

DOI: 10.13382/j.jemi.B2508524

融合 SSAD-Z-score 异常检测的 LFLSTM 水泥脱硝系统 NO_x 预测模型

陈薇¹ 何浩然¹ 李鑫¹ 邱亚¹ 褚彪²

(1. 合肥工业大学电气与自动化工程学院 合肥 230009; 2. 合肥水泥研究设计院有限公司 合肥 230051)

摘要:为解决水泥脱硝系统氮氧化物(nitrogen oxides, NO_x)浓度预测中异常数据干扰严重、时序建模困难的问题,提出了一种融合滑动窗口斜率异常检测-Z-score(sliding slope anomaly detection-Z-score, SSAD-Z-score)异常检测与多步时滞长短期记忆(lag features long short-term memory, LFLSTM)深度学习模型的复合预测方法。首先,基于NO_x生成与选择性非催化还原(selective non-catalytic reduction, SNCR)反应机理,采用相关系数筛选出窑尾烟室NO_x浓度和SNCR喷氨频率等核心输入变量。针对反吹等复杂异常,设计SSAD-Z-score算法,结合滑动窗口斜率变化与全局Z-score判据,实现对突发、趋势、局部扰动、反吹数据等多类型异常的自适应检测与修正。进一步在长短期记忆(long short-term memory, LSTM)模型中引入多步时滞特征(lag features, LF),充分利用历史时序信息,提升模型对NO_x浓度动态变化的建模能力。实验结果表明,SSAD-Z-score方法在异常检测覆盖率(47.67%)、召回率(0.844)等指标上,均显著优于传统Z-score方法、孤立森林、局部异常因子(local outlier factor, LOF)和小波变换。同时,SSAD-Z-score在修正后标准差方面表现卓越,为所有对比方法中最低,且平均修正幅度达8.81。参数灵敏度分析显示,在滑动窗口大小10、斜率阈值0.15、Z-score阈值1.65的中等参数设置下,SSAD-Z-score在修正后标准差(35.65)和平均修正幅度(8.81)等方面取得最佳平衡。将修正后的数据输入LFLSTM模型进行NO_x浓度预测,LFLSTM模型在平均单样本预测耗时方面仅为4.88 ms,远低于1 s的数据采集频率,决定系数(coefficient of determination, R²)为0.884 5,均方根误差(root mean squared error, RMSE)为2.026 5,平均绝对误差(mean absolute error, MAE)为1.798 5,预测性能均显著优于Transformer、Informer和门控循环单元(gated recurrent unit, GRU)等主流模型。最终,结合SSAD-Z-score修正的LFLSTM模型在整体系统性能上表现最佳,R²达到0.824 097, RMSE为4.305 981, MAE为3.297 620,均显著优于直接使用LFLSTM和Z-score-LFLSTM的预测模型。研究结果验证了所提方法在提升异常检测准确性和NO_x浓度预测精度方面的有效性,为水泥脱硝系统智能控制提供了坚实的数据基础和方法支撑。

关键词:水泥脱硝系统;异常检测;NO_x浓度预测;LFLSTM;时滞特征;标准分数;参数灵敏度分析

中图分类号: TP183; TN081

文献标识码: A

国家标准学科分类代码: 120.3030

LFLSTM NO_x prediction model for cement denitrification system integrated with SSAD-Z-score anomaly detection

Chen Wei¹ He Haoran¹ Li Xin¹ Qiu Ya¹ Chu Biao²

(1. School of Electrical and Automation Engineering, Hefei University of Technology, Hefei 230009, China;

2. Hefei Cement Research & Design Institute Corporation Ltd, Hefei 230051, China)

Abstract: To address the challenges of severe outlier interference and temporal modeling difficulties in predicting NO_x concentrations within cement denitrification systems, this paper proposes a composite prediction method integrating SSAD-Z-score anomaly detection with LFLSTM deep learning models. First, based on NO_x generation and SNCR reaction mechanisms, core input variables—including kiln tail flue gas chamber NO_x concentration and SNCR ammonia injection frequency—are selected using correlation coefficients. To handle complex anomalies like backwashing, the SSAD-Z-score algorithm is designed. It combines sliding window slope changes with global Z-score criteria to achieve adaptive detection and correction for multiple anomaly types, including sudden spikes, trends, local disturbances, and backwashing data. Furthermore, multi-step lag features are introduced into the LSTM model to fully leverage historical time-series information, enhancing the model's ability to capture dynamic changes in NO_x concentration. Experimental results

demonstrate that the SSAD-Z-score method significantly outperforms traditional Z-score methods, Isolated Forest, LOF, and Wavelet Transform in metrics such as anomaly detection coverage (47.67%) and recall (0.844). Additionally, SSAD-Z-score exhibits exceptional performance in corrected standard deviation, achieving the lowest value among all comparison methods with an average correction magnitude of 8.81. Parameter sensitivity analysis indicates that under moderate parameter settings—sliding window size 10, slope threshold 0.15, and Z-score threshold 1.65—SSAD-Z-score achieves optimal balance in terms of corrected standard deviation (35.65) and average correction magnitude (8.81). Inputting the corrected data into the LFLSTM model for NO_x concentration prediction yielded an average single-sample prediction time of just 4.88 milliseconds—significantly below the 1-second data acquisition frequency. The model achieved an R^2 of 0.884 5, RMSE of 2.026 5, and MAE of 1.798 5, demonstrating markedly superior predictive performance compared to mainstream models like Transformer, Informer, and GRU. Ultimately, the LFLSTM model combined with SSAD-Z-score correction demonstrated the best overall system performance, achieving an R^2 of 0.824 097, an RMSE of 4.305 981, and an MAE of 3.297 620. These metrics significantly outperformed prediction models using either LFLSTM alone or Z-score-LFLSTM. The research results validate the effectiveness of the proposed method in enhancing anomaly detection accuracy and NO_x concentration prediction precision, providing a robust data foundation and methodological support for intelligent control of cement denitrification systems.

Keywords: cement denitrification system; anomaly detection; NO_x concentration prediction; LFLSTM; lag features; Z-score; parameter sensitivity analysis

0 引言

随着城市化进程的加快,建筑行业发展迅速,水泥作为基础建材需求持续增长。然而,部分中小型水泥企业在追求产能的同时,忽视了环保要求,导致环保设施不完善,排放不达标。统计数据显示,水泥行业氮氧化物(nitrogen oxides, NO_x)排放量约占全国总排放量的10%,已成为大气污染治理的重点领域^[1]。为此,国家相继出台了《水泥工业大气污染物排放标准》等政策,对NO_x排放提出了更为严格的要求,如何在保证水泥产量的前提下降低NO_x排放,成为亟需解决的工程难题^[2]。同时,NO_x排放受到多种复杂因素的影响,如水泥原料成分、窑炉温度、气流速率以及脱硝过程等,如何建立一个准确的模型来预测NO_x排放同样面临挑战^[3]。

目前,NO_x浓度预测建模方法主要有机理建模和数据驱动建模两大类。

机理建模是通过建模对象的反应机理构建方程组,通过求解方程组的数值解构建估计模型。孙崇建等^[4]研究了水泥窑内的温度建模及预测, Yao等^[5]建立了基于选择性非催化还原烟气脱硝技术(selective non-catalytic reduction, SNCR)的混合脱硝工艺。机理建模虽然能够反应出过程变化,但由于水泥系统过程复杂,仅仅靠机理建模难以实现NO_x生成和脱硝过程的模型建立。

数据驱动建模是利用数据特性来映射变量间的关系,以此建立输入变量和输出变量关系模型。刘文彪等^[6]结合卷积神经网络(convolutional neural network, CNN)与双向长短期记忆网络(bi-directional long short term memory, Bi-LSTM)模型实现对轴承性能退化预测,虽然在时序数据预测方面较其他模型效果好,但预测误

差依然显著。Sun等^[7]以NO_x生成的影响变量作为输入,NO_x排放作为输出,构建窑炉改造前后NO_x排放预测模型,虽然在建模上具有较好地非线性拟合能力,但存在局部优化或过拟合等缺点。Yang等^[8]通过皮尔逊系数选取相关度高的辅助变量,并利用互信息方法确定变量间延迟时间,提出一种基于最小二乘支持向量机(least squares support vector machine, LSSVM)的NO_x浓度实时动态预测模型。虽然在非线性建模方面表现突出,但模型在多维时序变量数据的特征提取方面能力有限,限制了模型预测性能。Xie等^[9]通过分析NO_x生成和脱硝反应,利用互信息实现变量筛选,并引入新的注意力机制,提出一种能够拟合多变量耦合、非线性和大延迟系统的动态预测模型。He等^[10]使用卷积神经网络与长短期记忆网络相结合的方法对分解炉的NO_x排放量进行预测。Kang等^[11]利用动态联合互信息理论估计反应和系统延迟,结合Bi-LSTM模型实现燃煤锅炉NO_x排放浓度预测。

虽然模型结构相比传统的单层网络在多维变量数据特征提取能力和模型自适应学习能力有更好的优势,但忽略了变量在时间维度上的特征信息与变量数据的异常处理。

针对上述问题,本文提出了一种融合滑动窗口斜率异常检测与Z-score异常检测(sliding slope anomaly detection-Z-score, SSAD-Z-score)的异常检测方法,与多步时滞长短期记忆(lag features long short-term memory, LFLSTM)深度学习模型的复合预测方法。首先,基于NO_x生成与SNCR反应机理,采用相关系数筛选核心输入变量。针对反吹等复杂异常,设计SSAD-Z-score算法,结合滑动窗口斜率变化与全局Z-score判据,实现多类型异常的自适应检测与修正。通过参数灵敏度分析,确定

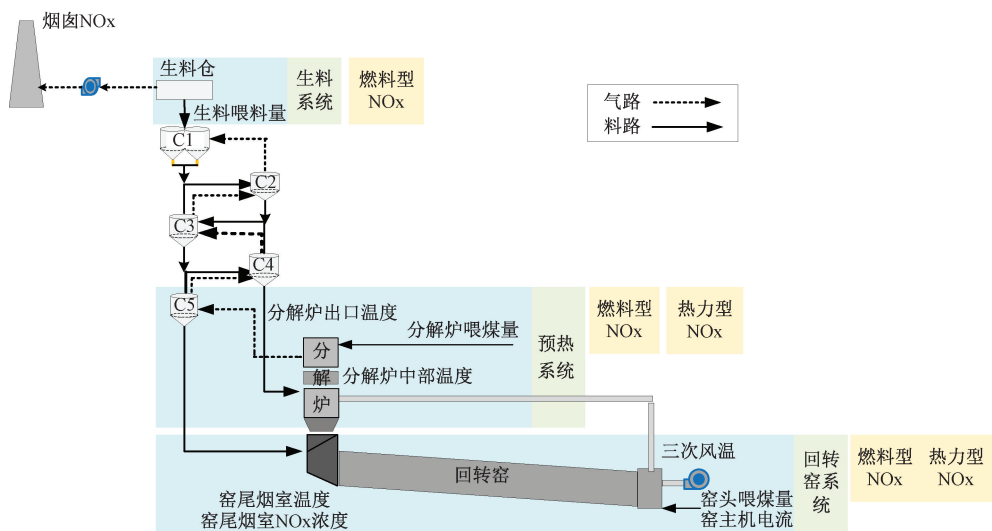


图 1 水泥生产系统 NO_x 生成设备与流程示意图

Fig. 1 Schematic diagram of NO_x generation equipment and process in cement production system

