

# 基于LSTM-Attention的卷烟产能预测研究

王 志, 张金珠\*, 白佳康

河北工业大学理学院, 天津

收稿日期: 2026年1月12日; 录用日期: 2026年1月26日; 发布日期: 2026年1月31日

---

## 摘要

针对智能制造产能预测中数据层级复杂的问题,采用某卷烟厂四年的多层级生产数据,经系统化预处理后确定最优建模粒度,并设计了基于班次分组的线性插值重构方法解决时序不连续性问题。在此基础上,提出了融合长短期记忆(long short-term memory, LSTM)与注意力机制的轻量化预测框架,并引入早停策略防止过拟合。实验结果表明,模型在预测准确性和稳定性方面优于对比方法,远超传统的经验计算方式。同时深入分析了多步预测误差累积、泛化能力下降等固有局限性的根本原因,并提出了系统性的改进策略。结论表明,LSTM-Attention融合框架在保证预测精度的同时,具备良好的工程适用性,为智能制造环境下的生产优化决策提供了重要参考。

---

## 关键词

工业产能预测, LSTM, 注意力机制, 轻量化模型, 时序预测

---

# Research on Industrial Productivity Prediction Based on Dynamic Attention Mechanism

Zhi Wang, Jinzhu Zhang\*, Jiakang Bai

School of Sciences, Hebei University of Technology, Tianjin

Received: January 12, 2026; accepted: January 26, 2026; published: January 31, 2026

---

## Abstract

To address data hierarchy complexity and computational redundancy in smart manufacturing capacity prediction, multi-level production data from a cigarette factory is preprocessed to determine optimal modeling granularity. A shift-grouping method with linear interpolation resolves time-

\*通讯作者。

**series discontinuity. A lightweight forecasting framework integrating Long Short-Term Memory (LSTM) and an attention mechanism are proposed, incorporating early stopping to prevent overfitting. Results demonstrate superior prediction accuracy and stability over benchmarks, significantly outperforming traditional empirical methods. Fundamental causes of limitations—including multi-step error accumulation and reduced generalization—were analyzed. Systematic improvement strategies are formulated based on this analysis. The LSTM-Attention framework ensures prediction precision while offering strong engineering applicability, providing vital reference for production optimization in smart manufacturing.**

## Keywords

**Industrial Capacity Prediction, Long Short-Term Memory (LSTM), Attention Mechanism, Light Weight Model, Time-Series Forecasting**

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在全球范围内推进工业 4.0、智能制造等国家战略的背景下，生产计划与控制成为制造业的核心环节 [1]，其效能依赖于准确的产能预测。现有产能预测研究多集中于能源、化工等流程工业[2]-[4]，对离散制造业关注相对不足[5] [6]。在烟草行业，现有智能化研究多偏向于制丝工艺优化与质量预测[7] [8]，对卷包机等核心设备的产能预测研究较少。卷烟行业产能预测面临三大特殊挑战：1) 多品牌频繁切换，单台卷包机需生产数十种牌号，每次切换导致产能剧烈波动；2) 时序数据非连续性，受订单驱动影响，生产活动呈“班次 - 机台 - 牌号”的碎片化结构，形成业务逻辑驱动的结构性缺失，这与标准时序数据集的随机缺失模式存在本质区别；3) 设备老化与原料差异，设备性能衰减和烟叶特性波动进一步增加预测复杂性。这些特征使得通用时序预测方法难以直接应用。近年来，深度学习方法在时序预测领域取得进展，但也面临挑战[9]：LSTM 等循环网络擅长捕捉时序依赖但难以处理长期模式；Transformer、Informer 等注意力模型在长序列上表现优异，但参数量大、计算复杂度高，在工业小样本场景和边缘设备部署时受限。因此，需要在预测精度与计算效率间取得平衡的方法。

针对上述问题，本研究提出一种适配卷烟行业的轻量化 LSTM-Attention 预测框架。主要贡献包括：

- 1) 提出基于班次分组的时序重构方法，通过业务驱动的分段线性插值解决数据非连续性问题，并量化分析了插值误差对预测性能的影响；
- 2) 设计轻量化 LSTM-Attention 模型，通过精简的注意力机制在保持预测精度的同时降低计算复杂度，适配工业现场实时预测需求；
- 3) 通过与 LSTM 等多种方法对比，以及系统化消融实验，验证了方法的有效性。基于某卷烟厂四年真实数据的实验表明，单步预测 MAPE 可控制在 7.36%，相比标准 LSTM 提升 5.0%，为生产计划制定、原料采购和设备维护提供数据支持。

