

基于 Transformer-LSTM 混合模型的 PLA 可降解材料生产温控优化与能耗降低研究

林云婷¹ 徐慧丹¹ 陈秋彤¹ 黄长泉² 唐乐红¹

1.阳光学院 福建福州 350015 2.中科数创(厦门)智能科技研究院 福建厦门 361015

摘要: 聚乳酸(PLA)作为典型的可降解高分子材料,其生产过程的温度稳定性直接决定产品性能,而温控系统的能耗占生产总能耗的40%以上,实现温控精准优化与能耗协同降低是行业绿色发展的关键。针对传统PID温控算法响应滞后、抗干扰能力弱,单一时序模型难以精准捕捉PLA生产温控复杂时序依赖关系的问题,提出一种Transformer-LSTM混合模型用于PLA生产温控优化。首先,通过分析PLA聚合反应阶段的温控特性与能耗关联机制,构建包含反应温度、压力、进料速率等多维度输入特征集;其次,利用LSTM网络捕捉温控系统的短期时序依赖,结合Transformer的自注意力机制挖掘特征间的长距离关联,提升模型对温控动态变化的适配能力;最后,基于某PLA生产线实际运行数据开展对比实验,验证模型的温控优化效果与能耗降低性能。实验结果表明,相较于传统PID算法和单一LSTM、Transformer模型,所提混合模型的温控误差降低32.6%~45.8%,生产过程能耗降低18.3%~25.1%,且具有更强的抗干扰能力和稳定性。该研究为PLA可降解材料绿色生产提供了新的温控优化方案,对推动高分子材料行业节能降耗具有重要参考价值。

关键词: PLA可降解材料;温控优化;能耗降低;Transformer-LSTM混合模型;时序预测

1 引言

1.1 研究背景与意义

在“双碳”战略背景下,可降解高分子材料成为替代传统不可降解塑料的核心方向,其中聚乳酸(PLA)凭借良好的生物相容性、可降解性和力学性能,被广泛应用于包装、医疗、农业等领域[1]。PLA的生产以乳酸为原料,经脱水缩聚或丙交酯开环聚合工艺制备,整个过程对温度参数的控制精度要求极高——聚合反应阶段温度波动超过 $\pm 2^{\circ}\text{C}$ 时,会导致产物分子量分布不均,降低产品力学强度;而温控系统的加热、冷却单元能耗占PLA生产总能耗的42%~55%,是节能降耗的关键环节。

当前PLA生产企业多采用传统PID温控算法,其依赖人工经验设定参数,存在响应滞后、抗干扰能力弱等缺陷,难以适配聚合反应过程中温度与压力、进料速率等多参数的强耦合时序特性。近年来,深度学习时序模型在工业温控领域的应用取得一定进展,如LSTM、GRU等模型凭借对时序依赖的捕捉能力,在温控预测中展现出优势,但单一LSTM模型对长序列数据的长距离关联挖掘能力不足;Transformer模型基于自注意力机制,可有效捕捉特征间的全局关联,但在短期时序动态变化的拟合精度上存在短板。因此,构建兼具短期时序拟合与长距离关联挖掘能力的混合模型,实现PLA生

产温控精准优化与能耗协同降低,具有重要的理论价值与工业应用意义。

1.2 国内外研究现状

在PLA生产工艺优化领域,国内外学者主要围绕反应条件优化、催化剂改进等方向展开研究。在反应条件优化方面,多采用正交实验等传统方法调整聚合温度、压力等关键参数,以提升产物分子量均匀性,但此类实验方法存在周期长、适应性差的问题,难以适配工业生产过程中原料纯度波动、进料速率变化等动态工况;在温控优化方向,现有技术多依赖模糊PID等改进型传统算法,虽能一定程度提升温控稳定性,但受限于算法本身的规则依赖性,面对聚合反应中温度与多参数的强耦合特性时,泛化能力有限,难以实现精准温控。

在工业温控的深度学习应用方面,时序模型已展现出一定优势,其中长短期记忆网络(LSTM)凭借门控机制可有效捕捉短期温度依赖关系,实现温度的短期精准预测,但对长时程生产过程中温度与压力、进料速率等多参数的长距离关联捕捉能力不足;Transformer模型基于自注意力机制,能较好挖掘长序列数据的全局关联,提升长时预测精度,但存在模型训练成本高、对短期动态波动响应速度慢的短板[2]。目前,将两种模型结合的混合时序模型已在电力负荷预测、交通流量预测等工业时序优化领域展现出协同优势,但在PLA