最优参数组合,提升异常检测的准确性和数据质量。最后,将修正后的数据输入 LFLSTM 模型进行 NO_x 浓度预测,仿真结果表明,所提方法在异常检测覆盖率、数据修正幅度和预测精度等方面均优于传统方法。本文的研究为水泥脱硝系统 NO_x 智能预测与控制提供了理论基础和工程参考。

1 水泥脱硝系统 NO_x 机理分析及辅助变量选取

1.1 水泥脱硝系统 NO_x 产生机理分析

在水泥生产过程中,NO_x 的生成是一个复杂的化学过程,主要包括 3 种类型:燃料型 NO_x、热力型 NO_x 和瞬时型 NO_x^[12]。其中,燃料型 NO_x 占 NO_x 总生成量的 60%~80%,热力型 NO_x 占 15%~25%,而瞬时型 NO_x 占比仅为 5%以下。

燃料型 NO_x 主要来源于煤中的有机氮化合物。当煤炭在高温下燃烧时,煤中的氮化合物与氧气反应生成 NO_x^[13]。燃料型 NO_x 的生成量与煤的氮含量、燃烧温度与氧气浓度密切相关。燃料型 NO_x 多产生于煤炭含氮化合物浓度高、含氧量多且易燃烧的区域,如回转窑和分解炉中。

热力型 NO_x 是在高温环境下,空气中的氮气与氧气发生氧化反应生成的。热力型 NO_x 多产生高温区、烟气停留时间长、含氧量相对多的地方^[14],如回转窑和分解炉等温度较高的区域生成,尤其是这些区域的氧化反应区。

瞬时型 NO_x 生成是由煤燃烧过程中产生的碳氢基团与氮气发生碰撞,快速反应生成 NO^[15]。通常发生在

低温缺氧的环境中,与温度关系弱,因生成量较少,所占比例较低,对 NO_x 排放的影响较小。

水泥生产系统中不同设备在 NO_x 的生成中起着重要作用。图 1 展示了水泥生产系统 NO_x 生成设备与流程。水泥生产系统中不同设备影响 NO_x 生成的变量如表 1 所示。

表 1 影响 NO _x 生成变量表		
Table 1 Variables affecting NO _x generation		
设备	影响变量	NO _x 类型
生料仓	生料喂料量	燃料型 NO _x
	分解炉出口温度	热力型 NO _x
分解炉	分解炉中部温度	燃料型 NO _x
	分解炉喂煤量	
	窑尾烟室温度	热力型 NO _x
回转窑	窑主机电流	
	窑头喂煤量	
	窑尾烟室 NO _x 浓度	

1.2 水泥脱硝系统 SNCR 机理分析

为了降低水泥生产过程中 NO_x 的排放,SNCR 被广泛应用。SNCR 是一种无需催化剂、仅通过还原剂降低 NO_x 排放浓度的脱硝技术^[16]。常用还原剂包括氨水和尿素,通过向烟气中喷射氨水或尿素,在合适的温度条件下使 NO_x 被还原为氮气跟水。

SNCR 脱硝的主要化学反应如下:

$$4\text{NH}_3 + 4\text{NO} + \text{O}_2 \rightarrow 4\text{N}_2 + 6\text{H}_2\text{O} \tag{1}$$

$$8\text{NH}_3 + 6\text{NO}_2 \rightarrow 7\text{N}_2 + 12\text{H}_2\text{O} \tag{2}$$

在实际应用中,分解炉完全燃烧区的温度一般控制在 850℃~1 000℃之间,符合 SNCR 脱硝所需的理想温度窗口条件^[17]。为确保高效脱硝,氨水喷入装置的设计考虑到烟气的流动特点,一部分设置在分解炉的炉尾,而

另一部分则喷入 C5 筒,以确保氨水能够与 NO_x 充分混合,并在适宜温度下完成还原反应。

在水泥生产过程中,回转窑产生的热力型 NO_x 进入分解炉的氧化反应区,与该区域生成的燃料型 NO_x 混合。烟气沿着分解炉烟道方向流动,在 SNCR 氨水喷射系统的作用下,氨水被雾化后与烟气充分接触,使 NO_x 还原为氮气和 水,从而降低 NO_x 的排放浓度。经过 SNCR 脱硝反应后的烟气含 NO_x 浓度较低,剩余的 NO_x 随烟气流向分解炉出口,进一步进入五级筒预热器^[18]。

SNCR 脱硝的最终结果通过烟囱处的 NO_x 浓度进行评估。烟囱 NO_x 浓度的变化可直接反映脱硝工艺的优化程度,合理控制氨水喷射量和喷射位置可以显著提高 NO_x 的去除效率,减少氨逃逸,优化脱硝成本。图 2 所示为 SNCR 脱硝系统在水泥生产流程中位置、氨水喷射系统及烟囱 NO_x 监测点。

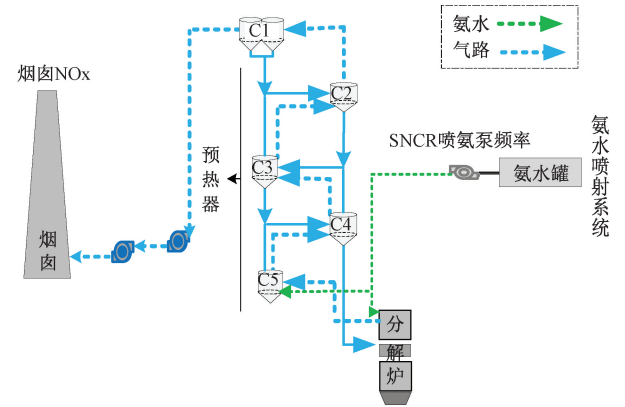


图 2 水泥生产系统 SNCR 脱硝工艺示意图

Fig. 2 Schematic diagram of SNCR denitrification process in cement production system

1.3 辅助变量选择

在构建水泥脱硝系统的烟囱 NO_x 预测模型时,合理选择输入变量至关重要。结合水泥生产线中采集到的多个连续时间段的特征变量数据,初步选取变量表如表 2 所示。

表 2 水泥生产线采集变量表

Table 2 Collected variables in cement production line		
变量点位置	变量名称	单位
窑头	窑头喂煤量	t/h
	窑尾烟室 NO _x 浓度	μL/L
窑尾	窑尾烟室温度	℃
	窑主机电流	A
分解炉	分解炉喂煤量	t/h
	分解炉中部温度	℃
	分解炉出口温度	℃
预热器	生料喂料量	t/h
	SNCR 喷氨泵频率	Hz
脱硝系统相关数据	烟囱 NO _x	mg/Nm ³

为了进一步确定辅助变量,采用皮尔逊相关系数(Pearson correlation coefficient)来度量辅助变量与烟囱 NO_x 浓度的相关程度^[19]:

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

式中: σ_X 为样本标准差; $\text{Cov}(X,Y)$ 为辅助变量与烟囱 NO_x 浓度之间的协方差。通过计算水泥脱硝系统中各个过程参量与烟囱 NO_x 浓度之间的相关性系数,结果如表 3 所示。

表 3 各辅助变量与烟囱 NO_x 浓度相关系数

Table 3 Correlation coefficients between auxiliary variables and chimney NO_x concentration

变量名	PEARSON 相关系数
窑头喂煤量	0.116 718
窑尾烟室温度	-0.708 75
窑尾烟室 NO _x 浓度	0.819 39
窑主机电流	0.079 105
分解炉喂煤量	-0.545 87
分解炉中部温度	-0.683 51
分解炉出口温度	0.669 249
生料喂料量	0.040 931
SNCR 喷氨泵频率	-0.827 05

通过工艺分析与相关性分析,本文最终选择相关系数绝对值大于 0.8 的变量作为烟囱 NO_x 浓度预测的辅助变量,其中包括窑尾烟室 NO_x 浓度和 SNCR 喷氨泵频率。

2 基于 Z-score 的数据异常检测算法及缺陷分析

2.1 脱硝系统异常数据分析

针对数据驱动的建模方法,在模型建立前的异常值检测和缺失值补全处理工作是后续精确建模的基础^[20]。为确保预测模型在工业实际应用中的可靠性与准确性,需要对采集到的原始数据进行异常检测和预处理。

在水泥脱硝系统的实际运行过程中,NO_x 浓度测量数据受到多种异常因素的干扰,主要包括反吹操作引发的突发性异常、传感器故障导致的漂移型异常以及工艺波动引起的局部扰动等。

异常行为中反吹异常是最为典型且影响显著的一类异常数据。反吹操作通常由袋式除尘器或烟气分析仪的自清洁程序定期触发,其周期一般为 50 min 左右,每次持续约 1 min。反吹期间,窑尾烟室 NO_x 浓度会在极短时间内骤降至异常低值,随后迅速恢复至正常水平,形成明显的“V”型数据特征。该过程不仅表现为瞬时突变,还常伴有短时的非线性回升和局部波动,具有明显的局

部性和周期性。图 3 所示为某段实际测量数据,其中红框所示即为典型的反吹过程。

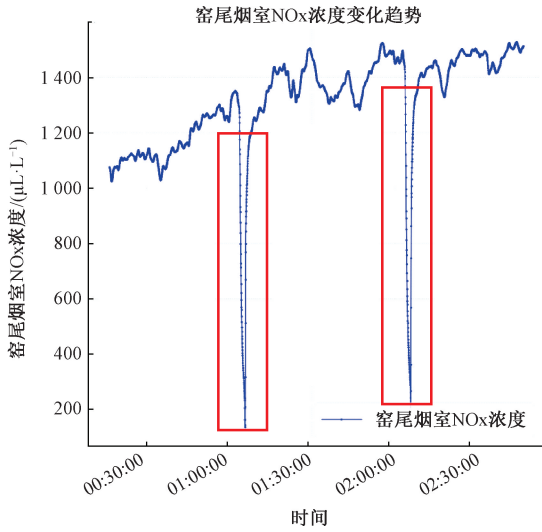


图 3 反吹异常数据

Fig. 3 Back-blowing anomaly data

此外,部分异常数据表现为传感器漂移或信号丢失,导致 NO_x 浓度出现缓慢偏移或长时间异常稳定,严重影响数据的真实性和连续性。统计分析显示,异常数据在原始采集序列中占比可达 2%~5%,若不加以处理,会导致后续建模过程中模型参数估计偏差、预测精度下降。

