

基于 SOP 先验与极值敏感损失的中药提取过程关键工艺参数解耦数字孪生建模

胡思学^{1,3}, 张欣^{2,4*}, 王钰淳^{1,3}, 刘洋^{2,3}, 章晨峰^{2,3}, 王团结^{2,4}

1. 江苏海洋大学工程学院, 江苏 连云港 222001

2. 江苏康缘药业股份有限公司, 江苏 连云港 222001

3. 江苏省海洋药物和现代中药创制重点实验室(江苏康缘药业股份有限公司/江苏海洋大学), 江苏 连云港 222001

4. 中药制药过程控制与智能制造技术全国重点实验室, 江苏 连云港 222001

摘要: **目的** 针对中药提取过程存在强耦合、长时滞及非线性特点, 以及传统深度时序模型在工况变化时易产生预测滞后和极值平滑等问题, 提出一种结合标准操作规程(standard operating procedure, SOP)先验的中药提取过程关键工艺参数解耦数字孪生建模方法 SOP 引导的通道独立分块时间序列 Transformer (SOP-guided channel-independent patch time series Transformer, SGC-PatchTST)。**方法** 该模型以 PatchTST 为基础架构, 从 3 个方面进行改进: 采用正交解耦后期融合机制, 将离散 SOP 指令转化为工艺意图特征, 并利用柔性门控进行融合以抑制工况切换产生的滞后; 构建主支路和副支路进行解耦的结构, 对回流温度进行独立建模, 以削弱多变量耦合造成的干扰; 设计极值敏感混合损失函数, 将局部统计与极值惩罚相结合, 在 Patch 分块低通滤波条件下增强模型的极值拟合与抗噪能力。所有基线模型均统一采用 OneCycleLR 学习率调度与早停策略。实验数据按批次独立划分训练集、验证集、测试集, 杜绝跨批次数据泄漏。研究使用中药提取车间实际生产数据构建预测模型并进行自回归多步推演评估。**结果** 在 50 批次分层划分下, SGC-PatchTST (优化超参数 $\tau=1.5\sigma$, $\lambda=2.0$) 的上、下温度单次评估平均绝对误差(mean absolute error, MAE)分别降至 0.36 °C 和 0.35 °C (较标准 PatchTST 模型分别大幅降低 40.0%和 50.2%), 且二者的 95%分位数误差(P95 error)分别控制在 1.31 °C 和 1.33 °C, 抗极值偏差能力显著提升。同时, 得益于主副解耦架构, 受外部干扰明显的回流温度 MAE 降至较低的 0.04 °C, 而核心安全参数(罐内压力)的微小波动预测亦保持在 0.000 6 MPa 的较高精度。此外, 在 3 次独立分层划分下, 核心上、下温度 MAE 均值分别保持在 0.33 °C (变异系数 6.8%)和 0.32 °C (变异系数 6.4%), 证实了较高的统计稳健性。6 路组合消融实验和 25 组超参数网格搜索分别验证了各模块的贡献和超参数鲁棒性。**结论** 该 SGC-PatchTST 方法在预测准确性和时效性上均达到了比较理想的程度, 具有较强的物理合理性, 可为后续基于模型的强化学习控制提供支持。

关键词: 中药提取; 数字孪生; PatchTST; 先验注入; 极值敏感损失; 多变量时序预测

中图分类号: TP18; R283.3 **文献标志码:** A **文章编号:** 0253-2670(2026)19-7739-12

DOI: 10.7501/j.issn.0253-2670.2026.19.019

Decoupled digital twin modeling of key process parameters for traditional Chinese medicine extraction process based on SOP priors and extrema-sensitive loss

HU Sixue^{1,3}, ZHANG Xin^{2,4}, WANG Zhengchun^{1,3}, LIU Yang^{2,3}, ZHANG Chenfeng^{2,3}, WANG Tuanjie^{2,4}

1. School of Electronic Engineering, Jiangsu Ocean University, Lianyungang 222001, China

2. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China

3. Jiangsu Key Laboratory of Marine Drugs and Modern Traditional Chinese Medicine Innovation (Jiangsu Kanion Pharmaceutical Co., Ltd./Jiangsu Ocean University), Lianyungang 222001, China

4. National Key Laboratory of Traditional Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacturing Technology, Lianyungang 222001, China

收稿日期: 2026-07-16

基金项目: 国家长三角科技创新共同体联合攻关项目(2023CSJGG1700); 连云港市重点研发计划(产业前瞻与关键技术)(CG2423)

作者简介: 胡思学, 男, 硕士研究生, 研究方向为人工智能。E-mail: 83533691@qq.com

*通信作者: 张欣, 男, 博士研究生, 硕士生导师, 研究方向为中药制造过程新技术。E-mail: zxtcm@126.com

Abstract: Objective To address the strong coupling, long time delays, and nonlinear characteristics inherent in the traditional Chinese medicine (TCM) extraction process, as well as the issues of prediction lag and extreme-value smoothing commonly exhibited by traditional deep time-series models during operating condition transitions, a decoupled digital twin modeling method of key process parameters for the TCM extraction process, integrating Standard Operating Procedure (SOP) priors, named SOP-guided channel-independent patch time series Transformer (SGC-PatchTST), is proposed. **Methods** Based on the PatchTST architecture, the proposed model introduces three major enhancements: 1. An orthogonal decoupling late-fusion mechanism is employed to transform discrete SOP instructions into process-intention features, utilizing a soft-gating mechanism for fusion to suppress the lag caused by operating condition switches; 2. A decoupled structure comprising primary and auxiliary branches is constructed to independently model the reflux temperature, thereby mitigating the interference caused by multivariable coupling; 3. An extrema-sensitive hybrid loss function is designed, combining local statistics with extreme-value penalties to enhance the model's extrema-fitting and noise-resistance capabilities under the low-pass filtering conditions of Patch-based tokenization. All baseline models uniformly adopt the OneCycleLR learning rate schedule and an early-stopping strategy. Experimental data are independently partitioned into training, validation, and test sets on a batch-by-batch basis to strictly eliminate cross-batch data leakage. Actual production data from a TCM extraction workshop are utilized to construct the predictive model and evaluate it via autoregressive multi-step inference. **Results** Under a 50-batch stratified split, the single-evaluation Mean Absolute Errors (MAEs) of the upper and lower temperatures for SGC-PatchTST (optimized hyperparameters $\tau=1.5\sigma$, $\lambda=2.0$) decrease to 0.36 °C and 0.35 °C, respectively (which are 40.0% and 50.2% lower than the standard PatchTST model, respectively). Its capability to resist extreme deviations is significantly enhanced, with the 95% percentile errors (P95 errors) for the upper and lower temperatures controlled at 1.31 °C and 1.33 °C, respectively. At the same time, benefiting from the decoupled primary and secondary architectures, the MAE of the reflux temperature, which is susceptible to external disturbances, is reduced to a lower value of 0.04 °C. The prediction of minor fluctuations in the core safety parameter (in-tank pressure) is also maintained at a relatively high precision of 0.000 6 MPa. Furthermore, under three independent stratified splits, the average upper and lower temperature MAEs remain at 0.33 °C (coefficient of variation, CV = 6.8%) and 0.32 °C (CV = 6.4%), respectively, confirming its good statistical robustness. A six-step combination ablation study and a 25-group hyperparameter grid search validate the contribution of each module and the hyperparameter robustness, respectively. **Conclusion** The proposed SGC-PatchTST method achieves an ideal level of prediction accuracy and timeliness while demonstrating strong physical plausibility, providing reliable support for subsequent model-based reinforcement learning control.

