

DOI: 10.11949/0438-1157.

基于时滞感知 Transformer 的加氢装置柴油收率预测方法

阮志坚, 李佳鹤, 李佳洋

(蓝卓数字科技有限公司, 浙江 杭州 310053)

摘要: 柴油收率受原料、催化剂及操作参数的非线性、异构性与时滞效应耦合影响, 传统数据驱动模型因忽略物理可解释的时序依赖而精度受限。针对上述问题, 提出一种基于时滞感知 Transformer (Time-Lag Aware Transformer, TLAT) 的柴油收率预测方法: 首先基于条件互信息量化各变量对收率的有效时滞窗口; 继而在 Temporal Fusion Transformer 基础上, 设计工艺先验引导的结构化注意力掩码, 约束模型仅在合理时滞范围内建模动态依赖; 同时引入静态-动态门控融合模块, 利用静态上下文生成可学习门控信号, 动态调节动态特征的权重, 实现静态变量对动态特征的自适应调制。基于某石化企业 140 万吨/年加氢裂化装置近 3 年 104 维高频率运行数据, TLAT 在测试集上 MAE 达 0.0182、 R^2 为 0.937, 显著优于 XGBoost、LSTM、Informer 和标准 TFT, 为智能工厂精益管控提供有效支撑。

关键词: 加氢; 柴油收率; 预测; 动态建模; 时滞感知; Transformer

中图分类号: TP183; TE624

文献标志码: A

文章编号: 0438-1157 (XXXX) XX-0001-12

A method for predicting diesel yield of hydrocracking unit based on time-lag-aware transformer

RUAN Zhijian, LI Jiahe, LI Jiayang

(Bluetron Digital Technology Co., Ltd., Hangzhou 310053, Zhejiang, China)

Abstract: Diesel yield is jointly influenced by the nonlinear, heterogeneous, and time-lag effects of feedstock properties, catalyst performance, and operational parameters. Conventional data-driven models suffer from limited prediction accuracy due to their inability to explicitly capture physically interpretable temporal dependencies. To address this issue, a diesel yield prediction method based on a Time-Lag Aware Transformer (TLAT) is proposed. First, conditional mutual information is employed to quantitatively identify the effective time-lag window of each process variable with respect to diesel yield. On this basis, a process-prior-guided structured attention mask is constructed within the Temporal Fusion Transformer (TFT) framework, which constrains the model to learn dynamic dependencies only within physically reasonable time-lag ranges. In addition, a static - dynamic gating fusion module is introduced, where static contextual information is utilized to generate learnable gating signals that dynamically adjust the contribution of dynamic features, enabling adaptive modulation of dynamic representations by static variables. The proposed TLAT model is validated using 104-dimensional high-frequency operational data collected over three years from a 1.4 Mt/a hydrocracking unit of a petrochemical enterprise. Experimental results

收稿日期: 2026-04-24 修回日期: 2026-06-05

通信作者: 李佳鹤 (1983—), 男, 硕士, 高级工程师, lijiahe@supos.com

第一作者: 阮志坚 (1985—), 男, 硕士, 高级工程师, ruanzhijian@supos.com

基金项目: 国家重点研发计划项目 (2024YFB3311905)

引用本文: 阮志坚, 李佳鹤, 李佳洋. 基于时滞感知 Transformer 的加氢装置柴油收率预测方法[J]. 化工学报, XXXX, XX(XX): 1–12

Citation: RUAN Zhijian, LI Jiahe, LI Jiayang. A method for predicting diesel yield of hydrocracking unit based on time-lag-aware transformer[J].

CIESC Journal, XXXX, XX(XX): 1–12

demonstrate that TLAT achieves a mean absolute error (MAE) of 0.0182 and a coefficient of determination (R^2) of 0.937 on the test set, significantly outperforming benchmark models including XGBoost, LSTM, Informer and the standard TFT. These results indicate that the proposed method provides effective technical support for refined monitoring and control in smart factory applications.

Keywords: hydroprocessing; diesel yield; prediction; dynamic modeling; time-lag-aware; Transformer

引 言

在现代炼油企业中,加氢裂化装置^[1]作为提升重油轻质化水平与优化产品结构的重要二次加工单元,其运行效率直接关系到企业经济效益与生产稳定性。其中,柴油收率^[2]是衡量装置运行水平的核心指标之一,不仅决定柴油产品产量与轻重组分分布,还直接影响氢耗、能耗以及全厂物料平衡状态。在140万吨/年级别加氢裂化装置中,柴油收率每波动0.5%,均可能导致装置日均数十吨产品结构变化,进一步带来原料利用率下降、氢气消耗增加及能耗成本上升等问题,单年经济效益影响可达数百万元。同时,在炼化企业生产过程中,柴油收率预测偏差还可能导致工艺参数调节滞后,影响装置稳定运行与优化决策效果。因此,实现柴油收率的高精度预测,对于生产优化、实时调控与精益运行具有重要工程意义。

然而,在实际工业场景中,柴油收率的形成过程高度复杂,不仅受到进料性质(如原料组成、硫氮含量、密度、初馏点等)和催化剂体系^[3](含Ni、Co、Mo、W等活性组分)的共同影响,还与空速、反应温度、系统总压、氢油比、回流比等多变量操作参数呈现强非线性、动态耦合关系。尤为关键的是,这些输入变量对最终收率的影响普遍存在显著的时滞效应——即变量变化与收率响应之间存在时间延迟,且不同变量的时滞特性呈现异构性。例如,反应温度扰动通常能够在数小时内快速影响产品分布,而催化剂失活则往往需要数天甚至数周才逐渐体现。若未对此类物理可解释的时序依赖规律进行有效建模,将会显著影响其预测模型的精度与泛化性能。

近年来,数据驱动方法凭借其在建模复杂非线性关系方面的优势,已逐渐应用于加氢裂化等复杂工业过程的关键指标预测^[4-5]。然而,在具有强动态耦合和显著时滞特性的工业时序场景中,现有方法整体上仍存在对长时序依赖刻画不足、对工艺时滞等领域先验信息利用不充分的问题,从而在预测精

度、稳定性及物理一致性方面受到一定限制。

针对上述问题,本文提出一种基于时滞感知Transformer(Time-Lag Aware Transformer, TLAT)的加氢裂化装置柴油收率预测方法,通过将工艺时滞先验与Transformer结构进行深度融合,实现复杂工业时序过程的物理一致性建模。本文主要贡献如下:

(1)提出一种“时滞先验—注意力结构”联合建模机制,将变量级工艺时滞约束显式嵌入Transformer注意力计算过程,实现具有物理一致性的动态依赖建模;

(2)设计静态-动态门控融合模块,通过静态工艺上下文自适应调节动态特征响应强度,增强模型对催化剂状态变化与工况切换的适应能力;

(3)基于某140万吨/年加氢裂化装置近3年运行数据开展验证,结果表明TLAT在预测精度、稳定性与可解释性方面均优于XGBoost、LSTM、Informer及标准TFT等主流方法。

本文研究为复杂炼油过程的高精度预测与智能优化提供了一种融合工艺机理与深度学习的有效建模方法,对工业智能预测与精益管控具有一定工程应用价值。

1 相关工作

1.1 深度时序建模在石化领域的应用

早期研究主要采用机器学习为主^[7],谢洋等人^[8]结合多层感知机(MLP)和随机森林(RF)对加氢裂化装置现场数据进行有效训练和预测。随着工业数据规模的持续增长,深度学习方法逐渐应用于石化装置运行状态建模与关键指标预测^[9]。曹跃等人^[10]针对加氢裂化装置开展工业数据驱动研究,实现了多工况切换过程的状态检测,为复杂工况下的时序建模提供了工程实践参考。Vieira等人^[11]引入XGBoost回归模型对石化领域氨生产装置的余热锅炉液位进行预测。随后长短时记忆(LSTM)^[12]、门控循环单元(GRU)^[13]被引入,用于刻画工业过程中的