## 2. 基本原理

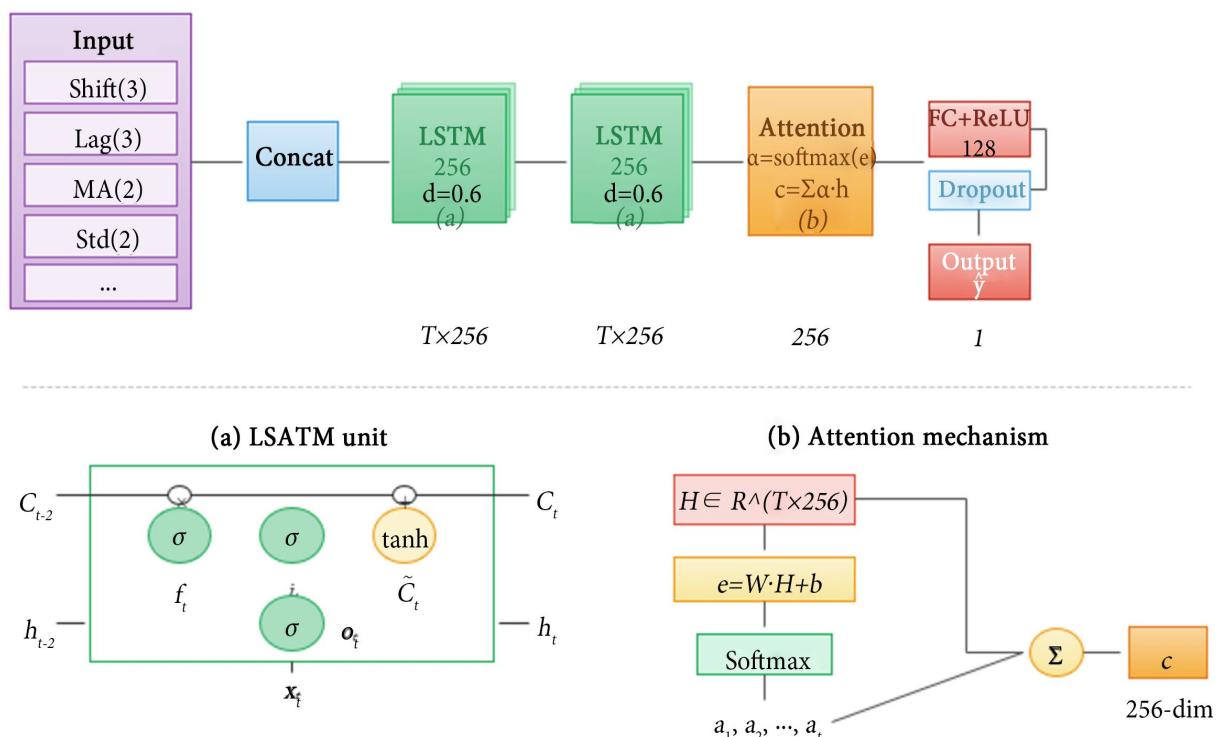
### 2.1. LSTM 循环神经网络模型架构

鉴于卷烟产能预测作为一个相对细分的研究领域，相关的深度学习应用研究较为有限[10] [11]，而

LSTM 作为处理时序预测任务的经典深度学习模型，已在众多领域中得到广泛应用与验证[12]。因此，本文选择 LSTM 作为基础模型来处理卷烟产能时序预测任务。LSTM 是一种改进的循环神经网络，通过门控机制有效解决了传统 RNN 的梯度消失问题，增强了对长时序列数据的处理能力，更适合处理产能数据的时序特征。LSTM 网络的基本神经元如图 1(a)所示。

## 2.2. 注意力机制

为实现卷烟产能的预测，本研究采用一种基于注意力机制的网络结构，以适应动态产能预测的输出要求。其核心在于聚焦相关信息区域，同时降低对无关信息的关注程度，从而增强模型的准确性。注意力机制通过采用概率分布赋予关键信息足够的关注度，有效补偿了 LSTM 因序列过长而造成的信息损失[13]。此外，该机制以概率分配的方式取代了随机赋予权重的做法。图 1(b)展示了注意力机制的结构。



**Figure 1.** Architecture of the LSTM-Attention model  
**图 1.** 长短期记忆网络-注意力机制模型结构

## 2.3. 损失函数

为实现卷烟产能的预测，本研究采用一种基于注意力机制的网络结构，以适应动态产能预测的输出要求。其核心在于聚焦相关信息区域，同时降低对无关信息的关注程度，从而增强模型的准确性。注意力机制通过采用概率分布赋予关键信息足够的关注度，有效补偿了 LSTM 因序列过长而造成的信息损失[13]。此外，该机制以概率分配的方式取代了随机赋予权重的做法。图 2 展示了注意力机制的结构。