Key words: traditional Chinese medicine extraction; digital twin; PatchTST; prior injection; extrema-sensitive loss; multivariate time series forecasting

中药提取工艺表现出较强的耦合性、较大的时滞和非线性特点,批次一致性问题中药现代化需要解决的主要问题^[1]。在过程分析技术(process analytical technology, PAT)发展和工业 4.0 背景下,建立高保真数字孪生体是实现智能化闭环控制的重要方式^[2-4]。

中药提取过程的建模方法可大致分为 3 类。机制模型基于传热、传质和相变的热力学平衡方程建立,能反映过程的本质物理规律,但参数(如传热系数、比热容)获取困难。统计模型如偏最小二乘回归和人工神经网络,依赖过程数据建立经验映射,但难以捕捉长时序依赖关系。深度学习模型如长短期记忆网络(long short-term memory, LSTM)和 Transformer 等近年来在时序预测方面展现出优势,但在工况切换时会出现相位滞后和极值平滑问题,且多变量耦合容易造成特征竞争和过拟合^[5-7]。分块时间序列 Transformer(patch time series

transformer, PatchTST)通过分块机制将长序列切分为局部子序列,在保留时序依赖的同时引入隐式低通滤波,但标准 PatchTST 仍缺乏对工艺先验和变量间物理约束的建模能力。

针对上述问题,本研究将标准操作规程(standard operating procedure, SOP)先验和解耦 PatchTST 方法进行结合^[8-9]。该方法借鉴物理信息神经网络和知识先验融合的思想^[10-12],采用正交解耦后期融合、主副支路独立建模及极值敏感混合损失函数的设计,以达到工艺意图引导和抗噪极值拟合的目的。与以往在输入端直接拼接工艺信息(即 $X_{\text{input}}=[X, S]$,其中 X 为物理变量, S 为 SOP 独热编码)的方法不同,本研究提出了一种特征层正交融合机制。在数学与机制层面,传统的直接拼接法极易导致低维且稀疏的离散 SOP 信号在深层高维网络中被连续的热力学特征淹没,引发特征竞争。本研究将物理状态特征热力学惯性延迟(H_{state})与工

艺意图特征 (H_{intent}) 在独立的正交子空间中隔离编码, 并利用柔性门控机制 $G = \sigma(W \cdot H_{\text{intent}})$ 在输出端进行乘性调制 $H_{\text{final}} = H_{\text{state}} \odot G$ 。其中, H_{final} 为经门控调制后的融合特征。

该设计的本质改进在于将多变量从“一体化黑箱建模”转变为“基于物理因果的解耦建模”。在机制上, 温度、压力是具有巨大热惯性的连续物理量, 而 SOP 步号切换是瞬间发生的离散指令。二者物理性质截然不同。正交融合机制允许骨干网络专注提取系统的 H_{state} , 而将 SOP 作为显式的“物理开关” (G) 在工况突变点处强制扭转预测轨迹, 完美契合了工业控制中“控制器 (SOP) 干预被控对象 (提取罐)”的因果机制。此外, 结合极值敏感损失函数, 有效解决了传统平均绝对误差 (mean absolute error, MAE) 固有的极值平滑缺陷。

1 中药提取工艺机制与数据分析

1.1 工艺特征与建模需求

中药提取过程中包含复杂热传导、质传输以及相变等现象, 而且其中多个因素相互影响紧密。本研究针对常压提取工艺开展相关工作, 在此基础上对其工艺原理、主要控制参数以及变化规律进行探讨, 为建立数字孪生体奠定基础。

1.2 常压沸腾提取机制

中药提取工段一般有 3 步: 加热、沸腾保温以及出液。在加热过程中, 底部夹套通入蒸汽用于加热提取罐内的药材及溶剂; 到一定温度时, 转为常压沸腾, 产生的二次蒸汽在顶部冷凝器中冷凝成液体流回。

该提取过程表现出两大显著的动态特征: 一方面系统有较大的时滞以及热惯性, 即蒸汽阀门动作后, 上、下温度传感器的变化会有所延迟; 另一方面, 在常压沸腾工况下, 罐内的压力维持在一个较小的标准偏差微正压范围内, 压力变化由内部沸腾状态驱动。因此要求模型不仅要有良好的跟踪效果, 而且要能较好地拟合在一个较小的标准偏差范围内变化的压力。提取工艺流程与传感器分布如图 1 所示。

1.3 变量解耦与因果重构

传统的时序模型一般采取所有变量一体化建模方法, 难以体现各个变量之间的联系以及容易造成负迁移等问题, 为此本研究对变量进行 2 方面的优化。

第一, 加入 SOP 先验状态。用“工艺步号”作

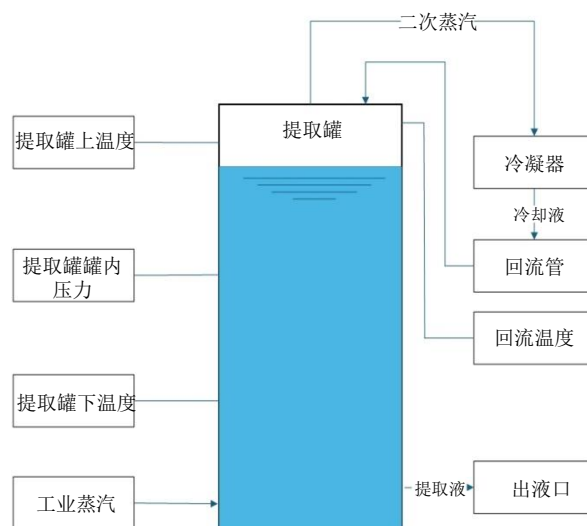


图 1 提取工艺流程与传感器分布示意图

Fig. 1 Schematic illustration of extraction process flow and sensor distribution

为硬先验加入到输入序列中, 让模型知道接下来的操作意图, 有利于其更好判断升温或者降温转折点。工艺步号取自批次记录的 SOP 工艺段标识, 本研究所涉常压提取工艺包含加热升温、沸腾保温、出液等阶段, 以独热编码 (one-hot) 形式表示为 12 维向量, 每个维度对应 1 个工艺步号, 当前执行步为 1, 其余为 0。该编码在批次生产前即可由工艺配方确定, 不依赖当前时刻之后的传感器数据, 因此不存在未来信息泄漏。第二, 分离状态变量结构。上温度、下温度以及罐内压力主要是由蒸汽进料决定, 是主要热力学参数; 而回流温度更多的是由冷却水量及起始温度决定, 外界干扰更大。基于系统热力学驱动机制的差异, 本研究使用“主-副”2 分支进行建模, 以削弱各变量之间的互相牵制。在自回归推演评估中, 各时间步使用的 SOP 步号取自该时间步对应的预定义工艺阶段, 外部设定值 (温度、压力) 使用目标批次中该时间步对应的已知值, 阀门开度沿用上一步动作, 被控变量由模型预测输出替换。该方式模拟了实际控制中设定值已知、被控变量需由模型推演的典型场景。

1.4 极值与高频噪声特征分析

在实际情况中, 由于工况变化而产生的真实极大值以及传感器噪声往往混在一起, 在实际生产的历史数据采样中发现, 在提取液沸腾以及液态水向气态水蒸气转化的相变期间, 温度传感器可能会因为气泡破裂或者局部湍流导致尖峰式跳变, 然而, 由中药提取系统巨大的热惯性可知, 吨级物料不可

能在很短时间内出现剧烈温度振荡违背物理常理，据此可判定该类尖峰跳变实为高频噪声干扰。

如果只把 MSE 当作唯一的优化目标来拟合，模型会在工艺突变点处更宜输出平滑的均值，并且容易对高频噪声出现过拟合。所以，理想的建模过程不但要拥有低通滤波的抗干扰能力，还要提高对真实极值的捕捉灵敏度，这是本研究设计混合损失函数的主要动机。

