

基于 CNN-BiGRU-Attention 的脱硫系统出口 二氧化硫动态预测模型^{*}

高 钾^{1,2} 田雪峰^{1,2} 蒋甲丁¹ 彭献永² 周怀春²

(1. 新疆工程学院能源工程学院 乌鲁木齐 830023; 2. 中国矿业大学低碳能源与动力工程学院 徐州 221116)

摘 要: 针对深度调峰下湿法烟气脱硫系统出口 SO_2 排放浓度难以精准测量, 提出一种基于动态建模的卷积神经网络(CNN)、双向门控循环单元(BiGRU)和注意力机制(Attention)相结合的 SO_2 浓度预测模型。基于核主成分分析对原始数据进行关键特征筛选得到 7 个特征变量作为输入特征。将注意力机制引入 CNN 和 BiGRU, 建立脱硫系统出口 SO_2 浓度预测模型, 并以某在役超临界 600 MW 机组的脱硫系统为研究对象进行了仿真实验。仿真结果显示, 本文所建立动态模型的平均绝对误差为 $0.706\ 4\ \text{mg}/\text{m}^3$ 、均方根误差为 $0.912\ 5\ \text{mg}/\text{m}^3$ 、平均相对误差为 6.27% , 与 CNN-BiGRU 相比, 分别下降了 25.07% 、 23.45% 和 17.28% , 更低于 CNN 和 BiGRU; 动态模型的决定系数为 96.74% , 与 CNN-BiGRU、BiGRU 和 CNN 模型相比, 分别提高了 3.91% 、 5.26% 和 9.66% 。表明基于 CNN-BiGRU-Attention 的动态模型具有较高的预测精度和学习能力, 能够准确的预测脱硫系统出口 SO_2 浓度变化趋势。

关键词: 湿法烟气脱硫系统; SO_2 浓度预测; 动态建模; 核主成分分析; CNN-BiGRU-Attention

中图分类号: X773; TN06 **文献标识码:** A **国家标准学科分类代码:** 470.30

Dynamic prediction model of sulphur dioxide concentration for WFGD system based on CNN-BiGRU-Attention

Gao Jia^{1,2} Tian Xuefeng^{1,2} Jiang Jiading¹ Peng Xianyong² Zhou Huaichun²

(1. School of Energy Engineering, Xinjiang Institute of Engineering, Urumqi 830023, China; 2. School of Low-Carbon Energy and Power Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: A dynamic prediction model of sulfur dioxide concentration is proposed to address the challenge of accurate measurement of sulfur dioxide emission concentration at the exit of a limestone-gypsum wet flue gas desulfurization system under deep peaking. The model integrates a convolutional neural network (CNN), a bidirectional gated recurrent unit (BiGRU), and an attention mechanism to predict the sulfur dioxide concentration. The model develops utilizes kernel principal component analysis to determine seven characteristic variables, which are then used as inputs to the model. The attentional mechanism is combined with CNN and BiGRU to construct a model for predicting SO_2 concentration at the outlet of the desulfurization system. A simulation experiment is conducted with the FDG system of an operating 600 MW supercritical unit as the research object. The simulation results demonstrate that the average absolute error MSE of the dynamic model established in this paper is $0.706\ 4\ \text{mg}/\text{m}^3$, the root mean square error RMSE is $0.912\ 5\ \text{mg}/\text{m}^3$, and the average relative error is 6.27% , which is 25.07% , 23.45% , and 17.28% lower compared with CNN-BiGRU, and even lower than CNN and BiGRU; The coefficient of determination of the dynamic model was 96.74% , which was 3.91% , 5.26% , and 9.66% higher than CNN-BiGRU, BiGRU, and CNN models respectively. This outcome indicates that the dynamic model based on CNN-BiGRU-Attention exhibits high prediction accuracy and learning ability, and can accurately predict the trend of SO_2 concentration at the outlet of the desulfurization system.

Keywords: wet flue gas desulfurization system; SO_2 concentration prediction; dynamic modeling; KPCA; CNN-BiGRU-Attention

0 引 言

为减轻环境污染,减少燃煤电厂排放的 SO_2 含量是电

力行业迫切需要应对的挑战^[1]。而且,随着“双碳”目标的推进,煤电正向调节型和安全保障型电源功能转变,机组负荷波动显著增加,导致湿法烟气脱硫(wet flue gas

收稿日期:2025-01-03

^{*} 基金项目:国家自然科学基金(72361033)、2024 新疆维吾尔自治区重大科技专项-全烧高碱煤下的锅炉快速变负荷研究(2024A01005-1)、中国华能集团公司科技项目(HNKJ23-HF31)资助

desulfurization, WFGD) 系统面临 SO_2 排放浓度难以精准调控的问题^[2-3]。因此,建立一种能够实时动态预测 SO_2 排放浓度的模型,对于实现前馈调节和优化脱硫系统运行具有重要意义。然而,由于炉膛燃烧过程的复杂性及 SO_2 浓度检测系统与煤质信息之间的非线性关系,基于传统机理模型的 SO_2 浓度预测方法面临挑战。

近年来,随着人工智能与大数据技术迅速发展,基于数据驱动的预测技术在燃煤发电厂污染物排放领域崭露头角。例如,马双忱等^[4]和金秀章等^[5]通过引入长短期记忆网络,专门针对脱硫系统的时序特性构建 SO_2 预测模型,有效缓解了因系统大惯性和延迟带来的挑战。郑茗友等^[6]采用了粒子群优化的套索算法,进一步提升了 SO_2 浓度预测的准确性。吴磊等^[7]和金秀章等^[8]利用改进粒子群算法优化长短期记忆神经网络模型,显著提高了脱硫出口 SO_2 浓度的预测精度。金秀章等^[9]提出了一种基于变量筛选和在线核极限学习机的出口 SO_2 浓度预测模型,性能优于极限学习机和核极限学习机算法。闫浩思等^[10]结合核极限学习机与特征选择策略,为 SO_2 排放监管提供了新思路。冯齐全等^[11]提出一种基于高斯混合模型-支持向量回归的 SO_2 浓度预测模型,各项评价指标较单模型最小二乘支持向量回归机有较大改善。孙坤等^[12]构建了卷积神经网络与双向门控循环单元相结合的深度学习模型,在 SO_2 预测上达到了均方根误差 0.699 6 和平均绝对误差 0.355 3,优于其他基准模型。

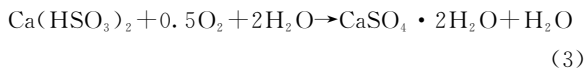
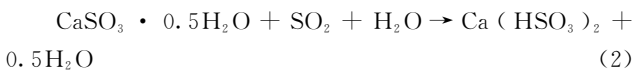
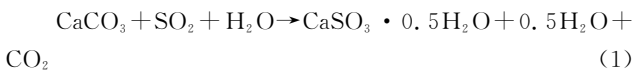
在国际研究方面,学者们也积极探索了基于数据驱动的 SO_2 浓度预测方法。例如, Yang 等^[13]利用最小二乘支持向量机方法建立了燃煤锅炉氮氧化物实时动态模型,取得了良好的效果。Ding 等^[14]基于电厂真实数据开发了基于梯度提升决策树的模型来预测氮氧化物的排放量,具有一定的工程价值。Li 等^[15]利用随机森林算法建立了静态模型、分段模型和动态模型,用于预测 SO_2 浓度,取得了良好的效果。Alnaim 等^[16]利用长短期记忆网络建立了 660 兆瓦四角切圆燃煤锅炉的氮氧化物动态排放预测,提高了预测精度。Yang 等^[17]利用主成分分析消除变量之间的耦合,并使用 Adam 策略优化器为长短期记忆网络建立氮氧化物排放模型,提升模型的泛化能力,提高了预测的精度和稳定性。尽管国内外的研究在提升预测精度方面取得了一定成就,但仍存在未能充分利用数据局部特征及模型收敛性不足的问题,从而限制了系统出口 SO_2 浓度预测精度的进一步提升。

为克服现有研究的局限性,本文提出了一种融合卷积神经网络(convolution neural network, CNN)、双向门控循环单元(bi-directional gated recurrent unit, BiGRU)以及注意力机制(Attention)的深度学习模型。该模型通过 CNN 自动捕获数据中的复杂非线性特征,利用 BiGRU 深入分析时间序列特性,并借助 Attention 机制甄别并加重对预测结果有显著贡献的特定时间点特征权重,从而实现对 WFGD 出口 SO_2 浓度的精准预测。