在卷烟产量预测任务中，为平衡模型对正常数据分布的敏感性与异常值的鲁棒性，本文采用 Huber 损失函数作为训练目标。其数学表达式为：

$$\text{Huber} = \begin{cases} \frac{1}{2}(\hat{y}_i - y_i)^2, & |y_i - \hat{y}_i| \leq \delta \\ \delta|\hat{y}_i - y_i|, & |y_i - \hat{y}_i| > \delta \end{cases} \quad (1)$$

式中,  $\hat{y}_i$  为模型输出的卷烟产量预测值,  $y_i$  为实际生产记录值,  $\delta$  为区分误差量级的阈值。该损失函数有效缓解了工业数据噪声对深度学习模型泛化能力的负面影响, 为稳定、高精度的卷烟产量预测提供了保障。

## 2.4. 早停策略

研究采用 LSTM-Attention 模型进行工业产能预测, 且原始数据经过必要的层级分类处理后, 部分子类别的样本规模显著缩减, 模型更容易在小样本场景下出现过拟合的现象。由于深度学习模型可以无限迭代下去, 其在训练过程中极易过度适应训练数据的特定模式, 导致泛化能力下降。为有效抑制这一现象, 本研究引入了早停(Early Stopping)策略。该策略通过持续监控模型在独立验证集上的性能表现, 在验证性能不再提升或开始下降时及时终止训练过程, 从而有效减少不必要的训练时间和计算资源消耗。

## 3. 实验与分析

本研究实验环境为 Windows 11; CPU 为 Intel i7-1065G7; GPU 为 Intel Iris Plus Graphics; Python 3.7; Pytorch 1.13。

### 3.1. 数据预处理

本研究采用某卷烟厂卷包机 2019 年 1 月 2 日至 2023 年 2 月 28 日的多层级生产数据集, 其原始数据包含日期、班别、机台、牌号等 11 个特征维度, 按整体层面、机台层面、机台 - 牌号层面、机台 - 牌号 - 班别层面四维度组织。研究发现粗粒度层级存在同一时间点多条记录的数据冗余问题, 当数据粒度细化至机台 - 牌号 - 班别维度时, 时间戳字段的缺失率呈现显著升高趋势, 经量化分析其缺失比例达到 17.3%。基于冗余率与数据完整性的平衡, 最终选取机台 - 牌号层级作为建模基准。在机台 60# 中筛选出 1835 条有效样本, 涵盖三类卷烟牌号。其中高频牌号“钻石(软红)”占比 72.5%, 符合行业主牌号集中生产规律, 故优先建立其独立预测模型。

由于选定的“机台 - 牌号”层级数据在时间维度上存在不连续性, 无法直接用于时间序列建模, 本研究采用基于班次分组的线性插值方法对缺失时间节点进行数据重构。具体而言, 首先依据数据起止时间构建完整的连续日期索引序列。随后, 将原始数据按班别分组, 并将每组数据分别与完整的日期序列进行左连接, 从而识别出缺失的时间点。对于识别出的缺失点, 采用线性插值方法对产量数值型变量进行插值估计, 并通过前向填充和后向填充方法处理序列端点的缺失值。经过处理, 数据集在时间维度上实现了完整重构, 数据规模显著增加, 各班次记录得以补全。

在完成核心的时序重构后, 本研究采用系统化数据预处理框架, 按照规范化流程依次执行数据清洗、特征工程与数据标准化。首先对缺失时序数据采用线性插值法填充, 随后基于四分位距准则(IQR 系数 1.5)检测并剔除异常值, 确保数据质量。针对离散型生产班别变量进行独热编码处理, 将分类信息转换为数值特征, 结果如表 1 所示。然后对时间序列数据进行系统性特征工程。具体构建的特征类型及方法如表 2 所示。经特征工程衍生的缺失值采用后向填充法处理以保持时间连续性, 剩余缺失值采用零值填充策略。最后对连续型数值特征采用 Z-score 标准化实现量纲统一, 并将数据集按照 8:2 的比例划分为训练集和测试集, 其中训练集包含前 80% 的时间序列数据, 测试集包含后 20% 的数据。进一步地, 从训练集中

分离出 10% 的数据作为验证集，用于模型训练过程中的性能监控和早停机制。

**Table 1.** One-Hot encoding results (First 5 samples as example)

**表 1.** 独热编码处理结果(以前 5 个样本为例)

Sample ID	Original Shift	Shift_Class A	Shift_Class B	Shift_Class C
0	Class A	1	0	0
1	Class B	0	1	0
2	Class C	0	0	1
3	Class A	1	0	0
4	Class B	0	1	0

