择关键时间点信息; 最后, 通过数值仿真以及某火电厂脱硫过程排放烟气 SO₂ 浓度的软测量验证所提出算法的预测效果.

关键词: 软测量; 门控循环单元; 辅助记忆调节门; 双信息流结构; 双阶段注意力; 深度学习

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2024.1326

引用格式: 廖开继, 隋璘, 熊伟丽. 基于双阶段注意力的双流记忆调节 GRU 软测量建模 [J]. 控制与决策, 2025, 40(9): 2848-2858.

Soft sensor modeling based on dual-stream memory-modulated GRU with dual-stage attention mechanism

LIAO Kai-ji, SUI Lin, XIONG Wei-li[†]

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: To address the issues of information redundancy and temporal information decay in modeling nonlinear dynamic industrial processes, a dynamic soft sensing algorithm based on a dual-stream memory-regulated gated recurrent unit (GRU) with an embedded two-stage attention mechanism is proposed. On one hand, a dual-stream information extraction structure is designed to capture both time-related and dynamic causal correlations by introducing time gates and causal gates into the GRU, thus forming complementary information flows and enhancing model prediction performance. On the other hand, feature attention and temporal attention mechanisms are introduced during the feature extraction and prediction output stages, respectively, to dynamically uncover the latent correlations between input and target features, capture key features, and assess the importance of different historical time points for the prediction target, thereby selecting critical time point information. Finally, the proposed algorithm's prediction performance is validated through numerical simulations and soft sensing of SO₂ concentration in flue gas emissions from a thermal power plant desulfurization process.

Keywords: soft sensor; gated recurrent unit; auxiliary memory regulation gate; dual information flow structure; dual-stage attention mechanism; deep learning

0 引言

在实际工业生产过程中, 实时监测某些难以直接测量的关键主导变量, 对于保障生产安全和提高产品质量至关重要^[1-2]. 软测量技术通过挖掘易测辅助变量和难测主导变量间的关系建立数学模型, 间接实现对主导变量的有效估计, 在现代工业过程检

测中发挥着重要作用^[3-4]. 软测量建模方法主要分为机理建模和数据驱动建模, 常见的数据驱动建模方法有偏最小二乘^[5]、支持向量回归^[6]和人工神经网络^[7]等. 其中, 人工神经网络凭借其强大的非线性建模和泛化能力受到广泛关注.

门控循环单元 (GRU) 网络是长短时记忆网络

收稿日期: 2024-11-14; 录用日期: 2025-02-20.

基金项目: 国家自然科学基金项目 (61773182); 国家外国专家项目 (H20240955).

责任编委: 魏秀琨.

[†]通信作者. E-mail: weili_xiong@jiangnan.edu.cn.

(LSTM) 的一种重要变体, 其具有更为简化的网络结构和门控机制, 不仅能有效捕捉序列中的时间依赖关系, 而且显著提高了模型训练过程的收敛速度, 被广泛应用于工业软测量领域^[8]. 然而, GRU 网络中固定的单门控更新机制导致其隐藏状态计算时存在线性约束关系, 限制信息流的充分通过, 从而影响模型预测性能^[9]. 为此, 文献 [9] 提出一种 p 范数门控方案, 通过调整超参数 p 控制历史信息流速, 以提高模型学习效率. 文献 [10] 通过在 GRU 网络隐藏状态中引入超参数 λ_1 和 λ_2 改变其固有的线性约束关系, 以便更好地控制信息的记忆与遗忘, 有效提高了模型的预测精度. 文献 [11] 在卷积 GRU 网络中引入调节参数 L 和 S , 分别实现对长期和短期时变延迟信息的捕捉, 从而增强模型的特征提取能力.

上述方法通过引入超参数, 并与更新门协同作用来更新隐藏状态, 有效增强了模型的学习能力. 然而, 额外的超参数不仅增加了模型调优过程的复杂度, 还可能使模型过度依赖于特定超参数配置, 从而削弱其泛化能力. 为解决此问题, 文献 [12] 提出一种快速循环网络 (FRU), 在 GRU 的基础上简化了门控学习机制, 并采用独立的遗忘门调节历史信息. 然而, FRU 中独立的门控不仅打破了 GRU 隐藏状态计算中的线性约束关系, 还间接改变了其固有的互斥调节策略. 相应实验表明, FRU 在训练和收敛速度方面优于传统 FRU, 但预测精度有所下降. 值得注意的是, 结合卷积网络进行特征提取能显著提高 FRU 的预测精度^[12]. 这表明, 改变 GRU 隐藏状态中的互斥调节策略可能会影响模型的特征提取能力. 因此, 在保留 GRU 互斥调节策略的同时, 规避单门控更新机制导致的线性约束关系, 对于提高模型学习效率具有重要意义.

在建模复杂动态时间序列时, 传统 GRU 难以有效捕捉复杂的时间依赖关系和动态特征. 注意力机制通过聚焦关键信息, 可显著提升模型训练速度和预测精度. 文献 [13] 将注意力机制嵌入 GRU 内部, 以提取候选隐藏状态中的关键信息, 增强了模型对长期依赖信息的记忆能力. 文献 [14] 提出轻量级注意力-卷积-GRU 架构, 利用轻量级注意力模块捕捉化学过程中的瞬态扰动信息, 实现了高精度动态建模. 文献 [15] 利用时间注意力机制聚焦于输入序列中的相关时间步, 并借助监督注意力机制在训练阶段差异化处理输出序列误差, 显著增强了模型的多步预测性能. 尽管这些研究均验证了注意力机制在提升模型预测性能方面的有效性, 但其主要聚焦于单一维度的信息提取, 未能充分捕捉时间序列中的

关键动态特征和全局依赖关系.

综上所述, 本文针对 GRU 网络因线性约束导致的信息传递不充分问题, 提出一种新型记忆调节门控循环单元 (MRGRU). 该网络充分考虑了隐藏状态中不同信息间的重要性差异, 在现有门控策略的基础上设计了两个辅助记忆调节门: 时间门和因果门, 分别用于对历史信息和当前信息进行精细调控; 并进一步构建时间相关和因果相关的信息提取结构, 生成互补信息流, 显著增强其在不同工业场景中的泛化能力和预测性能. 同时, 考虑到非线性动态工业过程建模中, 信息冗余和时序信息衰减问题会削弱模型对关键信息和长期依赖关系的捕捉能力, 在双流 MRGRU 软测量模型中嵌入双阶段注意力机制, 提出一种基于双阶段注意力机制的双流 MRGRU 软测量预测模型. 该模型在特征提取阶段利用特征注意力机制自适应提取关键特征变量, 抑制冗余信息影响, 提高模型的可解释性; 并在预测输出阶段采用时序注意力机制挖掘重要历史时刻信息, 提高模型对长时间序列的预测效果. 最后, 通过数值仿真与某火电厂烟气脱硫过程 SO_2 浓度预测验证了所提算法的有效性与优越性.

1 GRU 网络的基本原理

GRU 通过门控机制有效地解决了长时间序列建模中常见的梯度消失和爆炸问题^[16], 从而能够准确捕捉时间序列输入中的动态特征.

