DOI: 10.19666/j.rlfd.202511072

基于 LightGBM-Transformer 模型与 NSGA-II 的燃煤机组启停过程优化

丁鹏恺，陈　衡，黄子健，张肇清，刘文毅，潘佩媛
（华北电力大学能源动力与机械工程学院，北京　102206）

［摘　　要］【目的】提升燃煤机组深度调峰条件下启机升负荷与停机降负荷非稳态阶段的响应速度与经济性。【方法】基于 660 MW 机组 DCS 多月启停运行数据，首先采用滑动窗口范围阈值与稳态判据识别启停过程中的非稳态片段，随后采用最大信息系数筛选出与功率强相关的特征变量，并引入过程进度特征表征启停阶段的推进程度。针对跨月份工况差异，在各启停片段内按时间顺序分层划分数据集，并分别构建 LightGBM 与 Transformer 功率预测代理模型，采用贝叶斯优化完成超参数寻优。最后，将训练得到的预测模型嵌入 NSGA-II，以提升变负荷速率和降低燃料消耗为目标，在满足安全约束下进行寻优，输出可执行的给定值轨迹。【结果】非稳态片段预测模型在跨月份条件下仍保持较好的泛化性能，经 NSGA-II 优化后，启停非稳态片段耗时平均缩短约 5.34%，累计燃料量平均降低约 3.53%。【结论】提出的基于数据驱动代理建模与多目标寻优的启停优化方法，可在满足安全约束的前提下兼顾缩短启停时间与降低燃料消耗，为燃煤机组灵活调峰与在线运行辅助决策提供技术支撑。

［关　键　词］燃煤机组；深度调峰；快速启停；数据驱动；混合优化算法

［引用本文格式］丁鹏恺，陈衡，黄子健，等. 基于 LightGBM-Transformer 模型与 NSGA-II 的燃煤机组启停过程优化[J]. 热力发电, 2026, 55(8): 1.　Ding P K, Chen H, Huang Z J, et al. Optimization of start-up and shutdown processes for coal-fired units based on LightGBM-Transformer model and NSGA-II[J]. Thermal Power Generation, 2026, 55(8): 1.

Optimization of start-up and shutdown processes for coal-fired units based on LightGBM-Transformer model and NSGA-II

Ding Pengkai, Chen Heng, Huang Zijian, Zhang Zhaoqing, Liu Wenyi, Pan Peiyuan
(School of Energy, Power and Mechanical Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: [Objective] With the high penetration of renewable energy in power systems, coal-fired units are required to provide peak-shaving services with higher frequency and larger depth, making start-up load ramp-up and shutdown load ramp-down more time-critical and cost-sensitive. This study aims to improve the responsiveness and economic performance of non-steady-state operating phases during start-stop transitions under deep peak-shaving conditions, while ensuring compliance with operational safety constraints. [Methods] Multi-month DCS data from a 660 MW coal-fired unit sampled at 5 s were analyzed. Non-steady segments in start-up and shutdown were firstly identified by combining sliding-window range thresholds with steady-state criteria, yielding transient datasets aligned with the underlying operational transitions. Candidate variables were then screened using the maximal information coefficient to retain features most strongly associated with output power, supplemented with a process progress descriptor to represent phase evolution and improve discrimination across different load intervals. To accommodate cross-month operating variability, start-stop events were organized into source and target domains. Within each target-domain segment, a time-stratified split was adopted, with the first 10% used for validation and

收 稿 日 期：2025-11-21　修 回 日 期：2025-12-26　接 受 日 期：2026-01-06
基 金 项 目：国家自然科学基金面上项目（52276006）
Supported by: General Program of the National Natural Science Foundation of China (52276006)
第一作者简介：丁鹏恺（2003），男，硕士研究生，主要研究方向为燃煤机组运行优化，13204325001@163.com。
通信作者简介：陈衡（1989），男，博士，副教授，主要研究方向为多能互补集成与热力系统优化，heng@ncepu.cn。

the remaining 90% reserved for testing, thereby preserving temporal causality and avoiding leakage from future samples. Given the distinct dynamics of start-up load ramp-up and shutdown load ramp-down, two surrogate predictors were developed (LightGBM for tabular feature interactions and a Transformer for temporal dependency representation), and their hyperparameters were tuned via Bayesian optimization. After model selection, source-domain data were combined with the target-domain validation subset for final training. The trained surrogate models were embedded in an NSGA-II multi-objective optimization framework. Under safety constraints and ramp-rate limits, the optimization jointly improved load-ramping capability and reduced fuel consumption, producing executable set-point trajectories and load-change strategies suitable for operational guidance. [Results] The phase-specific surrogate models achieved stronger predictive performance than common baselines such as XGBoost and random forests across operating conditions, and maintained robust generalization in cross-month evaluation. When applied to representative non-steady segments, NSGA-II produced feasible set-point trajectories that satisfied the imposed constraints throughout the transient. Relative to baseline operation, the optimized strategies reduced the average duration of non-steady start-stop segments by approximately 5.34% and lowered cumulative fuel consumption by approximately 3.53%, indicating concurrent gains in speed and economy. [Conclusion] A data-driven hybrid modeling and multi-objective evolutionary optimization approach is developed for coal-fired unit start-up and shutdown under deep peak-shaving conditions. By coupling phase-specific power surrogates with constraint-aware optimization, the method generates implementable set-point trajectories that jointly accelerate transient response and reduce fuel use within safe operating envelopes, providing practical support for flexible operation and online decision assistance.

Key words: coal-fired unit; deep peak shaving; rapid start-up and shutdown; data-driven; hybrid optimization algorithm

全球能源转型的背景下，风电、光伏等可再生能源装机规模和发电占比不断提升。火电厂作为电力系统的重要支撑，提高火电厂调峰过程的灵活性与经济性，保障电力系统稳定具有重要意义[1-2]。然而，基于传统机理建模的运行优化方法存在建模成本高、求解效率低且难以精确表征复杂工况下机组动态特性等局限。随着大数据和人工智能技术的发展，数据驱动方法可通过整合工业现场海量运行数据、提取特征并训练模型，实现对真实系统的拟合[3]。