因此,针对脱硝系统中反吹等复杂异常数据的识别与修正,是保障 NO_x 浓度预测模型准确性和工程适用性的关键前提。

2.2 基于 Z-score 的数据异常检测算法

Z-score 是一种广泛应用于数据异常检测的方法,通过衡量数据点与数据集均值之间的偏离程度来判断数据点是否为异常^[21]。Z-score 的计算公式如下:

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

式中: X 是当前数据点的值; μ 是数据集的均值; σ 是数据集的标准差。当 Z-score 的绝对值超过预设的阈值时,认为该数据点为异常。

Z-score 异常检测的步骤包括数据预处理、计算均值和标准差、计算每个数据点的 Z-score、异常判定、异常修正等。

数据预处理:在应用 Z-score 之前,需要对数据进行预处理,以确保数据的质量。

计算均值和标准差:Z-score 的计算依赖于数据值的均值和标准差。

计算数据集的均值:

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (5)$$

式中: X_i 表示第 i 个数据点; n 为数据点总数。

计算数据集的标准差:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2} \quad (6)$$

均值和标准差反映了数据的整体分布特征,均值 μ 表示数据的中心位置,标准差 σ 衡量了数据的与均值的离散程度。

计算每个数据点的 Z-score:

$$Z_i = \frac{X_i - \mu}{\sigma} \quad (7)$$

Z-score 反映数据点与数据集之间的偏差程度,Z-score 越大,数据点越偏离均值。

异常判定:计算 Z-score 后,根据预设的阈值来判断数据点是否异常:

$$|Z_i| > Z_{\text{threshold}} \quad (8)$$

如果 Z-score 的绝对值 $|Z_i|$ 超过设定的阈值 $Z_{\text{threshold}}$,则该数据点被视为异常值,标记为需要修正的异常数据。

异常修正:对于被判定为异常的数据点,使用线性插值法进行修正。线性插值基于异常点前后的数据点,通过拟合其线性关系来计算异常点的修正值:

$$X_{\text{adjusted}} = X_a + \frac{(X_b - X_a)}{t_b - t_a} (t_{\text{current}} - t_a) \quad (9)$$

式中: X_a 和 X_b 是异常点前后的两个数据点; X_{adjusted} 是修正后的数据点值; t_a 和 t_b 分别是前后两个数据点的时间戳; t_{current} 是异常点的时间戳。利用前后数据点之间的线性变化趋势来修正异常数据。

2.3 基于 Z-score 的数据异常检测的缺陷

基于 Z-score 的数据异常检测方法的根本局限在于其全局性统计假设,仅依赖于整段数据的均值和标准差作为全局参考,无法感知数据局部结构和趋势变化。

图 4 所示为一段模拟数据,整个序列长度为 150 个点,正常平稳段 135 个点,值为 300;突发异常点 1 个,值为 500;具有反吹特征的复合异常数据段约 15 个点。

对于突发尖刺点,其对应的 Z-score 值计算为:

$$\mu = \frac{149 \times 300 + 500}{150} = 301.33 \quad (10)$$

$$\sigma = \sqrt{\frac{1}{150} \times [149 \times (300 - 301.33)^2 + (500 - 301.33)^2]} \approx 16.29 \quad (11)$$

$$Z_{55} = \frac{500 - 301.33}{16.29} \approx 12.23 \gg 1.65 \quad (12)$$

因此,该异常点能够被 Z-score 算法检测出来,验证了基于 Z-score 的数据异常检测算法对突发性异常数据具备良好的识别能力。

然而,对于反吹扰动数据段,该数据包含一个起始突

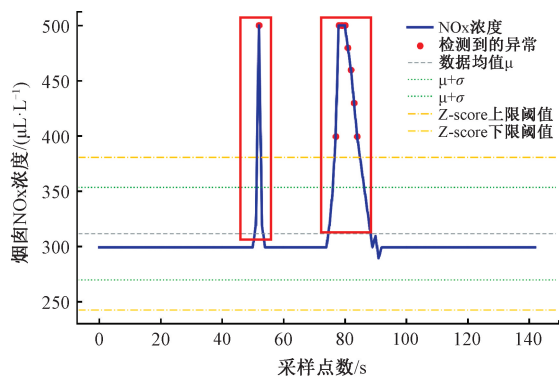


图4 异常数据

Fig. 4 Anomaly data

升、短暂高位保持、快速回落及末尾小幅振荡的结构。基于 Z-score 的数据异常检测算法仅仅能够识别其中几个数值偏离最大的点,而对于回落阶段以及带有非线性趋势的部分,检测结果出现明显不连续和遗漏。

第 76~90 个点构成反吹段,数据变化如下:

$$X = \begin{bmatrix} 300, 350, 400, 500, 500, 500, 480, \\ 460, 440, 420, 400, 390, 370, 360 \end{bmatrix}$$

将这 15 个点嵌入 150 长度的正常序列中,异常段总和约为 6 500,通过计算:

$$\mu = \frac{135 \times 300 + 6\,500}{150} = 313.33 \quad (13)$$

$\sigma =$

$$\sqrt{\frac{1}{150} \times [135 \times (300 - 313.33)^2 + \sum_{j=1}^{15} (X_j - \mu)^2]} \approx 31.2 \quad (14)$$

反吹段数据点的 Z-score 计算:

$$Z_{76} = \frac{500 - 313.33}{31.2} \approx 5.98 > 1.65 \quad (15)$$

$$Z_{78} = \frac{440 - 313.33}{31.2} \approx 4.05 > 1.65 \quad (16)$$

$$Z_{86} = \frac{360 - 313.33}{31.2} \approx 1.49 < 1.65 \quad (17)$$

通过计算可以看出,当数据在短时内连续上升或下降时,即便其变化幅度明显, $|X_i - \mu|$ 仍可能处于 1~1.5 倍标准差范围内,从而 $|Z_i| < 1.65$,难以越过异常判定阈值,造成趋势性、漂移型、局部扰动数据全部漏检。当异常段占比不足时,Z-score 计算的均值,标准差主要受到正常数据主导,导致该方法对时间局部窗口内的异常波动缺乏敏感性和响应性。

为了解决 Z-score 在全球统计特征的约束,提高敏感性和泛化能力,本文提出一种基于 SSAD-Z-score 异常检测方法,该方法结合了滑动窗口斜率分析和 Z-score 算法,能够同时处理局部异常、漂移型异常和趋势性异常数

据,通过增强对数据局部特征的敏感性,有效提高了反吹数据的异常检测精度。

3 基于 SSAD-Z-score 的数据异常检测算法

3.1 基于 SSAD-Z-score 的数据异常检测算法原理

SSAD-Z-score 异常检测算法是一种结合滑动斜率异常检测和 Z-score 异常检测方法的复合型算法,旨在增强对局部异常、漂移型异常、趋势性异常数据的检测能力,解决 Z-score 在检测这类异常的局限性。

滑动斜率异常检测 (sliding slope anomaly detection, SSAD) 是一种基于滑动窗口的异常检测方法,主要用于时间序列数据中的突变、趋势变化或者异常波动检测。该算法的核心思想是通过计算时间序列中每个滑动窗口的局部趋势斜率,并通过对相邻窗口斜率的变化进行比较,来识别数据中的异常行为或趋势突变。Z-score 算法则是标准化检测方法,用于衡量数据点相对于全局均值和标准差的偏离程度。

在 SSAD-Z-score 算法中,Z-score 主要用于对数据进行全局异常检测,确定哪些数据点明显偏离全局的趋势;SSAD 算法关注局部斜率变化,通过检测趋势变化较大的窗口,进一步加强异常点的识别。算法流程如图 5 所示。

SSAD-Z-score 算法主要包含 4 个步骤:数据预处理、局部斜率计算、全局异常检测、异常检测合并与修正、结果平滑与展示。

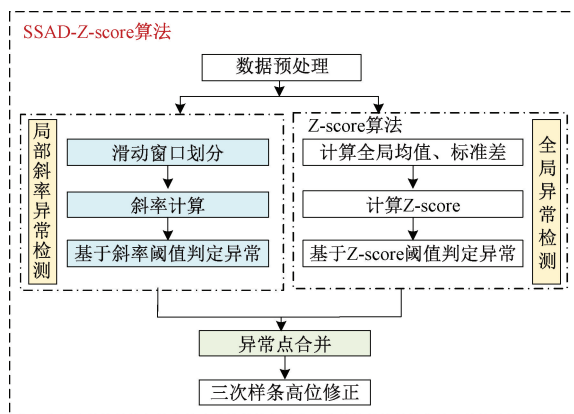


图5 算法流程

Fig. 5 Algorithm flowchart

1) 数据预处理:水泥生产线采集到的现场数据出现异常值、缺失值现象,对这些异常首先进行数据清洗,将非数值字符串 (“NS”) 转换为 NaN,并采用前向填充 (Forward Fill) 填补缺失值,以确保时间序列的连续性。

假设原始数据包含有缺失值 y_i , 通过前向填充方法,将缺失值替换为前一个有效数据点的值:

$$y_i = y_{i-1} \text{ (当 } y_i \text{ 为 NaN 时)} \quad (18)$$

计算复杂度: $C_{\text{preprocess}} = O(N)$, 其中, N 为数据序列的长度。

2) 局部斜率计算: 采用滑动窗口和线性拟合的方法, 提取每个窗口的数据的平均变化率, 即窗口的趋势信息。对于时间序列, 采用长度为 W 的滑动窗口, 步长为 R , 形成长度为 M 的子时间序列:

$$M = \frac{N - W}{R} + 1 \quad (19)$$

每个子序列的斜率 S_{window} 通过以下公式计算:

$$S_{\text{window}} = \frac{\sum_{i=1}^{n-1} (t_i - t_{\text{avg}}) (y_i - y_{\text{avg}})}{\sum_{i=1}^{n-1} (t_i - t_{\text{avg}})^2} \quad (20)$$

式中: t_{avg} 表示时间均值; y_{avg} 表示窑尾烟室 NO_x 均值。得到数据内部的趋势特征信息, 形成斜率序列 $\{S_1, S_2, \dots, S_n\}$, 每个斜率值代表了相邻时间点 NO_x 浓度变化率。