**Table 2.** Feature construction results

**表 2.** 特征构建结果

Feature Type	Specific Method	Feature Name	Purpose
Temporal Features	Date Decomposition	Year, month, day of week	Capture periodic and seasonal patterns
Lag Features	Historical Value Extraction	Production_lag_1, Production_lag_2, Production_lag_3	Model temporal dependencies and autocorrelation
Trend Features	Moving Average	Production_ma_3, Production_ma_7	Smooth short-term fluctuations and extract medium-term trends
Volatility Features	Rolling Standard Deviation	Production_std_3, Production_std_7	Quantify sequence volatility and uncertainty

### 3.2. 模型的训练与验证

本研究构建了基于 LSTM-Attention 的深度学习模型，该模型结合了 LSTM 的时序建模能力和注意力机制的特征选择优势。模型采用双层 LSTM 网络，隐藏单元数设为 256，层间 dropout 率为 0.6，用于提取时间序列的长短期依赖关系。在此基础上，引入基于全连接层的注意力机制，通过学习各时间步的重要性权重，动态聚合 LSTM 输出的隐藏状态，增强模型对关键时序信息的捕获能力。模型的输出层采用两层全连接网络，中间层应用 ReLU 激活函数和 Dropout 正则化，最终输出预测结果，所有预测结果均通过相应的标准化器进行逆变换，恢复至原始数据尺度进行评估，确保评估结果的实际意义和可解释性。

考虑到实际数据中可能存在异常值，本研究选用 Huber 损失函数( $\delta = 1.0$ )作为训练目标，该损失函数对异常值具有更强的鲁棒性。优化器采用 Adam 算法，初始学习率设为 0.001，权重衰减系数为  $1 \times 10^{-5}$ ，以防止过拟合现象。训练过程中实施 ReduceLROnPlateau 学习率调度策略，当验证集 R<sup>2</sup> 指标在 5 个 epoch 内无改善时，学习率按 0.5 的因子进行衰减，最小学习率限制为  $1 \times 10^{-6}$ 。同时，应用梯度裁剪技术(最大范数为 1.0)防止梯度爆炸问题，并采用基于验证集 R<sup>2</sup> 指标的早停机制，当验证集 R<sup>2</sup> 在连续 15 个 epoch 内未出现改善时，终止训练过程并保存性能最优的模型参数。训练过程最大 epoch 数设为 256。除基本的训练 - 验证 - 测试划分外，本研究还采用时间序列交叉验证(Time Series Split, k = 5)对模型的泛化性能进行进一步验证，确保模型的稳定性和可靠性。该验证策略充分考虑了时间序列数据的时序依赖特性，避免了传统交叉验证可能导致的数据泄露问题。

### 3.3. 预测模型评价指标

实验使用 4 个关键性能指标对模型进行评估，包括均方误差(Mean Squared Error, MSE)、平均绝对误

差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)及决定系数(Coefficient of Determination, R<sup>2</sup>)。计算公式如下：

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

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

其中，n 是样本数， $y_i$  是第 i 个样本的真实值， $\hat{y}_i$  是第 i 个样本的预测值， $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  为真实值的样本均值。

MSE、MAE、MAPE 都表示预测值与真实值之间的误差，值越小，表示模型的预测能力越好，精度越高。MSE 与 MAE 侧重绝对误差的分布特性，MAPE 揭示相对误差趋势，R<sup>2</sup> 衡量模型对目标变量变异性的解释能力，值越接近 1 表明拟合优度越好。四者共同为模型优化提供全面参考。

### 3.4. 消融实验

为验证所提方法中各组件的有效性，本研究设计了系统性消融实验，通过逐步移除或添加关键组件来量化其对模型性能的贡献。消融实验结构如表 3 所示。

引入特征工程后，模型的 MSE 从 1134.47 大幅降低至 705.65，降幅达 37.8%；MAE 从 27.85 降至 20.34，改善幅度为 27.0%；MAPE 从 10.17% 优化至 7.75%，提升 23.8%；R<sup>2</sup> 从 0.3017 提升至 0.5656，解释方差能力显著增强。这一结果验证了时间序列特征工程的重要作用。通过构建滞后特征、统计特征和时间特征等多维度信息，模型能够更好地捕捉卷烟产量等时间序列中的潜在规律和依赖关系，从而显著提升预测精度。

