

DOI: 10.19783/j.cnki.pspc.250930

基于数据驱动的可解释性低压台区理论线损率计算方法

蔡延华¹, 张浩博², 麦宇豪¹, 孟繁琪¹, 赵卓立², 彭显刚²

(1. 广东电网有限责任公司湛江供电局, 广东 湛江 524005; 2. 广东工业大学自动化学院, 广东 广州 510006)

摘要: 针对现有基于物理模型的理论线损率计算方法效率低难以在线应用, 而数据驱动方法可解释性不足、制约其工程推广应用等问题, 提出一种基于数据驱动的可解释性低压台区理论线损率计算方法。首先, 系统梳理了影响理论线损率物理机制的相关文献, 筛选出影响理论线损率的关联特征。其次, 构建以轻量级梯度提升机(light gradient boosting machine, LightGBM)为核心的计算模型, 采用改进灰狼优化算法(improved grey wolf optimizer, IGWO)进行超参数寻优, 以提升模型的计算精度。并引入 SHAP(shapley additive explanations)方法量化各特征对计算值的贡献, 揭示模型决策逻辑与理论线损率变化背后的物理损耗机理是否一致。最后, 基于真实台区数据集的算例验证所提方法的有效性。

关键词: 可解释性; 理论线损率计算; LightGBM; IGWO; SHAP

Data-driven interpretable method for calculating theoretical line loss rate in low-voltage distribution areas

CAI Yanhua¹, ZHANG Haobo², MAI Yuhao¹, MENG Fanqi¹, ZHAO Zhuoli², PENG Xiangang²

(1. CSG Guangdong Zhanjiang Power Supply Bureau, Zhanjiang 524005, China; 2. School of Automation, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: To address the limitations of existing theoretical line loss rate calculation methods, namely, the low efficiency and difficulty of online application in physics-based models, and the lack of interpretability in data-driven approaches that hinders their engineering deployment, this paper proposes an data-driven interpretable method for calculating theoretical line loss rates in low-voltage distribution areas. First, relevant literature on the physical mechanisms affecting theoretical line loss rate is systematically reviewed, and key influencing features are identified. Second, a calculation model is constructed with the light gradient boosting machine (LightGBM) as its core, and an improved grey wolf optimizer (IGWO) is employed for hyperparameter optimization to enhance model accuracy. Furthermore, the shapley additive explanations (SHAP) method is introduced to quantify the contribution of each feature to the calculated results, thereby revealing whether the model's decision logic aligns with the underlying physical mechanisms of line losses. Finally, case studies based on real-world distribution area datasets demonstrate the effectiveness of the proposed method.

This work is supported by the National Natural Science Foundation of China (No. 62273104).

Key words: interpretability; theoretical line loss rate calculation; LightGBM; IGWO; SHAP

0 引言

在现代电力系统中, 低压台区的电能损耗占配电网的 50%~60%^[1]。该损耗由技术线损主导, 其计算结果依赖于供电设备的物理参数、网架结构及

动态潮流负荷特性^[2]。准确量化由固有物理特性导致的损耗是分离非技术线损、客观评估运行经济性的基础, 也是制定降损策略及考核体系的先决条件^[3]。鉴于此, 开展理论线损率计算研究对提升供电企业精益化管理具有重要的工程应用价值^[4]。

现阶段, 理论线损率的计算方法主要分为两类^[5]: 基于物理模型的传统方法和基于人工智能的数据驱动方法。

基金项目: 国家自然科学基金项目资助(62273104); 广东电网有限责任公司科技项目资助(GDKJXM20231545)

基于物理模型的方法通过构建等效电路模型，依据基尔霍夫定律等原理进行潮流计算以解析损耗分布，如应用牛顿-拉夫逊法^[6]，或提出适应不同自动化水平的计算策略^[7]。然而，此类方法的有效性依赖于精确的物理建模。部分研究尝试量化三相不平衡度或负载率等特定指标对线损的影响^[8-9]，但在大规模、动态变化的配电网中，物理建模所依赖的逐时潮流迭代会产生计算负担，难以满足实时监测与高效治理等对时效性要求较高的应用场景^[10-11]。