斜率变化检测: 通过计算相邻滑动窗口之间的斜率变化 (ΔS), 判断是否存在异常。如果相邻窗口斜率变化的绝对值超过了预设的阈值, 则认为该数据点存在异常。斜率变化公式为:

$$\Delta S = |S_{\text{current}} - S_{\text{previous}}| \quad (21)$$

式中: S_{current} 为当前窗口斜率; S_{previous} 为前一个窗口的斜率。 $\Delta S > \text{阈值}$, 则认为该数据点存在异常。

计算复杂度: $C_{\text{slope}} = (N - W) \times O(W) = O(N \cdot W)$, 其中 W 为滑动窗口大小, 由于滑动窗口的大小是预设且相对较小的常数, 每次窗口内的线性回归操作开销是恒定的。所以复杂度简化为 $O(N)$ 。

3) Z-score 计算与全局异常检测: 首先计算数据集的全局均值和标准差:

$$Z_i = \frac{X_i - \mu}{\sigma} \quad (22)$$

其中, X_i 是第 i 个数据点的值。根据 Z-score 的值, 若某个数据点的 Z-score 绝对值大于设定的阈值, 则认为该数据点为全局异常。

计算复杂度: $C_{\text{stats}} = O(N)$, 在算法实际计算中, 统计量会通过高效的滚动更新机制进行维护, 使得每次更新的计算开销为 $O(1)$ 。

4) 异常检测合并与修正: 对于滑动窗口斜率检测和全局 Z-score 复合方法检测到的异常数据点, SSAD-Z-score 采用基于三次样条插值的高位修正策略进行修正。

过程分为 4 个步骤: 异常段识别与边界判断、三次样条拟合与评估、值域限制与噪声添加。

异常段识别与边界判断: 识别并聚合连续的异常点形成异常数据段。对于每个识别出的异常数据段 $[t_{\text{start}}, t_{\text{end}}]$, 会向前追溯评估当前窗口的数据点, 更精确的确定异常的实际起始点, 且融合脱硝系统反吹操作的已知持续时间为辅助判断右侧边界, 确保了反吹异常的完整

捕获和修正范围的准确界定。异常段识别计算复杂度为 $O(k)$, 确定异常边界计算复杂度为 $O(k \cdot N)$ 。

三次样条拟合与评估: 基于精准判定的异常段边界, 针对异常段采用三次样条高位修正策略生成新的修正数据, 保证修正曲线函数值、一阶导数和二阶导数的高阶连续性, 提供平滑、符合物理过程内在连续性的修正曲线。

修正时动态设置中间控制点和修正值域限制, 使能够自适应地根据异常段左右两侧的稳定正常数据点的范围进行修正, 并与期望的工艺正常波动区间进行融合。以避免过度简化与失真, 更加还原脱硝物理过程的真实趋势。

三次样条拟合计算复杂度与控制点数量 M 相关 $O(M)$ 。评估计算复杂度与段长度呈线性关系 $O(L_{\text{seg}})$ 。所有异常段总拟合和评估开销为 $O(N)$ 。

计算复杂度: $C_{\text{correct}} = O(k) + O(N \cdot k) + O(N)$

由于反吹异常段长度 k 相比整个数据段 N 来说很小, 因此整体计算复杂度为 $O(N)$ 。

值域限制与噪声添加: 为了确保修正后的数据具有物理上的合理性, 对三次样条插值结果进行值域限制与噪声添加处理。

SSAD-Z-score 算法的整体计算复杂度为 $O(N)$ 。

SSAD-Z-score 方法通过局部斜率的变化与全局 Z-score 的结合, 适应性强, 在窑尾烟室 NO_x 浓度数据存在周期性和趋势变化时, 能够准确检测到异常数据。通过对异常数据进行修正, 可以有效减少异常数据对后续分析和建模的干扰, 可以最终提高预测模型的准确性和稳定性。

3.2 异常修正对长期时序依赖的影响分析

任何形式的数据修正都可能在一定程度上改变原始数据的内在特征。修正策略设计不当或过于激进, 可能会导致系统状态变化的误判或延迟相应、破坏周期性模式、损失高频细节信息等负面影响。

在 SSAD-Z-score 异常检测中使用到的三次样条高位修正策略, 充分利用了水泥脱硝工业场景的数据特性, 能够有效规避上述风险, 将修正过程转化为提升长期时序依赖学习能力的积极因素。分析如下:

本文异常检测处理的核心目标是脱硝系统中由反吹操作引起的窑尾烟室 NO_x 浓度数据的反吹异常。在前文中分析过, 反吹异常具有瞬时性、局部性、周期性, 且在每个整点小时数据中所占比例小。SSAD-Z-score 算法主要针对此类异常进行设计, 其局部性的三次样条高位修正仅在受影响的狭窄区域内进行深度处理, 通过高阶连续性平滑重建异常段, 而非对整个数据序列进行全局性平滑, 从根本上避免了对数据长期趋势的影响。

其次, 反吹异常数据表现为极低的尖峰深谷, 幅值远超正常范围波动。这些异常点在数据序列中属于高幅值

噪声,会严重扰乱数据的统计特性,如均值和标准差,会在后续预测模型训练时被识别为有效模式,导致模型过拟合于噪声。SSAD-Z-score 算法修正反吹异常会使得后续预测模型能学习到数据中真正的长期趋势、周期性模式等时序依赖关系,增强模型对数据规律的捕捉能力,避免丢失趋势信息。

综上,SSAD-Z-score 异常检测与修正策略是基于对水泥脱硝系统工业场景数据特性的有针对性的设计。通过精准定位短时瞬态异常、采用修正技术进行局部平滑重构,使得算法能够有效规避过度平滑导致趋势信息丢失的风险。

3.3 基于 SSAD-Z-score 的数据异常检测参数调整策略

为提升 SSAD-Z-score 异常检测算法在多类型工况数据下的适应性与检测精度,本文提出了一种基于灵敏度估计的参数自适应调整策略。该策略以 Z-score 阈值、滑动窗口大小和斜率变化阈值 3 大关键参数为核心,结合异常类型对参数敏感性的差异,系统性地探讨参数对检测性能的影响机理,并实现针对性参数的动态优化配置。

Z-score 阈值决定了算法对偏离均值数据的容忍程度。滑动窗口大小影响统计平滑程度,控制了在进行局部斜率时所考虑的数据范围。斜率变化阈值控制了相邻窗口之间斜率变化的灵敏度。3 个参数协同作用,使模型能适应时序数据的动态特征。

为定量描述不同参数组合对检测灵敏度的非线性放缩作用,参考多参数非线性系统的响应建模思路,构建如下灵敏度估计函数:

$$Z_{scale} = \frac{k_0}{\alpha^{w_\alpha(n)} \cdot (1 + \gamma)^{w_\gamma(n)} \cdot \beta^{w_\beta(n)}} \quad (23)$$

式中: Z_{scale} 表示当前参数组合下检测器对异常数据的响应程度; k_0 为放缩常数; α 表示滑动窗口大小; β 表示斜率变化阈值; γ 表示 Z-score 阈值; $w_\alpha(n), w_\gamma(n), w_\beta(n)$ 表示针对异常类型 n 的参数灵敏度加权系数。该函数的构建基于参数与检测灵敏度之间的非线性关系,能够反映各个参数对检测性能的影响趋势。

1) 脱硝系统数据异常类型判断

为适应脱硝系统数据的复杂性,本文结合脱硝系统实际运行数据的统计特征和工程经验,将异常数据分为 5 大类:趋势性异常数据、漂移型异常数据、突发性异常数据、局部异常数据和反吹数据。每类异常数据在数据分布、变化速率和持续时间等方面具有显著差异,参数敏感性各不相同,通过不同指标的计算可以判断不同类型的异常,以提高后续异常检测的准确性。

针对检测到的疑似异常区间:

$$S = \{X_t\}_{t=s}^e \quad (24)$$

其中区间起止点分别为 s 和 e , X_t 表示区间内的第 t

个数据点。

(1) 区间均值:

$$\bar{X} = \frac{1}{e - s + 1} \sum_{t=s}^e X_t \quad (25)$$

(2) 区间极小值与极大值: $X_{\min}, X_{\max}$

(3) 区间极差:

$$D = X_{\max} - X_{\min} \quad (26)$$

(4) 区间标准差:

$$\sigma_x = \sqrt{\frac{1}{e - s + 1} \sum_{t=s}^e (X_t - \bar{X})^2} \quad (27)$$

(5) 区间斜率:

$$k = \frac{N \sum t_i X_{t_i} - \sum t_i \sum X_{t_i}}{N \sum t_i^2 - (\sum t_i)^2} \quad (28)$$

式中: t_i 表示区间内第 i 个样本点的序号, $i = 0, 1, \dots, N - 1$; N 表示区间长度。

(6) 区间一阶差分绝对值的 95% 分位数:

$$P_{95}(|\Delta X|) = \text{Percentile}_{95}(|X_{t+1} - X_t|) \quad (29)$$

式中: Percentile_{95} 表示取输入数据的第 95 百分位数。

(7) 全局均值与标准差:

$$\mu_g = \frac{1}{T} \sum_{t=1}^T X_t \quad (30)$$

$$\sigma_g = \sqrt{\frac{1}{T} \sum_{t=1}^T (X_t - \mu_g)^2} \quad (31)$$

式中: T 表示全局数据总长度。

不同异常类型所用统计指标如表 4 所示。

表 4 区间异常类型特点及判别准则

Table 4 Characteristics and discrimination criteria of interval anomaly types

异常类型	特点	判别准则
趋势性异常数据	数据持续上升或下降	$ k > 0.2$
漂移型异常数据	整体水平缓慢偏移	$0.05 < k < 0.2$
突发异常数据	短时剧烈波动	$\mu_g - X_{\min} > 2\sigma_g$ $D > 2\sigma_g$
局部异常数据	局部区间异常跳变	$P_{95}(\Delta X) > 2\sigma_x$
反吹数据	整体水平显著低于正常工况	$\bar{X} < 0.5\mu_g$

2) 数据异常检测参数调整策略

为进一步可视化各参数与异常类型的耦合关系,构建了三维响应曲面图如图 6 所示,展示了分别固定 β, γ, α 的条件下,不同异常类型在两个变量维度上的检测强度变化,表 5 总结了不同类型异常数据下对 3 类关键参数的配置级别。