GRU 结构如图 1 所示, 当前时刻输入 x_t 和前一时刻隐藏状态 h_{t-1} 为单元输入, 复位门 r_t 用于控制将多少 h_{t-1} 中的信息融入当前候选状态 \tilde{h}_t , 而更新门 z_t 则决定当前隐藏状态 h_t 需整合多少 \tilde{h}_t 的信息. 其状态信息更新方程为

$$r_t = \sigma(U_r x_t + W_r h_{t-1} + b_r), \quad (1)$$

$$z_t = \sigma(U_z x_t + W_z h_{t-1} + b_z), \quad (2)$$

$$\tilde{h}_t = \tanh(U_h x_t + W_h (h_{t-1} \odot r_t) + b_h), \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t. \quad (4)$$

其中: U 、 W 、 b 分别表示各门对应的权重矩阵和偏

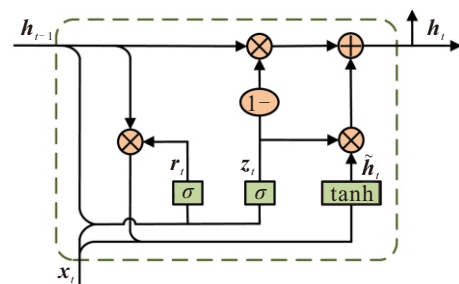


图1 GRU 结构

置向量, σ 为 sigmoid 函数, \tanh 为双曲正切函数, 符号 \odot 表示元素级乘法.

2 基于双阶段注意力机制的双流 MRGRU 软测量模型

2.1 双流 MRGRU 网络架构

本文提出一种基于双流结构的 MRGRU 网络, 通过引入时间门和因果门, 优化传统 GRU 隐藏状态的更新机制, 提升了模型对时间相关信息和动态因果关系的捕捉能力. 下面将详细阐述 MRGRU 的构建及其在双流结构中的应用.

2.1.1 MRGRU 的构建

由式 (4) 可知, GRU 在计算隐藏状态时利用 z_t 和 $1 - z_t$ 执行更新功能, 两个参数间的互斥关系深刻体现了模型在处理即时输入与长期记忆间的动态平衡策略, 促进损失函数在训练过程中快速收敛. 该机制使得 GRU 凭借更简洁的门控策略, 实现与 LSTM 相媲美的预测性能.

然而, 这种互斥性以线性约束的形式呈现: 当 z_t 趋近于 1 时, $1 - z_t$ 必然以相同速率趋近于 0. 此设计虽有利于简化模型结构和促进参数更新, 但在复杂非线性动态工业过程建模中, 可能成为制约信息有效传递的关键因素, 它限制模型充分捕捉多变且相互依赖的非线性特征, 进而影响其对复杂动态过程的准确建模能力.

本文在 GRU 隐藏状态中引入时间门 T_t 和因果门 C_t 优化网络结构, 并保留 z_t 和 $1 - z_t$ 作为互斥关系门, 构建新型 MRGRU, 其结构如图 2 所示. 时间门通过筛选前一时刻隐藏状态中的关键信息, 实现

对历史信息的记忆调控, 以有效捕捉与时间相关的信息特征; 因果门则用于选择当前候选状态中的重要信息, 调控当前信息对隐藏状态的影响程度, 以捕捉信息之间潜在的动态因果关系. 该设计在保留互斥关系的基础上, 有效改变了线性约束关系, 提升了模型的信息处理能力. MRGRU 的状态更新方程为

$$T_t = \sigma(U_T x_t + W_T h_{t-1} + b_T), \quad (5)$$

$$C_t = \sigma(U_C x_t + W_C h_{t-1} + b_C), \quad (6)$$

$$h_t = T_t \odot (1 - z_t) \odot h_{t-1} + C_t \odot z_t \odot \tilde{h}_t. \quad (7)$$

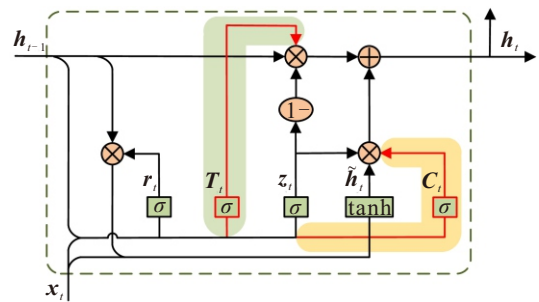


图2 MRGRU 结构

时间门和因果门的引入显著增强了门控策略的灵活性, 使得 MRGRU 能够对信息传递进行精细调节, 更有效地适应工业环境中的动态变化. MRGRU 可分化为两种形式, 仅具备时间门的为时间相关 GRU (TGRU), 仅具备因果门的为因果相关 GRU (CGRU).

2.1.2 基于双流结构的 MRGRU 设计

本文在 MRGRU 的基础上提出双流 MRGRU 网络模型 (DS-MRGRU), 如图 3 所示. 该模型由两部分组成: 一是由 TGRU 构成的时间相关流, 通过时间门 T_t 调控历史信息的记忆与遗忘, 从而专注于提取

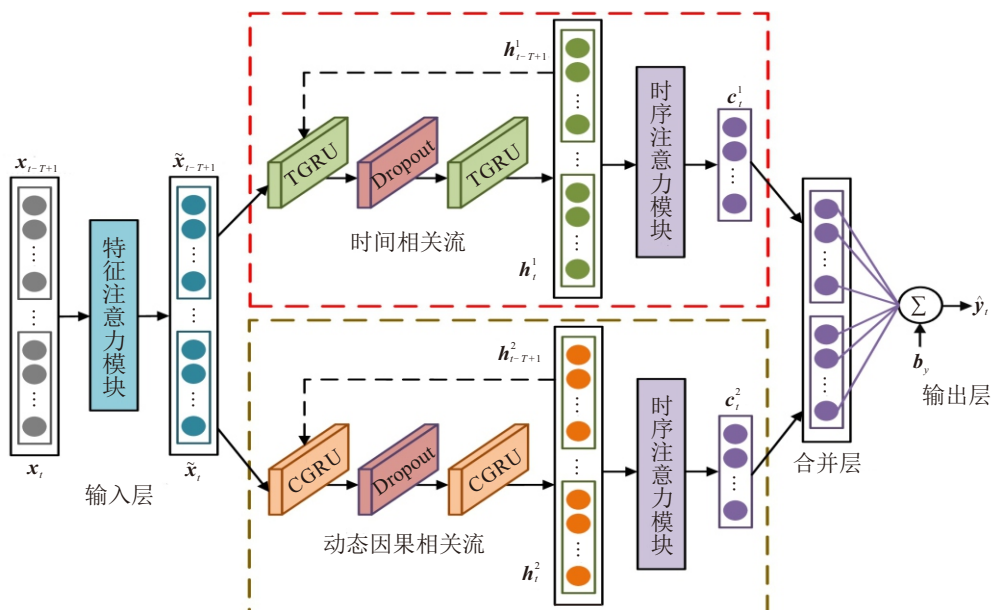


图3 FTA-DS-MRGRU 结构

时间序列中的时序关系; 二是由 CGRU 构成的动态因果相关流, 通过因果门 C_i 控制当前输入信息对预测结果的影响, 旨在捕捉当前输入与预测结果之间的动态因果关系, 确保模型能够准确响应当前环境的变化。