2 数据集描述与预处理

本研究所用数据来源于江苏康缘药业股份有限公司中药提取车间常压沸腾提取工艺的 50 个生产批次，共计 413 056 条采样记录。每条记录包含 32 维特征，可分为 5 类：（1）被控变量（4 维）：上温度、下温度、罐内压力、回流温度；（2）过程变量（6 维）：

蒸汽压力、蒸汽阀门开度、进液流量、出液流量及进/出液流量状态（进/出液阀门启停的二值指示量）；（3）工艺设定值（4 维）：设定蒸汽压力、设定上温度、设定下温度、设定加水量；（4）SOP 工艺步号（12 维）：以独热编码形式表示的当前工艺段标识；（5）机制派生特征（6 维）：温度变化率、距沸点距离、过热程度、温度加速度、压力变化率、饱和压差。各特征的统计描述如表 1 所示。上、下温度覆盖 20.8~105.0 °C 的宽动态范围，中位值约 88 °C，与常压沸腾工况相符；罐内压力波动极小（标准差仅 0.001 MPa），为微正压沸腾的典型特征。对于传感器采集过程中出现的缺失值，采用前向填充后再用后向填充的方式处理，剩余少量缺失值以 0 填充。对于气泡破裂或局部湍流引起的尖峰式高频噪声，保

表 1 数据集特征统计描述

Table 1 Statistical description of dataset features

类别	变量	单位	最小值	下四分位数	中位数	均值	上四分位数	最大值
被控变量	上温度	°C	20.8	45.5	88.1	75.3	100.6	105.0
	下温度	°C	19.3	44.8	87.2	74.6	100.0	104.5
	罐内压力	MPa	-0.017 0	-0.001 0	0.000 0	-0.000 1	0.000 0	0.027 0
	回流温度	°C	21.2	26.3	28.0	27.9	29.5	34.1
过程变量	蒸汽压力	MPa	-0.009	-0.002	-0.001	0.035	0.040	0.285
	蒸汽阀门开度	%	0.00	0.00	0.00	9.19	16.66	70.00
	进液流量	m ³ ·h ⁻¹	0.00	0.00	0.00	0.55	0.00	20.00
	出液流量	m ³ ·h ⁻¹	0.00	0.00	0.00	2.11	0.00	17.40
	流量状态（进液）	—	0.00	0.00	0.00	0.06	0.00	1.00
	流量状态（出液）	—	0.00	0.00	0.00	0.18	0.00	1.00
设定值	设定蒸汽压力	MPa	0.000	0.000	0.000	0.065	0.050	0.270
	设定上温度	°C	25.0	25.0	25.0	54.2	99.0	102.0
	设定下温度	°C	25.0	25.0	25.0	54.2	99.0	102.0
	设定加水量	m ³	0.00	0.00	0.00	1.22	3.15	3.15
SOP 先验	工艺步号（12 维独热编码）	—	0	0	0	0.015~0.454	0	1
派生特征	温度变化率	°C·步 ⁻¹	-70.6	-0.1	0.0	0.0	0.1	75.7
	距沸点距离	°C	-5.0	-0.6	11.9	24.7	54.5	79.2
	过热程度	°C	0.00	0.00	0.00	0.67	0.60	5.00
	温度加速度	°C·步 ⁻²	-76.1	-0.2	0.0	0.0	0.2	75.9
	压力变化率	MPa·步 ⁻¹	-0.017	-0.001	0.000	0.000	0.001	0.017
	饱和压差	MPa	-0.905	-0.778	-0.484	-0.448	-0.072	-0.002

统计量基于全部 50 批次共 413 056 条采样记录计算；流量状态为二值指示量（1 表示该时刻进/出液阀门开启，0 表示关闭）；SOP 工艺步号各维为独热编码，二者的均值均表示对应状态占全流程记录的比例；SOP 独热编码中，1 表示当前处于对应工艺步骤，0 表示不处于该步骤；均值范围 0.015~0.454 表示 12 个编码维度对应工艺步骤的采样记录占比分别介于 1.5%~45.4%。

Statistics are calculated based on a total of 413 056 sampling records across all 50 batches; The flow status is a binary indicator (1 means the inlet/outlet liquid valve is open at this moment, and 0 means it is closed); Each dimension of the SOP process step number adopts one-hot encoding, and the mean values of both represent the proportion of the corresponding state in the full-process records; For SOP one-hot encoding, 1 indicates that the corresponding process step is active, whereas 0 indicates that it is inactive, the mean range of 0.015—0.454 indicates that the sampling records associated with the respective process steps account for 1.5%—45.4% of all records across the 12 encoded dimensions.

留在原始数据中以测试模型的抗噪能力。

为防止跨批次数据泄漏，本研究按批次独立划分训练、验证和测试集，同一批次的所有采样点仅属于三者之一，不同批次的窗口之间不存在任何重叠。采用分层抽样策略，按冷启动批次（起始温度 < 40 °C，共 7 批）和热态标准批次（起始温度 ≥ 40 °C，共 43 批）进行分层，训练集、验证集、测试集分别包含 35、7、8 批，确保各集合均涵盖全工况类型。此外，为验证小样本条件下模型评估的可靠性，本研究在后续的核心性能验证中采用 3 个不同随机种子（42, 123, 456）进行 3 次独立的分层批次划分，以全面排查数据划分偶然性对评估结论的影响。

3 高保真数字孪生预测架构设计

针对中药提取过程中存在的多因素强耦合与相互牵制问题，本研究提出了一种融合 SOP 先验信息的解耦 PatchTST 预测架构。该架构包括 SOP 先验注入模块、主副解耦预测模型以及极值敏感损失函数 3 个部分。

在主预测支路中，本研究使用 Channel-Independent PatchTST 模型来对上温度、下温度、罐内压力进行长序列特征提取。副预测支路使用多层感知机（multilayer perceptron, MLP）单独对回流温度进行预测，并且利用后期融合机制把双分支的输出进行整合。该设计可以保证主要热力学参数的预测精度，同时也实现了对次要外部因素的有效建模。与传统的多输入多输出（multiple-input multiple-output, MIMO）端到端训练方法相比，该解耦结构减少了不同动态特性变量之间的相互干扰，并且也降低了负迁移问题。解耦数字孪生预测模型的整体架构如图 2 所示。

4 SOP 先验融合与极值敏感优化

4.1 正交时空解耦与后期融合

本研究利用 SOP 意图指导分支来处理工况变化造成的相位滞后问题。针对传统方法在输入端直接拼接工艺信息易导致的特征竞争与被淹没问题，本研究设计了一种基于特征层正交解耦的后期门控融合机制，该机制在数学与机理双重层面上构建了“物理状态特征 H_{state} ”与“工艺意图特征 H_{intent} ”的隔离与调制体系。

设输入的主导物理状态序列（如温度、压力）为 $X \in \mathbb{R}^{L \times C}$ ，其中 $L=168$ 为回溯窗口长度， C 为连续物理变量的维度。SOP 先验序列表示为 $S \in \mathbb{R}^{L \times K}$ ，其中 K 为工艺步号的独热编码维度。