对于参数的“低、中、高”分级,参考了脱硝系统历史数据的分布特性和多组实验结果。Z-score 阈值大小划分以正态分布理论为基础,滑动窗口大小划分以异常持

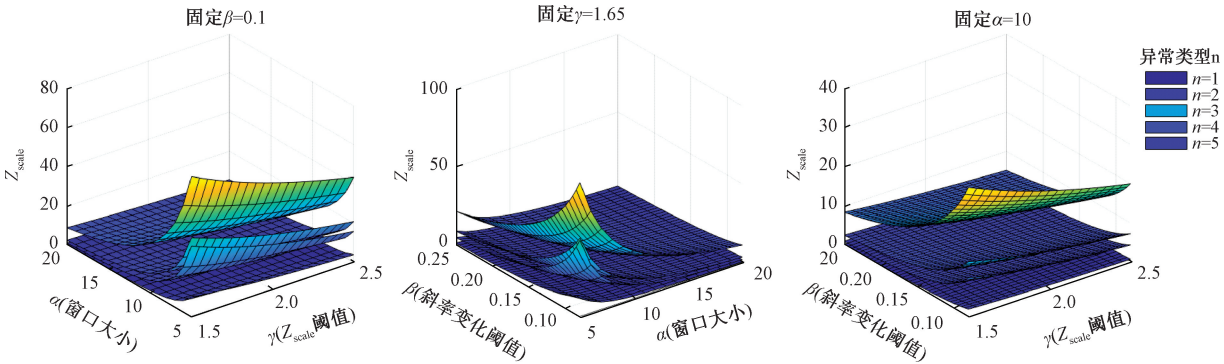


图 6 SSAD-Z-score 三维响应曲面图
Fig. 6 3D response surface plot of SSAD-Z-score

续时间和系统响应速度为基础,斜率变化阈值大小划分以系统正常波动范围为基础。

结果表明,SSAD-Z-score 异常检测方法在参数配置上具备良好的可调性与类型适配能力。其中中等参数的设定是通过大量实验对比不同参数组合下的检测覆盖率、误报率、修正幅度等指标后选取的最优解。不同异常类型在参数组合维度上的响应规律具有明显特征,基于灵敏度建模的调参方法不仅有助于实现面向异常类型的自适应参数配置策略,也为后续构建更加稳定、准确的数据预测模型提供了数据质量保障和理论支持。

表 5 不同类型异常数据 3 类关键参数配置设定表

Table 5 Configuration of three key parameters for different types of anomalous data

异常类型 编号	异常类型	Z-score 阈值 γ	滑动窗口 大小 α	斜率变化 阈值 β
$n=1$	趋势性异常数据	高设定	高设定	高设定
$n=2$	漂移型异常数据	中-高设定	中-高设定	高设定
$n=3$	突发异常数据	低设定	低设定	低设定
$n=4$	局部异常数据	中设定	中设定	低-中设定
$n=5$	反吹数据	中设定	中设定	中设定

4 基于 LFLSTM 的水泥脱硝系统 NO_x 预测模型

本文选择窑尾烟室 NO_x 浓度和 SNCR 喷氨泵频率作为预测模型的输入变量,同时在标准长短期记忆(long short-term memory, LSTM)模型的基础上引入烟囱 NO_x 前一时刻浓度值作为额外输入变量,以增强模型对时序数据的捕捉能力,提高预测精度。预测模型框架如图 7 所示。

4.1 预测模型输入变量处理

预测模型如下:

$$Y_{\text{NO}_x}(t+d) = aY_{\text{NO}_x}(t+d-1) + bX_{\text{NH}_3}(t) + cX_{\text{NO}_x}(t) \quad (32)$$

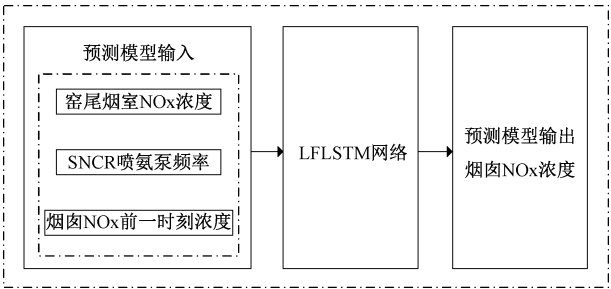


图 7 预测模型框架
Fig. 7 Prediction model framework

式中: d 代表预测步长; $Y_{\text{NO}_x}(t+d)$ 为未来 d 时刻的烟囱 NO_x 浓度; $X_{\text{NH}_3}(t)$ 为当前时刻的 SNCR 喷氨泵频率; $X_{\text{NO}_x}(t)$ 为当前时刻窑尾烟室 NO_x 浓度; $Y_{\text{NO}_x}(t+d-1)$ 为前一时刻烟囱 NO_x 浓度; a 为前一时刻烟囱 NO_x 浓度对预测结果的影响权重; b 为当前时刻 SNCR 喷氨泵频率对预测结果的影响程度; c 为当前时刻窑尾烟室 NO_x 浓度对预测结果的影响程度。变量时序匹配示意图如图 8(a) 所示。

然而,在实际工程应用中,在 t 时刻无法获得 $t+d-1$ 时刻的烟囱 NO_x 浓度数据,为此,通过时间序列分析和推算,可得最终预测模型如下:

$$Y_{\text{NO}_x}(t+d) = a^d Y_{\text{NO}_x}(t-d+1) + \sum_{i=0}^{d-1} a^i (bX_{\text{NH}_3}(t-i) + cX_{\text{NO}_x}(t-i)) \quad (33)$$

模型输出量: $t+d$ 时刻烟囱 NO_x 浓度;

模型输入量: $t-d+1$ 时刻、 $\cdots$ 、 $t-1$ 时刻、 t 时刻一号泵喷氨泵频率; $t-d+1$ 时刻、 $\cdots$ 、 $t-1$ 时刻、 t 时刻窑尾烟室 NO_x 浓度; t 时刻烟囱 NO_x 浓度。

式(33)表明, $t+d$ 时刻的烟囱 NO_x 浓度将由从 $t-d+1$ 时刻到 t 时刻的一号泵频率和窑尾烟室 NO_x 浓度,以及 t 时刻烟囱 NO_x 浓度 3 个变量作为模型输入进行预测。最终预测模型变量时序匹配示意图如图 8(b) 所示。

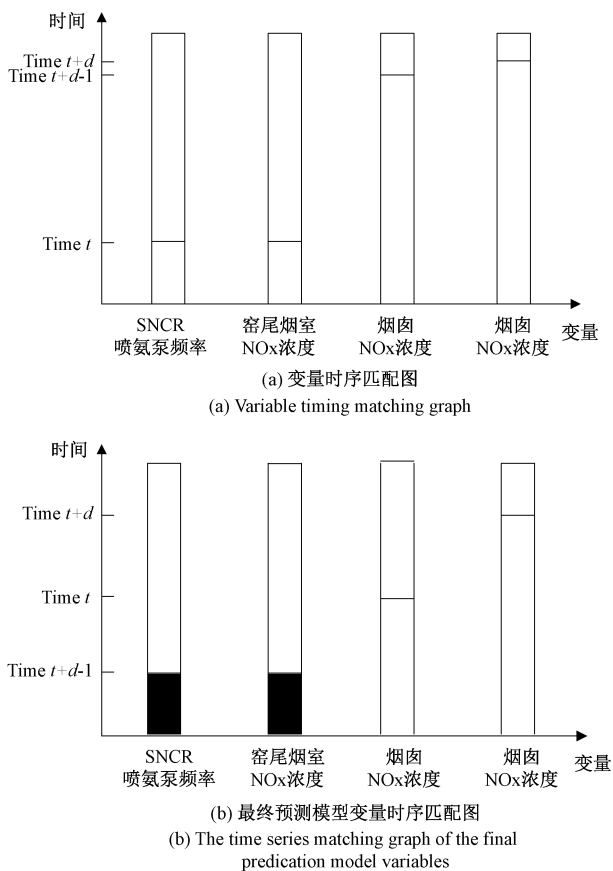


图8 变量时序匹配图

Fig. 8 Variable temporal matching diagram

4.2 LFLSTM 预测算法描述

LFLSTM 是一种结合滞后特征和长短期记忆网络的混合模型。LF 指的是将 SNCR 喷氨泵频率、窑尾烟室 NO_x 浓度、烟囱 NO_x 浓度数据作为输入特征,利用历史信息进行未来烟囱 NO_x 浓度预测。LSTM 在烟囱 NO_x 浓度的预测中能够学习到不同时间步之间复杂的非线性关系,使其能够从历史数据中推断出未来的 NO_x 浓度变化趋势。

LFLSTM 模型建立的主要步骤包括数据预处理与滞后特征构建、LFLSTM 模型构建与训练、模型评估等。

数据预处理与滞后特征构建:

在 LFLSTM 模型中,通过引入过去的时间步数据来帮助模型捕捉时间序列数据中的时序依赖关系。包括输入特征构建和目标输出构建。

输入特征构建是保证每一个样本的输入包括过去 lag 个时刻的喷氨泵频率和窑尾烟室 NO_x 浓度数据,和当前时刻的烟囱 NO_x 浓度。目标输出构建是保证目标输出为未来 lag 时刻的烟囱 NO_x 浓度。其中 lag 表示烟囱 NO_x 浓度值滞后于一号喷氨泵频率和窑尾烟室 NO_x 浓度的时间,根据实际工程工艺分析,认为 lag 为 180 s。

原始数据包含一号喷氨泵频率 $u_1(t)$ 、窑尾烟室 NO_x 浓度 $u_2(t)$ 、烟囱 NO_x 浓度 $u_3(t)$,其中 t 为时间索引。

构造输入特征如下:

$$X_t = \begin{bmatrix} u_1(t-lag), u_1(t-lag+1), \dots, u_1(t-1), \\ u_2(t-lag), u_2(t-lag+1), \dots, u_2(t-1), \\ u_3(t) \end{bmatrix} \quad (34)$$

其中, $lag=180$ s,表示将过去 180 个时间步(即过去 180 s)的一号喷氨泵频率、窑尾烟室 NO_x 浓度和当前时刻的烟囱 NO_x 浓度作为输入。

目标输出 Y_t 表示未来 180 s 的烟囱 NO_x 浓度。

$$Y_t = u_3(t+lag) \quad (35)$$

其中, $u_3(t+lag)$ 表示在 $t+lag$ 即未来 180 s 时的烟囱 NO_x 浓度。

并且对数据进行归一化,将输入特征和目标输出的值缩放到 $[0,1]$ 范围内。为提高评估模型的泛化能力,将数据集划分为训练集和测试集,测试集占数据集的 80%,测试集占 20%。

LSTM 模型构建与训练:

使用两层 LSTM 和 Dropout 层来构建模型,以提高网络捕捉历史和未来完整信息的能力^[22]。

输入层:接受形状为 $(time_steps, features)$ 的输入数据,输出形状为 $(\text{样本数}, time_steps, units)$ 的三维张量。其中 $units$ 为 LSTM 单元的个数, $units=128$ 。

$$h_t = \text{LSTM}(x_{t-1}, h_{t-1}; W_{\text{input}}, W_{\text{output}}) \quad (36)$$