在特征工程基础上，对比 LSTM 模型与 LSTM-Attention 模型的性能差异，结果显示注意力机制带来了进一步的性能提升。引入注意力机制后，MSE 从 705.65 进一步降低至 624.13，改善 11.6%；MAE 从 20.34 优化至 19.90，提升 2.2%；MAPE 从 7.75% 降至 7.36%，改善 5.0%；R<sup>2</sup> 从 0.5656 提升至 0.6158，模型拟合能力持续增强。注意力机制的有效性在于其能够动态分配时序信息的权重，使模型在预测过程中重点关注与当前预测目标最相关的历史时刻，从而提高预测的准确性和稳定性。

**Table 3.** Ablation experiment results

**表 3. 消融实验结果**

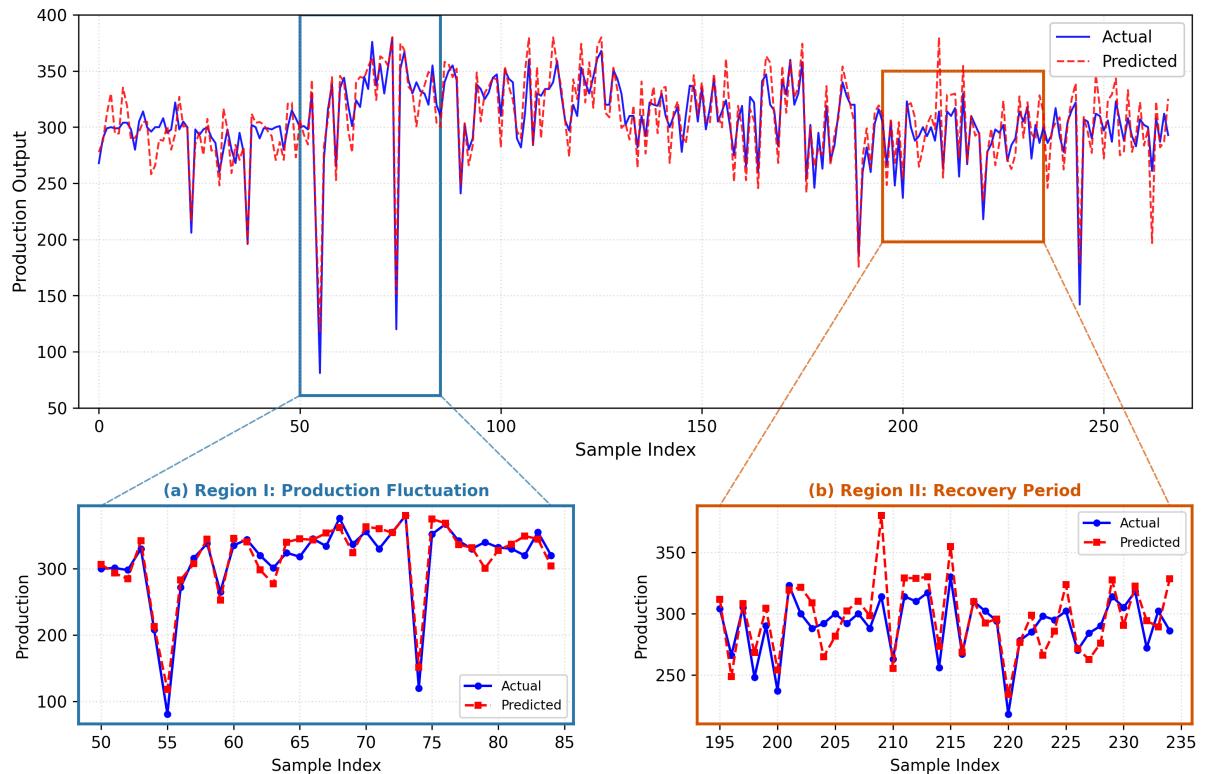
Model	MSE	MAE	MAPE%	R <sup>2</sup>
LSTM without Constructed Features	1134.47	27.85	10.17	0.3017
LSTM	705.65	20.34	7.75	0.5656
LSTM-Attention	<b>624.13</b>	<b>19.90</b>	<b>7.36</b>	<b>0.6158</b>