1 原理介绍

1.1 石灰石-石膏湿法脱硫系统

WFGD 系统中的烟气 SO_2 脱除过程是一个化学反应过程,主要在吸收塔中完成。如图 1 所示,烟气首先进入吸收塔底部,并向上流动与自上而下喷洒的石灰石浆液逆向相遇。石灰石浆液源自下方的浆液池,经由循环泵的推送抵达顶部的喷淋设备。期间,石灰石浆液内的主要成分 CaCO_3 与烟气里的 SO_2 发生化学作用,生成 CaSO_3 落入吸收塔底部的浆料罐中。然后,用氧化风扇吹出新鲜空气,迫使 CaSO_3 氧化成硫酸钙。最后,浆液罐底部的硫酸钙沉淀通过排出泵排出至脱水系统。硫酸钙沉淀物通过带式输送机脱水,形成石膏($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$)^[18-19],主要化学反应如下:



尽管 WFGD 系统在 SO_2 脱除方面表现出较高的效率,但在实际运行中, SO_2 浓度的精确测量和调控仍面临诸多挑战。以某 600 MW 超临界机组的脱硫系统为例,在深度调峰工况下,烟气流量和 SO_2 浓度波动显著,导致监测设备出现较大偏差^[20]。从图 2 可得:在负荷快速变化时段(08:00~12:00),监测值滞后于实际值,最大偏差达到 18.5%;在低负荷运行时段(00:00~06:00),监测值与实际值的偏差相对较小,但仍存在 5%~8% 的误差。

因此,开发一套能够实时、精准预测 WFGD 出口 SO_2 浓度的模型,对于优化脱硫系统运行、降低排放超标风险具有重要意义。本文提出的 CNN-BiGRU-Attention 动态预测模型,正是针对上述复杂工况下的关键挑战而设计,旨在提高 SO_2 浓度预测的精度。

1.2 混合模型

1) 卷积神经网络

卷积神经网络是一种专门用于处理具有网格状结构数据的神经网络,由输入层、卷积层、激活层、池化层和全连接层组成^[21-22],如图 3 所示。

(1) 卷积层

卷积层是 CNN 的核心部分,每个卷积层由多个卷积核组成,每个卷积核像一个滤波器,主要功能是对输入数据进行特征提取。借助卷积运算,卷积层能够有效捕获数据中的重要特征,并加强原始信号中的关键特征,同时有效减少特征的维度^[21]。如图 4 所示为一个 4×4 的输入通过一个 3×3 大小卷积核后输出结果的过程。卷积核每次向右移动 1 个单位,同时与数据窗口做内积。以第一个窗口为例,计算过程如式(4)所示。

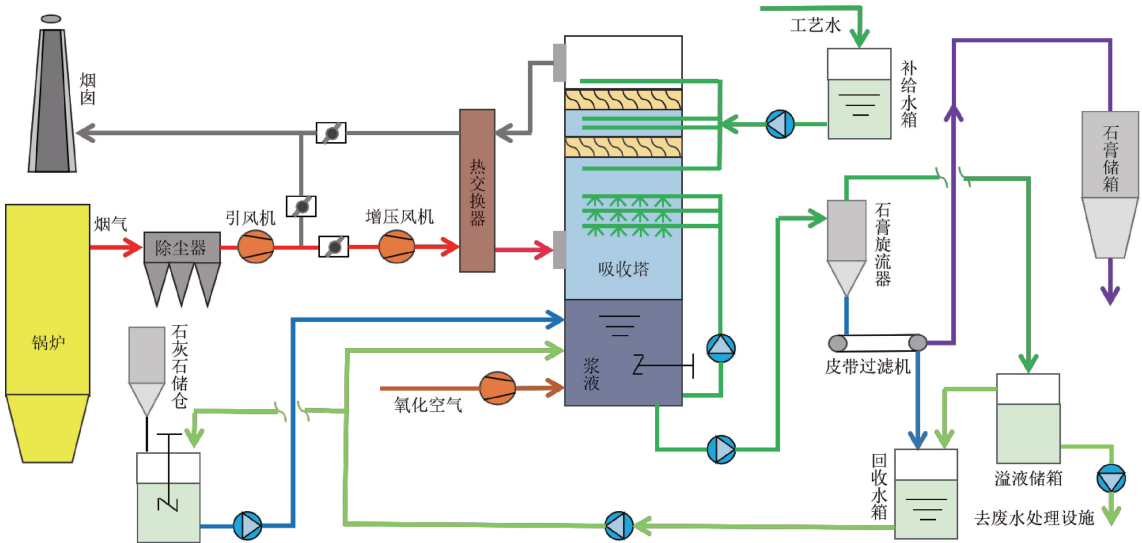


图 1 石灰石-石膏湿法脱硫工艺流程图

Fig. 1 Process flow diagram of Limestone-gypsum wet desulfurization

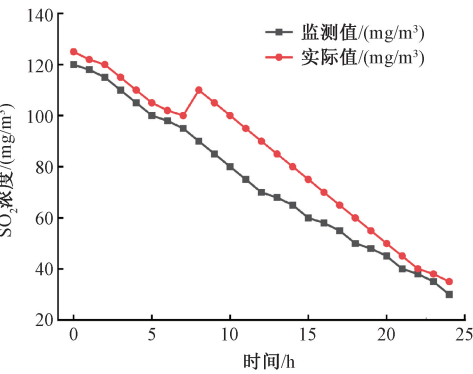


图 2 某 600 MW 机组系统 SO₂ 浓度监测值与实际值对比
Fig. 2 Comparison between monitored and actual SO₂ concentration in a 600 MW unit's desulfurization system

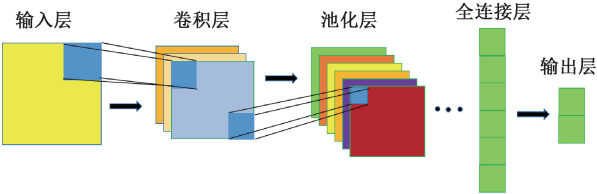


图 3 典型卷积结构图

Fig. 3 CNN model structure diagram

$$1 \times 1 + 0 \times 0 + 2 \times 1 + 3 \times 0 + 2 \times 1 + 0 \times 0 + 4 \times 1 + 2 \times 0 + 2 \times 1 = 11$$

(4)

在 CNN 中,每次卷积完一个数据矩阵内的部分数据之后,数据矩阵就会不停地平行运动,直到卷积完所有的数据。从图 3 可以看到,通过 3×3 的卷积核作用,原始二维数据从最初 4×4 变成了 2×2 的二维矩阵。

(2) 池化层

池化层一般连接在卷积层之后,主要作用是对数据进

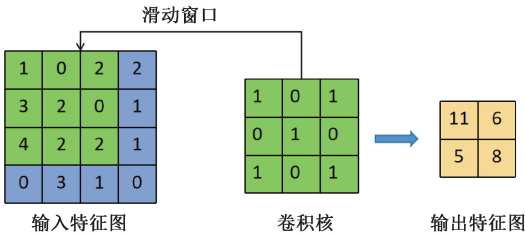


图 4 卷积运算的过程

Fig. 4 The process of convolution operation

行二次提取,减少预测所需的数据,从而降低网络的复杂度。池化操作主要有平均池化、最大池化等。平均池化是取出一个区域范围内的平均值,最大池化是取出一个区域范围内的最大值,如图 5 所示^[21-22]。

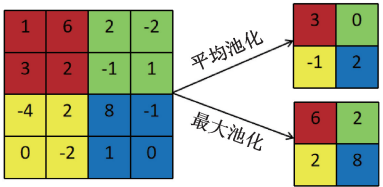


图 5 池化层卷积图

Fig. 5 Convolutional map of pooling layer

(3) 激活函数

激活层通过对全连接层的输出或卷积层的输出应用激活函数,使模型的非线性能力增强。常用的激活函数有 sigmoid 函数、ReLU 函数和 tanh 函数,公式如式(5)~(7)所示^[22]:

$$Sigmoid(x) = \frac{1}{1 + e^{-x}}$$

(5)

$$ReLU(x) = \max(0, x)$$

(6)

$$\tanh(x) = \frac{e^x + e^{-x}}{e^x - e^{-x}}$$

(7)

三者相比,ReLU 函数表现出更强的鲁棒性,具有速度快、稀疏性好、求导特性好等优点,故大部分 CNN 都使用 ReLU 函数作为激活函数。

(4)全连接层

全连接层位于 CNN 末端,和上层结构的神经元相接,实现数据维度变换,同时保留数据的有用信息。其主要作用是学习到前面层输出特征,进行全局整合,并映射到样本的标记空间。