模型中的时间相关流由两个级联的 TGRU 层构成, 层间插入失活层, 以增强模型的泛化能力并有效防止过拟合; 动态因果相关流则采用了两个级联的 CGRU 层。这种双流结构的设计不仅增强了模型对时间序列数据的理解分析能力, 也提升了其处理复杂预测任务时的鲁棒性与泛化性能, 且两个信息流并行组成, 避免了计算复杂度过度增加。

2.2 双阶段注意力机制

注意力机制通过动态调整输入信息各特征的权重分配, 精准提取有效信息。其注意力权重可直观量化各输入数据对模型预测结果的贡献, 有助于理解模型决策, 增强其透明度与可解释性。本文所提出软测量模型采用双阶段注意力机制: 在特征提取阶段引入特征注意力模块, 以捕捉关键特征并抑制无关特征影响; 在预测输出阶段引入时序注意力模块, 自适应关注历史数据中的关键时间点信息, 以提高模型捕捉长时间序列依赖关系的能力。

2.2.1 特征注意力机制

特征注意力机制通过为输入特征动态分配注意力权重, 挖掘出对模型预测结果具有重要影响的特征, 并加速模型训练收敛过程, 有效提高其学习效率。特征注意力结构如图 4 所示, 其计算公式为

$$q_{t1} = W_q^1 x_t + b_q^1; \quad (8)$$

$$e_t^i = V_1^i \sigma(W_1^i q_{t1} + U_1^i x_t^i + b_1^i), \quad 1 \leq i \leq n; \quad (9)$$

$$\alpha_t^i = \frac{|e_t^i|}{\sum_{m=1}^n |e_t^m|}, \quad 1 \leq i \leq n; \quad (10)$$

$$\tilde{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n). \quad (11)$$

其中: q_{t1} 为查询向量, e_t^i 为 t 时刻第 i 个变量对应的特征注意力得分, α_t^i 为特征注意力权重, \tilde{x}_t 为加权后的输入向量, n 为输入特征数量, W_q^1 、 b_q^1 、 W_1^i 、 U_1^i 、 b_1^i 、 V_1^i 为模型训练过程中的特征注意力参数。

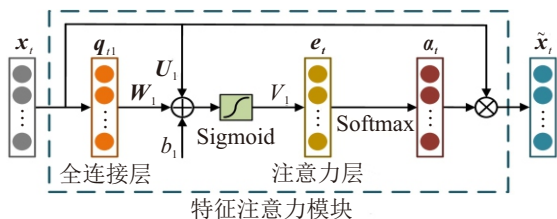


图4 特征注意力结构

2.2.2 时序注意力机制

在复杂工业系统中, 长采样管道因受信号传输路径时滞效应影响, 会在测量系统间引入未知时间延迟, 致使多源数据呈现出时序异步性; 同时, GRU 依赖隐藏状态递归传递建模, 在处理长序列输入时易受记忆衰减制约, 导致部分关键信息随网络深度的增加而逐步丢失。

为缓解此问题, 本文在双信息流结构中分别嵌入独立的时序注意力机制。该设计具有以下优势: 首先, 双时序注意力机制借助双流结构并行化训练的优势, 有利于减轻模型训练负担, 提升计算效率; 其次, 其能分别捕捉两条信息流中历史时刻隐藏状态的重要性, 有助于精准评估每条信息流中不同时间点对预测的贡献程度, 从而在最终预测阶段结合两条流的加权信息, 提高模型预测的准确性。时序注意力结构如图 5 所示, 其计算公式为

$$q_{t2} = W_q^2 h_t + b_q^2; \quad (12)$$

$$g_t^j = V_2^j \sigma(W_2^j q_{t2} + U_2^j h_{t-T+j} + b_2^j), \quad 1 \leq j \leq T; \quad (13)$$

$$\beta_t^j = \frac{|g_t^j|}{\sum_{k=1}^T |g_t^k|}, \quad 1 \leq j \leq T; \quad (14)$$

$$c_t = \sum_{j=1}^T h_{t-T+j} \beta_t^j. \quad (15)$$

其中: q_{t2} 为查询向量, g_t^j 为 t 时刻第 j 个时间步的隐藏状态对应的时序注意力得分, β_t^j 为时序注意力权重, c_t 为各隐藏状态加权求和后的向量, T 为时间步长, W_q^2 、 b_q^2 、 W_2^j 、 U_2^j 、 b_2^j 、 V_2^j 为模型训练过程中更新优化的时序注意力参数。

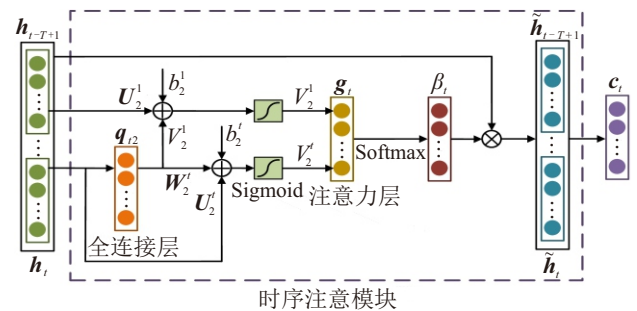


图5 时序注意力结构

2.3 基于双阶段注意力机制的 DS-MRGRU 软测量建模

本文提出一种基于双阶段注意力机制的 DS-MRGRU 软测量模型 (FTA-DS-MRGRU), 其融合了双阶段注意力机制与双信息流结构设计, 能高效捕捉复杂时间序列特征, 提高预测的准确性与鲁棒性。

首先,输入数据经归一化预处理后进入特征注意力模块,通过加权机制筛选出与预测任务紧密相关的特征;接着,两个信息流分别学习时间相关和动态因果相关特征,并进入时序注意力模块对各时刻点隐藏状态信息进行加权求和,从而生成一个包含关键历史时间点信息的综合向量;随后,通过合并层融合两个信息流提取的特征,得到更全面准确的特征表示;最后,经全连接层处理后输出最终预测结果.所提软测量模型的建模步骤如下.

step 1: 采用滑动窗口技术将时间序列数据转化为若干个长度为 T 的子序列,并将前 80% 作为训练集,剩余 20% 作为测试集;

step 2: 特征注意力模块为输入数据 $X = \{x_1, x_2, \dots, x_n\}$ 中的变量分配不同权重 $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$;

step 3: 每个子序列中 T 个时间步的输入数据分别输入到时间相关流和动态因果相关流,分别生成 T 个具有不同表示的隐藏状态信息 $\{h_{t-T+1}, h_{t-T+2}, \dots, h_t\}$;

step 4: 双信息流的隐藏状态信息进入时序注意力模块后,通过分配权重 $\beta = \{\beta_1, \beta_2, \dots, \beta_T\}$ 并求和的方式生成两个综合向量 c_{t1} 和 c_{t2} ,再利用合并层整合为最终向量 c_t ;

step 5: 通过双流结构训练多层 MRGRU,利用式 (8) ~ (15) 训练特征注意力权重 α 和时序注意力权重 β ,从而得到最优训练模型;

step 6: 将测试集数据输入经过训练的 FTA-DS-MRGRU 模型,获得预测结果,并通过评价指标对模型预测性能进行评估.