### 3.5. 模型预测对比分析

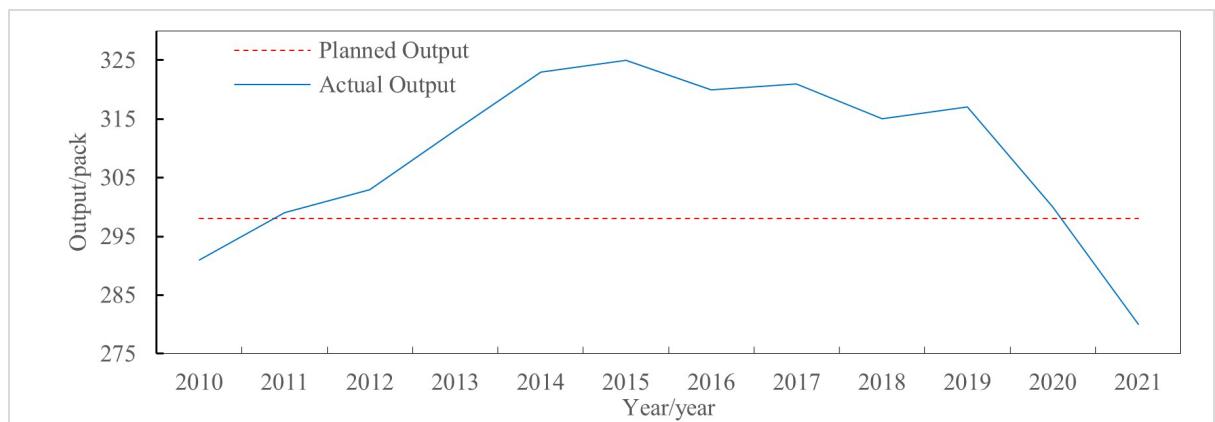
全面评估所提 LSTM-Attention 模型的性能优势，本研究选择了两类具有代表性的对比方法：1) 复杂

深度学习模型 ResNet-LSTM-Attention, 用于验证模型复杂度的适配性; 2) 传统统计与深度学习混合模型 ARIMA-LSTM, 用于对比不同建模范式的有效性。所有对比模型均采用相同的特征工程策略和实验设置, 确保结果的公平性和可比性。预测结果如图 2 所示, 远超传统经验计算方式(如图 3 所示), 性能结果如表 4 所示。

对比实验结果显示, ResNet-LSTM-Attention 模型的 MSE 上升至 772.03, 增长 23.7%。其性能下降的根本原因在于模型复杂度与数据规模的不匹配。ResNet 的深层网络结构在相对简单的卷烟产量预测任务中引入了过多参数, 在有限的训练数据下容易发生过拟合现象, 额外的网络深度未能带来性能增益反而



**Figure 2.** Single-step prediction results  
**图 2.** 单步预测结果



**Figure 3.** Traditional empirical calculation method  
**图 3.** 传统经验计算方式

**Table 4.** Single-step prediction: comparative experimental results of different models  
**表 4. 单步预测：不同模型的对比实验结果**

Model	MSE	MAE	MAPE%	R <sup>2</sup>
LSTM-Attention	<b>624.13</b>	<b>19.90</b>	<b>7.36</b>	<b>0.6158</b>
ResNet-LSTM-Attention	772.03	20.21	7.75	0.5248
ARIMA-LSTM	952.04	52.63	20.23	0.2348

降低了模型泛化能力。ARIMA-LSTM 混合模型表现最差, MSE 高达 952.04, 相比 LSTM-Attention 模型劣化 52.5%。其性能劣化源于两种建模范式的根本性冲突。ARIMA 基于线性自回归假设, 适用于平稳时间序列的建模, 而 LSTM 专长于捕捉非线性时序模式。在卷烟生产这一复杂非线性系统中, ARIMA 的线性假设与实际生产过程的非线性特性存在根本性冲突, 导致混合模型的性能劣化。对比实验结果表明, 所提 LSTM-Attention 模型在所有评价指标上均取得最优性能, 验证了模型复杂度与任务匹配度的重要性, 以及统一建模范式在时间序列预测中的优越性。

### 3.6. 模型泛化能力测试

为验证所构建预测模型的普适性, 本研究设计了跨机台、跨牌号的泛化能力测试实验。验证数据集选取 62#、63#机台的“钻石(软红)”牌号样本以及 60#机台的其他类牌号样本, 采用与训练数据相同的预处理流程, 确保数据处理的一致性。结果如表 5 所示。

**Table 5.** Model generalization performance results  
**表 5. 模型泛化能力结果**

Testing Scenario	MSE	MAE	MAPE%	R <sup>2</sup>
60#-Diamond	<b>624.13</b>	<b>19.90</b>	<b>7.36</b>	<b>0.6158</b>
60#-Hongjinlong	911.67	39.32	25.43	0.2730
62#-Diamond	823.24	35.03	21.57	0.3339
63#-Diamond	892.69	36.73	20.59	0.3743