2)双向门控循环神经网络

门控循环单元(gate recurrent unit,GRU)是循环神经网络(recurrent neural network,RNN)的变体,核心机制围绕两个关键门控运作:更新门与重置门^[23]。如图 6 所示,更新门决定过往时刻隐藏状态在当前状态中的保留比例,影响信息的延续性;重置门决定模型如何基于当前输入与前一状态进行信息的整合,从而适应序列中动态变化的需求。GRU 前向传递过程具体如式(8)~(11)所示。

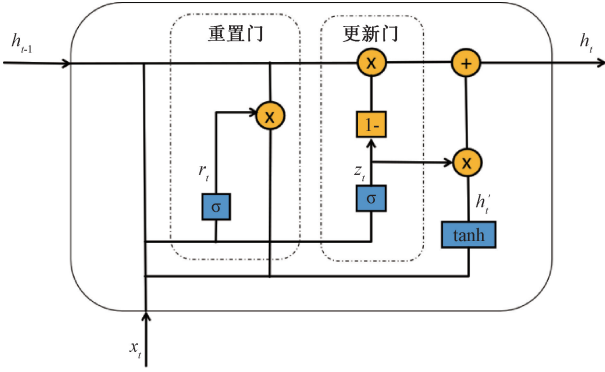


图 6 GRU 结构

Fig. 6 GRU structure

$$r_t = \sigma(W_r \cdot [X_t, h_{t-1}] + b_r) \quad (8)$$

$$z_t = \sigma(W_z \cdot [X_t, h_{t-1}] + b_z) \quad (9)$$

$$h_{xt} = \tanh(W_h \cdot [X_t, r_t \otimes h_{t-1}] + b_h) \quad (10)$$

$$h_t = (1 - z_t) \otimes h_{t-1} + z_t \otimes h_{xt} \quad (11)$$

式(8)~(11)中, r_t 为重置门, X_t 为当前输入, h_{t-1} 为上一时刻隐含信息, σ 为 sigmoid 函数, $[\]$ 表示矩阵堆叠, W_r 为重置门的权重, b_r 为重置门的偏置, z_t 为更新门, W_z 为更新门的权重, b_z 为更新门的偏置, h_{xt} 为候选隐藏状态内容, \tanh 为双曲正切函数, W_h 为候选隐藏状态的权重, b_h 为候选隐藏状态的偏置, h_t 为隐藏状态。

为能够有效捕获序列中过去和未来信息,双向门控循环单元被广泛应用。如图 7 所示,BiGRU 整合两个 GRU 单元,一个沿时间序列正向传播处理信息,另一个反向回溯,从而让模型同时拥有回顾过去与展望未来的能力。每个 BiGRU 单元在时刻 t 的输出,受到该时刻的偏置量、前向 GRU 和后向 GRU 输出的共同作用,公式如下^[23]:

$$\vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \quad (12)$$

$$\overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t+1}) \quad (13)$$

$$h_t = m\vec{h}_t + n\overleftarrow{h}_t + c_t \quad (14)$$

式中: $GRU(\cdot)$ 为 GRU 的计算过程, \vec{h}_t 为 t 时刻 GRU 前向隐藏层的输出值, \overleftarrow{h}_t 为 t 时刻 GRU 后向隐藏层的输出值, m 为 GRU 前向隐藏层输出权重, n 为 GRU 后向隐藏层输出权重, c_t 为 h_t 对应的隐藏层偏置量。

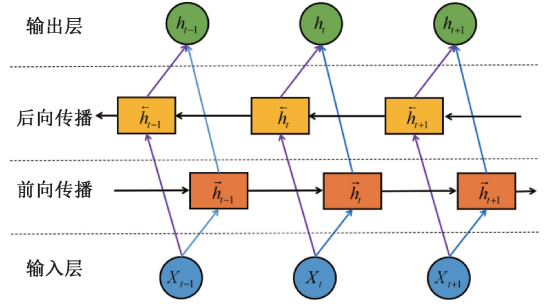


图 7 BiGRU 结构

Fig. 7 BiGRU structure

3)注意力机制

Attention 机制借鉴人类注意力集中处理信息的方式,如同大脑筛选信息的过程。基于注意力机制的深度学习,核心在于让机器学会去感知数据中的重要和不重要的部分,其结构如图 8 所示,计算步骤如下^[24]:

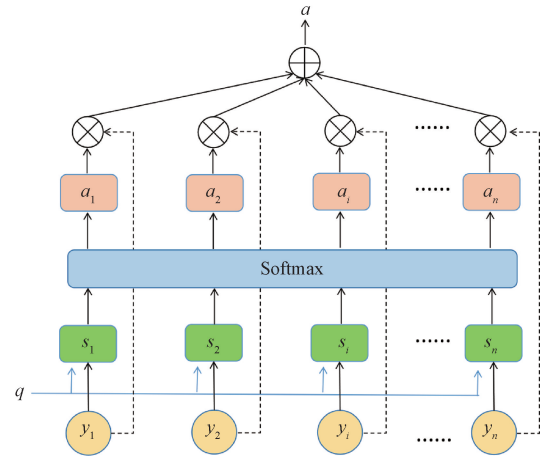


图 8 注意力机制

Fig. 8 Attention mechanism

(1)对输入序列 $[y_1, y_2, \dots, y_n]$ 进行编码,得到一组查询向量 q ;然后通过打分之间的相似度 $[s_1, s_2, \dots, s_n]$,打分函数如下:

$$s_i = V^T \tanh(W y_i + U q) \quad (15)$$

式中: V, W, U 均为参数待学习的矩阵。

(2)利用 $Softmax$ 函数进行归一化,得到每组输入向量 a_i ,公式如下:

$$a_i = \text{softmax}(s_i) = \frac{\exp(s_i)}{\sum_{i=1}^n \exp(s_i)} \tag{16}$$

(3)根据权重和对应的值向量计算加权和,得到最终的隐藏层状态值输出 a ,公式如下:

$$a = \sum_{i=1}^n a_i \tag{17}$$

4)混合模型介绍

CNN 在处理时间序列数据时存在长依赖问题,而 BiGRU 能有效解决这一难题,因此,CNN-BiGRU 模型能结合两者优势,但在处理大规模输入时,可能忽视个别时间点的关键特征,导致模型学习能力下降,影响预测性能。为此,引入 Attention 机制能够突出不同时间步特征对预测结果的贡献,从而提升模型精度。

基于 CNN-BiGRU-Attention 的预测模型结构如图 9 所示,包含输入、CNN、BiGRU、Attention 及输出层。接收特征数据后,CNN 层中的卷积层对输入的数据进行特征提取,池化层对特征数据进行降维处理;BiGRU 层利用双向单元捕捉时间序列特征;接下来,Attention 层根据各时间点特征对预测目标的关联度进行加权;最后,汇总加权特征并经输出层计算得出预测结果。

2 数据采集与预处理

2.1 数据采集

根据 SO₂ 生成和脱除机理分析^[9],与脱硫系统出口

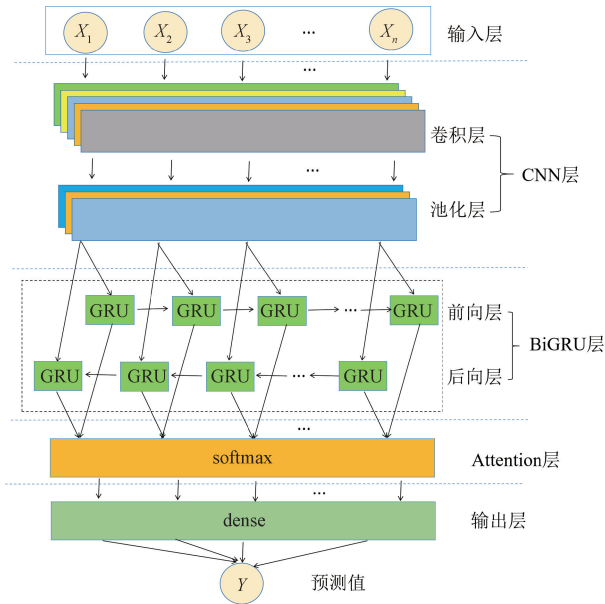


图 9 基于 CNN-BiGRU-Attention 的预测模型结构
Fig. 9 Structure of the prediction model based on CNN-BiGRU-Attention