时间依赖特性。Panjapornpon 等人^[14]针对乙烯生产中测量信号干扰,提出基于LSTM的补偿网络,减少了异常测量对能效预测的影响。王琳等人^[15]提出一种基于GRU的能耗预测模型并发现了加氢裂化过程中的关键因素。然而,该类基于循环神经网络的方法在面对长时间跨度依赖以及具有异构时滞特性的多变量工业过程时,通常仍依赖统一历史窗口进行建模,难以体现不同工艺变量对预测目标的物理作用时序差异。此外,卷积神经网络(CNN)^[16]及时间卷积网络(TCN)^[17]通过局部感受野建模动态模式,也在部分石化时序预测任务中得到应用。

Transformer^[18]架构凭借并行计算和长程依赖建模能力,在时序预测中得到广泛应用。杨晔等人^[19]将Transformer用于化工过程质量预测,实现了静态信息与动态时序特征的联合学习,提升了模型对复杂过程的拟合能力。张秦伟^[20]等人研究了一种时滞修正的Informer对制氢转化炉出口温度预测进行高精度预测,为转化炉运行优化与异常预警提供支撑。Maldonado-Cruz 和 Pyrcz^[21]基于TFT提出了一种融合静态与动态特征的多步预测模型,但其时间依赖仍通过统一历史窗口的自由注意力建模,未能明确区分不同变量的时间作用范围。

1.2 时滞建模与注意力机制研究进展

在时滞建模中,互信息^[22]可被广泛用于量化变量之间的时序依赖关系,是确定最优时滞窗口的有效手段。朱春梦等人^[23]借助多尺度符号转移熵(MSTE)分析变量时延与因果联系,为化工过程的时滞特性解析提供了可解释方法。Ji 等人^[24]提出一种时滞互信息(TDMI)方法,从过程数据中动态识别工业变量之间的因果关系,进而构建了故障传播路径。Li 等人^[25]提出了一种后向时滞变量选择方法(BTS),通过逐步筛选解释变量及其滞后项,自适应确定各变量的滞后阶数,避免人为设定统一固定滞后窗口。尽管上述方法能刻画时滞,但在高维噪声环境下仍存在不足,王雅琳等人^[26]针对加氢裂化流程提出的多重时滞辨识方法,通过趋势相似度分析实现了复杂工况下多变量时滞的精准捕捉,为后续注意力机制建模提供了可靠的时滞先验。为此研究者引入注意力机制以实现对时序影响强度的自适应建模^[27]。在数据预测中,注意力机制结合编解码器实现序列到序列建模,能够有效融合历史状态信息,从而提升预测精度和模型扩展性。He 等人^[28]提出一种基于差分进化优化的双层注意力LSTM模

型用于催化裂化(FCC)收率预测。郭子越等人^[29]结合门控循环单元(GRU)的时序建模能力与通道-空间混合注意力机制,融合发射光谱层析技术,实现了连续时间序列数据的预测与三维重建。Song 等人^[30]将结合时间和空间注意力与门控时间卷积网络,提高了能源负荷预测的准确性。此外,因果注意力机制^[31]通过约束非因果依赖,也被广泛应用于时序预测中。

1.3 现有方法的局限性分析

首先,时滞建模研究多以特征选择或统计分析形式存在,未能形成可端到端训练的深度学习结构;其次,静态工艺条件与动态操作变量之间的调制关系缺乏显式建模机制,限制了模型在工况变化下的泛化能力。现有方法对先验信息的利用仍以策略层或特征层融合为主,孙毅等人^[32]利用实时检测数据结合先验知识提出了一种应对时滞的张力控制策略。王笑辰等人^[33]提出融合先验知识与多源异构数据的风险预测方法,通过知识驱动标签构建、混合机器学习模型与可解释学习实现高精度预测及验证。但从整体来看,尽管已有研究通过数据分解与深度学习提升动态预测效果^[34],但仍缺乏将物理可解释时滞显式嵌入模型结构的系统性机制。

2 工艺时滞特性分析

2.1 加氢装置柴油收率预测的形式化定义

柴油收率预测可形式化为一个多变量时序回归问题。设系统在时刻 t 的观测输入为向量 $\mathbf{x}_t = [\mathbf{s}, \mathbf{z}_t] \in \mathbb{R}^d$,其中 $d = 104$ 为输入维度。根据物理属性,将输入变量划分为两类。静态变量 $\mathbf{s} \in \mathbb{R}^{d_s}$ ($d_s = 28$):在装置单次稳定运行周期内保持不变或变化极缓慢的参数,如催化剂型号(Ni-Mo/Co-Mo)、活性组分含量(Ni、Co、Mo、W等)、装填体积、床层高度等;动态变量 $\mathbf{z}_t \in \mathbb{R}^{d_z}$ ($d_z = 76$):随时间高频变化的操作与进料参数,如反应温度、系统总压、混合进料流量、硫/氮含量、密度、初馏点等,以小时级频率采集。

给定长度 T 的历史观测序列 $X_{1:T} = \{(\mathbf{s}, \mathbf{z}_1), (\mathbf{s}, \mathbf{z}_2), \dots, (\mathbf{s}, \mathbf{z}_T)\}$,目标是预测当前时刻 $t = T$ 的柴油收率 $y_T \in [0, 1]$ (以质量百分比表示)。因此,该任务可严格表述为学习一个非线性映射函数:

$$y_t = f(\mathbf{s}, \mathbf{z}_{t-\tau_{\max}}, \dots, \mathbf{z}_{t-1}, \mathbf{z}_t) + \varepsilon_t \quad (1)$$

式中, ε_t 为建模误差; $\tau_{\max}$ 为最大考虑时滞(本文取48小时)。由于各变量作用时滞不同且存在非线

性耦合,该问题本质上是一个异构输入、长依赖、非线性时序回归任务,对模型的时序建模能力与先验融合机制提出较高要求。

2.2 工艺变量时滞效应的物理机理

在加氢裂化过程中,从操作参数调整到柴油收率响应存在不可避免的时间延迟,其根源在于物理传递与化学反应动力学特性。主要时延来源包括:物料停留时间,原料从进料口进入反应器至产物流出需经历数小时的停留时间(典型值4–6 h),导致进料性质(如氮含量)变化对收率的影响滞后;热质延迟,反应温度设定值调整后,需通过加热炉、换热网络逐步传递至催化剂床层,存在热惯性;催化剂活性衰减,催化剂中毒或结焦是一个缓慢累积过程,其对收率的负面影响呈长期趋势,时滞可达数天甚至数周。

上述机制导致不同变量呈现异构时滞特性。例如:反应温度作为强调控变量,其扰动通常在1–3小时内反映于产品分布;进料氮含量因毒化作用需经多批次反应累积,有效时滞窗口常为6–24小时;催化剂装填量属静态变量,虽无“动态时滞”,但其对整个运行周期内动态响应的调制作用需在建模中显式体现。

若预测模型忽略此类物理规律,将导致两种典型错误:(1)将近期无关波动(如仪表噪声)误判为强相关因子;(2)遗漏远期关键影响(如一周前的进料劣化),从而在工况切换或异常工况下预测失效。因此,在建立柴油收率预测模型之前,需要首先识别各工艺变量对目标变量的有效作用时滞。

针对加氢裂化过程中的变量时滞辨识问题,现有方法中TDMI^[24]能够刻画变量间时序因果关系,但主要用于故障传播路径分析;MSTE^[23]适用于非平稳过程因果发现,其结果更多体现变量间因果强弱;BTS^[25]通过逐步筛选确定变量滞后阶数,适用于因果结构识别,但难以直接量化不同滞后时刻对目标变量的贡献。相比之下,条件互信息^[22](Conditional Mutual Information, CMI)能够在消除目标变量历史状态影响的基础上量化输入变量对柴油收率的独立时滞贡献,有效抑制冗余关联和间接影响,更符合加氢裂化过程多变量、多时间尺度耦合的动态特征,因此本文采用CMI实现时滞窗口量化。

2.3 基于条件互信息的时滞窗口量化

为科学、客观地量化各变量的有效时滞区间,本文采用条件互信息(CMI)作为度量准则。CMI能

有效剥离历史信息 $H_{t-1} = \{\mathbf{x}_1, \dots, \mathbf{x}_{t-1}\}$ 的干扰,评估变量 $x_{t-\tau}^i$ 对当前收率 y_t 的独立信息贡献,其数学定义为:

$$I(y_t; x_{t-\tau}^i | H_{t-1}) = \iint p(y_t, x_{t-\tau}^i | H_{t-1}) \log \frac{p(y_t, x_{t-\tau}^i | H_{t-1})}{p(y_t | H_{t-1}) p(x_{t-\tau}^i | H_{t-1})} dy_t dx_{t-\tau}^i dH_{t-1} \quad (2)$$