3 仿真实验

本文利用公开数值案例和实际火电厂脱硫过程软测量来评估所提出模型的预测性能.首先,通过参数选择实验分析时间步长 T 对模型性能的影响;其次,针对 GRU 存在的线性约束问题,与 p -Norm-GRU^[9]、TS-λGRUs^[10]、L/S-ConvGRU^[11] 和 CNN-FRU^[12] 进行对比实验;然后,通过量化分析注意力权重,验证双阶段注意力机制的有效性;最后,进行消融研究以验证每个模块的必要性和对整体模型性能的影响.

模型性能评价指标选取均方误差 (MSE)、平均绝对误差 (MAE) 和相关指数 (R^2) 如下:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2, \quad (16)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\mathbf{y}_i - \hat{\mathbf{y}}_i|, \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}}_i)^2}. \quad (18)$$

其中: N 为测试集样本总数, \mathbf{y}_i 为真实值, $\hat{\mathbf{y}}_i$ 为预测值, $\bar{\mathbf{y}}_i$ 为 \mathbf{y}_i 的平均值.

3.1 数值仿真

为模拟复杂工业过程普遍存在的非线性、动态性以及信息冗余特性,并考虑到长时间序列分析时可能面临的关键信息丢失问题,本节基于 Friedman 数值仿真构造一组仿真函数以验证模型性能^[17],数据集共 2 000 组样本数据.仿真函数定义为

$$\begin{aligned} X_1(t) &= 0.5x_1(t) + 0.3x_2(t-1)^2 + \\ &\quad 0.2x_2(t-3)^3 + e_1, \\ X_2(t) &= 0.7x_1(t-1) + 0.4x_2(t-3) - \\ &\quad 0.3x_3(t-2) + e_2, \\ X_3(t) &= 0.1x_2(t-2) + 0.4x_3(t-5)^2 + \\ &\quad 0.5x_4(t-11) + e_3, \\ X_4(t) &= 0.3x_3(t-5) + 0.3\sqrt{x_3(t-6)} + \\ &\quad 0.4x_4(t-4) + e_4, \\ X_5(t) &= 0.2x_5(t) + 0.3x_5(t-4) + \\ &\quad 0.5x_5(t-11)^2 + e_5, \\ y &= 10 \sin(\pi X_1(t)X_2(t)) + 20(X_3(t) - 0.5)^2 + \\ &\quad 10X_4(t) + 5X_5(t) + e_y. \end{aligned} \quad (19)$$

其中:输入变量 $x_1 \sim x_5$ 是在 $[0,1]$ 区间上均匀分布的随机数, y 为输出变量, $e_1 \sim e_5$ 为均值为零和标准差为 0.05 的高斯噪声, e_y 为均值为零和标准差为 0.1 的高斯噪声.此外,输入中引入 5 个冗余变量 $x_6 \sim x_{10}$,用于模拟实际系统中可能存在的非关键或无关因素,这些变量对输出变量无直接影响,但会增加模型训练负担与过拟合风险.

如式 (19) 所示,当前时刻的输出变量 y 不仅依赖于当前时刻和与之接近的 6 个历史时刻的输入信息,同时还依赖于第 11 个历史时刻的输入信息.该时刻信息经多个时间步的传递极有可能产生衰减,从而模拟实际复杂工业过程中长时间序列数据可能面临的信息衰减与丢失问题.

3.1.1 参数选择实验

为探究时间步长 T 对 FTA-DS-MRGRU 模型预测性能的影响,进行参数选择实验.每个参数在不同 T 值下进行 10 次仿真,采用 MSE 作为模型性能评价指标,并选取最接近平均值的结果作为评估依据.

如图 6 所示,当 $T = 12$ 时模型的 MSE 最小,这

与实验设置相符; 当 $T < 12$ 时, 输入缺失第 11 个历史时刻的重要信息, MSE 显著增大. 同时, 因第 7 到第 10 个历史时刻的输入为冗余信息, 当 T 由 12 降到 8 时, 冗余输入信息减少, 模型训练难度与过拟合风险降低, 进而使其预测性能提高. 当 $T > 12$ 时, 输入包含所有重要信息, 但随着 T 的增加, 冗余输入信息增多, 模型预测性能相应下降.

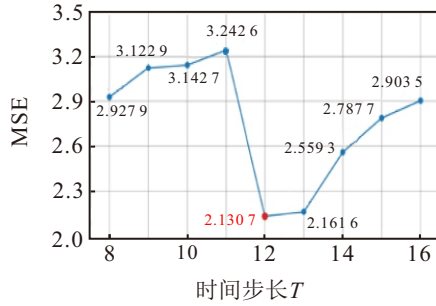


图6 MSE 随时间步长 T 变化曲线

3.1.2 对比实验

为验证 DS-MRGRU 模型的优越性, 与 5 种相关先进深度学习网络模型进行对比实验. 其中, p -Norm-GRU 模型的超参数 p 在区间 $[1, 5]$ 内以 0.1 为步长进行取值; TS- λ GRUs 模型的超参数 λ_1 和 λ_2 , 以及 L/S -ConvGRU 模型的超参数 L 和 S , 均从集合 $\{0.1, 0.2, \dots, 1\}$ 中选取. 经实验分析后, 最终选定的超参数为 $p = 1.6$, $\lambda_1 = 0.8$, $\lambda_2 = 0.7$, $L = S = 0.8$.

如表 1 所示, 与 GRU 相比, 其他模型预测性能均显著提高, 表明打破 GRU 隐藏状态的线性约束关系能有效提高其信息传递效率和学习能力.

表1 各模型数值仿真预测结果

算法模型	MSE	MAE	R^2
GRU	4.4181	1.6964	0.6755
p -Norm-GRU	4.1865	1.6559	0.6925
TS- λ GRUs	3.6059	1.4942	0.7351
L/S -ConvGRU	3.7738	1.5614	0.7228
CNN-FRU	3.6300	1.5154	0.7334
DS-MRGRU	3.4662	1.4613	0.7454

p -Norm-GRU、TS- λ GRUs 和 L/S -ConvGRU 因引入额外超参数, 其调优时间显著增加, 且与超参数设置精细度呈正相关. CNN-FRU 具有良好的预测性能, 但 CNN 的引入显著增加了计算复杂度和超参数优化难度. 因此, 在合理权衡模型调优效率与预测精度的同时, 往往难以对超参数进行精细调节, 制约了模型预测性能进一步优化.

与上述模型相比, DS-MRGRU 在不增加超参数的前提下, 利用双流信息提取结构显著提高了信息捕捉能力, 展现出较高的预测性能. 如图 7 所示, 所提出模型预测误差的分布范围明显更小, 表明其在

预测精度和稳定性方面具有显著优势.