近年来，预测模型结合智能优化算法在能源系统中逐步成为灵活性改造的有效工具。调峰灵活性改造研究主要围绕储能侧、可再生能源侧、电网与调度侧、负荷与需求侧和燃煤机组侧展开。在储能侧，Dong 等人[4]针对固体氧化物燃料电池与压缩空气储能的混合系统，采用人工神经网络配合 NSGA-II 进行多目标寻优，优化后㶲效率提高 10.8%，同时 CO_2排放降低 60.7 kg/(MW ℏ)，发挥系统级调峰作用；Du 等人[5]聚焦电池循环经济，以随机森林等 9 种算法建立预测模型，结合主成分分析、特征筛选及网格搜索调参，将计算时间缩至 7 min 且性能偏差小于 3%，提升了储能缓冲能力，提高了系统调峰灵活性。在可再生能源侧，Chen 等人[6]以多模态深度学习构建预测模型并耦合多目标优化，使光伏系统采光指标提升 1.2~3.4，把白天自发电显著抬升到负荷峰段，从而增强削峰能力。在电网与调度侧，Cestero 等人[7]构建 PINN 预测与 RL 优化框架，在满足电压约束下，训练加速约 50%，显著提升日内削峰与调峰灵活性。在负荷与需求侧，Wei 等人[8]构建物理模型库，训练多层感知机作为评估器，在多场景下 R^2 超过 0.95，并把约 1 308 kW ℏ 的整体削峰量汇聚成可调度弹性。在燃煤机组侧，陈东超等[9]通过训练 XGBoost 煤耗预测模型提取全工况最优给定值，结合预测模型和优化算法实现在线校准，在 1 000 MW 机组上，典型工况回归 R^2 达到 0.85，优化后机组实现节煤 0.84~5.91 g/(kW ℏ)、CO_2 减排 2.09~ 11.52 g/(kW ℏ)的目标，同时扩大了机组低负荷可达区。

上述研究在预测模型结合优化算法的效率与精度上已取得显著成效，但在应用上仍存在问题。首先，现有模型多基于短期样本，使其在跨月份工况下的泛化能力有限；其次，在燃煤机组侧的研究多聚焦于稳态寻优，启停过程中的非稳态区间在时间和能耗方面仍存在较大优化空间；最后，现有模型大多围绕常规运行工况展开，在一些极端工况下仍存在较大的偏差。

针对以上问题，本文提出了一种启停差异化的预测与优化方法，在统一的特征筛选与非稳态识别流程下，将启机升负荷与停机降负荷建模为 2 类动态特性不同的预测子任务，分别构建 LightGBM 与 Transformer 代理模型，并对 2 类模型独立开展超参数寻优与性能评估。先利用最大信息系数进行特征筛选，并基于非稳态数据筛选与跨月份的数据集划

分构建源域与目标域样本，分别对启机与停机非稳态区间训练预测模型，并嵌入贝叶斯优化提升预测精度，最终结合 NSGA-II 在变负荷速率与燃料消耗两目标之间搜索推荐解，加快启停速率并降低启停过程煤耗，同步提升燃煤机组参与调峰时的响应速度与经济性。

1 预测模型和优化算法

1.1 贝叶斯优化

在机器学习与深度学习算法中，模型超参数的设置决定了模型的性能。为在有限计算预算下高效搜索超参数，采用贝叶斯优化进行自动寻优[10]。

贝叶斯优化算法的核心由 2 部分组成：一方面，依据有限次函数评估结果，在参数空间中拟合一个概率代理模型，如高斯过程，从而给出目标函数在各候选点上的后验分布；另一方面，在该后验分布基础上构造采集函数，通过在参数空间内寻找采集函数的极值点来确定下一轮评估的位置。

贝叶斯优化过程的整体流程如图 1 所示。

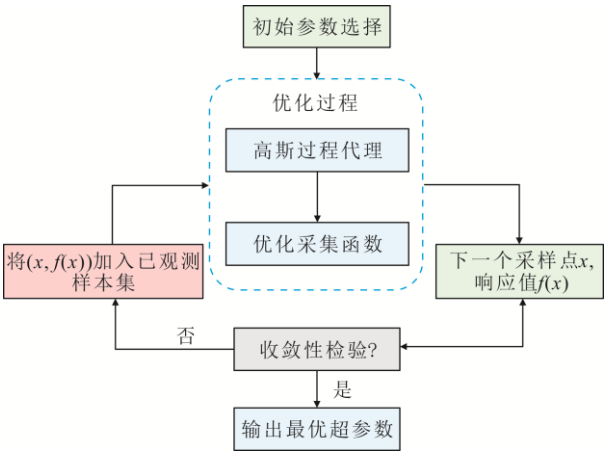


图 1 贝叶斯优化原理
Fig.1 Principles of Bayesian optimization

1.2 LightGBM 模型

LightGBM（light gradient boosting machine）是一种基于梯度提升的机器学习算法，具有训练效率高、内存消耗低等特点[11]。其采用直方图分裂与 leaf-wise 生长策略，并采用梯度单侧采样（GOSS）与互斥特征绑定（EFB）等机制，降低大规模样本与高维特征条件下的计算成本。

模型训练首先进行初始化，通常以常数项作为初始预测用于最小化经验风险。在每一轮迭代中，模型根据当前预测结果与真实值之间的误差构造新的回归树。在第 m 次迭代时，损失函数对当前预测值的梯度为[12]：

$$g_i(m)=-\left[\frac{\partial L(y_i\cdot F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right] \tag{1}$$

式中：$g_i(m)$为第 i 个样本在第 m 轮迭代中的梯度；x_i 为输入特征向量；y_i 为样本真实值；$F_{m-1}(x_i)$为上一轮迭代中对样本 i 的预测值。

算法利用梯度信息构建新的决策树 $h_m(x)$，用于拟合当前残差。根据新决策树的结果更新模型。经过 m 轮迭代后，最终的预测模型表示为：

$$F_M(x)=F_0(x)+\sum_{m=1}^{M}\eta h_m(x) \tag{2}$$

式中：$F_M(x)$为经过 M 轮迭代后的最终预测函数；$F_0(x)$为初始预测函数；x 为输入特征向量；η为学习率；M 为总迭代次数；$h_m(x)$为第 m 轮生成的回归树输出。

1.3 Transformer 模型

Transformer 是一类基于自注意力机制的序列建模结构，能够在统一表征空间内刻画多变量之间的非线性依赖关系，并通过注意力权重在不同时间步之间自适应聚合信息[13]。

为建立原始输入序列的位置信息，在嵌入层采用正弦和余弦的位置编码：

$$P_{o,2i}=\sin(o/10\ 000^{2i/d}) \tag{3}$$

$$P_{o,2i+1}=\cos(o/10\ 000^{2i/d}) \tag{4}$$

式中：o 为原始输入序列中的位置；i 为位置编码的维度索引；d 为嵌入向量的维度。

在编码器内部，单头缩放点积注意力计算过程为：

$$\boldsymbol{Q}=\boldsymbol{X}\boldsymbol{W}^{\boldsymbol{Q}} \tag{5}$$

$$\boldsymbol{K}=\boldsymbol{X}\boldsymbol{W}^{\boldsymbol{K}} \tag{6}$$

$$\boldsymbol{V}=\boldsymbol{X}\boldsymbol{W}^{\boldsymbol{V}} \tag{7}$$

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{8}$$

式中：$\boldsymbol{X}\in\boldsymbol{R}^{L\times d}$ 为编码器的输入矩阵；L 为输入序列长度；$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示查询、键与值矩阵；$\boldsymbol{W}^{\boldsymbol{Q}}$、$\boldsymbol{W}^{\boldsymbol{K}}$、$\boldsymbol{W}^{\boldsymbol{V}}$ 分别为查询、键与值的线性投影矩阵；d_k 为键向量的维度。