在实际计算中,采用Kraskov等人^[22]提出的K近邻算法(KNN)进行非参数估计。对每个动态变量 x^i ,在预设最大时滞 $\tau_{\max} = 48$ 小时范围内,以1小时为步长滑动计算 $\text{CMI}(\tau) = I(y_t; x_{t-\tau}^i | H_{t-1})$ 。有效时滞窗口 $[\tau_{\min}^i, \tau_{\max}^i]$ 定义为CMI值显著高于背景噪声的连续区间:

$$[\tau_{\min}^i, \tau_{\max}^i] = \{\tau \in [1, \tau_{\max}] | \text{CMI}(\tau) > \mu_{\text{noise}} + 2\sigma_{\text{noise}}\} \quad (3)$$

式中, $\mu_{\text{noise}}, \sigma_{\text{noise}}$ 均为 $\tau > 30$ 时CMI的均值与标准差,视为系统噪声基线。

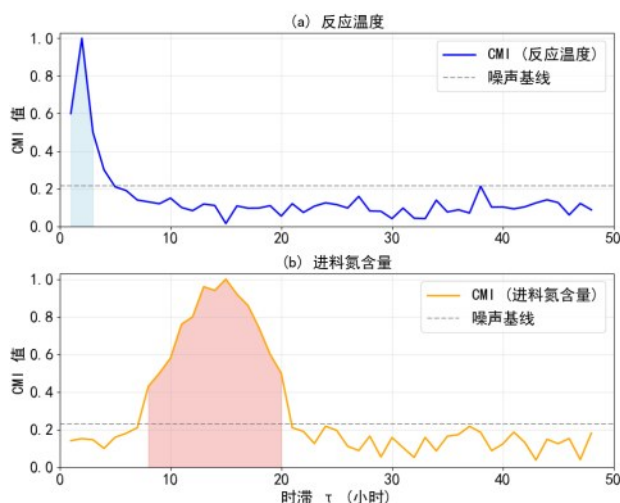


图1 两个典型变量的CMI归一化分布

Fig. 1 Normalized cmi distribution of two typical variables

该量化结果为后续设计结构化注意力掩码提供了物理可解释的先验依据,确保模型仅在合理时滞范围内建立依赖关系,避免过拟合虚假关联。

3 时滞感知Transformer(TLAT)模型架构

3.1 整体框架概述

本文提出的TLAT结合TFT的基础架构,针对加氢装置柴油收率预测中变量作用时滞异构、静态–动态耦合强烈等特点进行针对性改进。如图2所示,TLAT整体流程包含四个核心阶段:变量嵌入与位置编码→静态–动态门控融合→时滞感知注意

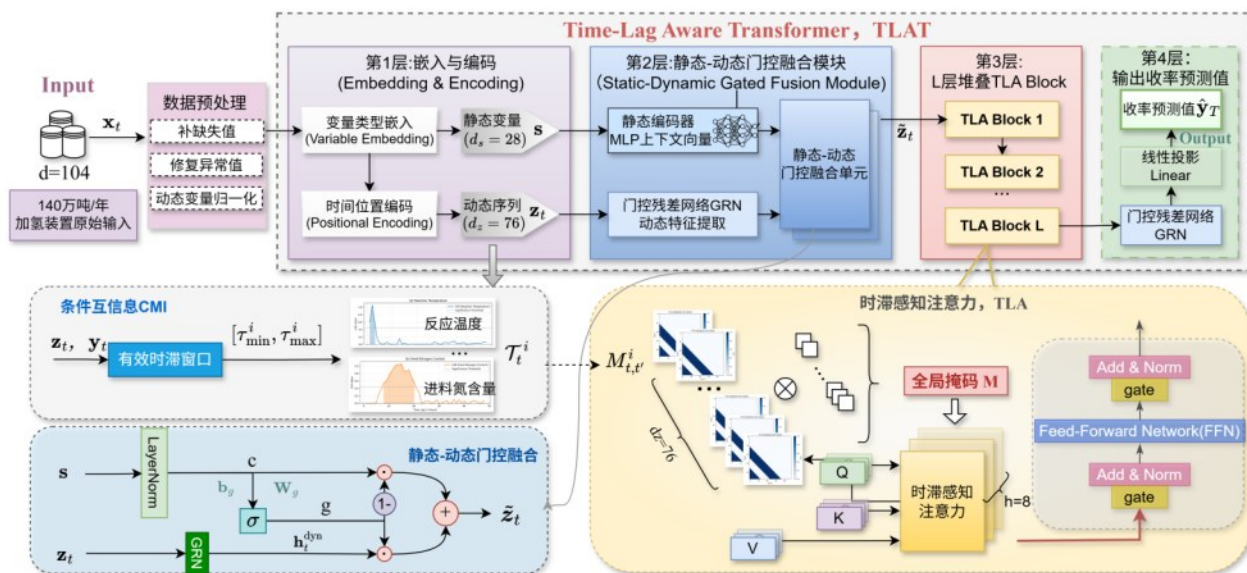


图2 网络架构图

Fig. 2 Network architecture diagram

力→输出收率预测值。

首先,原始输入经变量类型嵌入(Variable Embedding)和时间位置编码(Positional Encoding)后,分离为静态变量 s 与动态序列 z_t ;随后,通过静态-动态门控融合模块生成调制后的动态表示;接着,序列进入多层堆叠的时滞感知注意力块,每层均施加由工艺先验引导的结构化掩码;最终,通过门控残差网络(GRN)与线性投影(Linear)输出收率预测值 $\hat{y}_T$ 。相较于标准TFT,TLAT的核心改进在于:(1)将通用自注意力替换为物理约束的时滞感知注意力;(2)增强静态变量对动态特征的显式调制机制,而非简单拼接。这两项改进使模型在保持TFT强大时序建模能力的同时,显著提升物理可解释性与预测鲁棒性。

3.2 静态-动态门控融合模块

在加氢裂化装置中,催化剂等静态变量虽不随时间变化,却深刻调制着动态特征(如温度、压力)对柴油收率的影响强度。例如,高活性催化剂可放大温度扰动的收率响应,而失活催化剂则削弱其敏感性。标准时序模型(如TFT)通常将静态向量广播后与动态特征拼接,该策略无法显式建模静态变量对动态响应的非线性调制作用,导致模型在催化剂更换或催化剂活性衰减场景下泛化能力下降。

为此,本文设计静态-动态门控融合模块(Static-Dynamic Gated Fusion Module),其核心思想是:利用静态上下文生成可学习门控信号,动态调

节动态特征的权重,实现静态变量对动态特征的自适应调制。

静态编码器设计:静态向量 s 首先经MLP全连接网络映射为上下文向量:

$$c = \text{LayerNorm}(W_s s + b_s) \in \mathbb{R}^{d_c} \quad (4)$$

式中, $W_s \in \mathbb{R}^{d_m \times d_s}$, $b_s \in \mathbb{R}^{d_m}$ 为可学习参数; d_m 为模型隐藏维度。LayerNorm提升训练稳定性。

动态特征提取:动态序列 z_t 经门控残差网络(Gated Residual Network, GRN)提取时序特征:

$$h_t^{\text{dyn}} = \text{GRN}(z_t) = z_t + W_3(\text{ELU}(W_1 z_t + b_1) \odot \sigma(W_2 z_t + b_2)) \quad (5)$$

式中, $W_1, W_2, W_3 \in \mathbb{R}^{d_m \times d_z}$; $\sigma(\cdot)$ 为sigmoid函数; $\odot$ 为Hadamard乘积。GRN通过门控机制抑制冗余信息,保留关键时序模式。