式中: h_t 是 LSTM 的隐藏状态; $W_{\text{input}}, W_{\text{output}}$ 是输入和输出的权重矩阵; x_{t-1} 是前一时刻的输入。

第二层 LSTM:这层的输出形状是 $(\text{样本数}, units)$,即每个样本的最终状态。 $units=64$ 。

$$h_t = \text{LSTM}(h_{t-1}; W_{\text{input}}, W_{\text{output}}) \quad (37)$$

Dropout 层:在每一层 LSTM 后,加入 Dropout 层,使用 dropout 概率($p=0.3$)随机丢弃部分神经元,以减少过拟合。

$$h'_t = \text{Dropout}(h_t, p) \quad (38)$$

式中: h'_t 是应用 Dropout 后的隐藏状态。

输出层:使用一个全连接层输出预测值。输出层的激活函数为线性激活函数,输出形状为 $(\text{样本数}, 1)$ 。

$$y_t = \text{Dense}(h'_t; W_{\text{output}}) \quad (39)$$

式中: y_t 是模型预测的目标值; W_{output} 是输出层的权重矩阵。

选择均方误差 MSE 作为损失函数:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_{\text{true},i} - y_{\text{pred},i})^2 \quad (40)$$

式中: $y_{\text{true},i}$ 是真实值; $y_{\text{pred},i}$ 是预测值; N 是样本数。

使用 Adam 优化器进行训练:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} - \alpha \times \frac{m_t}{\sqrt{v_t} + \varepsilon} \tag{41}$$

式中： $\boldsymbol{\theta}$ 是模型参数； m_t 和 v_t 是动量和梯度平方的移动平均； α 是学习率； ε 是防止除零错误的常数。

4.3 算法性能评估

在完成多轮训练后,为评价模型的预测效果,检验模型的泛化能力。采用决定系数 (coefficient of determination, R^2), 均方根误差 (root mean squared error, RMSE), 平均绝对误差 (mean absolute error, MAE) 等评估指标来衡量模型的预测性能。

均方根误差 RMSE 是 MSE 的平方根:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{true,i} - y_{pred,i})^2} \tag{42}$$

式中: $y_{true,i}$ 是真实值; $y_{pred,i}$ 是预测值; N 是样本数量。

平均绝对误差 MAE 计算的是预测值和真实值之间的绝对差的平均值:

$$MAE = \frac{1}{N} |y_{true,i} - y_{pred,i}| \tag{43}$$

决定系数 R^2 反映了模型解释方差的比例:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{true,i} - y_{pred,i})^2}{\sum_{i=1}^N (y_{true,i} - \bar{y}_{true})^2} \tag{44}$$

式中: $\bar{y}_{true}$ 是真实值的均值。

5 仿真实验与结果分析

为进一步验证基于 LF-LSTM 的水泥脱硝系统 NOx 预测模型的有效性, 评估结合 SSAD-Z-score 算法的 LFLSTM 模型的实际部署可行性。本文以安徽省某水泥厂的分布式控制系统 DCS 为依托, 收集了该厂关键工艺参数的历史实际数据进行了一系列全面的仿真验证实验, 所用数据集包含了烟囱 NOx 浓度、窑尾烟室 NOx 浓度和一号喷氨泵频率等多个变量。所有对比实验均在统一的环境和条件下进行。

5.1 SSAD-Z-score 数据异常检测与修正算法性能对比评估

1) 不同异常检测算法对比实验

为验证 SSAD-Z-score 异常检测与修正算法在数据质量提升、异常检测精度及计算效率方面的有效性与优越性, 使用 SSAD-Z-score 算法对一段包含反吹异常数据的 2 800 个数据集进行异常检测, 将相同数据集同样输入到传统 Z-score、孤立森林、局部异常因子 (local outlier factor, LOF) 以及基于小波变换等多种异常检测方法中进行异常检测对比评估, 评估指标包括覆盖率、召回率、修正后标准差、平均修正幅度以及执行时间, 不同模型异

常检测效果如图 9~13 所示。表 6 列出了多种算法的详细性能指标对比。

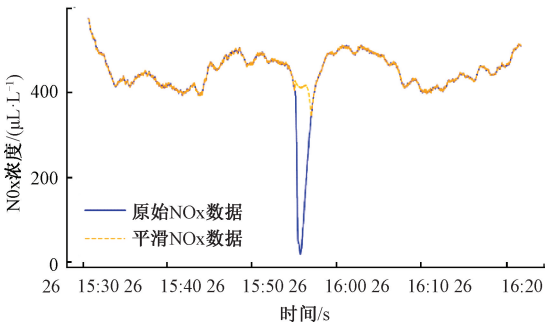


图 9 SSAD-z-score 异常检测与处理效果

Fig. 9 SSAD-Z-score anomaly detection and processing results

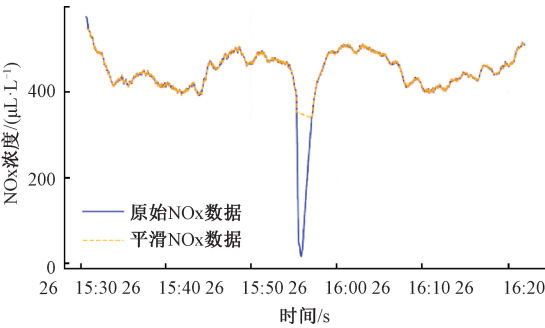


图 10 Z-score 异常检测与处理效果

Fig. 10 Z-score anomaly detection and processing results

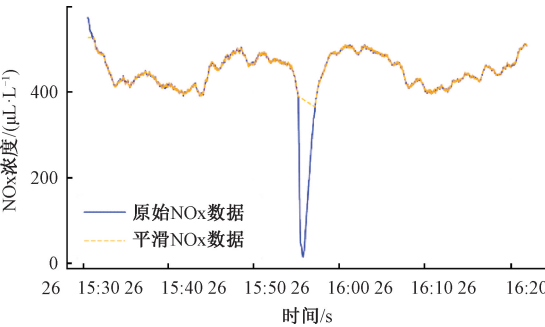


图 11 孤立森林异常检测与处理效果

Fig. 11 Isolated forest anomaly detection and handling visualization

从结果可以看出, SSAD-Z-score 以 47.67% 的最高覆盖率和 0.844 的最高召回率, 远超所有对比方法, 表明其能最准确地识别出水泥脱硝系统中的反吹异常, 确保了异常的低漏报率。修正后标准差为 35.65, 对比所有方法中最低的, 说明其卓越的数据平滑与去噪能力。尽管执行时间高于某轻量级方法, 但仍低于 1 s 的数据采集间隔, 完全满足水泥脱硝工业实时应用需求。相较之下, 其他方法在覆盖率和召回率上均表现不足, 修正效果有限。

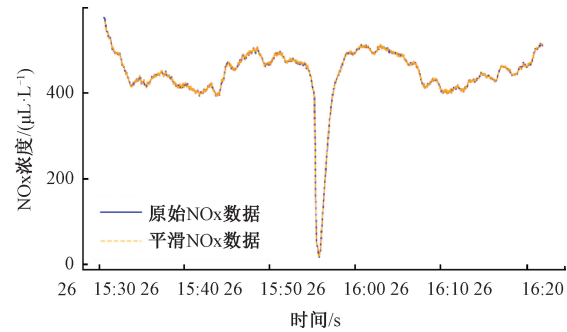


图 12 LOF 异常检测与处理效果

Fig. 12 LOF anomaly detection and handling effect diagram

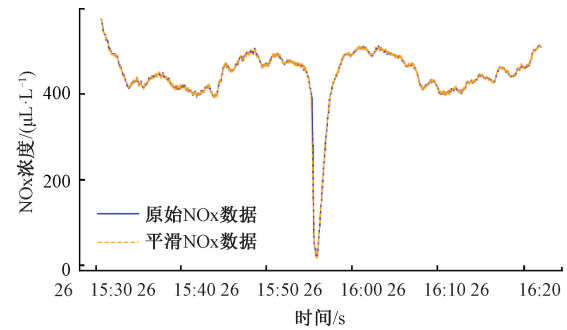


图 13 基于小波变换方法异常检测与处理效果

Fig. 13 Visualization of anomaly detection and handling based on wavelet transform method

表 6 不同异常检测算法性能对比

Table 6 Performance comparison of different anomaly detection algorithms

评估指标	SSAD-Z-score	基础 Z-score	孤立森林	LOF	小波变换
覆盖率/%	47.67	3.80	4.93	4.93	1.79
召回率	0.844	0.650	0.000	0.025	0.025
修正后标准差	35.65	39.51	80.59	105.21	105.02
平均修正幅度	8.81	6.70	7.96	0.17	0.38
执行时间/s	0.448 8	0.001 0	0.005 5	0.149 4	0.001 5

2) 不同参数下 SSAD-Z-score 算法检测性能对比实验
为验证 SSAD-Z-score 算法在不同参数配置下对实际工艺 NO_x 实时数据的应对效果,本节设计了一系列对比仿真实验,针对反吹异常数据分别考察滑动窗口大小 α 、Z-score 阈值 γ 、斜率变化阈值 β 对 SSAD-Z-score 异常检测性能的影响。