SO₂浓度有关系的变量有 26 个:机组负荷、总一次风量、总二次风量、WFGD 进出口烟气 O₂ 含量、入口烟气 SO₂ 浓度、WFGD 入口烟气流量、入口烟气温度、进出口粉尘含量等,如表 1 所示。

表 1 变量表

Table 1 List of variables

变量名称	单位	物理意义	变量名称	单位	物理意义
机组负荷	MW	表征电厂实时发电功率	入口烟气温度	℃	影响反应速率和脱硫效率
入口烟气 SO ₂ 浓度	mg/m ³	系统入口污染物负荷	进口烟气温度(喷淋前)	℃	用于监测系统异常工况
脱硫塔 pH 值	—	影响 SO ₂ 的反应效率	氧化风机流量	m ³ /h	影响 CaSO ₃ 氧化效率
吸收塔液位	m	反映浆液系统的稳定性	石灰石给浆密度	kg/m ³	影响 SO ₂ 脱除效率
脱硫效率	%	反映对 SO ₂ 的脱除效果	石灰石耗量	t/h	间接反映 SO ₂ 脱除负荷
石灰石给浆流量	m ³ /h	反映石灰石浆液的供应量	氧化风机功率	kW	影响 CaSO ₃ 氧化效率
入口烟气流量	m ³ /h	影响 SO ₂ 脱除负荷	氧化风机 4 频率	Hz	影响氧化空气供应量
煤质	—	燃煤的硫分、灰分等特性	再循环泵 1 电流	A	影响浆液循环量
煤量	t/h	与 SO ₂ 生成量正相关	再循环泵 2 电流	A	影响浆液循环量
总一次风量	m ³ /h	影响燃烧效率和 SO ₂ 生成	再循环泵 3 电流	A	影响浆液循环量
总二次风量	m ³ /h	影响燃烧效率和 SO ₂ 生成	再循环泵 4 电流	A	影响浆液循环量
进口烟气 O ₂ 含量	%	影响 SO ₂ 脱除效率	进口粉尘含量	mg/m ³	影响脱硫效率和设备磨损
出口烟气 O ₂ 含量	%	间接反映脱硫效率	出口粉尘含量	mg/m ³	用于评估除尘效果

本研究数据来源于浙江省某电厂 1 号机组 DCS 系统,采集周期为 2023 年 1 月~2024 年 7 月,覆盖了春夏秋冬四季及不同负荷工况,包括深度调峰、满负荷运行、计划性维护等阶段。原始数据采样频率为 10 s/次,累计数据量达 1 200 万条。为验证模型在短时动态预测中的性能,本文

选取 2024 年 7 月 18 日 00:00 至 7 月 19 日 00:00 的典型 24 小时数据(共 8 640 个样本)作为核心训练集。如图 10 所示,该时段包含升负荷、降负荷及负荷剧烈波动等多种工况,包含一个完成的变化周期,具有很强的代表性。

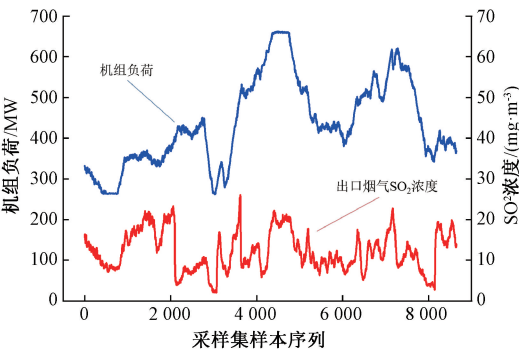


图 10 WFGD 出口 SO₂ 浓度与机组负荷的原始数据

Fig. 10 Plot of raw data of SO₂ concentration at WFGD outlet and unit loading

2.2 数据预处理

1)数据归一化

首先对输入数据进行清洗,并通过以下公式进行线性归一化处理,使数据尺寸标准化。

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{18}$$

式中: X_n 为归一化之后的数据, X 为当前待归一化的数据, X_{\min} 为待归一化总数据的最小值, X_{\max} 为待归一化总数据的最大值。

2)核主成分分析

主成分分析(principal component analysis,PCA)是一种线性降维方法,它通过将高维数据转换到低维空间来减少冗余,同时保留数据的最大方差^[25]。在实际应用中,PCA 面临非线性数据结构时可能引发信息失真。鉴于此,学者们发展了核主成分分析(kernel principal component analysis,KPCA)^[26]。KPCA 的关键创新在于整合了一个非线性变换 $\phi(x)$,它把原始数据集 $X = \{x_1, x_2, \cdots, x_n\}$ 从原始空间映射至一个更高维度的特征空间 H 中,如图 11 所示,不同的符号标记代表不同分类的样本点。此过程中采用的映射函数为:

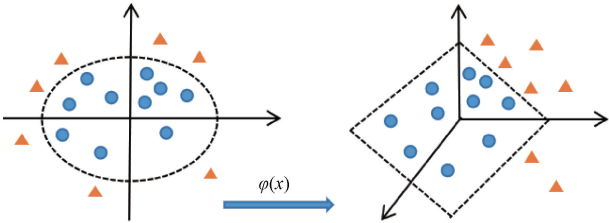


图 11 KPCA 方法基本思路示意图

Fig. 11 Schematic diagram of the basic idea of the KPCA method

$$\varphi: I^m \rightarrow H, x_r \rightarrow \varphi(x_r) \tag{19}$$

$$r = 1, 2, \cdots, n \tag{20}$$

在特征空间 H 内再使用线性的 PCA 对样本集合 $\phi(x_1), \phi(x_2), \cdots, \phi(x_n)$ 进行数据降维,这样问题就变成在特征空间内求解样本协方差矩阵的特征值与特征函数。

样本的协方差矩阵表示为:

$$C^H = \frac{1}{n} \sum_{r=1}^n \varphi(x_r) \varphi^T(x_r) \tag{21}$$

用 λ 表示协方差矩阵 C^H 的特征值, \mathbf{v} 表示对应的特征向量,则有:

$$C^H \mathbf{v} = \lambda \mathbf{v} \tag{22}$$

定义核函数:

$$K_{s,j} = \langle \varphi(x_s) \times \varphi(x_j) \rangle \tag{23}$$

特征向量 \mathbf{v} 在 $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_n)$ 构成的空间内,因此必定存在系数集 $a_s (s=1, \cdots, k)$,使得:

$$\mathbf{v} = \sum_{s=1}^k a_s \varphi(x_s) \tag{24}$$

式中: k 表示组成系数集的子集的数量。

以上公式联立,解得:

$$k \lambda K \mathbf{a} = K^2 \mathbf{a} \tag{25}$$

式中: $\mathbf{a} = [a_1, a_2, \cdots, a_k]^T$,因此,上式化简为:

$$k \lambda \mathbf{a} = K \mathbf{a} \tag{26}$$

经过映射后的数据并非零均值化的结果,因此用 \tilde{K} 代替 K ,对其进行中心化,得到:

$$\tilde{K} = K - I_n K - K I_n + I_n K I_n \tag{27}$$

式中: I_n 为 $n \times n$ 的矩阵且其中元素值均为 $1/n$ 。

3)基于 KPCA 的数据降维

本文选择 KPCA 对数据进行降维,其中核函数选取多项式。表 2 中列举了降维后得到的前 16 个主成分分量方差贡献率和累计贡献率的数值(后面的占比太小,为节省空间,不再列出)。

表 2 KPCA 降维后各主成分方差贡献率及累计贡献率
Table 2 Variance contribution and cumulative contribution of each principal component after dimension reduction in KPCA

主成分分量	方差贡献率	累计贡献率
KPCA1	0.340 471 079	0.340 471 079
KPCA2	0.285 367 538	0.625 838 618
KPCA3	0.063 658 048	0.689 496 665
KPCA4	0.046 265 351	0.735 762 017
KPCA5	0.036 008 721	0.771 770 737
KPCA6	0.026 025 511	0.797 796 249
KPCA7	0.018 066 936	0.815 863 185
KPCA8	0.013 787 487	0.829 650 672
KPCA9	0.013 256 256	0.842 906 928
KPCA10	0.012 519 173	0.855 426 101
KPCA11	0.011 265	0.866 691 101
KPCA12	0.008 915 113	0.875 606 214
KPCA13	0.007 445 899	0.883 052 113
KPCA14	0.006 837 02	0.889 889 133
KPCA15	0.006 618 82	0.896 507 953
KPCA16	0.006 054 069	0.902 562 022
⋮	⋮	⋮