静态-动态门控机制设计:利用静态上下文 c 生成动态调制门:

$$g = \sigma(W_g c + b_g) \in \mathbb{R}^{d_m} \quad (6)$$

式中, $W_g \in \mathbb{R}^{d_m \times d_c}$; $b_g \in \mathbb{R}^{d_m}$;门控值 $g_j \in [0,1]$ 表示第 j 个特征通道的保留比例。

融合逻辑:最终融合特征 $\tilde{z}_t$ 为动态特征与静态上下文的加权插值:

$$\tilde{z}_t = g \odot h_t^{\text{dyn}} + (1 - g) \odot c \quad (7)$$

通过融合逻辑,模型实现在动态特征与静态变量之间自适应平衡。若催化剂活性高,则 $g \rightarrow 1$,强调动态变化;反之 $g \rightarrow 0$,偏向静态背景。

梯度传播分析:对损失函数 L ,静态变量梯度经门控反向传播:

$$\frac{\partial L}{\partial \mathbf{c}} = \frac{\partial L}{\partial \mathbf{z}_t} \odot (\mathbf{h}_t^{\text{dyn}} - \mathbf{c}) \odot \mathbf{g}(1 - \mathbf{g}) \mathbf{W}_g^\top \quad (8)$$

$$\frac{\partial L}{\partial \mathbf{s}} = \left(\frac{\partial L}{\partial \mathbf{c}} \odot \gamma \right) \mathbf{W}_s^\top \quad (9)$$

式中, γ 为 LayerNorm 的缩放因子。该路径确保静态变量通过调制作用间接影响预测结果。

与传统拼接方法对比: 设传统拼接模型输出为 $\hat{\mathbf{y}}_{\text{concat}} = f([\mathbf{s}, \mathbf{z}_t])$, 其对静态变量的敏感度为:

$$\left\| \frac{\partial \hat{\mathbf{y}}_{\text{concat}}}{\partial \mathbf{s}} \right\| \approx \text{const} \quad (10)$$

即敏感度恒定。而本模块的敏感度为:

$$\left\| \frac{\partial \hat{\mathbf{y}}_{\text{gate}}}{\partial \mathbf{s}} \right\| \propto \|\mathbf{h}_t^{\text{dyn}} - \mathbf{c}\| \cdot \|\mathbf{g}(1 - \mathbf{g})\| \quad (11)$$

即随动态特征状态自适应调整, 更符合加氢裂化装置运行机理。

综上, 本模块通过显式门控机制, 精准刻画了催化剂等静态变量对动态过程的调制作用, 显著提升模型在装置状态变化下的适应能力与预测精度。

3.3 时滞感知注意力机制

标准 Transformer 中的自注意力机制允许任意历史时刻 $t' \in [1, T]$ 对当前 t 时刻产生影响, 其注意力权重计算为:

$$A_{t,t'} = \text{softmax} \left(\frac{\mathbf{q}_t \mathbf{k}_{t'}^\top}{\sqrt{d_k}} \right) \quad (12)$$

式中, d_k 为查询、键、值向量的维度; $\mathbf{q}_t, \mathbf{k}_{t'} \in \mathbb{R}^{d_k}$ 分别为查询与键向量。然而, 在加氢裂化等物理系统中, 变量作用受因果律与时滞约束, 例如进料氮含量的变化需经反应器停留时间 (4 ~ 6 h) 以后才能影响产品分布。若模型无约束地关注近期无关波动 (如传感器噪声), 将学习到非物理解, 导致泛化失败。

为此, 本文提出时滞感知注意力机制 (Time-Lag Aware Attention, TLA), 其核心是构建一个结构化硬掩码矩阵 $\mathbf{M} \in \{0, -\infty\}^{T \times T}$, 强制模型仅在物理合理的时滞窗口内建立依赖。对第 i 个动态变量 x^i , 设其有效影响时滞区间为 $[\tau_{\min}^i, \tau_{\max}^i]$ (由第 2.3 节 CMI 分析确定), 则其允许影响当前时刻 t 的历史时刻集合为:

$$\mathcal{T}_t^i = \{t' | \max(1, t - \tau_{\max}^i) \leq t' \leq t - \tau_{\min}^i\} \quad (13)$$

相应地, 变量级掩码定义为:

$$\mathbf{M}_{t,t'}^i = \begin{cases} 0, & t' \in \mathcal{T}_t^i \\ -\infty, & \text{otherwise} \end{cases} \quad (14)$$

由于 TLA 的输入 $\tilde{\mathbf{z}}_t$ 融合了 $d_z = 76$ 个具有异构时滞特性的动态变量, 不同变量具有异构时滞, 最

终掩码通过通道维度拼接融合为统一的全局掩码 $\mathbf{M}$ 。

具体而言, 静态-动态门控融合后的动态特征序列 $\tilde{\mathbf{z}}_t$, 被送入一个由 L 层 (本文实验选取 $L=4$) 深度堆叠而成的时滞感知注意力块 (TLA Block) 序列。设每层 TLA Block 的输入序列为 $\tilde{\mathbf{Z}} = [\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_T]^\top \in \mathbb{R}^{T \times d_m}$, 经线性投影得到查询、键、值矩阵:

$$\mathbf{Q} = \tilde{\mathbf{Z}} \cdot \mathbf{W}_Q, \mathbf{K} = \tilde{\mathbf{Z}} \cdot \mathbf{W}_K, \mathbf{V} = \tilde{\mathbf{Z}} \cdot \mathbf{W}_V \quad (15)$$

式中, $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d_m \times d_k}$ 为可学习参数矩阵, d_m 为模型隐藏维度, d_k 为查询、键、值向量的维度 (保证维度一致性)。为实现多变量异构掩码, 将 $\mathbf{Q}, \mathbf{K}$ 按动态变量维度拆分: $\mathbf{Q}^i \in \mathbb{R}^{T \times d_k^i}$ 表示与变量 $\tilde{z}^i$ 对应的查询子矩阵 (满足 $\sum_{i=1}^{d_m} d_k^i = d_k$), 则全局掩码通过块对角拼接构造:

$$\mathbf{M} = \text{blockdiag}(\mathbf{M}^1 \otimes \mathbf{I}_{d_1}, \dots, \mathbf{M}^{d_m} \otimes \mathbf{I}_{d_{d_m}}) \quad (16)$$

式中, $\otimes$ 为 Kronecker 积, $\mathbf{I}_{d_i}$ 为单位矩阵, $\mathbf{M}$ 为分块对角稀疏矩阵结构, 可确保掩码在每个变量对应的特征子空间内独立生效。

最终, TLA 的单头注意力输出为:

$$\text{head} = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} + \mathbf{M} \right) \mathbf{V} \quad (17)$$

由于无效位置的掩码值为 $-\infty$, 经 softmax 函数后, 这些位置的注意力权重被强制置为 0。这使得模型在每一层的特征交互过程中, 都严格遵守由工艺机理决定的因果时序约束, 从而实现了物理可解释、计算高效且抗干扰能力强的特征学习。

对于多头结构 (头数 h), 每个头独立计算后拼接并投影:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}_O \quad (18)$$

式中, $\mathbf{W}_O \in \mathbb{R}^{hd_k \times d_m}$ 。

多头注意力输出 $\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 随后经过 Add & Norm 和前馈网络 (FFN), 形成第 l 层的最终输出 $\mathbf{H}(l)$, 并作为下一层的输入。通过 L 层的堆叠, 模型能够逐层抽象并提炼出既符合数据规律又满足物理约束的高阶时空特征表示。

梯度传播特性: 由于采用硬掩码 ($-\infty$), 被屏蔽位置的梯度恒为零。设损失函数为 L , $T_t = U_t T_t^i$ 为联合有效时滞集, 则对于任意 $t' \notin T_t$, 有:

$$\frac{\partial L}{\partial A_{t,t'}} = 0, \quad \forall t' \notin T_t \quad (19)$$

$$\frac{\partial L}{\partial \mathbf{Q}_{t,:}} = \sum_{t' \in T_t} \frac{\partial L}{\partial A_{t,t'}} \cdot \frac{\partial A_{t,t'}}{\partial \mathbf{Q}_{t,:}} \quad (20)$$