多头注意力机制将上述过程应用于 h 个不同的头，对于每个头使用不同的权重矩阵 $\boldsymbol{W}_h^{\boldsymbol{Q}}$、$\boldsymbol{W}_h^{\boldsymbol{K}}$、$\boldsymbol{W}_h^{\boldsymbol{V}}$，各头输出拼接后经输出投影矩阵 $\boldsymbol{W}_{\mathrm{O}}$ 线性变换得到最终输出[14]：

$$\boldsymbol{M}(\boldsymbol{X})=\mathrm{Concat}\left[\mathrm{head}_1,...,\mathrm{head}_h\right]\boldsymbol{W}_{\mathrm{O}} \tag{9}$$

式中：$\boldsymbol{M}(\boldsymbol{X})$为多头注意力输出矩阵；$\mathrm{head}_i$为第 i 个注意力头的输出；$\boldsymbol{W}_{\mathrm{O}}$为输出投影矩阵；$h$ 为注意力头数量。

1.4 NSGA-II 多目标优化算法

NSGA-II（non-dominated sorting genetic algorithm II）是一种多目标进化算法，通过快速非支配排序与拥挤距离在每一代同时保证Pareto前沿的收敛性与多样性。算法以交叉与变异从父代中产生子代，将二者合并后进行非支配排序得到若干等级，在同级内按拥挤距离优先保留彼此间隔更大的个体，截取得到下一代，由此形成带精英保留的演化过程，逐步逼近Pareto前沿。

对解 x、y，若 x 在所有目标上不差于 y，且至少在一个目标上更优，则称 x 支配 y。约束采用可行性优先：先判可行与否，再比较违反量，最后比较目标值[15]。违反量定义为：

$$V(x)=\sum_j \max\{0,\ g_i(x)\}+\sum_k \left|h_k(x)\right| \tag{10}$$

式中：$g_i(x)\leqslant 0$ 与 $h_k(x)=0$ 分别为不等式与等式约束。

2 混合优化模型构建

2.1 数据集及预处理

本文所使用的数据来自某 660 MW 机组 DCS 的实时运行参数，包括 2024 全年和 2025 年前 2 个月数据，数据采样间隔为 5 s，每个月采样数据共计 50 余万组。

首先筛选出包含启停过程的 5 个月数据，随后分离出单次完整启停过程。本文中，启停过程定义为以一次启机或停机过程为中心的连续运行区间，自机组在稳定工况下达到零负荷或满负荷时刻起，至下一次在稳定工况下达到满负荷或零负荷时刻止。为反映启停的全过程特征以及不同负荷区段的参数变化规律，采样区间内除负荷爬升与回落外，还包含稳态保持及运行方式切换等阶段。把选取的 6 个启停过程依次编号为 E1—E6，各过程中机组输出功率按采样结果分析如表 1 所示。

2.2 特征变量选取

面向启停过程预测与优化，仅保留机组海量参数中与功率强相关的特征变量。为识别与功率最相关的候选特征，本文采用最大信息系数（maximal information coefficient，MIC）进行相关性筛选[16]。MIC 对非线性与非单调关系敏感，且较少依赖变量的尺度与分布假设，适合启停与大梯度变负荷下的复杂耦合。初选的 13 个特征变量关于机组输出功率的 MIC 按大小排序如表 2 所示。

表 1 输出功率采样结果分析

Tab.1 Analysis of output power sampling results

过程	编号	时间	采样时长/h	平均值/MW	最大值/MW
零至满负荷	E1	2024-04	10.99	332.60	602.75
	E2	2024-10	25.55	416.86	588.07
	E3	2025-02	14.86	376.12	600.30
满负荷至零	E4	2024-04	50.02	332.10	602.72
	E5	2024-09	16.28	319.49	564.83
	E6	2025-01	70.82	408.72	599.38

表 2 特征参数相关性分析

Tab.2 Correlation analysis of feature parameters

序号	特征变量	MIC 系数
1	燃料量	0.974
2	调节级压力	0.964
3	主给水流量	0.957
4	总风量	0.954
5	主给水压力	0.953
6	主蒸汽压力	0.936
7	主凝结水流量	0.779
8	空预器出口烟气含氧量	0.762
9	再热蒸汽温度	0.703
10	主蒸汽温度	0.543
11	空气预热器 B 出口烟温	0.500
12	空气预热器 A 出口烟温	0.485
13	闭式冷却水温度	0.337

把 MIC 系数大于 0.7 的变量输入到后续的预测模型中，变量筛选可以在提高预测精度的同时使优化算法在约束空间中更快收敛到可行优解。

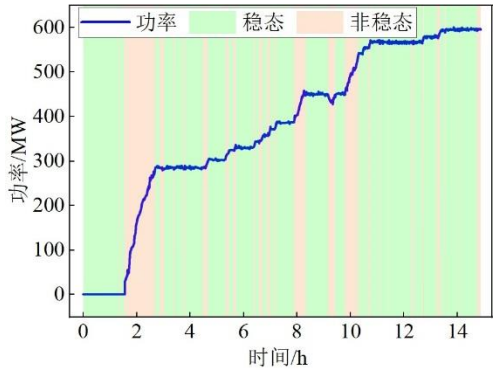
2.3 非稳态数据筛选

在完整的启停过程中，包括功率在内的各运行参数并非全程单调上升或下降。为优化启停过程，需要先从原始数据中识别出非稳态阶段，该阶段变负荷速率提升空间更大，因此更适合采用灵活的调节策略。本文采用滑动窗口方法进行非稳态识别，所用变量包括发电机功率、主蒸汽压力、主蒸汽温度、再热蒸汽温度和主给水流量。

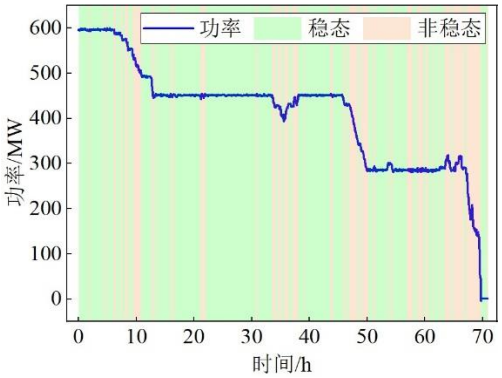
在每个窗口内计算各变量的极差，并设定各参数的变化阈值，窗口判定为稳态则整窗标记并前移，否则仅步进一个采样点。所有未被标记为稳态的数据，即构成非稳态数据集。

图 2 给出了待优化数据的非稳态划分结果。由图 2 可以看出，机组整个变负荷过程由间断分布的

稳态与非稳态区域交替构成。



a) 零至满负荷



b) 满负荷至零

图 2 启停过程稳态与非稳态分析

Fig.2 Analysis of steady and unsteady states during the start-up and shutdown processes

2.4 数据集划分

本文采用时间分层的源域与目标域方案。源域为启停过程 E1、E2、E4、E5，目标域分别为启机过程 E3 和停机过程 E6。

在跨月份的情况下，机组启停过程环境差异很大，把源域作为训练集后，直接采用目标域 E3 与 E6 的全量数据作为测试集会大幅降低预测精度，为增强模型跨月泛化能力，将 E3、E6 过程的数据在每个稳态与非稳态段内按时间顺序切分，每段前 10%作为验证集，后 90%作为测试集。完成超参数选择后，将源域和目标域 E3 与 E6 各段前 10%的验证样本合并进行最终训练，而后 90%样本只用于最终模型评估与运行优化。