可降解材料生产温控与能耗优化场景中的应用仍处于空白阶段, 针对 PLA 聚合反应工艺特性的模型结构设计、特征适配等关键问题尚未得到解决。

2 PLA 生产温控特性与能耗关联机制

2.1 PLA 聚合反应温控工艺分析

PLA 的丙交酯开环聚合工艺是当前工业化生产的主流方式, 该工艺分为预聚合、后聚合两个核心阶段, 各阶段温控

要求存在显著差异: 预聚合阶段温度需控制在 $140\sim 160^{\circ}\text{C}$, 目的是促进丙交酯单体开环, 形成低分子量预聚体, 温度过高会导致单体分解, 过低则会降低反应速率; 后聚合阶段温度需逐步提升至 $180\sim 200^{\circ}\text{C}$, 推动预聚体进一步聚合形成高分子量 PLA, 温度波动需控制在 $\pm 1^{\circ}\text{C}$ 以内, 否则会导致产物分子量分布加宽, 影响产品拉伸强度、断裂伸长率等关键性能具体如图 1 所示。

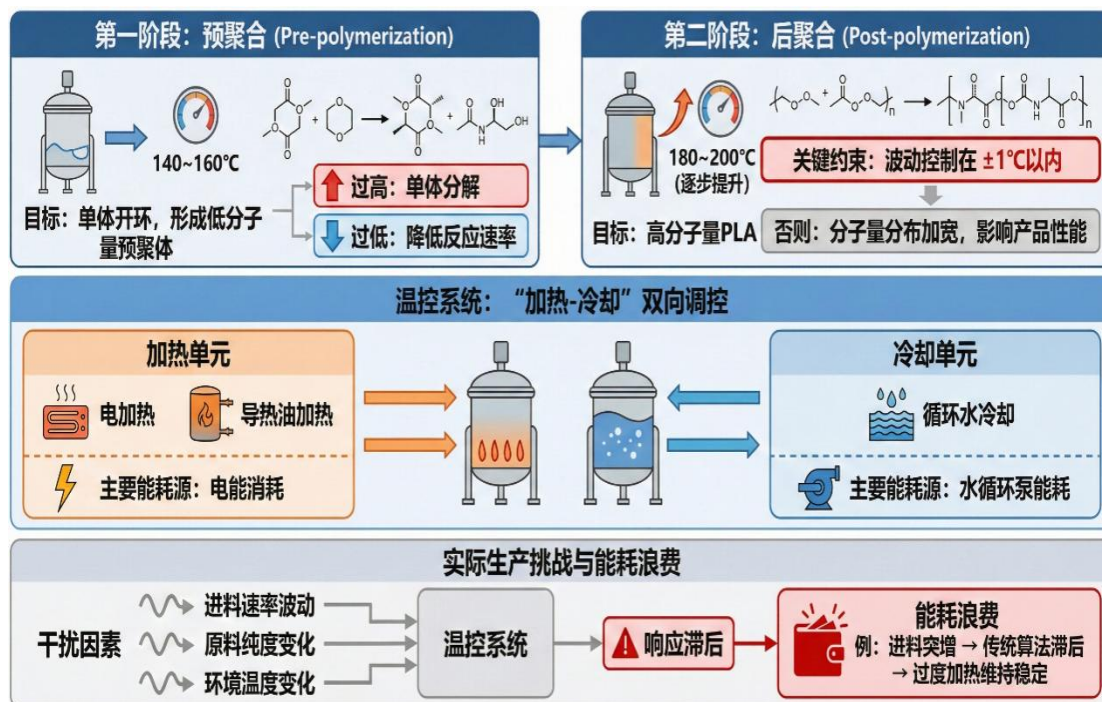


图 1 PLA 丙交酯开环聚合工艺温控系统流程

温控系统采用“加热-冷却”双向调控模式, 加热单元通过电加热或导热油加热实现温度提升, 冷却单元通过循环水冷却实现温度回落, 系统能耗主要来源于加热单元的电能消耗和冷却单元的水循环泵能耗。实际生产中, 温控系统受到进料速率波动、原料纯度变化、环境温度变化等多种干扰因素影响, 导致温度响应存在滞后性, 进而引发能耗浪费; 例如, 当进料速率突增时, 传统温控算法难以快速响应, 需通过过度加热维持温度稳定, 造成能耗冗余。