这确保训练过程严格遵循物理先验,避免参数更新偏离真实机理。

计算复杂度分析:设平均有效时滞长度为:

$$\bar{\tau} = \frac{1}{d_z} \sum_{i=1}^{d_z} (\tau_{\max}^i - \tau_{\min}^i + 1) \quad (21)$$

则单头注意力计算复杂度为:

$$\mathcal{O}(T\bar{\tau}d_k) + \mathcal{O}(T\bar{\tau}d_v) \quad (22)$$

该方法相较于标准注意力的 $\mathcal{O}(T^2d_k)$ 复杂度,在 $T = 48, \bar{\tau} \approx 12$ 时,理论加速比可达4倍。

TLA 与标准注意力对比的优势在于:1、物理一致性,TLA 严格限制注意力范围,消除非因果依赖;2、泛化能力,TLAT 通过关注稳定的时滞传递路径,在工况切换及异常扰动情况下表现出更强的鲁棒性;3、可解释性,注意力热力图可直接映射至 CMI 量化结果,便于工程师验证模型决策逻辑。

综上,TLA 通过先验引导的结构化稀疏性,在保持 Transformer 强大表达能力的同时,实现了数据驱动与物理机理的深度融合,为工业时序预测提供了高可信度建模范式。

3.4 模型训练与正则化

为全面优化模型性能,TLAT 采用复合损失函数,协同优化预测精度与整体拟合趋势。该损失由两部分加权构成:第一项为平均绝对误差(MAE),直接约束预测值 $\hat{y}_i$ 与真实收率 y_i 的绝对偏差;第二项为1减去拟合优度 $(1 - R^2)$,通过最大化 R^2 确保模型能有效捕捉数据的整体方差和变化趋势。

TLAT 复合损失函数具体定义如下:

$$L = \lambda_{\text{MAE}} \cdot \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| + \lambda_{R^2} \cdot \left(1 - \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \right) \quad (23)$$

式中, $\lambda_{\text{MAE}} = 0.7; \lambda_{R^2} = 0.3; \bar{y}$ 为真实收率均值。以优先保证点对点的准确性,同时兼顾全局拟合性能。此设计有效平衡了局部误差控制与全局模式学习,提升了模型在复杂工业场景下的泛化能力与稳定性。

优化器选用 AdamW, 初始学习率 $\eta = 1 \times 10^{-3}$, 配合余弦退火调度。训练采用早停策略:若验证集 MAE 在连续 15 轮未下降,则终止训练。模型训练环境设置在配备 NVIDIA A100 GPU 的服务器上,使用

PyTorch 框架进行端到端训练,批量大小 (Batch Size) 设为 64,总训练轮数不超过 100 轮。

关于注意力掩码的实现方式,本文采用硬约束 (即掩码值为 $-\infty$),而非可学习软掩码。理由在于:工艺时滞窗口由条件互信息离线量化,具有明确物理意义;硬约束能严格保证模型不违反先验知识,避免因数据噪声导致掩码学习偏离真实机理。实验表明 (见第 4 节),该策略可显著提升模型性能。

综上,TLAT 通过时滞感知注意力与静态-动态门控融合两大创新,实现了深度学习模型与加氢裂化装置运行机理的深度融合,为高精度、可解释的柴油收率预测提供有效解决方案。

4 实验设计与结果分析

4.1 数据集描述

本研究基于某大型石化企业 140 万吨/年加氢裂化装置的真实运行数据。该装置采用固定床反应器,主产柴油,副产石脑油与液化气。数据覆盖 2023 年 1 月至 2025 年 12 月共 36 个月,采样频率为 1 小时/点,总计约 26,280 条样本。输入变量共 104 维,其中静态变量 28 维 (包括催化剂型号 Ni-Mo/Co-Mo、活性组分含量、装填体积、床层高度等),动态变量 76 维 (包括反应温度、系统总压、混合进料流量、硫/氮含量、密度、初馏点 IBP、氢油比、循环氢纯度、床层压降等)。目标变量为柴油收率 $y_i \in [0,1]$,由在线分析仪结合物料平衡校正获得,精度达 $\pm 0.3\%$ 。

对原始样本进行数据预处理。首先,对缺失值采用线性插值填补;其次,对超出 3σ 范围的异常值使用滑动窗口中位数滤波进行平滑,以抑制传感器噪声同时保留工艺趋势;最后,对动态变量 $\mathbf{z}_t$ 进行 Min-Max 归一化至 $[0,1]$ 区间,静态变量则保留原始单位以维持物理可解释性。

数据集严格按照时间顺序划分,训练集为 2023 年 1 月至 2024 年 12 月 (约 17,520 条),用于模型训练;验证集为 2025 年 1 月至 6 月 (约 4,380 条),用于超参调优与早停;测试集为 2025 年 7 月至 12 月 (约 4,380 条),用于最终评估。数据集完整覆盖催化剂不同活性周期、季节变化及原料波动等典型工况,确保模型泛化能力评估具有工程代表性和长期可靠性。

4.2 对比模型与评估指标

为全面验证本文所提 TLAT 模型的有效性,本

文选取了六类具有代表性的基线模型进行对比分析。具体包括:(1)典型机器学习方法 XGBoost^[35],配置为树数量 500、最大深度 8、学习率 0.1;(2)多层感知器 MLP^[36],采用 3 层全连接结构,每层含 256 个神经元,使用 ReLU 激活函数;(3)长短期记忆网络 LSTM^[12],包含 2 层循环单元,隐藏维度为 128,并引入 0.2 的 dropout 以缓解过拟合;(4)基于稀疏注意力的长序列模型 Informer^[37],其编码器与解码器各设 2 层,多头注意力头数为 8;(5)工业时序建模 SOTA 方法 TFT^[6],设置网络层数为 3,门控残差网络(GRN)宽度为 160;(6)本文提出的 TLAT 模型,同样采用 3 层结构、8 个注意力头,模型维度 $d_m = 160$ 。所有深度学习模型统一使用 48 小时的历史窗口长度(即 $T = 48$),优化器选用 AdamW(初始学习率 $\eta = 1 \times 10^{-3}$),并设置早停机制(耐心轮数=15)。所有超参数均通过在验证集上进行网格搜索确定,确保公平比较。

为进一步评估时滞建模对预测性能的影响,本文基于第 2.3 节 CMI 分析结果,对各基线模型引入变量级时滞信息,仅保留各变量有效时滞窗口内的历史输入,在保持模型结构与超参数不变的条件下,分别构建 Lag-XGBoost、Lag-MLP、Lag-LSTM 和 Lag-Informer 模型。通过与原始基线模型进行对比,可进一步分析时滞先验与模型结构设计对预测性能提升的贡献。

评估指标方面选用四类广泛认可的回归性能度量指标,从多个角度评估模型的预测性能。

MAE (Mean Absolute Error) 衡量预测值与真实值的平均绝对偏差。

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (24)$$

RMSE (Root Mean Squared Error) 对大误差更敏感,反映误差的离散程度。

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (25)$$

MAPE (Mean Absolute Percentage Error) 以百分比形式体现相对误差,便于跨工况比较。

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (26)$$

R^2 (决定系数) 表示模型对目标变量方差的解释比例,越接近 1 表示拟合效果越好。

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad \text{其中 } \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (27)$$

式中, y_i 为实际收率; $\hat{y}_i$ 为收率预测值; N 为测试集数量。

为进一步验证性能提升的统计显著性,采用 Wilcoxon 符号秩检验^[38] (双尾,显著性水平 $\alpha = 0.05$),对 TLAT 与次优模型在测试集上的逐点绝对误差分布进行非参数假设检验。

4.3 性能对比分析

为系统评估 TLAT 模型的有效性,本文在加氢裂化装置数据集上开展了全面的对比实验。主性能对比结果如表 1 所示,TLAT 在所有四项评估指标上均显著优于全部对比模型,展现出卓越的预测精度与泛化能力。

表 1 实验主性能对比

Table 1 Core performance comparison of experiments

模型	MAE	RMSE	MAPE	R^2
XGBoost	0.0287	0.0361	2.15	0.842
Lag-XGBoost	0.027	0.0349	2.03	0.856
MLP	0.0263	0.0340	2.01	0.861
Lag-MLP	0.0249	0.0325	1.89	0.874
LSTM	0.0254	0.0328	1.92	0.876
Lag-LSTM	0.0236	0.0315	1.76	0.891
Informer	0.0231	0.0302	1.78	0.894
Lag-Informer	0.0219	0.0294	1.68	0.906
TFT	0.0215	0.0285	1.65	0.912
TLAT	0.0182	0.0241	1.38	0.937