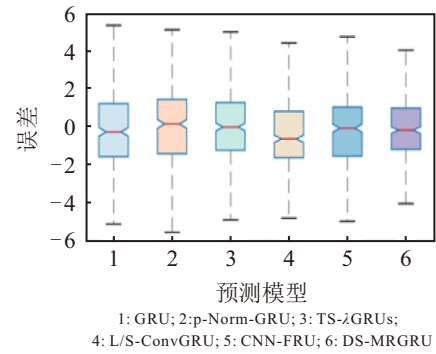


图7 各模型预测误差箱线图

3.1.3 注意力权重量化分析

所提出模型通过双阶段注意力机制动态调整注意力权重, 以实现空间和时序维度有效性的有效评估. 本文通过可视化模型内部注意力权重分布, 定量分析不同特征和各时刻隐藏状态对模型预测性能的具体贡献. 双阶段注意力权重依据式 (8) ~ (15) 进行前向传播计算, 并通过网络反向传播过程不断优化, 以提升模型预测精度和泛化性能.

如图 8 所示, 输入变量 $x_1 \sim x_5$ 相较于冗余变量 $x_6 \sim x_{10}$ 具有更高注意力权重, 表明其在预测过程中发挥着关键作用, 该结论与数值仿真预设契合, 验证了输入变量重要性评估的合理性和准确性.

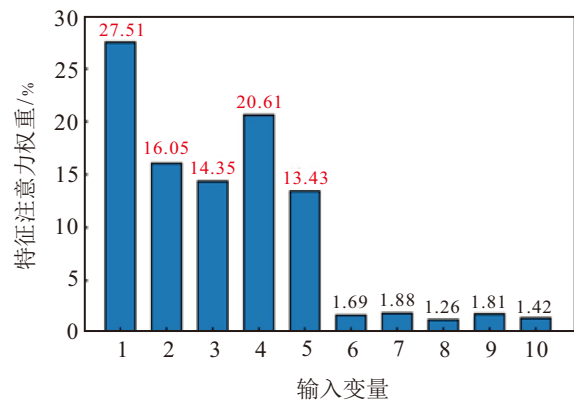


图8 特征注意力权重

为验证时序注意力能否有效过滤长时刻历史冗余信息, 将时间步长 T 设为 13. 图 9 展示了时序注意力权重的可视化结果, 其中横坐标 0 表示当前时刻, 12 表示最远历史时刻. 首先, 双信息流对近 7 个时刻的隐藏状态均赋予较高权重, 表明其能准确捕捉关键信息; 其次, 第 7 时刻 ~ 第 10 时刻虽未直接包含关键信息, 但因隐藏状态的递推计算特性, 其仍保留部分第 11 时刻的重要信息, 因此获得了一定的权重分配; 然后, 时序注意力机制通过为第 11 时刻状态分配适当权重, 有效缓解了长时序依赖中的信息衰减问题; 最后, 模型对第 12 时刻赋予极低权重, 准确

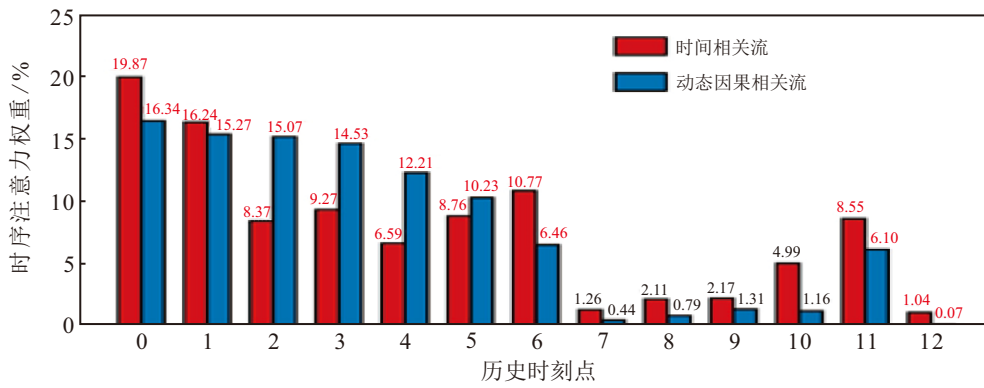


图9 时序注意力权重

识别并过滤了冗余信息. 这些结果验证了时序注意力机制在长时序建模中的有效性.

3.1.4 消融实验

为评估所提出模型中各子模块的独立贡献, 本文进行了消融实验. 表2的预测性能指标数据及图10的极坐标误差分布表明, MRGRU相较于GRU在各项性能指标上表现更优, 其经辅助记忆门的调控, 有效提升了信息捕获能力. DS-MRGRU利用双信息流提取结构, 丰富了信息流的传递, 进一步提高了预测

表2 消融实验中数值仿真预测结果

算法模型	MSE	MAE	R^2
GRU	4.4181	1.6964	0.6755
MRGRU	3.7035	1.5293	0.7280
DS-MRGRU	3.4662	1.4613	0.7454
TA-DS-MRGRU	2.8591	1.3312	0.7900
FA-DS-MRGRU	2.9313	1.3472	0.7847
FTA-DS-MRGRU	2.1616	1.1356	0.8412

性能. 此外, FTA-DS-MRGRU具有最低的MSE、MAE和最高的 R^2 , 且其预测误差幅值最小, 进一步验证了其预测准确性和鲁棒性. 综上所述, 实验结果充分验证了FTA-DS-MRGRU模型的优越性.