机组收到新的负荷指令时，可以将前一阶段的运行数据作为模型学习的参考并入原始训练集中，后一阶段的数据进行在线预测与优化。

2.5 模型评估指标

本文采用平均绝对误差（mean absolute error，MAE）δ_{MAE}、均方根差（root mean square error，RMSE）δ_{RMSE} 和拟合优度（R^2）作为模型的评价指标。各评价指标计算公式如下[17]：

$$\delta_{\mathrm{MAE}}=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{11}$$

$$\delta_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{12}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \tag{13}$$

式中：n 为样本数量；y_i 为实际值；$\hat{y}_i$ 为预测值；$\bar{y}_i$ 为实际值的平均值。

2.6 优化目标

由预测模型对当前工况变量 x_i 预测出的机组功率 $P(x_i)$，定义相邻步的变负荷速率为：

$$r_i=\frac{P(x_i)-P^*(x_{i-1})}{\Delta t} \tag{14}$$

$$r_{0,i}=\frac{P_0(x_i)-P_0(x_{i-1})}{\Delta t} \tag{15}$$

式中：$P^*(x_{i-1})$为第 $i-1$ 个时刻优化后的功率，MW；$P(x_i)$为第 i 个时刻的候选控制量对应的预测功率，MW；$P_0(x_i)$和 $P_0(x_{i-1})$分别为第 i 个和第 i–1 个时刻原始轨迹对应的预测功率，MW；r_i 为该时刻优化后的变负荷速率，MW/min；$r_{0,i}$ 为第 i 个时刻原始轨迹对应的变负荷速率，MW/min；Δt 为相邻采样时刻的时间间隔，min。

以启机升负荷过程为例，速率和燃料目标为：

$$f_1=\begin{cases}-\dfrac{r_i-r_{\min}}{r_{\max}-r_{\min}}, & r_{0,i}>0\\[2ex] \dfrac{|r_i-r_{0,i}|}{r_{\max}-r_{\min}}, & r_{0,i}\leqslant 0\end{cases} \tag{16}$$

$$f_2=\frac{F(x_i)}{F_{0,i}} \tag{17}$$

式中：$r_{\min}$、$r_{\max}$ 分别为优化开始后各时刻变负荷速率的最小值和最大值；$F(x_i)$为该时刻优化后的燃料量，t/h；$F_{0,i}$ 为该时刻的原始燃料量，t/h；f_1、f_2 分别为速率函数和燃料量函数，两目标共同最小化。

为保证优化轨迹在工程上的可执行性与安全性，在优化过程中添加安全约束。设第 i 个时刻的可控给定值为燃料量 F_i、总风量 A_i 与给水量 W_i，对应原始轨迹为 $F_{0,i}$、$A_{0,i}$、$W_{0,i}$。考虑执行机构能力与控制回路裕度，三类可控给定值相对原始轨迹的偏移幅度限制为±5%，如式(18)所示：

$$\begin{cases} 0.95F_{0,i} \leqslant F_i \leqslant 1.05F_{0,i} \\ 0.95A_{0,i} \leqslant A_i \leqslant 1.05A_{0,i} \\ 0.95W_{0,i} \leqslant W_i \leqslant 1.05W_{0,i} \end{cases} \tag{18}$$

同时，为保证优化结果不过于偏离原运行工况，将变负荷速率约束纳入可行域，限制优化后的速率不超过原始速率的 105%，并保持与原过程同向，如式(19)所示：

$$\begin{cases} 0 \leqslant r_i \leqslant 1.05r_{0,i}\ , \quad r_{0,i} > 0 \\ 1.05r_{0,i} \leqslant r_i \leqslant 0\ , \quad r_{0,i} \leqslant 0 \end{cases} \tag{19}$$

每一步优化结束，从当代可行解中按归一化双目标空间的加权欧氏距离选解，若当代无可行解，则选取总违反度最小的个体。被选中个体对应的预测功率记为 $P^*(x_i)$，并将其写回作为下一时刻 $P^*(x_{i-1})$ 的递推输入，由此形成逐时滚动优化过程，直至机组进入稳态。

2.7 模型建立过程

本文在监测数据上完成稳态与非稳态分段，提取关键特征参数，进行特征筛选后通过 LightGBM 与 Transformer 分别建立启机过程与停机过程的预测模型并通过贝叶斯优化选择模型最优超参数，在保证预测精度的同时提升泛化能力。

在预测模型基础上优化可控变量，以变负荷速率、燃料消耗为目标，同时设置安全约束，采用 NSGA-II 搜索可行的控制量组合并形成 Pareto 解集。最终输出可执行的给定值轨迹以提高机组调峰的灵活性与经济性。模型整体框架如图 3 所示。