模型在 62#、63#机台上的性能衰减反映了设备间的本质差异对预测模型的影响。从系统工程角度分析, 不同机台的机械精度、传动系统特性、控制算法参数等硬件差异导致了相同工艺条件下产量响应的系统性偏移。更深层次的原因在于设备老化程度的差异。生产设备在长期运行过程中, 关键部件(如切丝刀具、滤棒传送带、成形器等)的磨损程度不同, 导致设备的动态响应特性发生漂移。这种设备状态的非线性演化过程在训练数据中难以完全覆盖, 造成了模型在新设备上的泛化能力下降。从数据分布角度分析, 不同机台的数据分布存在协变量偏移现象。即使输入特征的统计特性相似, 但特征与目标变量间的映射关系在不同设备上存在差异, 这种域适应问题是深度学习模型在工业应用中面临的普遍挑战。

60#机台在“红金龙(软精品)”牌号上的性能衰减揭示了原料物性差异对生产过程的深层影响。不同牌号卷烟的叶丝配方、含水率、填充密度等物理参数存在显著差异, 这些差异通过复杂的物理化学过程影响生产效率。具体而言, “红金龙”牌号的高含水率特性会增加物料的黏附性, 影响传送带的运行效率; 其较低的填充密度要求更精确的计量控制, 对设备的响应速度提出了更高要求。这些物理特性的变化改变了设备的动态响应特征, 而模型在训练过程中主要学习了“钻石(软红)”牌号的生产规律, 难以适应新的物理约束条件。从控制系统角度分析, 不同牌号的生产往往需要调整设备的运行参数, 这些参数调整会改变系统的传递函数特性, 导致原有的输入 - 输出映射关系失效。模型的泛化能力衰减本质上反

映了静态学习模型在动态工业系统中的适应性限制。

针对上述泛化能力衰减问题,可从算法层面采取以下改进策略:一是基于设备嵌入的特征增强方法,将机台标识作为可学习的嵌入向量引入模型,与原始特征拼接后输入LSTM网络,使模型能够自动学习不同机台的隐式特征表示,捕捉设备磨损程度、机械精度偏差等难以直接量化的隐性因素;二是基于迁移学习的跨设备适应方法,采用“预训练-微调”范式,首先利用数据充足的源机台训练基础模型,随后在目标机台的少量样本上微调参数,实现跨设备迁移;三是基于领域自适应的分布对齐方法,通过对抗训练策略学习到领域不变的特征表示,降低数据分布差异对预测的影响。上述策略可根据实际应用场景灵活或组合使用,其具体实现与验证作为后续研究的重点方向。

### 3.7. 多步预测

为满足生产管理中的前瞻性决策需求,本研究采用递归预测策略进行多步预测,即将前一步的预测值作为下一步的输入特征。多步预测能够提供更长时间窗口的生产状态预估,为生产调度优化、维护计划制定等提供重要参考。基于多步预测实验结果(表6),多步预测性能随步长增加呈系统性衰减,主要源于迭代过程中的误差累积放大、设备状态长程依赖建模不足以及超越训练工况覆盖的数据稀疏性制约。尽管15步预测的MAPE增至9.51%,5步预测仍保持8.95%的实用精度,可为短期生产调度提供有效参考。这一结果为管理者确定了预测时效边界。在实际应用中,可以5步预测作为可靠预测的时间限度,用于排产的精细化调度,超过该时限的结果预测仅作为趋势性参考,用于日常的粗粒度规划。

**Table 6.** Multi-step prediction results

**表6. 多步预测结果**

Prediction Steps	MSE	MAE	MAPE%	R <sup>2</sup>
5	897.42	23.53	8.95	0.4620
10	958.76	24.93	9.36	0.4209
15	991.93	25.25	9.51	0.3907

## 4. 结论

1) 本研究构建的工业产能预测模型结构简洁、易于部署,能够有效利用MES系统中记录的多机台、多牌号历史生产数据,为智能制造环境下的生产决策提供重要支撑。在生产优化层面,模型可预测不同机台生产各牌号卷烟的产量表现。这为管理者制定高效的生产安排策略提供了科学依据。在供应链管理方面,模型的前瞻性预测能力有助于物流中心优化原材料储备策略,避免因材料储备不足导致的停机问题。就企业数字化转型而言,本研究为推进企业数字化进程、加快智慧工厂建设提供了技术支撑。模型具有良好的工程使用性和预测精度,验证了深度学习方法在复杂制造环境中的应用潜力,为后续的生产智能化改造奠定了理论基础和实践经验。尽管模型在跨机台、跨牌号场景下存在一定的泛化性能衰减,但在单步预测任务中表现出的优异性能,以及在短期多步预测中的实用精度,充分证明了该方法在实际生产环境中的有效性和可靠性,此外,本研究已从算法层面提出了设备嵌入、迁移学习和领域自适应等改进策略,为后续提升模型泛化能力指明了方向。