据表 2 可知,前 7 个核主成分的累计贡献率为 0.815 863 185,累计方差贡献率大于 80%,说明前 7 个核主成分基本上涵盖了原始数据的主要信息。另外从方差贡献率的柱状图(图 12)看出,从第 8 个乃至以后的主成分贡献率的图像接近水平,累计方差贡献率的图像也趋近于平缓,说明从第 8 个核主成分开始的贡献率很低,所包含的原始信息量很少。

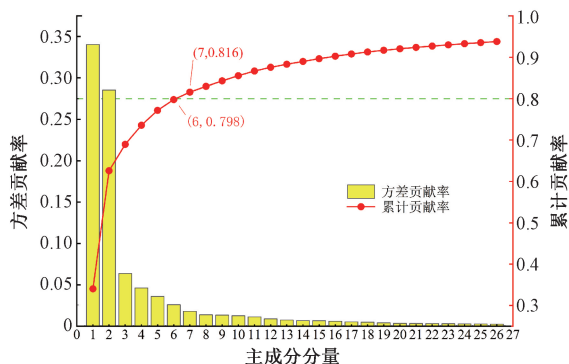


图 12 各主成分方差贡献率及累计贡献率

Fig. 12 The variance contribution rate and the cumulative contribution rate of each principal component

3 预测模型设计

3.1 模型构建

本文构建基于 CNN-BiGRU-Attention 的混合模型对 WFGD 出口 SO_2 浓度进行预测,模型运行流程如图 13 所示。步骤如下:

- 1)对数据进行清洗、归一化处理,并采用核主成分降维,然后划分数据为训练集和测试集,两者比例为 8:2;
- 2)构建 CNN-BiGRU-Attention 的混合模型。另外,该模型整合 Dropout 策略来抑制过拟合,采纳 Adam 优化算法调整模型参数;
- 3)对网络模型参数进行调整,反复训练,直至达到预期效果,保存模型;
- 4)将测试集样本数据输入优化后的模型,获取预测结果,然后利用评价函数对预测结果进行分析和评价,验证模型的有效性。

3.2 模型训练与优化策略

模型使用 MATLAB R2023b 编程实现,所使用的电脑搭载 AMD Ryzen5 3500U 处理器,主频为 1.6 GHz,配备 12.0 GB 内存,显卡的型号 Radeon Vega,16 GB。输入层节点个数为 7,输出层节点个数为 1,隐藏层层数设置为 17,LearnRateDropFactor 为 0.5,初始学习率设为 0.005,批处理量设为 200,LearnRateDropPeriod 为 150,且采用自适应学习率的 Adam 优化算法。调试过程如图 14 所示。

图 14(a)中,当 LearnRateDropFactor 数值为 0.5 时,训练集和测试集上误差都小于其他 4 个参数值对应的误

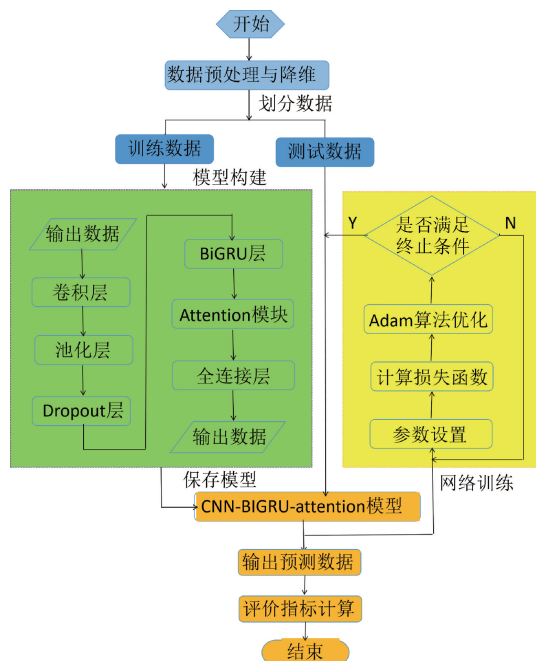


图 13 基于混合模型的 SO_2 预测模型流程图

Fig. 13 Flowchart of the SO_2 prediction model based on hybrid model

差,所以选定 LearnRateDropFactor 参数值为 0.5。图 14(b)中,对比不同 MaxEpochs 值所对应的误差,当 MaxEpochs 为 200 时,训练集和测试集效果都较好,故选取 MaxEpochs 参数值为 200。如图 14(c)所示,当 InitialLearnRate 参数选取 0.005 时,训练集和测试集的误差相比于其他参数时较小,则选取 0.005 为 InitialLearnRate 参数值。从图 14(d)可以看出,当 LearnRateDropPeriod 为 150 的时候,训练集和测试集的误差较小,则选取 150 为 LearnRateDropPeriod 参数值。

4 实验与结果分析

4.1 实验评价指标

本文预测模型的评价指标采用平均绝对误差(MAE)、均方根误差(RMSE)、平均相对误差(MAPE)和决定系数 R^2 ,公式如下^[27]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (28)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (29)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (30)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (31)$$

式中: y_i 为 SO_2 的真实值; \hat{y}_i 为 SO_2 的预测值; \bar{y}_i 为 SO_2 的平均值; n 为预测集数据个数。

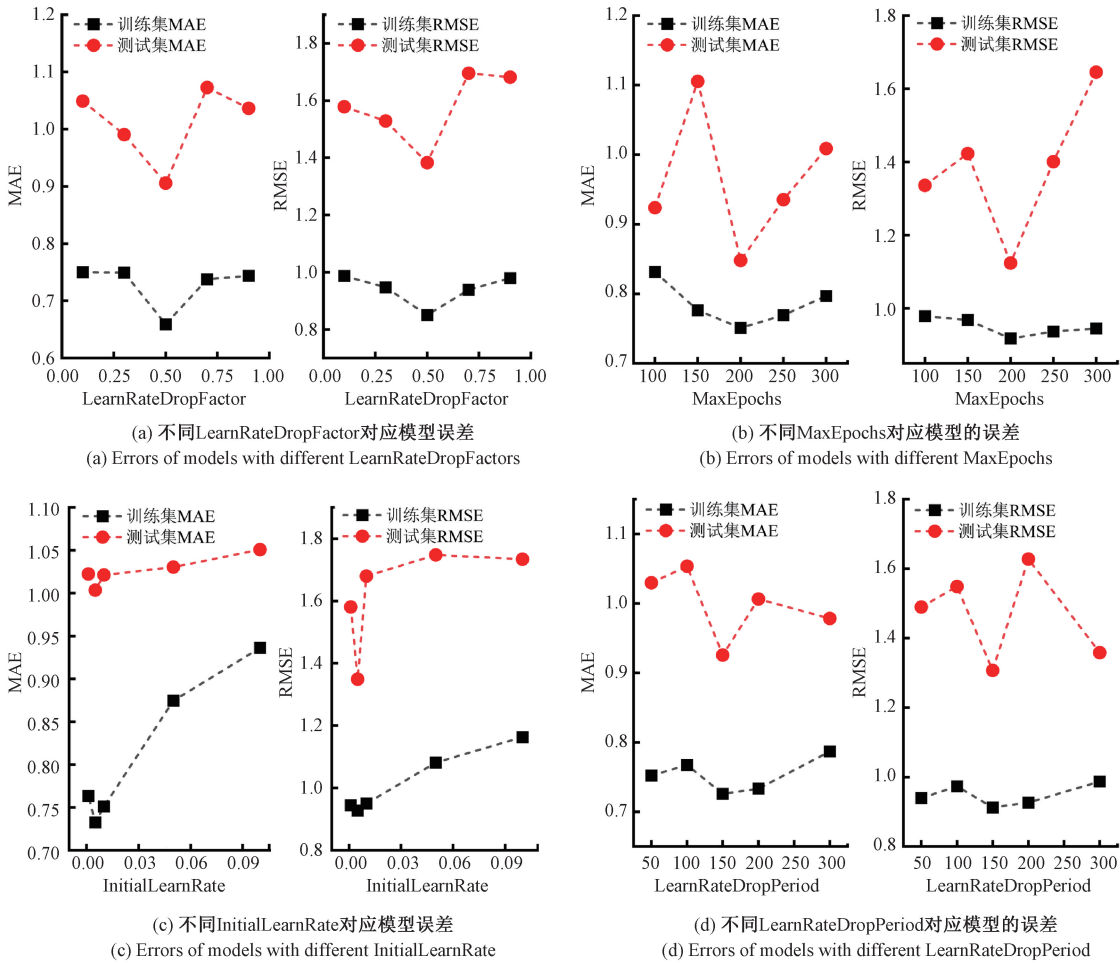


图 14 不同参数对应的模型的误差

Fig. 14 Errors of models corresponding to different parameters

4.2 模型结果分析

1) 特征值降维对预测结果的影响

为验证 KPCA 算法对特征值降维的有效性,现建立对比实验:分别以全部 26 个特征变量和经过降维的 7 个特征值为模型输入,采用以上超参数并用混合模型分别运行,预测 WFGD 出口 SO_2 浓度,得到预测结果对比(图 15)、误差(图 16)、评价指标值(表 3)。