固定斜率变化阈值为 0.1、Z-score 阈值为 1.65,逐步调整滑动窗口大小,效果如图 14 所示。

固定滑动窗口大小为 10、Z-score 阈值为 1.65,逐步调整斜率变化阈值大小,效果如图 15 所示。

固定滑动窗口大小为 10、斜率变化阈值为 0.1,逐步调整 Z-score 阈值大小,效果如图 16 所示。

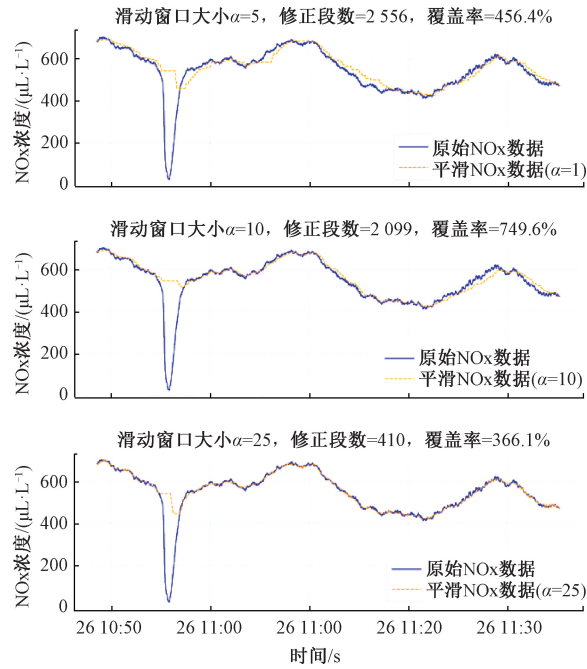


图 14 不同滑动窗口大小下异常检测效果

Fig. 14 Anomaly detection results with different sliding window sizes

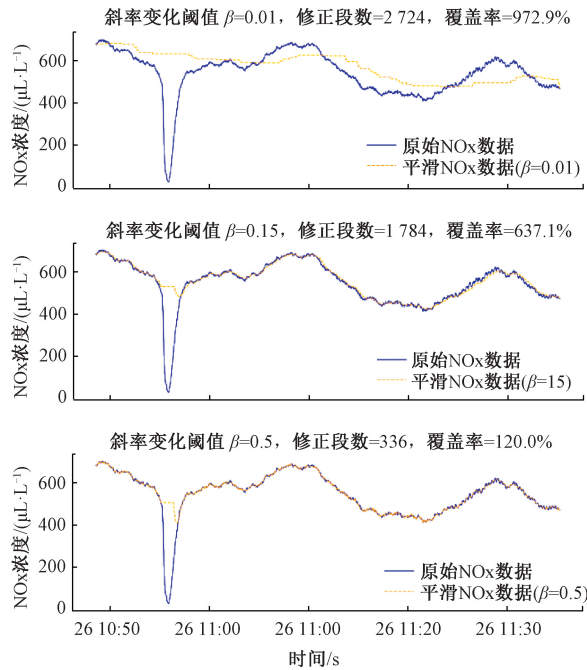


图 15 不同斜率阈值大小下异常检测效果

Fig. 15 Anomaly detection results with different slope thresholds

表 7~9 可以看出中等参数设置 ($\alpha=10$ 、 $\beta=0.15$ 、 $\gamma=1.65$) 在异常检测覆盖率、平均修正幅度和数据相关性 3 方面均取得了最佳平衡,能有效识别反吹数据中的复合

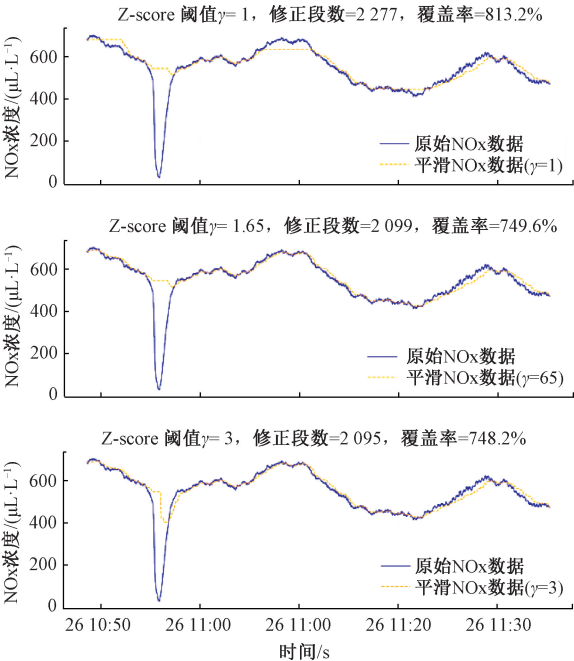


图 16 不同 Z-score 阈值大小下异常检测效果
Fig. 16 Anomaly detection results with different Z-score thresholds

异常,又能避免过度修正导致的数据失真,是反吹数据异常检测的最优选择。

表 7 不同斜率变化阈值 β 的性能指标对比

Table 7 Performance comparison of different slope change thresholds β

斜率阈值	异常检测覆盖率	平均修正幅度	与原始数据相关性
0.01	973.51	52.95	0.370
0.15	631.84	14.68	0.778
0.5	117.47	9.40	0.813

表 8 不同滑动窗口大小 α 的核心性能指标对比

Table 8 Core performance comparison of different sliding window sizes α

滑动窗口大小	异常检测覆盖率	平均修正幅度	与原始数据相关性
5	298.68	11.09	0.812
10	117.47	9.40	0.813
25	51.86	7.80	0.795

表 9 不同 Z-score 阈值 γ 的核心性能指标对比

Table 9 Core performance comparison of different Z-score thresholds γ

Z-score 阈值	异常检测覆盖率	平均修正幅度	与原始数据相关性
1	324.92	15.12	0.803
1.65	117.47	9.40	0.813
3	110.97	8.40	0.836

仿真验证了关于反吹类数据对应的灵敏度函数响应规律,关于 $n=5$ 的反吹类复合异常数据,采用中等设置的滑动窗口大小、斜率变化阈值、Z-score 阈值时,检测灵敏度与误判控制效果达成良好平衡,能够准确识别反吹过程中的异常数据。

5.2 LFLSTM 模型预测性能对比评估

1) 不同预测模型对比实验

为验证 LFLSTM 算法的可行性与优越性,将处理后数据输入到 LFLSTM 模型、Transformer 模型、Informer 模型、门控循环单元 (gated recurrent unit, GRU) 模型中进行训练。使用拟合度 R^2 、RMSE、MAE 进行对比。所有模型均在相同的数据集、相同的训练集/测试集划分比例、相同的滞后步长、相同的训练周期和批处理大小下进行训练和评估。烟囱 NO_x 浓度的预测值与实际值的对比预测结果如图 17~20 所示,多种模型指标如表 10 所示。

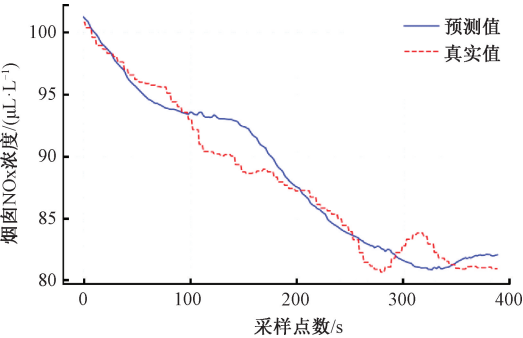


图 17 LFLSTM 模型预测效果图

Fig. 17 LFLSTM model prediction visualization

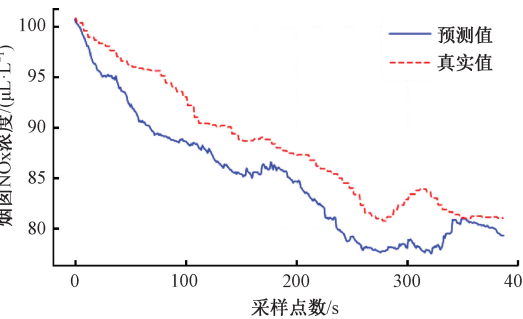


图 18 Transformer 模型预测效果图

Fig. 18 Transformer model prediction visualization

从结果可以看出,LFLSTM 模型 R^2 高达 0.884 5,远超过 GRU、Transformer 和 Informer。LFLSTM 的 RMSE 和 MAE 均为最低,表明预测结果最接近真实值,具有最高的稳定性和准确性。GRU 和 Transformer 虽然在平均单样本预测耗时上略有优势,但 LFLSTM 的 4.88 ms 也远低于水泥脱硝工业系统 1 s/点的数据采集频率,实时性完全达标。

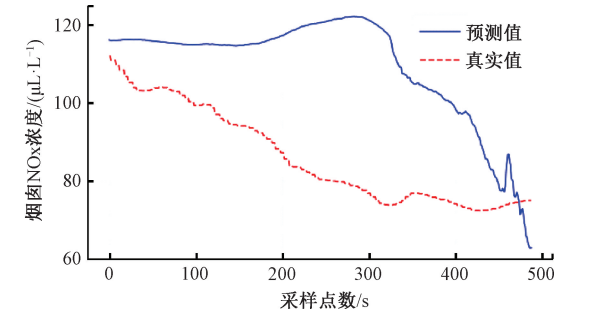


图 19 Informer 模型预测效果
Fig. 19 Informer model prediction visualization

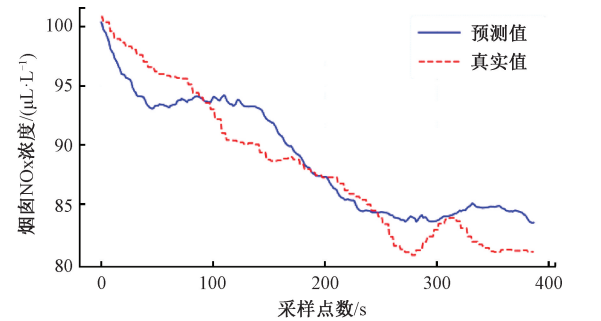


图 20 GRU 模型预测效果
Fig. 20 GRU model prediction visualization

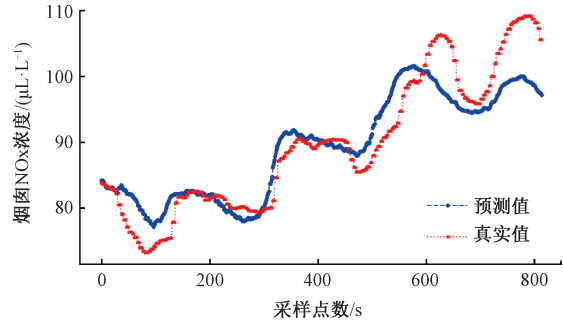


图 21 结合 SSAD-Z-SCORE 模型预测效果
Fig. 21 Prediction results using the SSAD-Z-score model

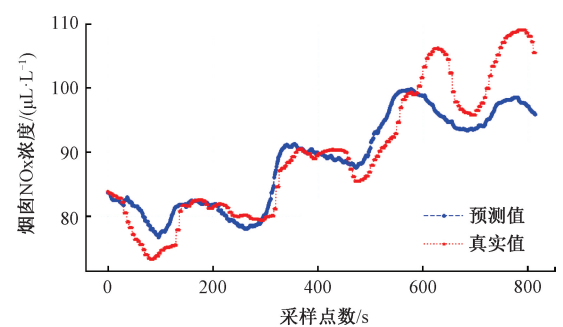


图 22 结合 Z-score 模型预测效果
Fig. 22 Prediction results using the Z-score model

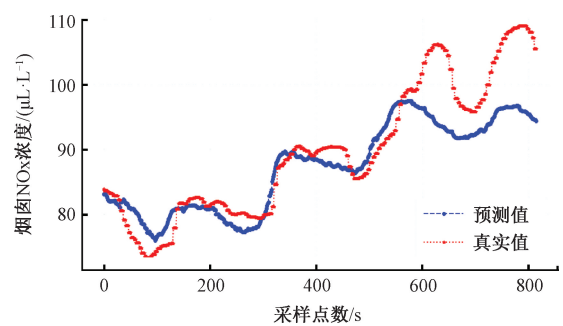


图 23 未进行异常数据处理模型预测效果
Fig. 23 Prediction results without anomaly data processing

表 10 不同预测模型的核心性能指标对比

Table 10 Comparison of core performance metrics across different predictive models				
评估指标	LFLSTM	Transformer	Informer	gru
平均单样本预测				
耗时/ms	4. 88	7. 29	8. 43	1. 80
R ²	0. 884 5	0. 433 4	-2. 039 2	0. 786 9
RMSE	2. 026 5	4. 489 4	21. 146 6	2. 382 4
MAE	1. 798 5	3. 910 2	19. 060 0	5. 506 0