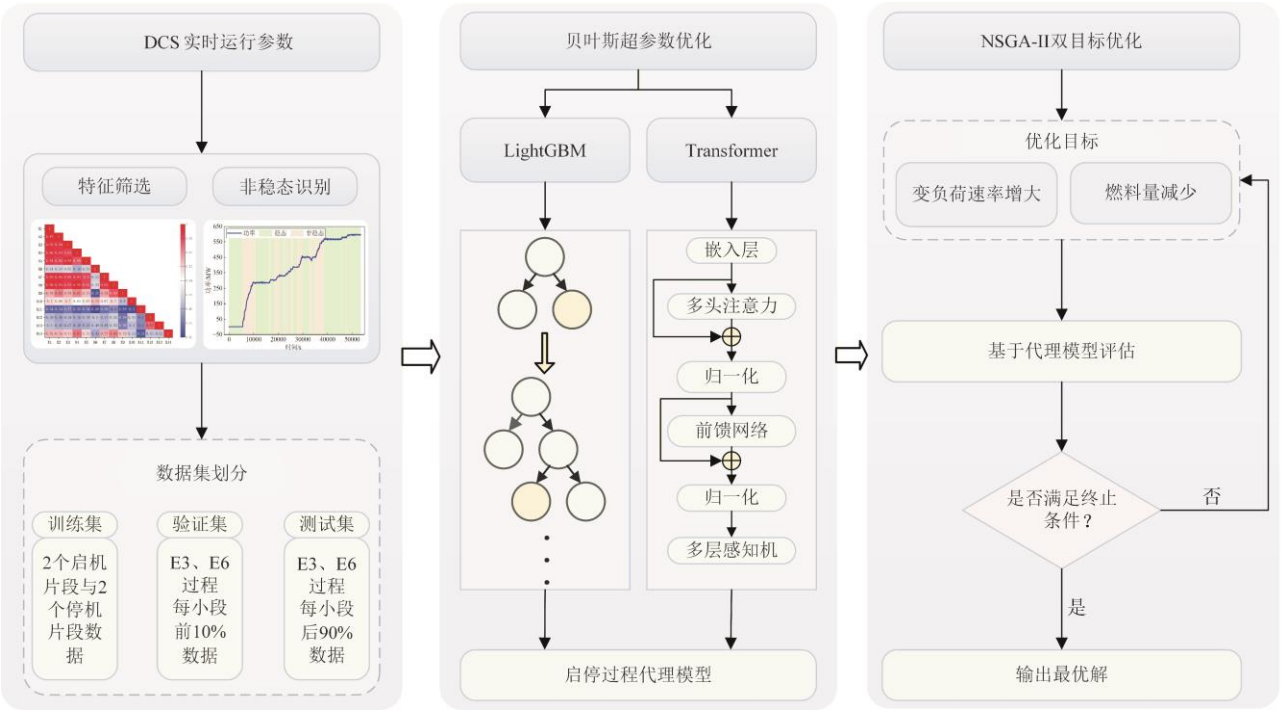


图 3 LightGBM-Transformer 模型与 NSGA-II 整体架构

Fig.3 Overall architecture of the LightGBM-Transformer model and NSGA-II

3 结果与讨论

3.1 超参数优化

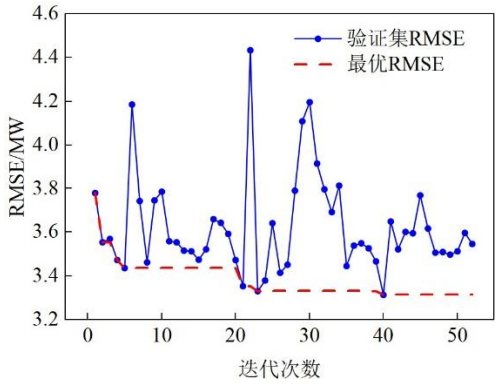
采用贝叶斯优化对 LightGBM 和 Transformer 两模型进行超参数选取，先进行随机探索 12 次，随后迭代 40 次，共 52 次寻优。两模型最终都以验证集 RMSE 最小为目标。

对于 Transformer 模型，输入序列长度 L 与位置编码共同决定其可建模的时间跨度与时序顺序信息表达能力，L 过小会削弱对燃烧与传热滞后效应的覆盖，L 过大则易引入冗余并增加训练复杂度。

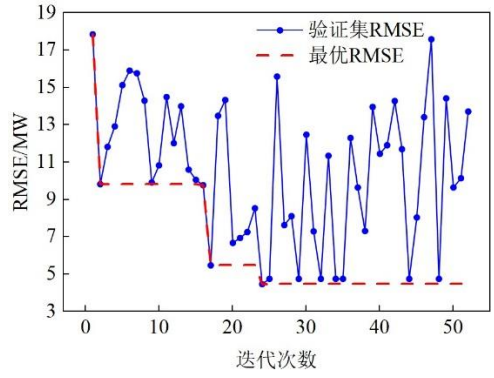
本文预先设定 L =64 的滑动序列作为输入，并引入位置编码表征窗内时间顺序；对于 LightGBM 模型，预测性能主要受树结构复杂度与叶节点划分策略影响。两模型的关键超参数配置与优化结果如表 3 所示，迭代过程如图 4 所示。从图 4 中可以看出，两模型在迭代过程中均快速收敛，RMSE 最终稳定在较低值。

表 3 超参数优化结果
Tab.3 Results of hyperparameter optimization

模型	超参数	搜索范围	优化结果
LightGBM	决策树数量	[200, 1 500]	978
	学习率	[0.01, 0.10]	0.043 6
	最大深度	[3, 10]	6
	叶子比例系数	[0.50, 1.00]	0.852
	子样本率	[0.60, 1.00]	0.859
Transformer	隐层维度	[96, 256]	147
	多头数	[2, 10]	7
	层数	[2, 6]	3
	学习率	$[10^{-4}, 3\times10^{-3}]$	0.001 17
	丢弃率	[0, 0.20]	0.01
	批量大小	[320, 640]	376



a) LightGBM 超参数寻优



b) Transformer 超参数寻优

图 4 模型超参数优化结果
Fig.4 Results of model hyperparameter optimization

为证明贝叶斯优化的优越性，本文在保持搜索空间边界、迭代次数、数据划分及训练策略完全一致的条件下，引入随机搜索与网格搜索作为对比，不同优化方法的性能如表 4 所示。

从表 4 中可以看出，在同样搜索次数下，贝叶斯优化相比于其他超参数搜索方法可以获得更低的 RMSE，且在迭代过程中可以更快收敛。

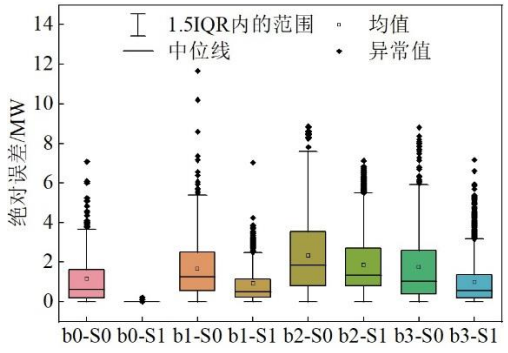
表 4 不同超参数搜索方法对比
Tab.4 Comparison of different hyperparameter search methods

模型	方法	前 20 次最优验证 RMSE/MW	全程最优验证 RMSE/MW	测试集 RMSE/MW
LightGBM	贝叶斯优化	3.437	3.314	2.539
	随机搜索	3.528	3.494	2.895
	网格搜索	3.773	3.516	2.746
Transformer	贝叶斯优化	5.484	4.477	3.017
	随机搜索	6.562	4.993	4.454
	网格搜索	6.331	5.052	4.647

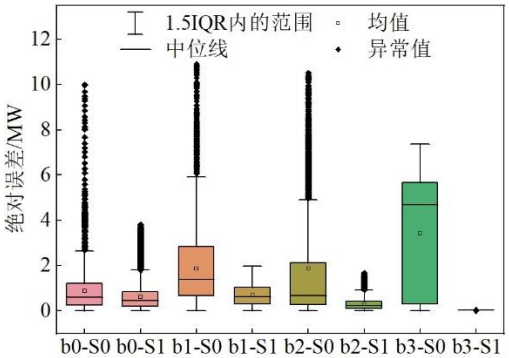
3.2 机组输出功率预测

机组在所有采样点中功率整体变化范围基本维持在 0~600 MW 附近，为表征机组在升降负荷过程中的相对进度，构建分段进度特征。将功率轨迹归一化并离散为 4 个连续的阶段，编号为 0~3，其中前 3 段基于 0~450 MW 等分，第 4 段为 450 MW 至实际最大功率的区间。把进度特征与经 MIC 筛选的数值特征共同用于预测机组输出功率。