从表 1 可以看出,引入时滞建模后,各基线模型预测性能均得到不同程度提升。其中, Lag-XGBoost、Lag-MLP、Lag-LSTM 和 Lag-Informer 的 MAE 分别较对应原模型下降 5.9%、5.3%、7.1% 和 5.2%,说明工艺时滞先验对于柴油收率预测具有重要价值。本文提出的 TLAT 取得最佳预测性能,其 MAE 达到 0.0182, RMSE 为 0.0241, MAPE 为 1.38%, R^2 达到 0.937,显著优于所有对比模型。这表明本文性能提升并非仅来源于时滞窗口筛选,而是得益于时滞先验与 Transformer 结构的深度融合机制。通过将变量级时滞约束显式嵌入注意力计算过程, TLAT 能够在特征交互过程中持续利用工艺先验,从而实现更准确、更稳定的柴油收率预测。

为进一步直观评价模型对柴油收率变化趋势、局部波动及工况切换过程的跟踪能力,本文补充测试集真实值与预测值时间序列对比,如图 7 所示。结果表明, TLAT 预测曲线与真实柴油收率变化趋势高度一致,能够较好跟踪测试集期间的缓慢波

动、局部峰谷及阶段性变化。相比次优模型 Lag-Informer 和 TFT, TLAT 在波动区间的预测滞后更小, 局部偏差更低。

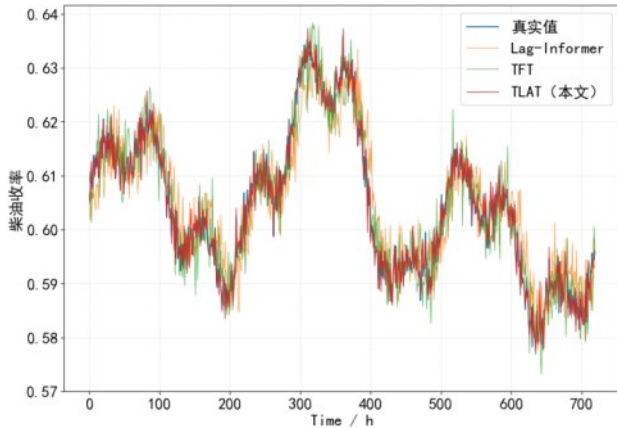


图3 测试集真实值与预测值全局时间序列对比图

Fig. 3 Time series comparison of actual and predicted diesel yield

进一步选取典型工况切换区间进行局部放大分析,如图8所示。可以看出,Lag-Informer在收率快速变化阶段存在一定滞后,TFT虽能够较好拟合柴油收率的整体变化趋势,但在局部峰谷区域存在一定程度的预测波动放大现象,而TLAT能够更准确捕捉收率变化的上升、下降及峰值位置,说明时滞感知注意力机制有助于提升模型对工业动态过程的跟踪能力。

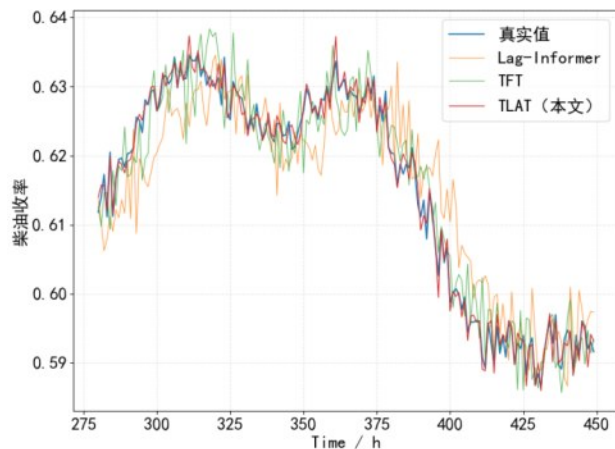


图4 典型工况切换区间局部放大对比图

Fig. 4 Local comparison under typical operating-condition transition

为进一步验证性能提升的统计显著性,采用 Wilcoxon 符号秩检验对 TLAT 与次优模型 TFT 在测试集上的逐点绝对误差进行双尾检验,结果如图4所示,所得P值小于0.001,表明差异高度显著,非随

机波动所致。

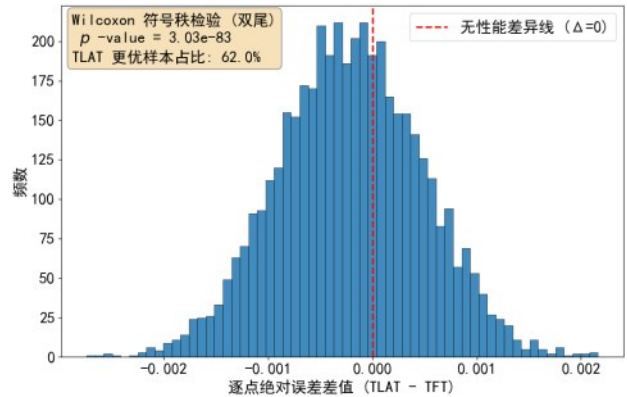


图5 TLAT与TFT模型逐点绝对误差双尾检验

Fig. 5 Two-tailed test of point-wise absolute errors between TLAT and TFT models

为验证模型中两个核心模块——时滞感知注意力掩码与静态-动态门控融合机制的必要性,设计了三组消融实验,结果如表2所示。变体A1移除时滞掩码,MAE升至0.0208(+14.3%);A2将门控融合替换为传统拼接,MAE升至0.0197(+8.2%);A3同时移除这两个模块,性能最差(MAE=0.0221, $R^2=0.901$)。结果充分证明,本文提出的两大创新模块均对性能提升具有不可替代的作用,且存在协同增益效应。

表2 消融实验结果

Table 2 Ablation experiment results

模型变体	MAE	R^2
TLAT(完整)	0.0182	0.937
A1:无时滞掩码	0.0208	0.918
A2:无门控	0.0197	0.925
A3:两者皆无	0.0221	0.901

此外,为验证 TLAT 模型参数设置的合理性,本文进一步开展参数敏感性分析,重点考察时滞窗口尺度、历史窗口长度 T 、注意力头数 h 、网络层数 L 以及模型维度 d_m 对模型性能的影响。

对于时滞窗口尺度,将 CMI 量化窗口分别缩放为 0.5×、1.0× 和 2.0×,发现原始尺度(1.0×)下 MAE 最低(0.0182);窗口过窄(0.5×)遗漏长程依赖,过宽(2.0×)则引入无关噪声,验证了所提时滞估计策略的合理性。

对于历史窗口长度 T ,分别测试 $T=\{24, 36, 48, 72\}$ h。实验结果表明,当 $T=48$ h 时模型性能最佳。窗口过短会遗漏长程动态依赖,而窗口过长则易引入冗余噪声。

对于注意力头数 h ,分别测试 $h=\{2, 4, 8, 16\}$ 。结果表明, $h=8$ 时模型达到最佳性能。头数过少会限制长程依赖建模能力,而头数过多则会导致注意力分散并增加训练复杂度。

对于TLA Block层数 L ,分别设置 $L=\{1, 2, 3, 4, 5\}$ 。随着层数增加,模型性能逐渐提升,但当 $L>4$ 后性能提升趋于饱和。

进一步分析模型维度 d_m 对性能的影响,分别测试 $d_m=\{64, 128, 160, 256\}$ 。实验结果表明,当 d_m 较小时,模型难以充分表征复杂工业变量之间的非线性耦合关系;而 d_m 过大则会增加参数规模与过拟合风险。综合比较发现,当 $d_m=160$ 时模型取得最佳综合性能与训练稳定性,因此本文最终采用 $d_m=160$ 作为模型隐藏维度配置。