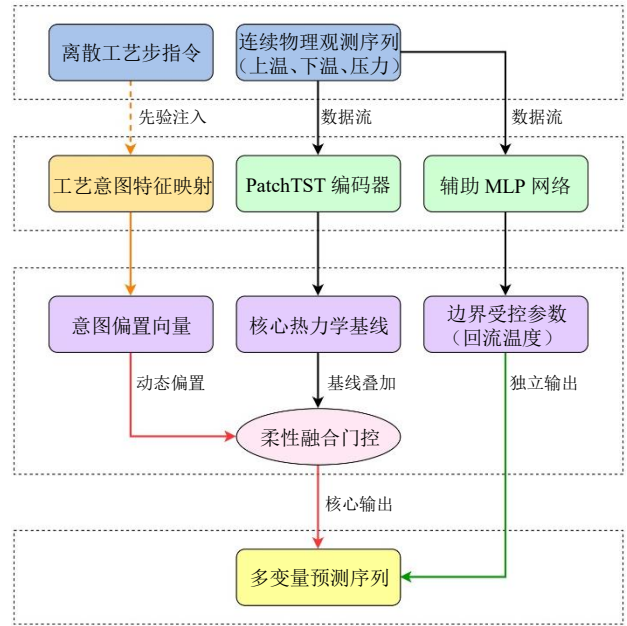


图 2 解耦数字孪生预测模型整体架构

Fig. 2 Overall architecture for decoupled digital twin predictive model

首先，主预测模型利用 Patch 分块机制对高维序列 X 进行降维与特征提取。设定分块长度 $P=12$ ，步长 $S_p=6$ ，原始序列被划分为 N 个片段。

$$N=(L-P)/S_p+1 \quad (1)$$

通过 Channel-Independent 的自注意力编码器，提取出纯粹由热力学惯性驱动连续物理状态特征 H_{state} 。

$$H_{\text{state}}=\text{PatchTST}(X) \in \mathbb{R}^{P_{\text{out}} \times C} \quad (2)$$

其中 $P_{\text{out}}=24$ 为预测视距。从信号处理的角度来看，Patch 分块机制引入了隐式低通滤波效应，有效抑制了气泡破裂等高频噪声，使 H_{state} 能够平稳刻画低频演变趋势。

同时，在 SOP 意图指导分支中，为避免离散指令与连续物理量产生特征纠缠，本研究在独立的正交子空间中对 S 进行时空变换。在时间维度上对齐预测视距以提取热惯性延迟，在通道维度上将工艺意图映射至多变量空间，获得意图特征 H_{intent} 。

$$H_{\text{intent}}=S \cdot W_{\text{time}} \cdot W_{\text{channel}} \in \mathbb{R}^{P_{\text{out}} \times C} \quad (3)$$

W_{time} 、 W_{channel} 为可学习的线性映射矩阵

在网络输出端，为实现物理因果逻辑的强绑定，本研究摒弃了简单的加性融合，设计了非线性柔性门控后期融合单元。该单元利用 H_{intent} 生成动态调制权重，对底层物理预测结果 H_{state} 施加乘性干预。

$$G=\sigma(W_g \cdot H_{\text{intent}}+b_g) \quad (4)$$

$$\hat{Y} = H_{\text{state}} + H_{\text{state}} \odot G \quad (5)$$

$\sigma(\cdot)$ 为 Sigmoid 激活函数; W_g 和 b_g 分别为门控映射的可学习权重矩阵和偏置向量; G 为取值范围在 $(0, 1)$ 的门控矩阵; $\odot$ 表示哈达玛积 (Hadamard product), 即对应元素相乘; $\hat{Y}$ 为最终的主支路预测输出

该设计的数学本质与物理机制可以概括如下: 当提取过程处于恒温沸腾等稳态工况时, SOP 意图未发生改变, 门控矩阵 G 趋于零或常数, 式 (5) 退化为 $\hat{Y} \approx H_{\text{state}}$, 模型高度依赖自回归热惯性输出, 保证稳态精度; 当遭遇加水或出液等工况切换时, 离散的 SOP 突变迅速激发门控 G 产生大幅阶跃, 作为高优先级的“物理偏置”通过乘性调制途径强制扭转热惯性轨迹。此机制在数学与物理双重层面上有效抑制了纯数据驱动模型在动态工况切换时普遍存在的预测滞后问题。

4.2 极值敏感型混合损失函数

传统的均方误差 (mean squared error, MSE) 损失容易使得模型向均值收敛, 在工艺突变点附近会造成极值被抹平的现象, 因此本研究提出一种新的 Extrema-Aware MixLoss, 把局部统计特征、绝对误差以及变化率进行结合作为优化的目标。

该损失函数包括基础误差项、一阶差分防滞后项以及 Z-score 动态极值惩罚算子 3 部分。对于每个时间点 (i), 计算真实的目标序列 (Y) 在预测窗口内均值 μ 和方差 σ , 然后得到每个时间点上的偏离程度 Z_i 。

$$Z_i = \frac{|y_i - \mu|}{\sigma + \epsilon} \quad (6)$$

其中 ϵ 为极小常数, 用于防止分母为零。为了提升模型对极值点的敏感度, 引入非线性放大的权重 W_i 。

$$W_i = 1.0 + \lambda \cdot Z_i \cdot I(Z_i > \tau) \quad (7)$$

其中, $I(\cdot)$ 为指示函数, 当括号内条件成立时取 1, 否则取 0。之所以采用指示函数进行选择激活 (而非将 Z_i 直接作为连续权重), 是因为工艺稳定阶段的数据点占绝大多数, 其 Z_i 值集中在 1.0 附近, 若不加以区分地施加连续权重, 等同于对全部样本提高拟合要求, 既增加计算开销, 又可能使模型在平稳区段产生不必要的过拟合。通过设置阈值 τ , 仅对偏离程度超过 τ 的少数极值点施加额外惩罚, 实现“正常点正常拟合, 极值点重点拟合”的设计目标; τ 为极值触发阈值, λ 为极值惩罚系数。本研究取 $\tau=1.5\sigma$ 、 $\lambda=2.0$, 选取依据见“5.7”项下超参数敏

感性分析。

最终的混合损失函数 L_{total} 定义如下。

$$L_{\text{total}} = \frac{\alpha}{M} \sum W_i (y_i - \hat{y}_i)^2 + \frac{\beta}{M} \sum W_i |y_i - \hat{y}_i| + \gamma \cdot L_{\text{diff}} \quad (8)$$

$\hat{y}_i$ 是模型预测值, α 、 β 、 γ 是超参数权重

$$L_{\text{diff}} = \frac{1}{M-1} \sum |(\hat{y}_{i+1} - \hat{y}_i) - (y_{i+1} - y_i)| \quad (9)$$

其中, L_{diff} 为一阶差分惩罚项, 用来使预测路径变化趋势接近真实路径变化趋势, 从而减少相位滞后。该设计可以在工况稳定情况下有较小误差, 在工况改变时能更快速地响应极值以及斜率的变化, 从而提高模型对于变化过程的鲁棒性。

5 实验验证与结果讨论

本研究基于提取车间真实数据开展自回归多步推演评估。对比基线模型包含线性映射模型 (Linear)、一维卷积神经网络 (convolutional neural network-1D, CNN-1D)、LSTM、标准自注意力模型 (Vanilla Transformer) 以及标准 PatchTST (使用 Channel-Independent PatchTST 架构和 MSE 损失, 无 SOP 融合且无主副解耦)。

5.1 实验设置与工业级评价体系重构

实验使用滑动窗口生成序列数据, 在过去的时间段内有 $L=168$ 个时间步, 向前预测未来 $P=24$ 个时间步的物理状态序列。为确保对比公平性, 所有模型 (含标准 PatchTST 及后续消融变体) 统一使用相同的输入维度、预测目标变量和预测视距。

模型训练采用 AdamW 优化器, 初始学习率设为 3×10^{-3} , 采用 OneCycleLR 学习率调度策略, 预热比例设为 0.3。批次大小设置为 128, 权重衰减系数为 1×10^{-3} , 梯度裁剪阈值设为 1.0。训练集滑动采样步长为 4, 验证集滑动采样步长为 24。早停机制中耐心值设为 6 个评估周期, 若连续 6 个评估周期内验证损失未下降则提前终止训练。PatchTST 骨干网络的隐层维度 $d_{\text{model}}=32$, 注意力头数 $n_{\text{head}}=4$, 编码器层数 $n_{\text{layers}}=2$, 随机失活 (Dropout) 比率为 0.25。