根据按进度与状态进行分组后的测试集，绘制箱线图，其中 b0—b3 表示进度分段，S0 与 S1 分别表示非稳态与稳态，箱体的边界分别表示下四分位数和上四分位数，超出 1.5 倍四分位距范围的点记为异常值，具体如图 5 所示。



a) 启机升负荷过程 LightGBM 模型绝对误差分布



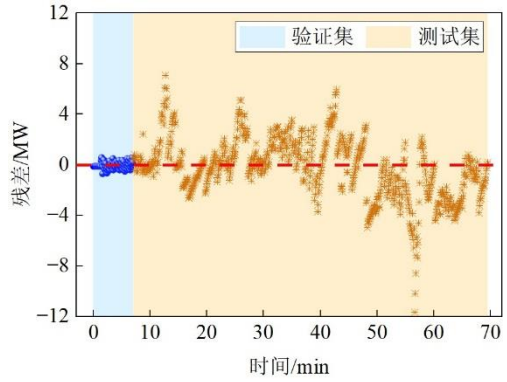
b) 停机降负荷过程 Transformer 模型绝对误差分布

图 5 启停不同进度与状态下的绝对误差箱线图
Fig.5 Box plots of absolute error under different progress and states of start-stop operation

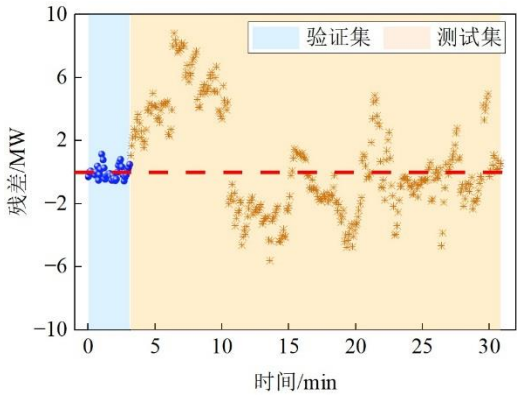
从图 5 可见：稳态部分误差更小且更集中，功率仅有小幅波动且其余运行参数稳定；而非稳态部分受参数变化与时序滞后的影响更大，误差不可避免地偏高。

因此重点优化非稳态阶段的数据，缩短启停时间的实际收益主要来自非稳态时段的压缩，稳态区间基本不受此类优化影响。在启停过程中分别选取耗时最长的两段非稳态区间进行分析，其在验证集和测试集上的残差分布如图 6 所示。

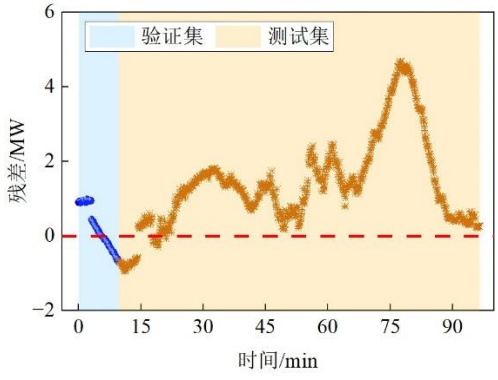
图 6 所示的非稳态区间中，预测基本贴合真实轨迹，残差基本维持在零线附近，说明模型精度足以作为启停过程运行优化的代理模型。



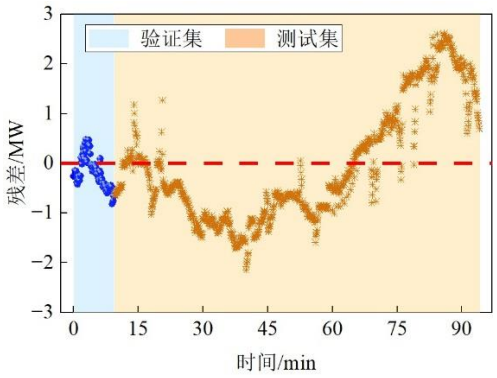
a) 启机升负荷 0~283 MW



b) 启机升负荷 449~536 MW



c) 停机降负荷 431~345 MW



d) 停机降负荷 552~496 MW

图 6 启停升降负荷过程残差随时间分布

Fig.6 Distribution of residuals over time during start-stop load process

为验证目标域分段划分策略的必要性，在相同模型结构、特征集合与超参数配置下，设置 2 种评估方案进行对比。方案 A 按稳态与非稳态连续区间分段，方案 B 不进行分段而直接按时间顺序划分验证集和测试集。2 种方案评估结果如表 5 所示。

表 5 不同数据集划分方法评估指标对比

Tab.5 Comparison of evaluation metrics for different dataset partitioning methods

方案	源域	目标域	过程	MAE/MW	RMSE/MW	R^2
方案 A	E1、E2	E3	启机	1.849	2.539	0.986
	E4、E5	E6	停机	1.838	3.017	0.959
	E1、E3	E2	启机	1.974	2.718	0.982
	E4、E6	E5	停机	2.097	3.347	0.953
	E2、E3	E1	启机	2.128	2.963	0.978
	E5、E6	E4	停机	2.189	3.583	0.948
方案 B	E1、E2	E3	启机	4.674	7.873	0.928
	E4、E5	E6	停机	4.238	8.267	0.883
	E1、E3	E2	启机	4.857	8.214	0.923
	E4、E6	E5	停机	4.419	8.617	0.878
	E2、E3	E1	启机	4.091	8.517	0.918
	E5、E6	E4	停机	4.712	8.963	0.869

表 5 对比结果表明，方案 A 在预测误差上显著优于方案 B，说明目标域内不同运行阶段的非平稳性与分布差异会导致跨工况混合误差，而分段划分能够提升跨月泛化能力并更贴近在线预测场景。

随后构建 XGBoost、LightGBM、RF（random forest）、Transformer 4 个模型，分别针对启停过程建模。4 个模型均采用贝叶斯优化确定超参数，使用一致的数据划分与特征集合，最终得到非稳态部分的整体预测效果对比如表 6 所示。由表 6 可知，LightGBM 在启机升负荷过程中的误差最低，MAE 和 RMSE 分别为 1.849、2.539 MW；Transformer 在

停机降负荷过程中的误差最低，MAE 和 RMSE 分别为 1.838、3.017 MW。

表 6 不同预测模型非稳态评估指标对比
Tab.6 Comparison of non-steady-state evaluation metrics across different prediction models

过程	模型	MAE/MW	RMSE/MW	R^2
启机升负荷	LightGBM	1.849	2.539	0.986
	XGBoost	1.985	2.728	0.972
	RF	2.955	4.178	0.952
	Transformer	3.620	5.067	0.937
停机降负荷	LightGBM	3.210	6.032	0.940
	XGBoost	3.216	7.584	0.911
	RF	3.393	6.507	0.929
	Transformer	1.838	3.017	0.959

结果表明，不同启停阶段对应的最优预测模型存在差异。启机阶段负荷及主要参数变化相对集中，LightGBM 更能刻画能功率与关键特征之间的分段非线性关系；停机阶段持续时间更长，受热惯性影响更明显，Transformer 更适于表征长时序依赖与滞后特征。因此，采用启停差异化建模能够获得更优预测性能。