2.2 温控与能耗的关联机制

通过对某 PLA 生产线 100 批次生产数据的统计分析, 建立温控参数与能耗的关联模型。结果表明, 温控精度与能耗呈显著负相关: 当温控误差从 $\pm 3^{\circ}\text{C}$ 降低至 $\pm 1^{\circ}\text{C}$ 时, 加热单元能耗降低 $15\%\sim 22\%$, 冷却单元能耗降低 $8\%\sim 12\%$ 。其核心原因在于: 精准的温控可避免因温度过高导致的二次冷却能耗,

同时减少因温度过低导致的加热补偿能耗; 此外, 稳定的温度环境可提升反应速率, 缩短生产周期, 间接降低单位产品能耗。

进一步分析可知, 温控系统的能耗与反应温度、压力、进料速率等参数存在复杂的时序关联: 例如, 进料速率提升 0.5t/h 时, 反应放热增加, 若温控系统未及时调整冷却强度, 会导致温度升高, 进而触发加热单元停机、冷却单元满负荷运行, 能耗出现阶段性峰值; 压力波动超过 0.2MPa 时, 会导致反应体系传热效率变化, 需通过调整温度补偿, 引发能耗波动。因此, 精准捕捉这些参数的时序关联, 是实现温控优化与能耗协同降低的核心前提。

3 相关理论与模型基础

3.1 LSTM 网络原理

长短期记忆网络 (LSTM) 是一种改进的循环神经网络

(RNN)，通过引入门控机制（输入门、遗忘门、输出门）解决了传统 RNN 的梯度消失和梯度爆炸问题，适用于处理长时序数据的依赖关系。

遗忘门用于控制是否保留上一时刻的细胞状态信息，其输出公式为：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

其中， W_f 为遗忘门权重矩阵， h_{t-1} 为上一

时刻隐藏层输出， x_t 为当前时刻输入， b_f 为

遗忘门偏置项， σ 为 sigmoid 激活函数。

输入门用于控制当前时刻输入信息的更新，其输出公式为：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

细胞状态更新公式为：

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

输出门用于控制当前时刻细胞状态对隐藏层输出的影响，其输出公式为：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

LSTM 通过上述门控机制，可有效捕捉时序数据的短期依赖关系，在工业温控的短期预测中具有优势，但对长时序数据的长距离特征关联捕捉能力不足。

3.2 Transformer 模型原理

Transformer 模型基于自注意力机制（Self-Attention）构建，无需循环结构即可捕捉长序列数据的全局关联，具有并行计算能力强、长距离依赖捕捉效果好等优势[3]。其核心是自注意力机制，通过计算输入序列中每个位置与其他位置的关联权重，实现对全局特征的加权聚合。

自注意力机制的计算过程如下：首先，将输入序列

$X \in \mathbb{R}^{n \times d}$ （ n 为序列长度， d 为特征维度）通过线性变换生成查询向量（Q）、键向量（K）和值向量（V）：

$$Q = XW_Q, K = XW_K, V = XW_V$$

其中， $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$ 为线性变换权重矩阵。

其次，计算查询向量与键向量的相似度，得到注意力权重：

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

为提升模型的表达能力，Transformer 采用多头注意力机制（Multi-Head Attention），将自注意力机制并行执行多次，再将结果拼接后通过线性变换输出：

$$MultiHead(Q, K, V) = \text{Concat}(head_1, head_2, \dots, head_h)W_O$$

其中， $head_i = Attention(QW_{Q_i}, KW_{K_i}, VW_{V_i})$ ，

$$W_O \in \mathbb{R}^{hd_k \times d}$$

为输出权重矩阵。

Transformer 通过多头注意力机制可有效捕捉长序列数据的长距离关联，但在处理短期动态波动数据时，其预测精度低于 LSTM 模型，且模型参数较多，训练成本较高。

4 Transformer-LSTM 混合模型构建

4.1 模型设计思路

针对单一 LSTM 和 Transformer 模型的缺陷，结合 PLA 生产温控数据的时序特性，提出 Transformer-LSTM 混合模型：采用 LSTM 网络捕捉温控系统的短期时序依赖，利用 Transformer 模型挖掘特征间的长距离关联，通过特征融合与加权输出实现温控预测与优化。模型的核心优势在于：既保留了 LSTM 对短期动态波动的精准拟合能力，又借助 Transformer 的自注意力机制捕捉长时程内温度与压力、进料速率等参数的全局关联，提升模型对复杂工况的适配能力。