4.4 可解释性与工程价值分析

时序预测模型在加氢裂化装置柴油收率预测等工业场景中的落地不仅依赖预测精度,更需具备可解释性以建立工艺操作人员的信任。本文通过多维度分析验证了TLAT的物理合理性与决策透明度。图5展示了典型工况下的时滞感知注意力热力图,横轴为历史时刻($t-48$ 至 t),纵轴为关键工艺变量。结果显示,模型对反应温度的注意力集中在最近 $t-1$ ~ $t-3$ 小时,符合反应器热惯性小、响应迅速的特性;而对进料氮含量的关注则显著分布在 $t-8$ ~ $t-20$ 小时区间,与基于条件互信息(CMI)的时滞分析结果高度一致,表明模型准确捕捉了氮化物在催化剂表面累积毒化的延迟效应。这一发现有力验证了所提时滞感知机制的物理合理性。

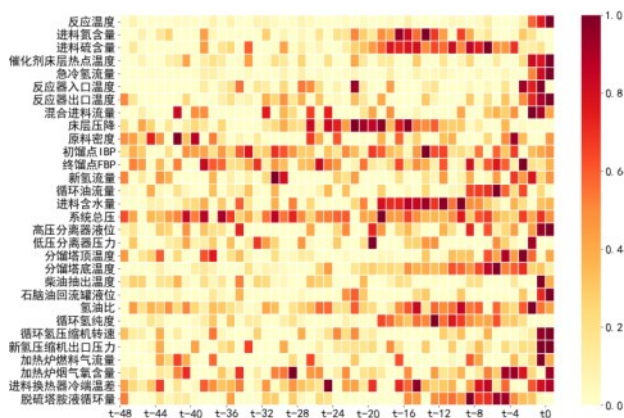


图6 时滞感知注意力热力图

Fig. 6 Time-lag aware attention heatmap

为进一步量化各变量对预测的贡献,采用Gradient×Input方法^[39]计算特征重要性。如图6所

示,排名前五的变量依次为:反应温度(23.1%)、进料氮含量(18.7%)、氢油比(15.2%)、催化剂Ni含量(12.4%)和系统压力(9.8%)。该排序与加氢裂化工艺机理高度吻合——温度主导反应速率,氮含量影响催化剂活性,氢油比决定加氢深度,而金属组分与压力则是关键操作边界条件,充分体现了模型与领域知识的一致性。

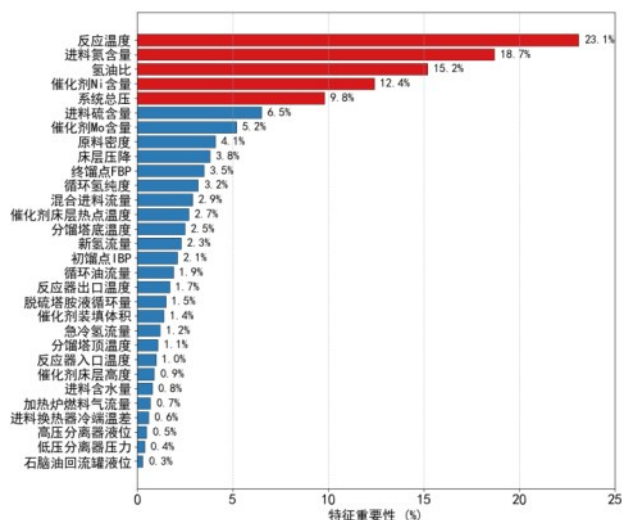


图7 变量重要性分析

Fig. 7 Variable importance analysis

5 结论与展望

本文针对柴油收率预测中原料、催化剂与操作参数间复杂的非线性、异构性及时滞耦合效应,提出了一种时滞感知Transformer(TLAT)模型。通过条件互信息精准识别各输入变量对收率的有效时滞窗口,并在TFT架构中嵌入工艺先验引导的结构化注意力掩码,有效约束模型在物理合理的时序范围内建模动态依赖;同时,设计静态-动态门控融合模块,显式表征催化剂等静态因素对过程动态的调制机制。基于140万吨/年加氢裂化装置3年高维高频运行数据的实验表明,TLAT在测试集上MAE低至0.0182、 R^2 达0.937,显著优于XGBoost、LSTM、Informer及标准TFT等主流方法,验证了其在精度、可解释性与工程实用性方面的综合优势。未来将突破当前离线时滞估计的限制,研究发展在线自适应时滞学习机制以应对工况切换;另外,可拓展TLAT至多产品质量协同预测场景,并探索其与数字孪生、实时优化(RTO)系统的深度集成,构建“感知—预测—调控”一体化智能决策闭环,为石化企业迈向高阶智能制造提供关键技术支撑。

符号说明

$A_{t,t'}$	—— t 时刻对历史 t' 时刻的自注意力权重
b	——偏置向量
c	——静态上下文向量
d_m	——模型隐藏维度
d_k	——查询、键、值向量的维度
g_j	——第 j 个特征通道的保留比例
H_{t-1}	——时刻 t 之前全部历史观测信息集合
h	——注意力头数
I	——单位矩阵
$k_{t'}$	—— t 时刻查询键
L	——TLA Block 层数
L	——损失函数
M	——全局掩码
q_t	—— t 时刻查询向量
s	——静态变量, 在装置单次稳定运行周期内保持 不变或变化极缓慢的参数
T	——历史窗口长度
W	——权重矩阵
x_t	——系统在时刻 t 观测的输入向量
$\hat{y}_{\text{concat}}$	——传统拼接模型输出
y_T	——预测当前时刻 $t = T$ 的柴油收率
z_t	——动态序列, 随时间高频变化的操作与进料 参数
$\tilde{z}_t$	——最终融合特征
ε_t	——建模误差
γ	——LayerNorm 的缩放因子
μ	——均值
$\tau_{\max}$	——最大时滞
$\bar{\tau}$	——平均有效时滞长度
σ_{noise}	——标准差
$\sigma(\cdot)$	——sigmoid 函数
$\odot$	——Hadamard 乘积
$\otimes$	——Kronecker 积
下角标	
g	——门控层
s	——全连接层

参考文献

- [1] Zhou J Z, Shi T, Qian Q M, et al. Protocol for the design and accelerated optimization of a waste-to-energy system using AI tools[J]. STAR Protocols, 2023, 4(4): 102685.
- [2] 延炜. 提高柴油收率的优化措施[J]. 炼油与化工, 2021 (1): 29-31.
Yan W. Optimization measures for improving diesel yield[J]. Refining and Chemical Industry, 2021(1): 29-31.
- [3] 聂红, 张锐, 陈文斌, 等. 柴油质量升级加氢催化剂与工艺技术的开发及应用[J]. 化学反应工程与工艺, 2025, 41(5): 587-595.
- [4] Nie H, Zhang R, Chen W B, et al. Development and application in upgrading diesel quality with hydrogenation catalysts and process technology[J]. Chemical Reaction Engineering and Technology, 2025, 41(5): 587-595.
- [5] Pang Z Y, Huang P, Lian C, et al. Data-driven prediction of product yields and control framework of hydrocracking unit[J]. Chemical Engineering Science, 2024, 283: 119386.
- [6] Zheng Q Q, Fan Y J, Zhou Z, et al. Research on product yield prediction and benefit of tuning diesel hydrogenation conversion device based on data-driven system[J]. Energies, 2023, 16(14): 5332.
- [7] Lim B, Arık S Ö, Loeff N, et al. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4): 1748-1764.
- [8] Gmeinbauer J, Ramakrishnan C, Reichhold A. Prediction of FCC product distributions by means of feed parameters[J]. Oil Gas European Magazine, 2004, 120(3): 29-33.
- [9] 谢洋, 陆新建, 邵晓雯, 等. 基于 MLP-RF 组合模型预测加氢裂化装置柴油产量[J]. 化学工程, 2024, 52(11): 66-70.
Xie Y, Lu X J, Shao X W, et al. Prediction of productivity of diesel in hydrocracking unit based on MLP-RF combined model[J]. Chemical Engineering, 2024, 52(11): 66-70.
- [10] Yang F, Xu M, Lei W Q, et al. Artificial intelligence methods applied to catalytic cracking processes[J]. Big Data Mining and Analytics, 2023, 6(3): 361-380.
- [11] 曹跃, 余冲, 李智, 等. 工业数据驱动的加氢裂化装置多工况切换过渡状态检测[J]. 化工学报, 2023(9): 3841-3854.
Cao Y, Yu C, Li Z, et al. Industrial data driven transition state detection with multi-mode switching of a hydrocracking unit[J]. CIESC Journal, 2023(9): 3841-3854.
- [12] Vieira C A, Victor dos Santos P, dos Santos Lima H O, et al. Application of machine learning models for predicting variables of industrial processes: a case study in a petrochemical industry[C]// 2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI). November 13-15, 2024, Bogota D. C., Colombia. IEEE, 2024: 1-6.
- [13] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [14] Zheng W D, Chen G. An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization[J]. IEEE Transactions on Cybernetics, 2022, 52(12): 13902-13914.
- [15] Panjapornpon C, Bardeeniz S, Hussain M A. Improving energy efficiency prediction under aberrant measurement using deep compensation networks: a case study of petrochemical process[J]. Energy, 2023, 263: 125837.
- [16] 王琳, 张永渝, 金涛, 等. 基于稀疏特征提取和门控循环单元的炼化装置能耗预测方法研究[J]. 现代化工, 2025, 45(10): 253-258.
Wang L, Zhang Y Y, Jin T, et al. Refinery energy consumption prediction method based on sparse feature extraction and gated recurrent unit[J]. Modern Chemical Industry, 2025, 45(10): 253-258.
- [17] Geng Z Q, Zhang Y H, Li C F, et al. Energy optimization and prediction modeling of petrochemical industries: an improved convolutional neural network based on cross-feature[J]. Energy,