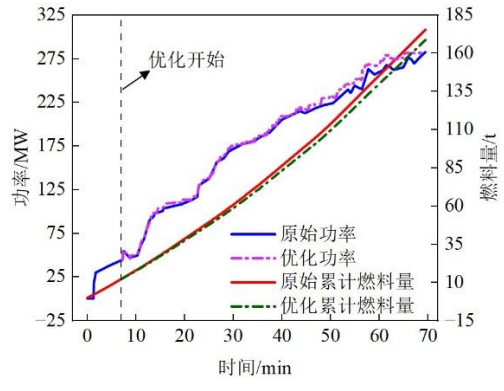
3.3 启停过程优化

由于数据集划分将每个片段后 90%作为测试集，因此优化从该时刻开始执行。实际运行中独立可控变量主要为燃料量、总风量与给水量，允许其取值在原始轨迹±5%范围内，其余被动变化参数仅允许在原始轨迹±1%附近小幅波动，并约束变负荷速率不超过原始速率的 105%。从 4 个过程中各取 2 个时刻，选取时刻的优化结果如表 7 所示。

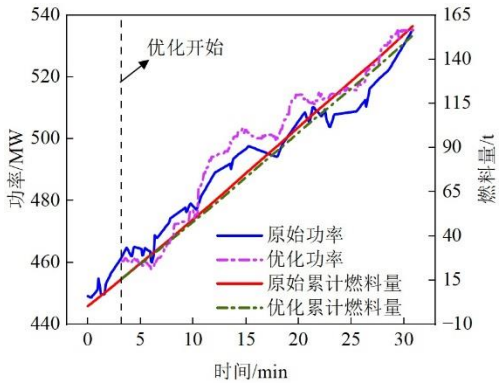
从表 7 中可以看出选取时刻的速率水平明显提高，燃料消耗明显下降，把每个时间点的数据向前滚动优化，按照得到的优化值轨迹输入预测模型中，可得到如图 7 和表 8 所示的整体优化结果。

表 7 启停升降负荷部分数据点优化结果
Tab.7 Optimization results of selected data points for start-stop variable load part

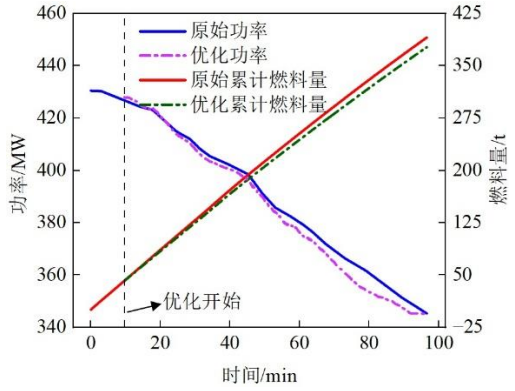
过程	所属片段	时刻	数据	总风量/(m^3 h^{-1})	给水量/(t h^{-1})	燃料量/(t h^{-1})	功率/MW	变负荷速率/(MW min^{-1})
启机升负荷	0~283 MW	第 37.25 min	实际值	1 280.35	909.49	156.40	190.25	11.53
			优化值	1 343.90	936.18	148.58	191.43	11.98
		第 65.25 min	实际值	1 388.63	816.19	190.74	269.34	10.33
			优化值	1 457.71	856.46	181.21	272.26	10.81
	449~536 MW	第 18.75 min	实际值	2 011.77	1 564.73	314.36	499.32	10.39
			优化值	1 976.96	1 503.04	298.65	503.52	10.90
		第 22.92 min	实际值	2 020.55	1 551.06	308.57	503.93	9.12
			优化值	1 966.69	1 474.86	293.14	510.22	9.55
停机降负荷	431~345 MW	第 48.25 min	实际值	1 699.10	1 269.80	240.79	393.10	–3.02
			优化值	1 705.31	1 220.27	229.38	391.35	–3.12
		第 66.75 min	实际值	1 631.40	1 215.00	233.04	372.81	–3.21
			优化值	1 648.25	1 226.07	221.77	369.18	–3.37
	552~496 MW	第 46.50 min	实际值	1 958.52	1 612.44	321.05	516.16	–2.74
			优化值	2 003.96	1 624.36	305.13	513.70	–2.87
		第 66.25 min	实际值	1 951.62	1 604.86	318.77	513.20	–1.84
			优化值	1 999.73	1 630.11	306.31	508.99	–1.92



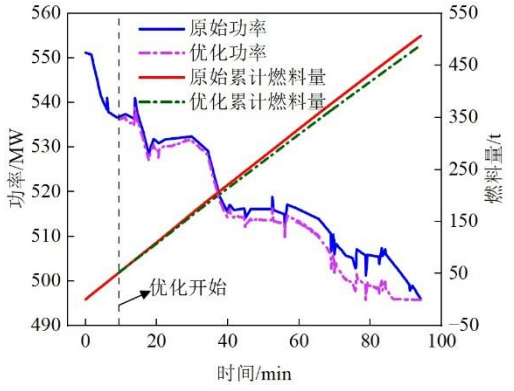
a) 启机升负荷 0~283 MW



b) 启机升负荷 449~536 MW



c) 停机降负荷 431~345 MW



d) 停机降负荷 552~496 MW

图 7 启停升降负荷整体优化结果
Fig.7 Overall optimization results for start-stop variable load

表 8 NSGA-II 整体优化结果
Tab.8 Overall optimization results of NSGA-II algorithm

过程	片段	原始时长/min	优化后时长/min	时间缩短/%	原始累计燃料量/t	优化后累计燃料量/t	燃料节省/%
启机升负荷	0~283 MW	69.42	65.58	5.53	175.44	168.77	3.80
	449~536 MW	30.83	29.92	2.95	158.85	153.62	3.29
停机降负荷	431~345 MW	96.58	91.83	4.92	390.45	376.54	3.56
	552~496 MW	94.08	86.58	7.97	506.48	488.83	3.48

从图 7 和表 8 中可以得出，机组按照优化后的运行参数轨迹运行可以实现启停升降负荷时间缩短和燃料消耗降低。其中升负荷两片段时间平均缩短 4.24%，燃料平均节省约 3.55%，降负荷两片段时间平均缩短约 6.45%，燃料平均节省 3.52%，进一步证明了该优化方法在工程中的实际意义。

3.4 工程应用讨论

本文给出的启停时间缩短比例和燃料消耗降低幅度，均基于机组 DCS 历史运行数据的离线仿真结果，反映了在既有运行边界内对启停方式进行调整所能获得的潜在优化空间。由于启停调峰相比于其他调峰方式频次更低，且受电网调度与安全运行要求约束较强，因此基于运行数据的离线仿真分析通常作为启停优化方法工程应用前的必要阶段。

在实际 DCS 中，可执行的给定值优化轨迹可作为运行决策的辅助参考使用，实时读取 DCS 关键测点与机组状态后，利用训练得到的预测模型对功率响应进行快速计算，并由 NSGA-II 在给定的安全约束内生成未来短时域的推荐给定值轨迹，但启停与深度调峰过程中仍存在多参数强耦合与运行特性波动现象，后续研究可通过开展长期运行监测更新模型，提升预测与优化系统的稳定性，在保证运行可靠性的前提下逐步验证节能与提速效果。