为全面且真实地反映模型在工业实际应用中的性能, 本研究摒弃了常规的单一指标评价, 重构了多维度的工业级评价体系。首先, 在评估方式上, 杜绝了传统时序预测中依赖真实标签的单步前向推演, 全面采用自回归多步推演, 以此真实模拟工业闭环控制中设定值已知、被控状态需由模型长期推演的典型应用场景。其次, 在量化指标上构建了

涵盖全局、极值与微观维度的 3 层评估矩阵：在全局稳态层面，采用 MAE、MSE 和决定系数 (R^2) 来评估温度等大范围跨度变量的整体拟合精度；在鲁棒极值层面，针对工况突变易引发的预测滞后与极值平滑问题，特别引入 95%分位数误差 (P95 error)，即满足概率 $P(|y_i - \hat{y}_i| \leq P95) = 0.95$ 的误差临界值，以此量化模型在追踪瞬态工艺突变时的抗最大偏差能力；在微观安全边界层面，针对罐内压力等绝对波动极小 ($\sigma \approx 0.003$ MPa) 的核心参数，鉴于常规评估指标在此失效，本研究进一步引入误差概率密度分布、回归误差特征 (regression error characteristic, REC) 曲线及绝对误差箱线图，从而深度刻画模型在微小波动工况下的局部跟踪高保真度。

各量化指标的计算公式如下。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2 \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n |y_i - \hat{y}_i|^2}{\sum_{i=1}^n |y_i - \bar{y}|^2} \quad (12)$$

$$P(|y_i - \hat{y}_i| \leq e_{95}) = 0.95 \quad (13)$$

$$E_{\max} = \max_i |y_i - \hat{y}_i| \quad (14)$$

n 为样本点数； y_i 与 $\hat{y}_i$ 分别为第 i 个采样点的真实值与模型预测值； $\bar{y}$ 为真实值均值； e_{95} 为满足式 (13) 的 95%分位数误差临界值，即 P95 error； $E_{\max}$ 为最大绝对误差 (MaxError)，用于超参数敏感性评估

5.2 全局预测结果

各模型在真实批次测试集上的全局性能统计如表 2 所示。定量评估结果表明，纯数据驱动方法在强耦合热力学系统中鲁棒性不好。就温度而言，LSTM 因不能感知工艺切换，容易在工况突变处产生较大偏差，导致其上温度和下温度的 MSE 分别达到了 3.476 8 和 3.841 7。Vanilla Transformer 虽然加入了注意力机制，但是其对下温度的拟合效果依然较差，MAE 达到了 1.019 0 °C，MSE 高达 4.055 8。标准 PatchTST 通过 Patch 分块机制将上和下温度的 MAE 分别降至 0.603 6、0.697 0 °C，整体已优于 LSTM 和 Vanilla Transformer；但是其 MSE 依然处于较高水平，且回流温度 MAE (0.059 8 °C) 仍明显高于 SGC-PatchTST 的 0.041 8 °C，这表明单纯的 Patch 分块不足以彻底解决多变量耦合和工况切换带来的拟合滞后问题。相比之下，SGC-PatchTST 通过 SOP 先验注入与解耦架构，将上、下温度的

表 2 多模型全局预测性能对比

Table 2 Comparison of global prediction performance among multiple models

变量	模型	MAE	MSE	R^2
上温度/°C	Linear	1.059 2	4.317 0	0.995 4
	CNN-1D	1.053 2	4.744 2	0.994 9
	LSTM	0.687 9	3.476 8	0.996 3
	Vanilla Transformer	0.867 3	3.434 4	0.996 3
	PatchTST	0.603 6	2.722 7	0.997 1
	SGC-PatchTST	0.362 0	2.521 2	0.997 3
下温度/°C	Linear	1.215 9	5.244 3	0.994 4
	CNN-1D	1.066 3	5.127 7	0.994 5
	LSTM	0.764 9	3.841 7	0.995 9
	Vanilla Transformer	1.019 0	4.055 8	0.995 6
	PatchTST	0.697 0	2.984 0	0.996 8
	SGC-PatchTST	0.347 0	2.176 0	0.997 7
罐内压力/MPa	Linear	0.000 6	—	—
	CNN-1D	0.000 6	—	—
	LSTM	0.000 6	—	—
	Vanilla Transformer	0.000 6	—	—
	PatchTST	0.000 6	—	—
	SGC-PatchTST	0.000 6	—	—
回流温度/°C	Linear	0.117 5	0.043 7	0.993 9
	CNN-1D	0.115 9	0.046 7	0.993 5
	LSTM	0.116 1	0.044 9	0.993 7
	Vanilla Transformer	0.100 7	0.038 8	0.994 6
	PatchTST	0.059 8	0.020 5	0.997 1
	SGC-PatchTST	0.041 8	0.011 6	0.998 4

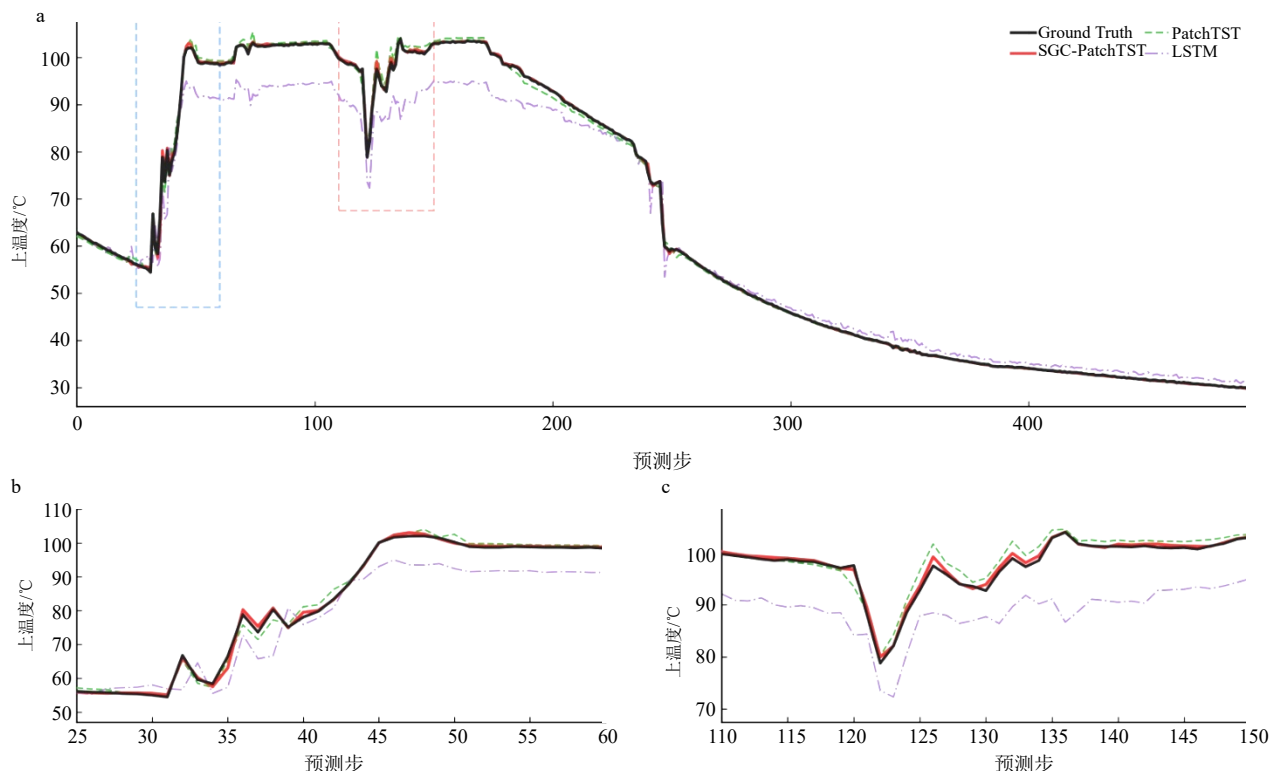
所有模型均在单次分层划分 (随机种子 42) 下训练评估，以确保对比的绝对公平性；罐内压力因绝对波动极小 ($\sigma \approx 0.003$ MPa)，MSE 与 R^2 失去评价意义，故不适用；“—”表示该项指标未列出。All models were trained and evaluated under a single stratified split (random seed = 42) to ensure absolute fairness in the comparison; due to the extremely small absolute fluctuation of the tank pressure ($\sigma \approx 0.003$ MPa), MSE and R^2 lose their evaluation significance and are therefore not applicable; “—” indicates that the corresponding metric is not reported.