- 2020, **194**: 116851.
- [17] 刘浩宇, 王晨, 倪元, 等. 信号分解与TCN-LSTM构建及在石化企业外排水水质预测中的应用[J]. 环境化学, 2025, **44**(12): 4809-4818.
- Liu H Y, Wang C, Ni Y, et al. Research on signal decomposition and the construction of TCN-LSTM applied to the prediction of wastewater quality at the discharge outlets of petrochemical plants [J]. Environmental Chemistry, 2025, **44**(12): 4809-4818.
- [18] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]// Proc of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2017: 6000-6010.
- [19] 杨晔, 卢建刚. 基于融合Transformer的门尼系数预测建模研究[J]. 化工学报, 2025, **76**(1): 266-282.
- Yang Y, Lu J G. Mooney viscosity prediction modeling based on fusion Transformer[J]. CIESC Journal, 2025, **76**(1): 266-282.
- [20] 张秦伟, 张菁雯, 郑斌, 等. 基于MIC-CCF-Informer的制氢转化炉出口温度预测[J]. 现代化工, 2025, **45**(12): 229-234.
- Zhang Q W, Zhang J W, Zheng B, et al. Prediction of temperature at outlet of hydrogen production reformer based on MIC-CCF-Informer network[J]. Modern Chemical Industry, 2025, **45**(12): 229-234.
- [21] Maldonado-Cruz E, Pyrcz M J. Multi-horizon well performance forecasting with temporal fusion transformers[J]. Results in Engineering, 2024, **21**: 101776.
- [22] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information[J]. Physical Review E, 2004, **69**(6): 066138.
- [23] 朱春梦, 李增, 柳楠, 等. 基于自编码器和多尺度符号转移熵的FCC沉降器跑剂故障检测[J]. 化工学报, 2025, **76**(9): 4512-4523.
- Zhu C M, Li Z, Liu N, et al. Fault detection of catalyst loss in FCC disengager based on autoencoder and multi-scale symbolic transfer entropy[J]. CIESC Journal, 2025, **76**(9): 4512-4523.
- [24] Ji C, Ma F Y, Wang J H, et al. Real-time industrial process fault diagnosis based on time delayed mutual information analysis[J]. Processes, 2021, **9**(6): 1027.
- [25] Li M, Xu X H, Hu L. Multivariate conditional Granger causality analysis based on feature selection and Backward Time-lagged Variable Selection algorithm[C]//2024 14th Asian Control Conference (ASCC). July 5-8, 2024, Dalian, China. IEEE, 2024: 1879-1884.
- [26] 王雅琳, 夏海兵, 袁小锋, 等. 基于趋势相似度分析的多重时滞辨识及其在加氢裂化流程中的应用[J]. 化工学报, 2018, **69**(3): 1149-1157.
- Wang Y L, Xia H B, Yuan X F, et al. Multi-delay identification by trend-similarity analysis and its application to hydrocracking process[J]. CIESC Journal, 2018, **69**(3): 1149-1157.
- [27] 杨晨, 彭小龙, 朱苏阳, 等. 一种基于时序注意力动态卷积的油气井产量预测方法[J]. 油气藏评价与开发, 2025, **15**(6): 1046-1055.
- Yang C, Peng X L, Zhu S Y, et al. An oil and gas well production prediction method based on temporal attention and dynamic convolution[J]. Petroleum Reservoir Evaluation and Development, 2025, **15**(6): 1046-1055.
- [28] He G, Luo L, Zhou L, et al. Deep learning prediction of yields of fluid catalytic cracking via differential evolutionary dual-stage attention-based LSTM[J]. Fuel, 2024, **370**: 131826.
- [29] 郭子越, 金莹, 朱孙泳, 等. GAST-Net: 基于门控循环单元和注意力机制的发射光谱层重建预测[J]. 光学学报, 2025, **45**(16): 198-211.
- Guo Z Y, Jin Y, Zhu S Y, et al. GAST-net: reconstruction and prediction by emission spectral tomography based on gated recurrent units and attention mechanism[J]. Acta Optica Sinica, 2025, **45**(16): 198-211.
- [30] Song C R, Yang H D, Cai J Y, et al. Multi-energy load forecasting via hierarchical multi-task learning and spatiotemporal attention [J]. Applied Energy, 2024, **373**: 123788.
- [31] Zhu Y Q, Yang F, Torgashov A. Causal-Transformer: Spatial-temporal causal attention-based transformer for time series prediction[J]. IFAC-PapersOnLine, 2024, **58**(14): 79-84.
- [32] 孙毅, 邵远楫, 单继宏, 等. 基于先验时滞特征的线缆捻制丝胶张力均衡控制方法[J]. 计算机集成制造系统, 2024, **30**(9): 3174-3182.
- Sun Y, Shao Y J, Shan J H, et al. Strands tension balance control method of cable twisting based on characteristics of priori time-delay[J]. Computer Integrated Manufacturing Systems, 2024, **30**(9): 3174-3182.
- [33] 王笑辰, 李世娟, 黄文胜, 等. 融合先验知识与多源异构数据的临床误诊风险预测与特征识别研究[J/OL]. 数据分析与知识发现, 2026: 1-18. (2026-01-02)[2026-01-02]. <https://link.cnki.net/urlid/10.1478.g2.20260113.1308.002>.
- Wang X C, Li S J, Huang Wensheng, et al. Clinical misdiagnosis risk prediction and feature identification based on prior knowledge and multi-source heterogeneous data[J/OL]. Data Analysis and Knowledge Discovery, 2026: 1-18. (2026-01-02)[2026-01-02]. <https://link.cnki.net/urlid/10.1478.g2.20260113.1308.002>.
- [34] 张京新, 何皎洁, 蔡庆旺, 等. 基于二次分解和BiLSTM的污水厂出水COD浓度预测[J]. 化工学报, 2025, **76**(6): 2859-2871.
- Zhang J X, He J J, Cai Q W, et al. Prediction of COD concentration in wastewater treatment plant effluent based on secondary decomposition and BiLSTM[J]. CIESC Journal, 2025, **76**(6): 2859-2871.
- [35] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. ACM, 2016: 785-794.
- [36] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, **323**(6088): 533-536.
- [37] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, **35**(12): 11106-11115.
- [38] Fay M P, Brittain E H, Shih J H, et al. Causal estimands and confidence intervals associated with Wilcoxon-Mann-Whitney tests in randomized experiments[J]. Statistics in Medicine, 2018, **37**(20): 2923-2937.
- [39] Shrikumar A, Greenside P, Kundaje A B. Learning important features through propagating activation differences[C]// International Conference on Machine Learning., 2017