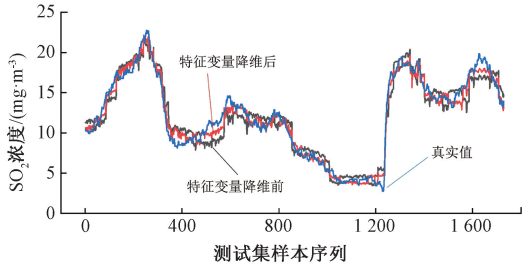


图 15 输入特征变量降维前后的预测结果

Fig. 15 Prediction results before and after dimensionality reduction of input feature variables

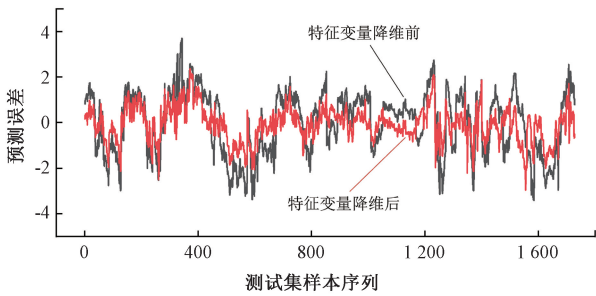


图 16 输入特征变量降维前后的预测误差

Fig. 16 Prediction error before and after dimensionality reduction of input feature variables

从图 15~16 和表 3 可知,输入变量经过降维后,模型预测精度和运行速度均获得了提升:与降维前模型预测结果相比,降维后的 MAE 下降了 25.07%,RMSE 下降了 26.91%,MAPE 下降了 17.60%, R^2 提高了 5.59%,运行时间下降了 18.41%。输入变量经过降维后,预测值波动幅度变小,稳定性增强:降维后误差稳定在区间 $(-2, 2)$,

表 3 特征变量降维前后预测结果评价指标值和运行时间

Table 3 Evaluation indicator values and running time of prediction results before and after screening

指标	MAE	RMSE	R ²	MAPE	用时/s
降维前	0.982 8	1.248 5	0.915 5	0.088 23	239
降维后	0.736 4	0.912 5	0.966 7	0.072 7	195

个别超过±2.5,而降维前误差频繁突破±3,个别情况还突破±3.5;当 WFGD 出口 SO₂ 浓度缓慢变化时,预测值与实际值能实现紧密跟随,但出现骤升或骤降时,降维前预测值总是变化缓慢且幅度偏小。实验结果显示,原始数据特征变量数量过多,且存在相互影响和冗余关系,这会

降低模型的预测精度。因此,通过有效的降维方法可以在一定程度上简化模型,提高预测准确性,并减少模型的运行时间。

2)不同模型对预测结果的影响

为了验证本文所建立的 CNN-BiGRU-Attention 模型在 WFGD 出口 SO₂ 浓度预测上的优越性,本文在使用同一个数据集的基础上,分别采用 CNN、BiGRU、CNN-BiGRU 和 CNN-BiGRU-Attention 模型仿真实验,综合模型结果如图 17 所示,各模型对比结果如图 18,各模型的预测误差如图 19,图 20 为各模型误差对比箱形图,各模型评价指标值如表 4 所示。

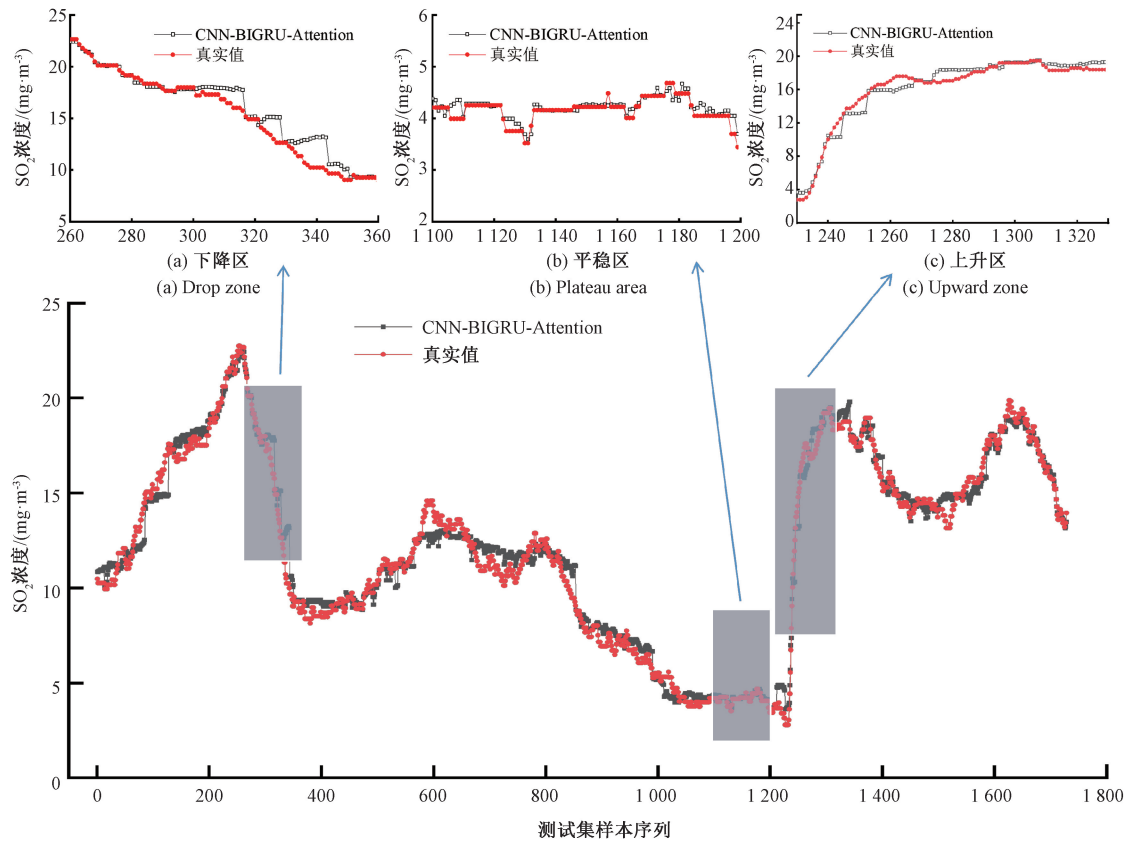
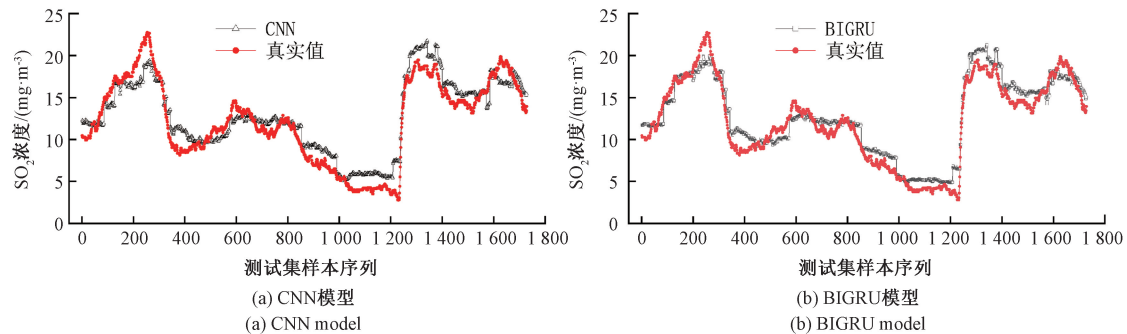