MAE 分别大幅降至 0.362 0、0.347 0 °C，MSE 降至全组最低，决定系数 R^2 分别提升至 0.997 3 和 0.997 7。在受边界扰动明显的回流温度预测中，主副解耦架构把多变量干扰进行隔离。和端到端 MIMO 建模相比，独立副支路把回流温度 MAE 进一步降低到 0.041 8 °C， R^2 提高到 0.998 4。罐内压力方面，SGC-PatchTST 模型的 MAE 保持在 0.000 6 MPa 的极低水平，表现出极高的高保真监测精度，完全满足工业安全边界的监控需求。

5.3 动态极值追踪与相位滞后分析

为了从微观上定量分析各模型在复杂热力学工况变化下的瞬态响应性能,选取 50 批次分层测试集中极具代表性的批次 Z251163-1 (温度跨度 30~

104 °C, 涵盖加水急降与急剧升温 2 种典型非稳态工况) 的自回归推演结果进行局部放大对比 (图 3)。图中“预测步”表示预测结果的绘图评价点序号,相邻评价点通常间隔约 2 min,并非单次预测视距内的步号。



a-上温度全局预测轨迹; b-升温段局部放大 (预测步 25~60); c-温度急降及恢复段局部放大 (预测步 110~150)。

a-overall upper-temperature prediction trajectories; b-local magnification of temperature-rise segment (prediction steps 25—60); c-local magnification of temperature-drop and recovery segment (prediction steps 110—150).

图 3 上温度自回归多步推演多工况局部放大对比

Fig. 3 Locally magnified comparison of autoregressive multi-step inference for upper temperature under multiple operating conditions

图 3-a 给出了该批次完整的自回归推演轨迹。在全局尺度上,SGC-PatchTST 的预测曲线与真实物理轨迹几乎完全重合 (全批次 MAE=0.19 °C); 标准 PatchTST 在工况切换段出现了肉眼可辨的跟踪滞后 (全批次 MAE=0.56 °C), 而 LSTM 在升降温过程中产生了大幅偏离和振荡 (全批次 MAE=3.50 °C)。

图 3-b 展示了升温段的局部预测结果 (预测步 25~60, 共 36 个评价点)。真实上温度由区间起点的 56.12 °C 经短时波动上升, 在预测步 48 处达到 102.03 °C, 随后回落并维持在 98~100 °C。在该区间内,SGC-PatchTST、标准 PatchTST 和 LSTM 的 MAE 分别为 0.558、1.151、5.609 °C, 表明 SGC-PatchTST 在该升温段具有较小的平均预测误差。

图 3-c 展示了温度急降及恢复段的局部预测结果 (预测步 110~150, 共 41 个评价点)。真实上温度由区间起点的 99.75 °C 下降, 在预测步 122 处达到 78.80 °C, 随后经波动回升, 在预测步 136 处达到 103.91 °C。SGC-PatchTST、标准 PatchTST 和 LSTM 在该区间的 MAE 分别为 0.492、1.379、8.719 °C, 最大绝对误差分别为 1.729、4.211、17.243 °C。SGC-PatchTST 与标准 PatchTST 的预测谷值均位于预测步 122 处, 而 LSTM 的预测谷值出现在预测步 123 处, 且谷值温度低估。上述结果表明,SGC-PatchTST 在该温度急降及恢复区间具有较小的平均误差和最大偏差。

上述局部对比表明,SGC-PatchTST 在图 3-b 升温段和图 3-c 温度急降及恢复段的 MAE 均低于标

准 PatchTST 和 LSTM, 并在图 3-c 区间取得较小的最大绝对误差。这些结果体现了模型对不同动态温度变化过程的跟踪能力。各设计组件对预测性能的贡献将在后续消融实验中进一步分析。

5.4 局部低通滤波与抗噪鲁棒性分析

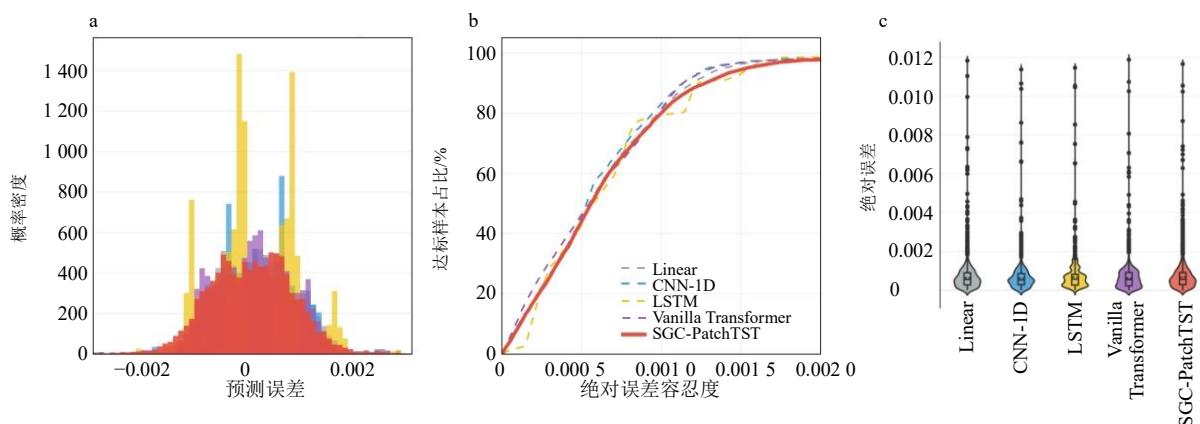
温度在长时间自然降温、恒温沸腾阶段时, 温度传感器采集的真实信号里含有较高频率的锯齿形噪声。传统线性模型对这类噪声较为敏感, 预测值会出现较大波动, 甚至出现单步超过 $1\text{ }^{\circ}\text{C}$ 的非物理跳变——在实际工业应用中, 若将带有高频噪声的预测信号直接用于底层执行机构的闭环控制, 会引发调节阀门频繁启停, 大幅增加硬件疲劳与设备故障风险。

对比之下, 本研究运用的 Patch 分块机制在将序列输入骨干网络之前, 先对局部 Token 进行聚合, 从

信号处理角度引入了隐式低通滤波效应, 使预测曲线呈现出以热惯性为主的平滑变化特征。结合上述分析, 极值敏感损失中 $\tau=1.5\sigma$ 的阈值设置进一步确保了高频随机噪声 ($Z_i < \tau$) 被排除在极值惩罚范围之外, 使模型在平稳生产阶段自动退化为标准 MSE 的平滑优化模式, 从而在“抗噪鲁棒性”与“极值敏感度”之间取得了较好的理论平衡。定量统计结果表明, 在恒温沸腾阶段, Linear 模型预测值的标准差达到真实值的 1.8 倍, 而 SGC-PatchTST 预测值的标准差与真实信号的宏观波动基本一致, 充分验证了 Patch 低通滤波与极值门限机制协同作用下的抗噪能力。

5.5 罐内压力稳态监测验证

罐内压力作为提取过程的核心安全边界参数, 对其预测性能的全面评估具有重要工程意义, 多维度评估结果如图 4 所示。



a-误差概率密度分布; b-回归误差特性曲线; c-绝对误差分布。

a-probability density distribution of error; b-regression error characteristic curve; c-distribution of absolute error.