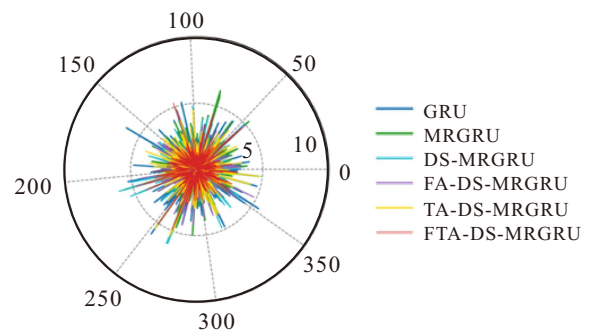


图10 数值仿真极坐标预测误差分布

3.2 火电厂脱硫工艺案例

以山东某电厂脱硫过程为案例, 利用所提出软测量模型预测烟气脱硫过程排放的SO₂浓度. 该电

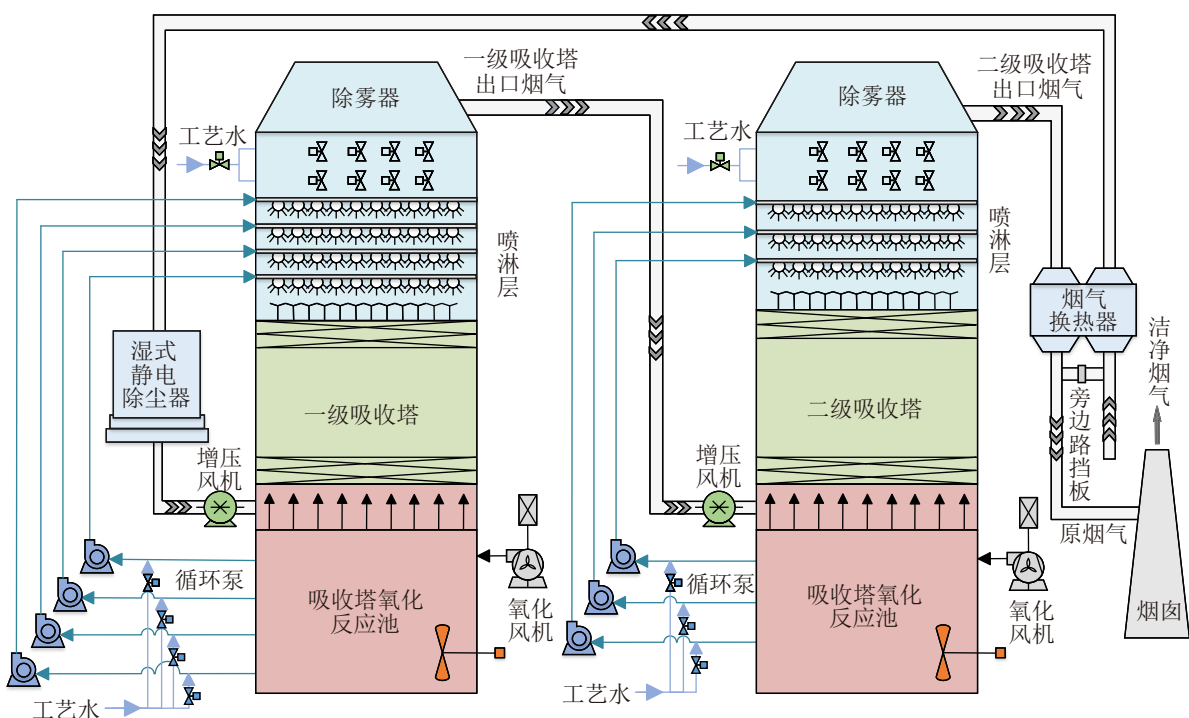


图11 烟气吸收脱硫工艺流程

厂采用石灰石-石膏脱硫技术: 首先, 石灰石经处理形成石灰石浆液, 并进入吸收塔的氧化反应池; 随后, 该浆液与工艺水结合, 形成循环浆液进入喷淋层, 与烟气接触生成石膏浆液; 最后, 净化处理后的烟气经烟囱排出^[18]. 烟气吸收脱硫工艺如图 11 所示. 洁净烟气中 SO₂ 浓度直接反应了脱硫效果, 精确监测该参数对评估脱硫质量至关重要.

该电厂采用的双塔脱硫结构在高效去除烟气中 SO₂ 的同时, 也显著增加了系统的复杂性. 一方面, 两级吸收塔间存在较多相互耦合的过程变量, 且各变量重要性各异, 若未经重要性评估而直接建模, 则将大幅降低模型的预测精度. 另一方面, 较长的采样管道导致测量系统间存在未知时间延迟, 使得重要信息在传递过程中逐渐衰减^[19]. 因此, 合理评估特征重要性并考虑时序信息的衰减, 对提高脱硫过程软测量建模的预测性能至关重要.

所采数据集共 2676 组样本, 采样周期为 1 min, 包括吸收塔出口烟气 SO₂ 浓度、石灰石浆液吸收塔流量、循环浆泵电流等 18 个过程变量和 1 个质量变量——洁净烟气中 SO₂ 浓度.

3.2.1 参数选择实验

为分析火电厂脱硫工艺中时间步长 T 对 FTA-DS-MRGRU 模型预测性能的影响, 进行参数选择实验. 如图 12 所示, $T = 7$ 时模型的 MSE 最小, 预测性能最佳, $T = 3$ 为局部最优点. 当 T 较小时, 有效输入信息不足, 模型难以捕捉到复杂的历史依赖关系. 当 T 取值过大时, 输入包含更多长时刻历史冗余信息, 增加了模型的训练难度和过拟合风险, 导致其预测性能下降.

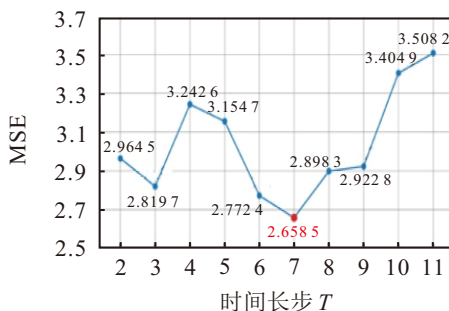


图12 MSE 随时间步长 T 变化曲线

3.2.2 对比实验

为验证 DS-MRGRU 模型在火电厂脱硫数据集的有效性, 与相关模型进行对比实验. 各模型超参数选择范围与数值仿真案例相同, 最终选择结果为 $p = 2$, $\lambda_1 = 0.6$, $\lambda_2 = 0.8$, $L = S = 0.7$.

如表 3 所示, 其他模型的预测性能相比 GRU 均

有所提升, 且 DS-MRGRU 展现出最优性能指标. 此外, 尽管其他模型的预测性能与所提出模型接近, 但其额外超参数难以精细调节, 增加了调优过程复杂度, 限制其预测性能达到最佳.

表3 各网络模型 SO₂ 浓度预测结果

算法模型	MSE	MAE	R ²
GRU	4.5930	1.7045	0.7756
p -Norm-GRU	4.0142	1.6201	0.8039
TS- λ GRUs	3.9228	1.6071	0.8084
L/S -ConvGRU	3.8515	1.5795	0.8119
CNN-FRU	3.9543	1.6295	0.8068
DS-MRGRU	3.7965	1.5703	0.8146

各模型预测散点如图 13 所示, 所提出模型的数据点更紧密地聚集于 45 度线附近, 表明其预测结果更贴近 SO₂ 浓度真实值.

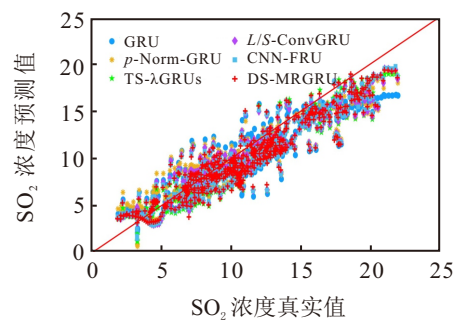


图13 各模型预测结果散点图

3.2.3 注意力权重量化分析

本文通过可视化双阶段注意力权重, 深入探究 SO₂ 浓度预测过程中各输入变量及不同时刻隐藏状态的重要性. 图 14 量化了各输入变量在预测 SO₂ 浓度时的重要程度. 其中, 第 2、第 3、第 11 和第 13 个变量的注意力权重占比均超过 10%, 分别对应一级吸收塔出口烟气 SO₂ 浓度、一级吸收塔石膏浆液 pH 值、炉烟囱入口净烟气压力以及机组脱硫效率中间值. 这些变量在脱硫过程中影响显著: 吸收塔出口