2) 不同数据集下 LFLSTM 预测性能评估实验

为了评估修正窑尾烟室 NO_x 浓度数据是否影响 LFLSTM 模型的预测效果,将未处理的数据、使用基础 Z-score 算法处理后的数据以及使用 SSAD-Z-score 算法处理后的数据输入到 LFLSTM 模型中进行预测评估对比。使用多个评估指标,包括 R²、RMSE、MAE。烟囱 NO_x 浓度的预测值与实际值的对比如图 21~23 所示。

3 种模型评估指标如表 11 所示。

表 11 评估指标数据表

Table 11 Evaluation metrics data table			
指标	LFLSTM	Z-score-LFLSTM	SSAD-Z-score-LFLSTM
R ²	0. 731 911	0. 763 511 3	0. 824 097
RMSE	5. 315 882	4. 966 535 5	4. 305 981
MAE	3. 873 885	3. 530 417 9	3. 297 620

可以看出,SSAD-Z-score-LFLSTM 在各项评估指标上均优于 LFLSTM 和 Z-score-LFLSTM。

具体来说,SSAD-Z-score-LFLSTM 的 R² 值为 0. 824 097,显著高于 LFLSTM 和 Z-score-LFLSTM,表明其拟合效果最佳,能够更准确地拟合烟囱 NO_x 浓度数据。SSAD-Z-score-LFLSTM 的 RMSE 为 4. 305 981,较 LFLSTM 和 Z-score-LFLSTM 更低,显示出较小的误差和更高的预测稳定性。SSAD-Z-score-LFLSTM 的 MAE 为 3. 297 620,低于其他两个模型,说明其预测值与实际值的偏差最小。

综上,SSAD-Z-score-LFLSTM 在准确性、稳定性和误差控制方面表现更优,具有较高的应用价值。

6 结 论

本文围绕水泥脱硝系统 NO_x 浓度的智能预测与异常数据处理,提出并实现了一套涵盖机理分析、变量筛选、异常检测与深度学习建模的系统性技术路线。通过构建 SSAD-Z-score 算法,结合滑动窗口斜率变化与全局 Z-score 判据,实现了对多类型复杂异常数据的高效识别与自适应修正,显著提升了原始数据的质量和后续建模的可靠性。在参数灵敏度分析中,系统评估了不同参数组合对不同异常检测性能的影响,明确了合理参数配置对于平衡检测灵敏度与数据保真性的关键作用。在建模阶段,将多步时滞特征引入 LSTM 模型,充分挖掘了 NO_x 浓度的时序演化规律,提升了模型对动态变化的刻画能力。实验结果表明,通过与传统方法的对比,所提出的复合方法在异常检测准确性和 NO_x 浓度预测精度方面有明显提升。未来工作将进一步优化特征工程和模型结构,探索更高效的异常检测与预测方法,并加强数据去噪与模型泛化能力研究,以提升工业过程的智能化水平。

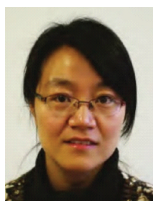
参考文献

- [1] 贾福强. 水泥工业大气污染物及防治技术[J]. 绿色环保建材, 2021(9): 36-37.
JIA F Q. Atmospheric pollutants and control technologies in cement industry[J]. Green Environmental Protection Building Materials, 2021(9): 36-37.
- [2] LIU H, HONG Q, LIU H, et al. Effects of temperature and additives on NO_x emission from combustion of fast-growing grass[J]. Frontiers in Energy Research, 2021, 9: 772755.
- [3] YAN A, CAO S. A NO_x concentration prediction model based on a sparse regularization stochastic configuration network[J]. Instrumentation, 2024, 11(3): 13-22.
- [4] 孙崇建. 水泥回转窑机理建模及预测控制研究[D]. 济南: 济南大学, 2021.
SUN CH J. Research on mechanism modeling and predictive control of cement rotary kiln [D]. Jinan: University of Jinan, 2021.
- [5] YAO Y F, CHEN Y X. Technical route to achieve ultra-low emission of nitrogen oxides with predictive model of nitrogen oxide background concentration[J]. Processes, 2020, 8(9): 1104-1112.
- [6] 刘文彪, 段礼祥, 耿帆, 等. 基于 CNN-BLSTM 网络的轴承性能退化预测 [J]. 电子测量与仪器学报, 2021, 35(2): 80-86.
LIU W B, DUAN L X, GENG F, et al. Prediction of bearing performance degradation based on CNN-BLSTM network [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(2): 80-86.
- [7] SUN B M, WANG D H, YANG B, et al. Prediction model for the boiler NO_x emission with material properties based on the artificial neural network [J]. Advanced Materials Research, 2013, 676: 40-45.
- [8] YANG T, MA K, LYU Y, et al. Real-time dynamic prediction model of NO_x emission of coal-fired boilers under variable load conditions [J]. Fuel, 2020, 274: 117811-117820.
- [9] XIE P, GAO M, ZHANG H, et al. Dynamic modeling for NO_x emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network [J]. Energy, 2020, 190: 116482.
- [10] HE W, LI J F, TANG Z H, et al. A novel hybrid cnn-lstm scheme for nitrogen oxide emission prediction in FCC unit [J]. Mathematical Problems in Engineering, 2020: 1-12.
- [11] KANG J J, NIU Y G, HU B, et al. Dynamic modeling of SCR denitration systems in coal-fired power plants based on a bi-directional long short-term memory method [J]. Process Safety and Environmental Protection, 2021, 148: 867-878.
- [12] 徐清泉. 基于 CT-DBN 的水泥煅烧过程烟气脱硝系统模型预测控制研究[D]. 秦皇岛: 燕山大学, 2023.
XU Q Q. Research on model predictive control of cement calcination flue Gas denitrification system based on CT-DBN [D]. Qinhuangdao: Yanshan University, 2023.
- [13] 宋智星. 基于深度信念网络的水泥窑烟气脱硝预测控制研究[D]. 秦皇岛: 燕山大学, 2022.
SONG ZH X. Research on predictive control of cement kiln flue Gas denitrification based on deep belief network [D]. Qinhuangdao: Yanshan University, 2022.
- [14] 闫欢. 基于 LSTM 神经网络的水泥烟气 NO_x 排放浓度预测模型研究[D]. 秦皇岛: 燕山大学, 2023.
YAN H. Research on prediction model of NO_x emission concentration in cement flue Gas based on LSTM neural network [D]. Qinhuangdao: Yanshan University, 2023.
- [15] 李宁, 丁雪松, 徐雨红, 等. 循环流化床锅炉燃烧中 NO_x 生成机理与预测研究进展 [J]. 能源与节能, 2021(8): 74-76.
LI N, DING X S, XU Y H, et al. Progress in mechanism and prediction of NO_x formation in circulating fluidized bed boiler combustion [J]. Energy and Energy Conservation, 2021(8): 74-76.
- [16] 倪旭光, 王涛, 蒋宝庆. SNCR-SCR 混合法烟气脱硝技术在水泥行业中的应用 [J]. 水泥, 2021(2): 61-63.
NI X G, WANG T, JIANG B Q. Application of SNCR-SCR hybrid flue Gas denitrification technology in cement

- industry[J]. Cement, 2021(2):61-63.
- [17] 凌绍华,景长勇,马婧,等. 选择性非催化还原法烟气脱硝工业试验[J]. 化工环保,2013,33(4):304-307.
LING SH H, JING CH Y, MA J, et al. Industrial test of selective non-catalytic reduction flue GAS denitrification[J]. Chemical Industry and Environmental Protection, 2013, 33(4):304-307.
- [18] 邸银露. 基于 MT-BiLSTM 算法的水泥窑烟气脱硝模型预测控制研究[D]. 秦皇岛:燕山大学,2023.
DI Y L. Research on model predictive control of cement kiln flue Gas denitrification based on MT-BiLSTM Algorithm [D]. Qinhuangdao: Yanshan University, 2023.
- [19] 陈薇,解俊哲,康志伟,等. 基于改进的 SE-LSTM 水泥熟料 f-Cao 预测模型[J]. 电子测量与仪器学报,2025, 39(1):225-233.
CHEN W, XIE J ZH, KANG ZH W, et al. Improved SE-LSTM based prediction model for f-Cao in cement clinker [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(1):225-233.
- [20] 李玲. 燃煤电站 SCR 脱硝系统数据驱动建模与复合优化控制研究[D]. 北京:华北电力大学(北京),2022.
LI L. Data-driven modeling and composite optimization control of SCR denitrification system in coal-fired power plant[D]. Beijing:North China Electric Power University (Beijing), 2022.
- [21] 刘涛. 基于峰噪比和改进 Z-score 算法的音频异常检测方法[D]. 天津:南开大学,2022.
LIU T. Audio anomaly detection method based on peak signal-to-noise ratio and improved Z-score algorithm[D]. Tianjin:Nankai University, 2022.

- [22] 陈佳琳,尚志武,张雷. 基于 IDRSN-BiLSTM 的铣削加工表面粗糙度预测方法[J]. 仪器仪表学报,2024, 45(4):27-36.
CHEN J L, SHANG ZH W, ZHANG L. Prediction method of milled surface roughness based on IDRSN-BiLSTM[J]. Chinese Journal of Scientific Instrument, 2024, 45(4):27-36.

作者简介



陈薇,2002 年于中国科学技术大学获得学士学位,2007 年于中国科学技术大学获得博士学位,现为合肥工业大学副教授,主要研究方向为复杂系统建模与控制和非线性预测控制。

E-mail: windy02@hfut.edu.cn

Chen Wei received her B. Sc. degree from University of Science and Technology of China in 2002, Ph. D. degree from University of Science and Technology of China in 2007, respectively. Now she is an associate professor in Hefei University of Technology. Her main research interests include complex systems modeling and nonlinear predictive control.



何浩然(通信作者),2023 年于合肥工业大学获得学士学位,现为合肥工业大学硕士研究生,主要研究方向为复杂系统建模与预测控制。

E-mail: 1171269253@qq.com

He Haoran (Corresponding author) received his B. Sc. degree from Hefei University of Technology in 2023. Now he is a M. Sc. candidate in the Hefei University of Technology. His main research interests include complex system modeling and predictive control.