图 4 罐内压力性能评估

Fig. 4 Performance evaluation of tank pressure

鉴于常压沸腾工况下罐内压力波动幅度极小, 单一的全局误差指标及传统折线图难以直观反映模型的局部跟踪性能。为此, 本研究引入误差概率密度分布、REC 曲线及绝对误差箱线图, 构建了多维度的综合评价体系。

可视化与定量评估结果表明, SGC-PatchTST 的整体预测误差极小 (MAE 仅为 0.0006 MPa)。由图 4-a 可知, 该模型的误差概率密度在零点附近呈现出高度集中的尖峰分布, 说明其预测值与真实轨迹高度贴合; 图 4-b 的 REC 曲线进一步显示, 在极其微小的绝对误差容忍度下, SGC-PatchTST 能最快收敛并达到极高的预测达标率, 展现出卓越的稳态跟踪性能。此外, 误差箱线图 (图 4-c) 显示, 传统

基线方法普遍存在明显的长尾效应 (即容易出现偶发的非物理突刺), 而 SGC-PatchTST 的误差分布极为紧凑, 对局部较大偏差实现了强效抑制。综上所述, SOP 先验意图的物理引导与 Patch 分块机制的低通滤波效应相辅相成, 切实提升了核心安全边界参数在微小波动工况下监测的高保真度与工程实用性。

5.6 消融实验分析

为验证 SOP 意图引导、主副解耦架构及极值敏感损失函数的有效性, 本研究设计了 6 路组合消融实验, 各变体定义如下。M0: 标准 PatchTST (无 SOP、无解耦、MSE 损失); M1: 在 M0 基础上加入 SOP 融合 (MSE); M2: 在 M1 基础上加入主副

解耦 (MSE); M3: 完整 SGC-PatchTST (在 M2 基础上替换为极值损失, 优化超参数 $\tau=1.5\sigma$, $\lambda=2.0$); M4: 在 M3 基础上去掉 SOP 融合 (仅保留

解耦与极值损失); M5: 在 M3 基础上去掉主副解耦 (仅保留 SOP 与极值损失)。具体对比结果如表 3 所示。

表 3 六路递进消融实验结果

Table 3 Results of six-step progressive ablation study

变量	指标	M0	M1	M2	M3	M4	M5
上温度/℃	MAE	0.603 6	0.449 4	0.403 3	0.362 0	0.371 7	0.437 2
	MSE	2.722 7	2.569 5	2.590 5	2.521 2	2.682 4	3.044 5
	R^2	0.997 1	0.997 2	0.997 2	0.997 3	0.997 1	0.996 7
	P95 error	1.645 5	1.399 4	1.435 2	1.314 2	1.455 4	1.569 2
下温度/℃	MAE	0.697 0	0.479 4	0.389 6	0.347 0	0.363 5	0.441 7
	MSE	2.984 0	2.470 1	2.330 1	2.176 0	2.330 5	2.773 2
	R^2	0.996 8	0.997 3	0.997 5	0.997 7	0.997 5	0.997 0
	P95 error	1.805 4	1.408 8	1.514 6	1.332 1	1.470 4	1.552 6
罐内压力/MPa	MAE	0.000 6	0.000 6	0.000 6	0.000 6	0.000 6	0.000 6
回流温度/℃	MAE	0.059 8	0.057 2	0.052 6	0.041 8	0.054 5	0.047 1
	MSE	0.020 5	0.019 1	0.017 3	0.011 6	0.018 2	0.014 4
	R^2	0.997 1	0.997 2	0.997 5	0.998 4	0.997 4	0.997 8

所有模型均在单次分层划分 (随机种子 42) 下训练评估, 以确保对比的绝对公平性; 罐内压力因绝对波动极小 ($\sigma \approx 0.003$ MPa), MSE 与 R^2 失去评价意义, 故不适用。粗体表示各温度变量对应指标的最优结果, 其中 MAE、MSE 和 P95 error 越小越好, R^2 越大越好; 罐内压力 MAE 在所示精度下相同, 故不作粗体区分。

All models were trained and evaluated under a single stratified split (random seed=42) to ensure absolute fairness in the comparison; due to the extremely small absolute fluctuation of the tank pressure (0.003 MPa), MSE and R^2 lose their evaluation significance and are therefore not applicable; Bold values indicate the best result for each metric of the temperature variables. Lower MAE, MSE, and P95 error and higher R^2 indicate better performance; tank-pressure MAE values are identical at the reported precision and are therefore not distinguished in bold.

由表 3 的量化结果可见, 完整架构 SGC-PatchTST (M3) 的上、下温度 MAE 分别降至 0.362 0 °C 和 0.347 0 °C, 且代表抗最大偏差能力的 P95 error 分别大幅降至 1.314 2 °C 和 1.332 1 °C, 在所有变体中表现最佳。

深入对比各核心组件的独立贡献发现: 移除 SOP 意图引导后 (M4), 不仅上温度 MAE 发生反弹 (较 M3 升高 2.7%), 其上、下温度的 P95 error 更是显著升高至 1.455 4、1.470 4 °C。这从微观残差层面强有力地验证了 SOP 前瞻性物理开关在极值损失协同下, 对于捕捉工况突变、抑制极端预测误差的不可替代性; 而移除主副解耦后 (M5), 上温度 MAE 显著升至 0.437 2 °C (升幅约 20.8%), P95 error 升至 1.569 2 °C, 证实了解耦架构在削弱多变量热力学耦合干扰方面具有关键作用。此外, 各改进组件之间展现出了显著的协同效应, 特别是在受外部干扰明显的回流温度预测中, 传统的直接拼接法 M1 带来的提升极小, 只有当“SOP 门控”与“双支路解耦”同时存在时 (M3), MAE 才断崖式下降至 0.041 8 °C。

为验证核心模型的统计稳健性, SGC-PatchTST 在 3 次独立分层划分下的上、下温度 MAE 均值分别保持在 0.33 °C (变异系数=6.8%) 和 0.32 °C (变异系数=6.4%), 与单次测评结果高度一致, 充分佐证了完整架构在小样本工业数据下依然具备显著的综合优势与可靠性。

5.7 超参数敏感性分析

极值敏感损失函数中 2 个关键超参数极值触发阈值 τ 和极值惩罚系数 λ 的选择直接影响到极值拟合与抗噪性能之间的平衡。本研究以网格搜索法分析 $\tau \in \{0.8\sigma, 1.0\sigma, 1.2\sigma, 1.5\sigma, 2.0\sigma\}$ 和 $\lambda \in \{1.0, 2.0, 3.0, 5.0, 10.0\}$ 共 25 种组合对上温度预测性能的影响, 每种组合训练 20 epoch 后评估验证集 MaxError 和 MAE。结果如表 4 所示。所有 25 组参数组合的 MaxError 均落在 6.31~6.36 的极窄区间内, 变异系数仅为 0.15%, 说明模型对 τ 和 λ 的微小变动不敏感, 超参数的选择具有极好的鲁棒性。当选取 $\tau=1.5\sigma$, $\lambda=2.0$ 时, MaxError 为 6.332 0, MAE 为 0.041 6, 处于误差曲面的平滑低谷区。当 λ 超过 5 时 MAE 开始明显上升 (如 $\tau=1.5\sigma$, $\lambda=5.0$ 时 MAE