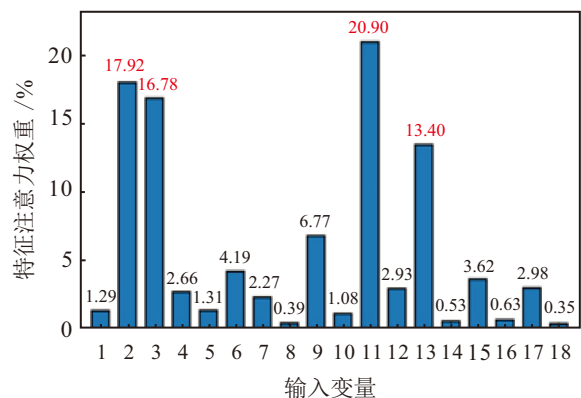


图14 特征注意力权重

烟气 SO₂ 浓度直接影响脱硫效果; 石膏浆液 pH 值关系到化学反应环境和脱硫效率; 净烟气压力反映烟气流动状态与系统负荷; 脱硫效率中间值则直接体现实时脱硫效果. 因此, 这 4 个变量对 SO₂ 浓度预测具有关键影响, 是优化脱硫工艺和提高预测准确性的重要考虑因素.

根据参数选择实验结果, 将时间步长 T 设为 7. 图 15 量化了各历史时刻点隐藏状态在预测 SO₂ 浓度时的重要程度, 横坐标 0 表示当前时刻点, 6 为最远历史时刻点. 由图 15 可知, 越靠近当前时刻点的隐藏状态获得越高注意力权重, 表明其携带的关键信息更多, 对预测结果影响更大. 这既验证了时序注意力捕捉时间序列关键信息的有效性, 也说明在电厂脱硫过程中, 近几个时刻的采样数据更能准确反映系统的动态变化和运行状态.

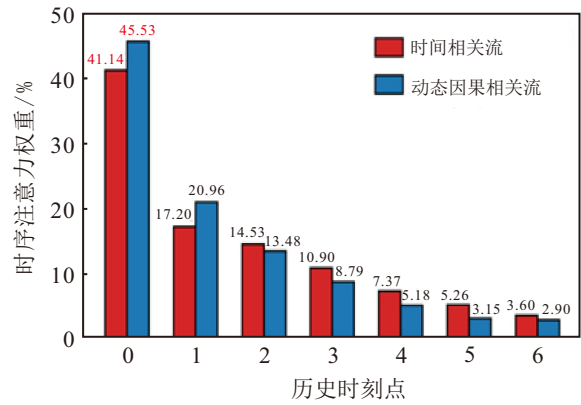


图15 时序注意力权重

3.2.4 消融实验

为验证所提出模型的优越性, 开展消融实验. 图 16、图 17 和表 4 分别展示了 SO₂ 浓度预测拟合曲线、误差频次分布以及性能指标数据. 图 17 中标

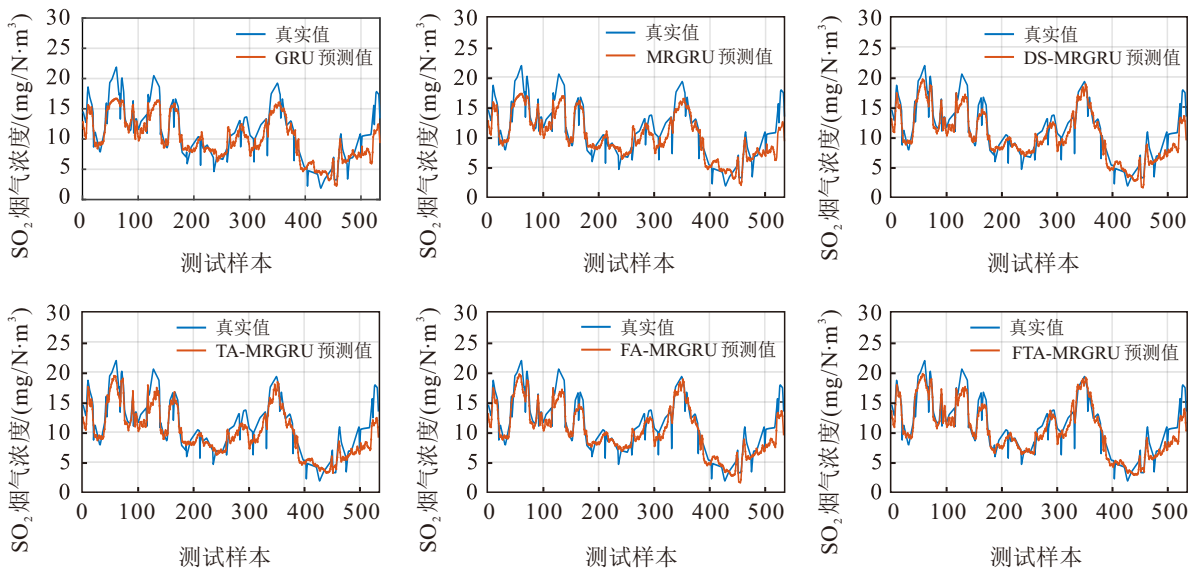


图16 消融实验中 SO₂ 浓度预测拟合曲线

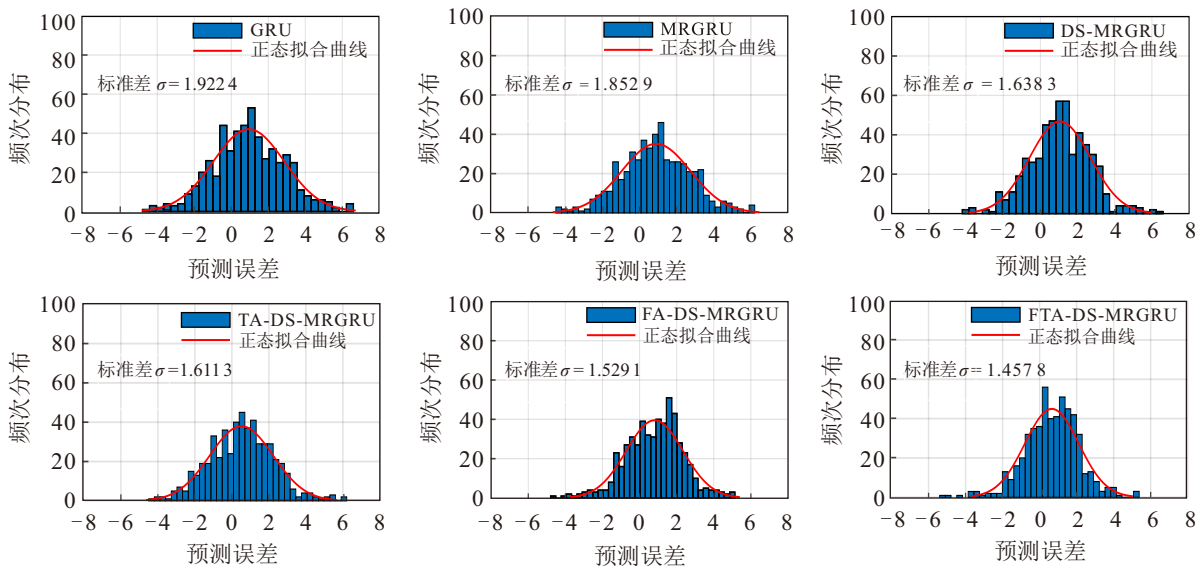


图17 消融实验中 SO₂ 浓度预测误差频次分布

准差 σ 反应了预测误差的离散程度, 其值越小表示模型的稳定性和可靠性越高。

表4 消融实验中 SO_2 浓度预测结果

算法模型	MSE	MAE	R^2
GRU	4.5930	1.7045	0.7756
MRGRU	4.2388	1.6403	0.7929
DS-MRGRU	3.7965	1.5703	0.8146
TA-DS-MRGRU	3.2122	1.4313	0.8431
FA-DS-MRGRU	2.9951	1.4017	0.8537
FTA-DS-MRGRU	2.6585	1.2765	0.8701