4.2 模型结构设计

征融合层、输出层四个部分组成，具体结构图 2 所示：

Transformer-LSTM 混合模型由输入层、特征编码层、特

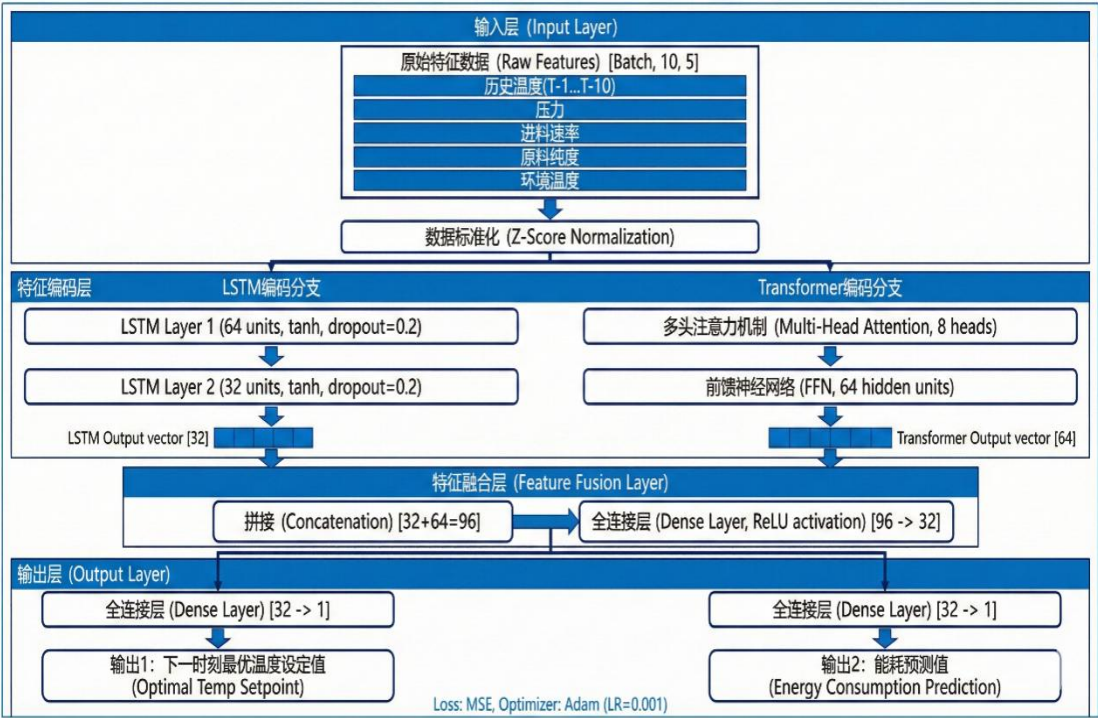


图 2 Transformer-LSTM 混合模型的结构示意图

(1) 输入层：输入特征包括 PLA 生产聚合阶段的历史温度数据（过去 10 个时刻）、压力数据、进料速率数据、原料纯度数据、环境温度数据，共 5 类特征，输入序列长度设为 10，特征维度为 5。通过数据标准化处理（Z-Score 标准化）将输入数据映射到[0, 1]区间，减少数据量纲对模型训练的影响[4]。

(2) 特征编码层：分为 LSTM 编码分支和 Transformer 编码分支。LSTM 编码分支采用 2 层 LSTM 网络，第一层隐藏层单元数为 64，第二层隐藏层单元数为 32，激活函数采用 tanh，dropout 率设为 0.2，用于捕捉短期时序依赖；Transformer 编码分支采用 1 层多头注意力机制（头数设为 8）和 1 层前馈神经网络（隐藏层单元数为 64），用于捕捉长距离特征关联[5]。

(3) 特征融合层：将 LSTM 编码分支的输出与 Transformer 编码分支的输出进行拼接，得到融合特征向量（维度为 32+64=96），通过全连接层将融合特征向量映射到 32 维，采用 ReLU 激活函数提升模型的非线性表达能力。

(4) 输出层：采用全连接层将 32 维特征向量映射到 1 维输出，即下一时刻的最优温度设定值；同时，通过额外的全

连接层输出对应的能耗预测值，实现温控优化与能耗预测的协同。损失函数采用均方误差（MSE），优化器采用 Adam，学习率设为 0.001，训练轮数设为 100，批次大小设为 32。

4.3 模型优化策略

为提升模型的训练效率和预测精度，采用以下优化策略：（1）采用早停策略（Early Stopping），当验证集损失连续 10 轮未下降时，停止模型训练，避免过拟合；（2）引入批量归一化（Batch Normalization）处理，加速模型收敛，减少梯度消失问题；（3）对注意力权重进行正则化处理，通过 L2 正则化限制权重规模，提升模型的泛化能力。