2) 决定系数(R<sup>2</sup>)用于评估模型对因变量变异性的解释程度,其值越接近1,表明模型解释力越强。然而,本研究中构建的模型R<sup>2</sup>值最高为0.6,表明所建模型仅能解释约60%的目标变量变异。尽管本研究已充分利用卷烟厂提供的所有特征,并通过系统化特征工程构建了新的预测变量,但仍有约40%的变异无法被现有特征解释。这一结果揭示了工业数字化转型过程中的关键问题:数据采集质量与生产管理流

程的协同不足。为进一步提升模型性能，实现数据驱动决策的循环，管理层面应优化生产排班制度，强制操作人员通过 MES 等系统记录设备异常细节，实现跨部门数据完整性核查机制；技术层面需在关键设备加装多源状态传感器，通过振动频谱分析捕捉磨损特征，利用声发射技术监测故障信号，并将高频传感数据与 MES 工单信息实时关联。

## 参考文献

- [1] Bertolini, M., Mezzogori, D., Neroni, M. and Zammori, F. (2021) Machine Learning for Industrial Applications: A Comprehensive Literature Review. *Expert Systems with Applications*, **175**, Article ID: 114820. <https://doi.org/10.1016/j.eswa.2021.114820>
- [2] 赵海峰, 诸立凯, 刘长松, 等. 基于注意力机制的 CNN-GRU 煤层气产能预测方法研究[J]. 煤矿安全, 2023, 54(12): 11-17.
- [3] 金毅, 郑晨晖, 宋慧波, 等. 基于神经网络模型的煤层气产能预测研究[J], 河南理工大学学报(自然科学版), 2025, 44(1): 46-56.
- [4] Yuan, X., Li, L., Shardt, Y.A.W., Wang, Y. and Yang, C. (2021) Deep Learning with Spatiotemporal Attention-Based LSTM for Industrial Soft Sensor Model Development. *IEEE Transactions on Industrial Electronics*, **68**, 4404-4414. <https://doi.org/10.1109/tie.2020.2984443>
- [5] Xia, Z., Wang, Y., Yang, K. and Jin, L. (2022) Production Capacity Prediction of Thyristor Based on Fuzzy Neural Network. *Journal of Advances in Information Technology*, **13**, 100-105. <https://doi.org/10.12720/jait.13.1.100-105>
- [6] Cai, H., Wan, J. and Chen, B. (2024) Digital Twin-Driven Multi-Factor Production Capacity Prediction for Discrete Manufacturing Workshop. *Applied Sciences*, **14**, Article No. 3119. <https://doi.org/10.3390/app14073119>
- [7] 张家刚, 阴艳超, 易斌, 等. 融合多尺度卷积网络与 TransGRU 的流程生产过程质量多步预测方法[J]. 计算机集成制造系统, 2025, 31(12): 4760-4773.
- [8] 李悦, 阴彦磊, 易斌. 融合双重注意力机制与 TCN 的复杂生产过程工艺质量精准预测[J]. 计算机集成制造系统, 2025, 31(12): 4724-4738.
- [9] Ji, C. and Sun, W. (2022) A Review on Data-Driven Process Monitoring Methods: Characterization and Mining of Industrial Data. *Processes*, **10**, Article No. 335. <https://doi.org/10.3390/pr10020335>
- [10] 陈天丽, 江志凌, 邵力波, 等. 基于深度学习的卷包设备清洁保养质量判别系统设计[J]. 中国烟草学报, 2022, 28(6): 20-29.
- [11] Ji, Q., Li, M., Xu, G. and Zhang, W. (2024) Quality Improvement of Cigarette Manufacturing Process Driven by Deep Learning: Application of Wasserstein Distance Algorithm. *IEEE Access*, **12**, 140472-140482. <https://doi.org/10.1109/access.2024.3467989>
- [12] Hennig, M., Grafinger, M., Hofmann, R., Gerhard, D., Dumss, S. and Rosenberger, P. (2021) Introduction of a Time Series Machine Learning Methodology for the Application in a Production System. *Advanced Engineering Informatics*, **47**, Article ID: 101197. <https://doi.org/10.1016/j.aei.2020.101197>
- [13] Wan, A., Chang, Q., AL-Bukhaiti, K. and He, J. (2023) Short-Term Power Load Forecasting for Combined Heat and Power Using CNN-LSTM Enhanced by Attention Mechanism. *Energy*, **282**, Article ID: 128274. <https://doi.org/10.1016/j.energy.2023.128274>