相较于传统 GRU, MRGRU 通过构建时间门和因果门, 实现了信息传递的精确调控, 有效提升了预测性能. DS-MRGRU 在此基础上构建并行的时间相关流和动态因果相关流, 显著增强了特征提取能力. FTA-DS-MRGRU 进一步融合双阶段注意力机制: 特征注意力优化辅助变量权重分配, 时序注意力机制专注提取重要历史时刻信息, 显著提升了模型的预测精度和鲁棒性. 实验结果充分展示了所提模型的优越性.

4 结论

本文提出一种基于双阶段注意力的 DS-MRGRU 软测量模型, 以解决实际工业过程中易产生冗余信息与长序列信息衰减问题. 首先, 在 GRU 网络中引入时间门和因果门, 构建 MRGRU 网络以有效改变隐含状态的线性约束关系, 并通过构建双信息流结构提升其信息提取能力; 其次, 在特征提取和预测输出阶段分别引入特征和时序注意力机制, 强化关键变量提取并捕捉重要历史信息, 从而提高模型的预测精度. 数值仿真与某火电厂烟气脱硫过程 SO_2 浓度预测实验表明, 所提出模型较现有深度学习方法具有更优的预测性能, 能够为实际工业过程的优化控制提供有力的技术支撑. 后续研究应着重关注模型的超参数调优和过拟合问题, 以期在实际应用中实现更高的预测精度和更强的鲁棒性.

参考文献 (References)

[1] 熊伟丽, 孙文心, 马君霞. 带自相关约束的 NARX 动态软测量模型[J]. 控制与决策, 2020, 35(4): 816-822. (Xiong W L, Sun W X, Ma J X. Autocorrelation constrained NARX dynamic soft sensing model[J]. Control and Decision, 2020, 35(4): 816-822.)

[2] Yuan X F, Li L, 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, 2021, 68(5): 4404-4414.

[3] 孙凯, 隋璘, 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[J]. 控制理论与应用, 2023, 40(1): 83-93. (Sun K, Sui L, Zhang F F, et al. Dynamic soft sensor algorithm based on nonnegative garrote and long short-term memory neural network[J]. Control Theory & Applications, 2023, 40(1): 83-93.)

[4] 金怀平, 王建军, 董守龙, 等. 基于深度学习特征提取与多目标优化集成修剪的选择性集成学习软测量方法[J]. 控制与决策, 2023, 38(3): 738-750. (Jin H P, Wang J J, Dong S L, et al. Selective ensemble learning for soft sensor development based on deep learning for feature extraction and multi-objective optimization for ensemble pruning[J]. Control and Decision, 2023, 38(3): 738-750.)

[5] 韩鹏东, 阎高伟, 任密峰, 等. 基于迁移子空间学习的偏最小二乘回归软测量方法[J]. 控制与决策, 2023, 38(11): 3147-3155. (Han P D, Yan G W, Ren M F, et al. Partial least squares regression soft sensor method based on transfer subspace learning[J]. Control and Decision, 2023, 38(11): 3147-3155.)

[6] Wu J, Cui X C, Meng J H, et al. Data-driven transfer-stacking-based state of health estimation for lithium-ion batteries[J]. IEEE Transactions on Industrial Electronics, 2024, 71(1): 604-614.

[7] Yuan X F, Qi S B, Wang Y L. Stacked enhanced auto-encoder for data-driven soft sensing of quality variable[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(10): 7953-7961.

[8] Sui L, Sun W, Liu W, 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.

[9] Pham T, Tran T, Phung D, et al. Faster training of very deep networks via p-norm gates[C]. The 23rd International Conference on Pattern Recognition. Cancun, 2016: 3542-3547.

[10] Xie R M, Hao K R, Huang B, et al. Data-driven modeling based on two-stream λ gated recurrent unit network with soft sensor application[J]. IEEE Transactions on Industrial Electronics, 2020, 67(8): 7034-7043.

[11] Sun C, Zhang Y X, Huang G L, et al. A soft sensor model based on long&short-term memory dual pathways convolutional gated recurrent unit network for predicting cement specific surface area[J]. ISA Transactions, 2022, 130: 293-305.

[12] Sun C, Zhang Y, Zhang Y X, et al. A soft sensor model for predicting cement-specific surface area based on convolution fast recurrent unit network[J]. IEEE Sensors

- Journal, 2023, 23(1): 425-433.
- [13] Niu Z Y, Zhong G Q, Yue G H, et al. Recurrent attention unit: A new gated recurrent unit for long-term memory of important parts in sequential data[J]. Neurocomputing, 2023, 517: 1-9.
- [14] Li Y, Li N, Ren J Z, et al. An interpretable light attention-convolution-gate recurrent unit architecture for the highly accurate modeling of actual chemical dynamic processes[J]. Engineering, 2024, 39: 104-116.
- [15] An N, Hong S, Kim Y, et al. Dual attention-based multi-step ahead prediction enhancement for monitoring systems in industrial processes[J]. Applied Soft Computing, 2023, 147: 110763.
- [16] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [17] Sui L, Sun K, Ma J X, 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: 3537611.
- [18] Sun K, Wu X L, Xue J Y, 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.
- [19] 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]. 控制理论与应用, DOI: [10.7641/CTA.2024.30565](https://doi.org/10.7641/CTA.2024.30565).
(Sui L, Ma J X, Xiong W L. Gated recurrent unit network based on attention garrote and its application for industrial soft sensors[J]. Control Theory & Applications, DOI: [10.7641/CTA.2024.30565](https://doi.org/10.7641/CTA.2024.30565).)

作者简介

廖开继 (2001-), 男, 硕士生, 主要研究方向为复杂工业过程建模, E-mail: 15150241883@163.com;

隋璘 (1996-), 男, 博士生, 主要研究方向为机器学习、软测量建模, E-mail: suilin359@163.com;

熊伟丽 (1978-), 女, 教授, 博士, 博士生导师, 主要研究方向为智能软测量技术及应用、基于数据的工业过程监测, E-mail: weili_xiong@jiangnan.edu.cn.