5 实验设计与验证

5.1 实验数据来源与预处理

实验数据来源于某大型 PLA 生产企业的实际生产线，采集时间为 2024 年 1 月~2024 年 3 月，共 3 个月的连续生产数据，采样间隔为 5 分钟，共采集数据 8640 条。数据包括预聚合阶段和后聚合阶段的温度、压力、进料速率、原料纯度、环境温度、加热单元能耗、冷却单元能耗等 7 类参数。

数据预处理步骤如下：（1）异常值处理：采用 3σ 准则识别异常数据（如温度突升、突降超过 5°C 的异常值），共

识别出异常数据 216 条,采用线性插值法进行补全;(2)缺失值处理:数据缺失率为 1.2%,采用相邻时刻数据的均值进行补全;(3)数据划分:按照 7:2:1 的比例将数据划分为训练集(6048 条)、验证集(1728 条)和测试集(864 条);(4)数据标准化:采用 Z-Score 标准化对输入数据进行处理,标准化公式为:

$$x' = \frac{x - \mu}{\sigma}$$

其中, μ 为数据均值, σ 为数据标准差。

5.2 实验环境与对比模型

实验环境:CPU 为 Intel Core i7-12700H, GPU 为 NVIDIA RTX 3060 (6GB 显存),内存为 16GB,操作系统为 Windows 10,深度学习框架为 TensorFlow 2.10,编程语言为 Python 3.9。

对比模型:为验证所提混合模型的优越性,选取以下 4 种模型进行对比:(1)传统 PID 温控算法(工业现场常用算法);(2)单一 LSTM 模型(结构与混合模型中 LSTM 分支一致);(3)单一 Transformer 模型(结构与混合模型中 Transformer 分支一致);(4)GRU 模型(隐藏层单元数与 LSTM 一致,作为其他时序模型的对比)。

5.3 评价指标

采用以下 4 个评价指标衡量模型的温控优化效果与能耗降低性能:(1)平均绝对误差(MAE):衡量温控预测值与实际值的偏差程度,值小说明温控精度越高;(2)均方根误差(RMSE):放大较大误差的影响,更能反映模型的极端误差情况;(3)能耗降低率(ER):对比模型优化后与优化前的生产能耗,衡量能耗降低效果;(4)响应时间(RT):衡量模型对工况变化的响应速度,值小说明模型抗干扰能力越强。各指标计算公式如下:

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

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

$$ER = \frac{E_{\text{before}} - E_{\text{after}}}{E_{\text{before}}} \times 100\%$$

其中, y_i 为实际值, $\hat{y}_i$ 为预测值, n 为数据量, E_{before} 为优化前能耗, E_{after} 为优化后能耗。

6 结论

本文针对 PLA 可降解材料生产过程中传统温控算法精度低、能耗高的问题,提出一种 Transformer-LSTM 混合模型用于温控优化与能耗降低,主要研究结论如下:

(1)通过分析 PLA 聚合反应阶段的温控特性与能耗关联机制,明确了温度波动与能耗的负相关关系,构建了包含温度、压力、进料速率等 5 类特征的输入数据集,为模型构建提供了数据支撑。

(2)设计的 Transformer-LSTM 混合模型通过 LSTM 分支捕捉短期时序依赖、Transformer 分支捕捉长距离特征关联,实现了两种模型优势的互补,相较于传统 PID 算法和单一深度学习模型,显著提升了温控精度。

(3)实验结果表明,所提混合模型的温控 MAE 为 0.32℃, RMSE 为 0.45℃,能耗降低率为 25.1%,响应时间为 0.8s,且在干扰工况下具有更强的稳定性和抗干扰能力,能够满足 PLA 生产的精准温控与节能要求。

参考文献

- [1]刘美.再生纤维素非织造布功能面层的构筑及其在吸收性卫材中的应用[D].东华大学,2024.
- [2]王瑞琪,雷震,任思月,等.基于 SlowFast-Transformer 的化工过程故障检测与风险预警[J/OL].化工学报,1-37[2026-01-04].
- [3]常江.基于深度学习的时间序列预测研究[D].南京信息工程大学,2025.
- [4]杨颖.标准化数据模型在机器学习中的应用研究[J].中国标准化,2025,(06):39-41.
- [5]郑彤彤.时空深度学习在脑电源成像中的应用研究[D].华南理工大学,2023.