图 17 CNN-BiGRU-Attention 模型预测结果

Fig. 17 Prediction results based on CNN-BiGRU-Attention model



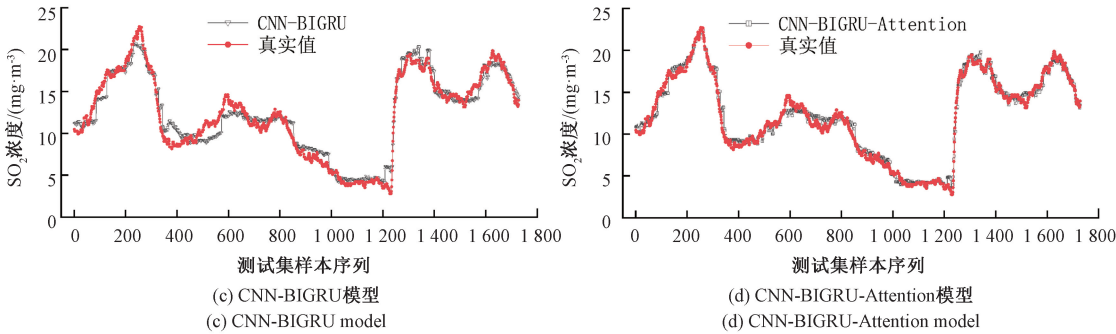


图 18 各模型预测值与真实值对比图

Fig. 18 Comparison of predicted and real values of each model

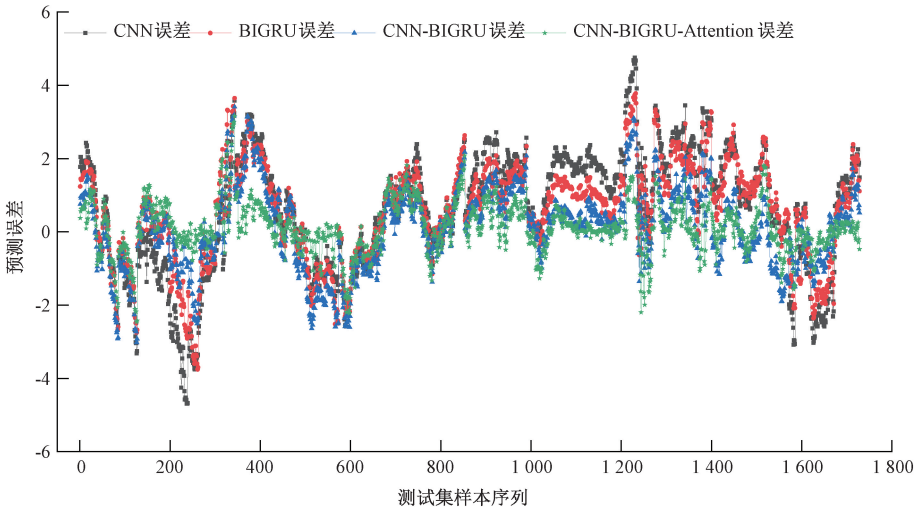
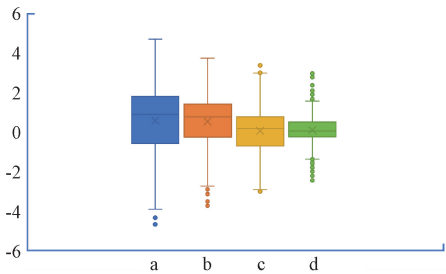


图 19 各模型预测误差图

Fig. 19 Prediction error for each model



注:(a)CNN; (b)BiGRU; (c)CNN-BiGRU;
(d)CNN-BiGRU-Attention

图 20 各模型预测误差箱形图

Fig. 20 Box plots of prediction errors for each model

从图 17~19 可以看出,在 WFGD 出口 SO_2 浓度预测模型中,当目标值走向比较平稳时,各种模型预测值大都能实现紧密跟随且波动幅度比较小,但 CNN-BiGRU 效果好于 CNN 和 BiGRU,CNN-BiGRU-Attention 效果最好,且稳定性高,误差稳定在区间 $(-2, 2)$;当 SO_2 浓度变化幅度大时,CNN-BiGRU-Attention 模型预测值与实际值紧密跟随,且变化后迅速保持稳定,而其他模型预测值有时

表 4 不同模型预测结果对比

Table 4 The comparative analysis between the proposed prediction model and other baseline models

模型	评价指标			
	MAE	RMSE	MAPE	R^2
CNN	1.419 6	1.710 7	9.82%	88.22%
BiGRU	1.177 8	1.410 7	8.19%	91.91%
CNN-BiGRU	0.942 8	1.192 0	7.58%	93.10%
CNN-BiGRU-Attention	0.706 4	0.912 5	6.27%	96.74%

滞后实际值,且波动幅度大,尤其是 CNN 模型,误差经常突破 ± 3 ,甚至个别突破 ± 4 。对比 4 种模型预测结果,CNN-BiGRU 模型的预测效果好于 CNN 和 BiGRU,表明它能集合 CNN 挖掘数据的单点纵向特征和 BiGRU 提取数据的双向时序横向特征的优点,有较好的稳定性和准确度。加入 Attention 后,CNN-BiGRU-Attention 模型比 CNN-BiGRU 模型预测效果更好,这一改善验证了 Attention 机制能够通过对于隐藏层特征施加不同权重的高效分配,优化了网络训练过程,确保关键时刻的高价值的

信息得到重视,克服了信息遗漏的问题,增强了模型的整体预测精确度。

图 20 对比了不同模型的误差分布箱线图。结果表明,模型 CNN-BiGRU-Attention 的误差中位数最低且分布最为集中,说明其预测性能最优且稳定性最高。相比之下,模型 CNN 的误差分布较广且中位数高,表明其预测精度和稳定性均较差。模型 CNN-BiGRU 和 BiGRU 的性能介于模型 CNN 和模型 CNN-BiGRU-Attention 模型之间,但是 CNN-BiGRU 比 CNN 误差分布较为集中,且中位数略低,表明前者优于后者。由表 4 可以得出,CNN-BiGRU-Attention 模型预测效果最好,CNN-BiGRU 模型随其后,CNN 和 BiGRU 预测效果最差。CNN-BiGRU-Attention 模型的 MAE 为 0.706 4、RMSE 为 0.912 5、MAPE 为 6.27%,与 CNN-BiGRU 相比,分别下降了 25.07%、23.45%、17.28%,更超过 CNN 和 BiGRU 模型。CNN-BiGRU-Attention 模型的 R^2 为 96.74%,与 CNN-BiGRU、BiGRU 和 CNN 模型相比,分别提高了 3.91%、5.26%和 9.66%。这说明 CNN-BiGRU-Attention 模型集合 CNN、BiGRU 和 Attention 优点,进一步提高了模型训练的精度和稳定性。

5 结 论

WFGD 的运行机理复杂,影响出口 SO_2 浓度因素众多,且相互影响,存在复杂的非线性关系,本文采用 KPCA 进行数据降维,取得了较好的效果。

本文提出了基于 CNN、BiGRU 和 Attention 相结合的混合深度网络结构的 WFGD 出口 SO_2 浓度动态预测模型,且运行 CNN-BiGRU-Attention 模型输出预测值,各项评价指标都优于 CNN、BiGRU 和 CNN-BiGRU 网络。

结果表明,CNN-BiGRU-Attention 模型能够结合 CNN、BiGRU 和 Attention 模型三者的优势,充分挖掘数据自身的时间序列属性,获取深层次的时间相关性,有较高的预测精度和稳定性,具有一定的工程应用价值。