4 结 论

本文针对燃煤机组在深度调峰背景下启停响应慢、能耗高的难题，提出并验证了一种融合数据驱动建模与优化算法的方法，主要得出以下结论。

1）基于滑动窗口识别非稳态片段，并采用MIC筛选关键特征，分别构建启机升负荷LightGBM与停机降负荷Transformer预测模型并结合贝叶斯优化寻优。启机升负荷LightGBM对功率的MAE和RMSE分别为1.849、2.539 MW，停机降负荷Transformer对功率的MAE和RMSE分别为1.838、3.017 MW。

2）对比不同数据划分与不同模型结果表明，按非稳态区间分段的目标域评估策略可显著提升跨月泛化能力；同时，启机升负荷中LightGBM较XGBoost对功率的MAE和RMSE分别降低6.85%和6.93%，较RF分别降低37.43%和39.23%；停机降负荷中Transformer较LightGBM对功率的MAE和RMSE分别降低42.74%和49.98%，相比XGBoost分别降低42.85%与60.22%。说明选取的模型更适合对机组输出功率的预测。

3）将预测模型嵌入NSGA-II进行变负荷速率与燃料量双目标优化，离线仿真结果显示，在安全约束下可实现启停升降负荷过程时间平均缩短约5.34%，燃料消耗平均降低约3.53%，显著提升了机组调峰

的灵活性与经济性。

[参 考 文 献]

[1] 冀帅宇, 段立强, 王远慧, 等. 典型燃煤机组灵活调峰策略及性能研究[J]. 热力发电, 2023, 52(9): 94.
Ji S Y, Duan L Q, Wang Y H, et al. Research on flexible peak load regulation strategy and performance of typical coal-fired units[J]. Thermal Power Generation, 2023, 52(9): 94.

[2] 杜小泽, 张璇, 庞力平, 等. 燃煤发电机组灵活调峰下机炉安全状态监测与调控研究综述[J]. 中国电机工程学报, 2024, 44(18): 7178.
Du X Z, Zhang X, Pang L P, et al. Review on the monitoring and regulation of boiler and turbine safety states in coal-fired power generation units under flexible peak shaving[J]. Proceedings of the CSEE, 2024, 44(18): 7178.

[3] 马恒瑞, 袁傲添, 王波, 等. 基于深度学习的负荷预测研究综述与展望[J]. 高电压技术, 2025, 51(3): 1233.
Ma H R, Yuan A T, Wang B, et al. Review and prospect of load forecasting based on deep learning[J]. High Voltage Technology, 2025, 51(3): 1233.

[4] Dong C, Gunina M G. ANN-driven optimization and dynamic performance assessment of a hybrid energy system with enhanced SOFC and compressed air energy storage[J]. Energy, 2025, 328: 136448.

[5] Du H, Zhou X, Nord N, et al. Development of a robust data-driven surrogate model to improve energy flexibility of an integrated district heating system with a thermal storage system[J]. Energy, 2025, 336: 138398.

[6] Chen Z, Cui Y, Ren Z, et al. Data-driven approach of atrium skylights-integrated photovoltaic systems design based on multimodal deep learning: Considering different regions in China[J]. Energy, 2025, 335: 138099.

[7] Cestero J, Delle F C, Muro S K, et al. Optimizing energy management of smart grid using reinforcement learning aided by surrogate models built using physics-informed neural networks[J]. Applied Energy, 2025, 401: 126750.

[8] Wei Z, Tang Z, Chen S, et al. Scalable evaluation of demand response potential of HVAC systems: establishing comprehensive room-centric model library and surrogate models[J]. Building Simulation, 2025, 18(9): 2475.

[9] 陈东超, 吕佳乐, 祁琳, 等. 基于机器学习的燃煤发电机组运行参数寻优[J]. 工程热物理学报, 2025, 46(7): 2148.
Chen D C, Lü J L, Qi L, et al. Operating parameters optimization for coal-fired power unit based on machine learning algorithms[J]. Journal of Engineering Thermophysics, 2025, 46(7): 2148.

[10] 毛明轩, 冯心营, 陈思宇, 等. 基于贝叶斯优化卷积神经网络的路面光伏阵列最大功率点电压预测方法[J]. 中国电机工程学报, 2024, 44(2): 620.
Mao M X, Feng X Y, Chen S Y, et al. A novel maximum power point voltage forecasting method for pavement photovoltaic array based on Bayesian optimization convolutional neural network[J]. Proceedings of the CSEE, 2024, 44(2): 620.

[11] 毕光均, 龙世兴, 罗明兴, 等. 基于 SURF-BoVF 特征与主动学习 EQB-LightGBM 的发电机局部放电模式识别方法[J]. 高压电器, 2025, 61(10): 159.
Bi G J, Long S X, Luo M X, et al. Partial discharge pattern recognization method for generator based on SURF-BoVF features and active learning model EQB-LightGBM[J]. High Voltage Apparatus, 2025, 61(10): 159.

[12] 时帅, 张皓, 黄冬梅, 等. 考虑爬坡特征量的海上风电短期分区功率预测[J]. 太阳能学报, 2024, 45(12): 258.
Shi S, Zhang H, Huang D M, et al. Short-term zoning power prediction of offshore wind power considering climbing feature quantities[J]. Acta Energiae Solaris Sinica, 2024, 45(12): 258.

[13] Song A, Guo X, Wen W, et al. Improving accuracy and generalization in single kernel oil characteristics prediction in maize using NIR-HSI and a knowledge-injected spectral tab transformer[J]. Artificial Intelligence in Agriculture, 2025, 15(4): 802.

[14] 杨婷婷, 李浩千, 陈晓峰, 等. 基于 WPT-Transformer 的磨煤机故障预警研究[J]. 热力发电, 2023, 52(12): 180.
Yang T T, Li H Q, Chen X F, et al. Research on fault early warning of coal mill based on WPT and Transformer[J]. Thermal Power Generation, 2023, 52(12): 180.

[15] 孙宇翔, 陈丽, 龙波, 等. 复杂载荷下核级管路支架的多力学目标优化布置[J]. 动力工程学报, 2025(1): 45.
Sun Y X, Chen L, Long B, et al. Multi mechanical objective optimization layout for nuclear safety-related piping supports under complex loads[J]. Journal of Chinese Society of Power Engineering, 2025, 45(1): 45.

[16] 许越, 李强, 崔晖. 基于 MIC-EEMD-改进 Informer 的含高比例清洁能源与储能的电力市场短期电价多步预测[J]. 电网技术, 2024, 48(3): 949.
Xu Y, Li Q, Cui H. Short-term multi-step price prediction for the electricity market with a high proportion of clean energy and energy storage based on MIC-EEMD-improved Informer[J]. Power System Technology, 2024, 48(3): 949.

[17] 刘宇航, 顾煜炯, 郑庆帅, 等. 基于数据挖掘的直接空冷机组背压预测及优化[J]. 热力发电, 2023, 52(5): 127.
Liu Y H, Gu Y J, Zheng Q S, et al. Prediction and optimization of back pressure of direct air-cooled unit based on data mining[J]. Thermal Power Generation, 2023, 52(5): 127.

（责任编辑　邓玲惠）