表 4 超参数网格搜索结果
Table 4 Results of hyperparameter grid search

τ	指标	$\lambda=1.0$	$\lambda=2.0$	$\lambda=3.0$	$\lambda=5.0$	$\lambda=10.0$
0.8 σ	MaxError	6.318 0	6.326 0	6.331 0	6.316 0	6.336 0
	MAE	0.040 2	0.040 3	0.042 4	0.041 7	0.043 2
1.0 σ	MaxError	6.314 0	6.339 0	6.327 0	6.336 0	6.340 0
	MAE	0.041 0	0.041 7	0.041 5	0.043 2	0.044 3
1.2 σ	MaxError	6.325 0	6.327 0	6.343 0	6.355 0	6.337 0
	MAE	0.040 2	0.042 7	0.042 4	0.045 2	0.046 4
1.5 σ	MaxError	6.332 0	6.332 0	6.347 0	6.345 0	6.340 0
	MAE	0.040 9	0.041 6	0.045 6	0.054 4	0.047 8
2.0 σ	MaxError	6.329 0	6.326 0	6.331 0	6.328 0	6.338 0
	MAE	0.040 1	0.041 7	0.042 4	0.044 5	0.050 1

MaxError 与 MAE 为模型在归一化尺度（均值 0、标准差 1）下的相对误差量纲，旨在评估超参数敏感性的相对差异，与表 2、3 中的真实温度误差不具有直接数值可比性；经逆归一化后，此处归一化误差与物理温度误差的换算比例为 1 : 8~1 : 10（即归一化误差 1.0 $\approx$ 物理温度误差 8~10 $^{\circ}\text{C}$ ，具体取决于当前批次的温度标准差）。

MaxError and MAE represent relative error metrics on a normalized scale (mean = 0, standard deviation = 1) aiming to evaluate the relative differences in hyperparameter sensitivity and they are not directly numerically comparable to the actual physical temperature errors reported in tables 2 and 3; after inverse normalization, the conversion ratio between the normalized error and the physical temperature error is approximately 1 : 8—1 : 10 (i.e., a normalized error of 1.0 $\approx$ a physical temperature error of 8—10 $^{\circ}\text{C}$, depending on the temperature standard deviation of the specific batch).

达 0.054 4)，表明过大的极值惩罚系数会造成过拟合从而损害整体拟合质量；而当 $\tau=2.0\sigma$ 时，MAE 普遍高于 $\tau=1.0\sigma\sim1.5\sigma$ ，说明过高的触发阈值可能导致模型遗漏真实的工艺突变。综合考量极值拟合与全局精度的平衡，本研究最终确定 $\tau=1.5\sigma$ ， $\lambda=2.0$ 为最优超参数组合。

6 结论

针对中药提取过程中存在的时延大、多变量强耦合和高频噪声干扰等难题，本研究提出了一种结合 SOP 工艺先验的中药提取过程关键工艺参数解耦数字孪生建模方法 SGC-PatchTST。该方法在架构上通过正交时空解耦与后期门控融合，有效克服了离散指令与连续特征生硬拼接所产生的特征竞争，减轻了工况动态切换导致的预测滞后。同时，主副双支路解耦模式实现了核心热力学参数与受外界扰动变量的物理隔离，避免了多变量间的负迁移干扰，进而大幅提升了回流温度等关键参数的预测精度。此外，极值敏感混合损失函数与 Patch 分块机制相协同，在过滤传感器高频锯齿噪声的同时，增强了模型对真实工艺突变点的拟合敏感度，保障了全局预测轨迹的平稳性。基于车间真实批次数据的量化评估结果表明，该模型不仅将核心上、下温度的单次评估 MAE 分别降低至 0.36、0.35 $^{\circ}\text{C}$ （较基准 PatchTST 分别降低 40.0%和 50.2%），并在多次分层验证中分别保持了 0.33、0.32 $^{\circ}\text{C}$ 的较高水

平稳健均值；同时，得益于解耦架构的抗干扰能力，模型将回流温度 MAE 降至较低的 0.04 $^{\circ}\text{C}$ ，更在罐内压力的微小波动监测中保持了 0.000 6 MPa 的较高精度，展现出良好的多变量工程监测稳健性。综上所述，SGC-PatchTST 模型在预测准确性与动态跟踪时效性上均达到了工业级高保真数字孪生环境的构建要求，具有较强的物理机制合理性，可为后续引入多智能体前沿技术、探索深度强化学习与模型预测控制相结合的无人化与智能化闭环控制体系提供坚实的模型支撑^[13-18]。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 肖伟, 张欣, 徐芳芳, 等. 以功效物质为质量目标的中药生产过程质量控制研究与实践 [J]. 南京中医药大学学报, 2024, 40(10): 1030-1038.

[2] 管咏梅, 刘乐灵, 杨明, 等. 基于过程分析技术的中药提取工艺放大的关键问题与应用进展 [J]. 中草药, 2026, 57(2): 767-777.

[3] 陶飞, 张贺, 张辰源, 等. 从数字孪生到数智孪生: 回顾与展望 [J]. 计算机集成制造系统, 2026, 32(1): 1-17.

[4] 陶飞, 刘蔚然, 张萌, 等. 数字孪生五维模型及十大领域应用 [J]. 计算机集成制造系统, 2019, 25(1): 1-18.

[5] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting [J]. Proc AAAI Conf Artif Intell, 2021, 35(12): 11106-11115.

- [6] 陈嘉俊, 刘波, 林伟伟, 等. 基于 Transformer 的时间序列预测方法综述 [J]. 计算机科学, 2025, 52(6): 96-105.
- [7] 鱼欢欢, 于慧铖, 伍振峰, 等. 基于数智技术的中药固体制剂在线检测与动态评价发展趋势 [J]. 中草药, 2026, 57(13): 4949-4963.
- [8] Nie Y Q, Nguyen N H, Sinthong P, *et al.* A time series is worth 64 words: Long-term forecasting with transformers [EB/OL]. *arXiv* (2022-11-27) [2026-09-20]. <https://arxiv.org/abs/2211.14730>
- [9] Huang X Y, Tang J, Shen Y M. Long time series of ocean wave prediction based on PatchTST model [J]. *Ocean Eng*, 2024, 301: 117572.
- [10] Karniadakis G E, Kevrekidis I G, Lu L, *et al.* Physics-informed machine learning [J]. *Nat Rev Phys*, 2021, 3(6): 422-440.
- [11] 姜翠, 焦建彬. 融合知识先验的车辆轨迹预测深度学习算法 [J]. 中国科学院大学学报, 2026, 43(2): 209-217.
- [12] 张橙, 李雪, 叶茂, 等. 物理信息神经网络在两相流中的应用 [J]. 化工学报, 2024, 75(11): 3835-3856.
- [13] 张颖伟, 高鸿瑞, 张鼎森, 等. 基于多智能体的数字孪生及其在工业中应用的综述 [J]. 控制与决策, 2023, 38(8): 2168-2182.
- [14] Spielberg S, Tulsyan A, Lawrence N P, *et al.* Toward self-driving processes: A deep reinforcement learning approach to control [J]. *AIChE J*, 2019, 65(10): e16689.
- [15] Li H R, Qiu T, You F Q. AI-based optimal control of fed-batch biopharmaceutical process leveraging deep reinforcement learning [J]. *Chem Eng Sci*, 2024, 292: 119990.
- [16] Sutton R S, Barto A G. *Reinforcement Learning* [M]. 2nd ed. Cambridge, MA: MIT Press, 2018: 159-165.
- [17] Wang Z M, Gu C Y, Chen B, *et al.* Model predictive control for continuous pharmaceutical manufacturing with mass retention constraints [J]. *Comput Chem Eng*, 2025, 203: 109332.
- [18] Abou El Qassime M, Shokry A, España A, *et al.* Development of approximate scheduling-adaptive controllers for multi-products continuous chemical processes using deep learning techniques and model predictive control [J]. *Comput Chem Eng*, 2026, 204: 109359.

[责任编辑 潘明佳]