参考文献

- [1] TANG L, QU J B, MI Z F, et al. Substantial emission reductions from chinese power plants after the introduction of ultra-low emissions standards[J]. Nature Energy, 2019, 4(11): 929-938.
- [2] 曾庆华,冉鹏,董坤,等. 基于大数据平台的 SO_2 排放 GWO-N-BEATS 预测算法[J]. 热能动力工程, 2024, 39(3): 125-131.
ZENG Q H, RAN P, DONG K, et al. SO_2 emission prediction by GWO-N-BEATS algorithm based on big data platform[J]. Journal of Engineering for Thermal Energy and Power, 2024, 39(3): 125-131.
- [3] ZHANG G M, WANG W, CHEN Z Y, et al. Modeling and optimal dispatch of a carbon-cycle integrated energy system for low-carbon and economic operation[J]. Energy, 2022, 240: 122795.
- [4] 马双忱,林宸雨,周权. 基于深度神经网络的脱硫系统预测模型及应用[J]. 化工进展, 2021, 40(3): 1689-1698.
MA SH CH, LIN CH Y, ZHOU Q. Prediction model and application of desulfurization system based on deep neural network [J]. Advances in Chemical Engineering, 2021, 40(3): 1689-1698.
- [5] 金秀章,刘岳,于静,等. 基于变量选择和 EMD-LSTM 网络的出口 SO_2 浓度预测[J]. 中国电机工程学报, 2021, 41(24): 8475-8483.
JIN X ZH, LIU Y, YU J, et al. Export SO_2 concentration prediction based on variable selection and EMD-LSTM network[J]. Proceedings of the CSEE, 2021, 41(24): 8475-8483.
- [6] 郑茗友,王伟,赵文杰,等. 基于 PSO-Lasso 算法的电站湿法脱硫出口 SO_2 浓度预测[J]. 华北电力大学学报(自然科学版), 2022, 49(1): 112-117.
ZHENG M Y, WANG W, ZHAO W J, et al. Prediction of SO_2 concentration at the outlet of wet flue gas desulfurization in power plants based on PSO-Lasso algorithm[J]. Journal of North China Electric Power University (Natural Science Edition), 2022, 49(1): 112-117.
- [7] 吴磊,康英伟. 基于改进粒子群优化长短时记忆神经网络的脱硫系统 SO_2 预测模型[J]. 热力发电, 2021, 50(12): 66-73.
WU L, KANG Y W. The SO_2 prediction model for desulfurization system based on improved particle swarm optimization long short-term memory neural network [J]. Thermal Power Generation, 2021, 50(12): 66-73.
- [8] 金秀章,李京. 基于互信息 PSO-LSTM 的 SO_2 浓度预测[J]. 控制工程, 2022, 29(11): 1928-1932.
JIN X ZH, LI J. SO_2 concentration prediction based on mutual information PSO-LSTM [J]. Control Engineering, 2022, 29(11): 1928-1932.
- [9] 金秀章,陈佳政,张瑾. 基于变量筛选和 OS-KELM 的出口 SO_2 浓度预测[J/OL]. 华北电力大学学报(自然科学版), 1-10 [2025-03-19]. <http://kns.cnki.net/kcms/detail/13.1212.TM.20231017.1732.004.html>.
JIN X ZH, CHEN J ZH, ZHANG J. Prediction of SO_2 concentration at outlet based on variable screening and OS-KELM [J/OL]. Journal of North China Electric Power University (Natural Science Edition), 1-10 [2025-03-19]. <http://kns.cnki.net/kcms/>

- detail/13, 1212, TM, 20231017, 1732, 004, html.
- [10] 闫浩思, 赵文杰. 基于 MIC 和 MPA-KELM 的脱硫出口 SO₂ 浓度预测[J]. 计量学报, 2023, 44(2): 271-278.
- YAN H S, ZHAO W J. Prediction of SO₂ concentration at desulfurization outlet based on MIC and MPA-KELM[J]. Acta Metrologica Sinica, 2023, 44(2): 271-278.
- [11] 冯齐全, 乔宗良, 司风琪. 基于 GMM-LSSVR 模型的脱硫出口 SO₂ 浓度预测方法研究[J]. 热能动力工程, 2023, 38(10): 79-85.
- FENG Q Q, QIAO Z L, SI F Q. Research on the prediction method of SO₂ concentration at desulfurization outlet based on GMM-LSSVR model[J]. Journal of Engineering for Thermal Energy and Power, 2023, 38(10): 79-85.
- [12] 孙坤, 尹晓红. 基于数据去噪和 CNN-BiGRU 的 SO₂ 排放预测[J]. 电子测量技术, 2023, 46(13): 66-72.
- SUN K, YIN X H. SO₂ emission prediction based on data denoising and CNN-BiGRU [J]. Electronic Measurement Technology, 2023, 46(13): 66-72.
- [13] YANG T T, MA K F, LU Y, et al. Real-time dynamic prediction model of NO_x emission of coal-fired boilers under variable load conditions[J]. Fuel, 2020, 274: 12.
- [14] DING X, FENG C, YU P, et al. Gradient boosting decision tree in the prediction of NO_x emission of waste incineration[J]. Energy, 2023, 264: 126174.
- [15] LI K, LI Q, FAN H, et al. Soft sensing of SO₂ emission for ultralow emission coal-fired power plant with dynamic model and segmentation model [J]. Fuel, 2023, 332: 125921.
- [16] ALNAIM A, SUN ZH, TONG D N. Evaluating machine learning and remote sensing in monitoring NO₂ emission of power plants[J]. Rem Sens, 2022, 14(3): 35.
- [17] YANG G T, WANG Y N, LI X L. Prediction of the NO_x emissions from thermal power plant using long-short term memory neural network [J]. Energy, 2020, 192: 13.
- [18] YIN X H, SUN K, LI S Y, et al. Enhancing deep learning for the comprehensive forecast model in flue gas desulfurization systems[J]. Control Engineering Practice, 2023, 138: 105587.
- [19] 李瑞连, 曾德良, 张光明, 等. 基于机理和数据混合模型的 SD-WFGD 系统 SO₂ 浓度动态预测[J]. 动力工程学报, 2024, 44(7): 1118-1128.
- LI R L, ZENG D L, ZHANG G M, et al. Dynamic prediction of SO₂ concentration in SD-WFGD system based on mechanism and data mixing model [J]. Journal of Power Engineering, 2024, 44 (7): 1118-1128.
- [20] 张明, 李娜, 王鹏. 深度调峰工况下脱硫系统 SO₂ 浓度预测模型研究[J]. 热能动力工程, 2021, 36(4): 78-85.
- ZHANG M, LI N, WANG P. Predictive modeling of SO₂ concentration in desulfurization system under deep peaking condition [J]. Journal of Engineering for Thermal Energy and Power, 2021, 36(4): 78-85.
- [21] NIU D X, YU M, SUN L J, et al. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism[J]. Applied Energy, 2022, 313: 118801.
- [22] WANG X, LIU X, BAI Y. Prediction of the temperature of diesel engine oil in railroad locomotives using compressed information-based data fusion method with attention-enhanced CNN-LSTM [J]. Applied Energy, 2024, 367: 123357.
- [23] 王长江, 张千龙, 姜涛, 等. 基于图卷积网络和双向门控循环单元的电力系统主导失稳模式辨识[J/OL]. 中国电机工程学报, 1-12[2025-03-19]. <https://doi.org/10.13334/j.0258-8013.pcsee.240123>.
- WANG CH J, ZHANG Q L, JIANG T, et al. Power system dominant instability mode identification based on graph convolutional networks and bidirectional gated cyclic units[J/OL]. Proceedings of the CSEE, 1-12[2025-03-19]. <https://doi.org/10.13334/j.0258-8013.pcsee.240123>.
- [24] 刘宇龙, 崔宪阳, 袁丁, 等. 基于多级注意力机制融合的电质量扰动点分类及时间定位方法研究[J]. 中国电机工程学报, 2024, 44(11): 4298-4310.
- LIU Y L, CUI X Y, YUAN D, et al. Research on multi-level attention mechanism optimized method for point classification and time interval identification of power quality disturbances [J]. Proceedings of the CSEE, 2024, 44(11): 4298-4310.
- [25] 江兵, 李响, 巢一帆, 等. 基于 KPCA-CGSSA-KELM 的变压器故障识别方法[J]. 电子测量与仪器学报, 2024, 38(5): 139-147.
- JIANG B, LI X, CHAO Y F, et al. Transformer fault identification method based on KPCA-CGSSA-KELM[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5): 139-147.
- [26] FAN ZH, XU F Y, LI C, et al. Application of KPCA and ada-boost algorithm in classification of functional magnetic resonance imaging of alzheimer's disease[J].

Neural Computing and Applications, 2020, 32: 5329-5338.

[27] 刘建军, 赵旭, 张卫东, 等. 基于 CNN(1D)-LSTM 模型的电站锅炉 SCR 入口 NO_x 浓度预测[J]. 电子测量技术, 2023, 46(13): 59-65.

LIU J J, ZHAO X, ZHANG W D, et al. Prediction of NO_x concentration in SCR inlet of power station boiler based on CNN (1D)-LSTM model [J]. Electronic Measurement Technology, 2023, 46(13):

59-65.

作者简介

高钾, 博士研究生, 讲师, 主要研究方向为锅炉燃烧模拟、监测与优化控制, 热辐射分析及智能发电。

E-mail: db23130006p41@cumt.edu.cn

彭献永(通信作者), 博士, 副教授, 主要研究方向为能源动力系统建模、优化控制, 智能发电与综合智慧能源。

E-mail: xianyong.peng@cumt.edu.cn