为规避对精确物理参数与复杂迭代计算的依赖，数据驱动方法利用海量运行数据，学习电气特征与线损率间的非线性映射关系^[12-14]，从而绕过显式物理建模需求。无论是基于时序分析的长短期记忆网络^[13]，还是处理图像化数据的卷积神经网络^[14]，均展现出良好的计算精度，相较于高成本的物理潮流计算，训练后的模型仅需一次前向传播即可完成在线推算，计算速度上具备天然优势，适用于大规模工程场景。然而，此类模型具有“黑箱”属性，内在决策逻辑的不透明导致可解释性不足^[15-16]，限制了其在需要物理或因果依据支撑的降损决策及电网规划中的应用。

在上述背景的基础上，本文提出一种融合可解释性技术的数据驱动理论线损率计算方法，旨在协同提升理论线损率计算效率与可解释性。首先，通过对影响理论线损物理机制相关文献的梳理，筛选出影响理论线损率水平的关联特征；其次，采用改进灰狼优化算法(improved grey wolf optimizer, IGWO)对轻量级梯度提升机模型(light gradient boosting machine, LightGBM)进行超参数寻优，捕捉数据与理论线损率间的非线性映射关系以构建高效计算模型；最后，引入 SHAP(shapley additive explanations)方法揭示在物理损耗机理驱动下的理论线损率变化规律，对模型的决策过程进行归因。通过真实台区数据集的算例验证表明，所提方法在计算精度上优于与之对比的 7 种机器学习模型。此外，SHAP 可视化分析验证了模型的决策逻辑符合由线损物理成因驱动理论线损率变化的客观规律。

1 关联特征的选取及处理

1.1 关联特征选取

理论线损率为根据配电网设备参数、运行方式、潮流分布以及负荷情况等计算得出的理论线损电量占供电量的百分比^[2]，理论线损电量(分子)与供电量(分母)的非线性耦合关系间接决定理论线损率的大小。为构建数据驱动的理论线损率计算模型，本文通过梳理物理损耗传导机制，选取了 9 类能够

有效表征“损耗-供电量”比例变化的关联特征。这些特征通过影响线路阻抗、电流幅值及分布特性作用于线损水平，按时变特性可分为如下两类。

1) 动态运行变量：对应电网的实时运行状态与负荷特性，包括负载率^[16-18]、负荷形状系数^[18-19]、三相不平衡度^[20-22]、功率因数^[5,23]、供电量^[24-25]及首末端压降^[26-27]。

2) 静态物理属性：对应电网的固有设备与结构参数，包括供电半径^[28]、网架结构^[29-30]和用户总数^[24,31]。

特征与理论线损率的物理对应关系详见表 1。

表 1 关联特征指标体系

Table 1 Indicator system of association feature

| 特征 | 物理意义 |
|--------|--|
| 负载率 | 表征线路可变损耗。基于损耗与电流的平方关系，负载率升高时损耗电量(分子)增长速率快于供电量，导致理论线损率随负载率升高呈加速上升的非线性趋势。 |
| 负荷形状系数 | 描述负荷波动特性。总供电量相同时，负荷波动越剧烈，电流均方根值越大，导致线损电量(分子)增加，使得波动负荷台区的理论线损率高于平稳台区。 |
| 三相不平衡度 | 衡量负荷对称性的指标。不平衡工况引发的零序与负序电流会增加线路的电能损耗，但不产生有效供电量。在同等供电水平下，不平衡度越高，无效损耗占比越大，进而增大理论线损率。 |
| 功率因数 | 衡量有功传输效率。功率因数降低会增加传输相同有功电量所需的视在电流。受损耗与电流平方关系影响，无功分量增加直接提升物理损耗(分子)，导致理论线损率升高。 |
| 供电量 | 理论线损率的分母基准。由于电能损耗与电流的平方成正比，供电量增加会导致线路电流增大，进而使损耗的增长速率高于供电量。除轻载工况外，理论线损率通常随供电量增加而呈现上升趋势。 |
| 首末端压降 | 数值上近似等于电能损耗率。较大压降反映线路阻抗在电能传输过程中产生了较高的能量损失(对应分子增加)，直接体现了传输效率的降低及理论线损率的升高。 |
| 供电半径 | 决定理论线损基础水平的几何参数。供电半径直接决定线路电阻，在供电量及电流密度不变时，半径越长则等效电阻越大，导致物理损耗(分子)增加，进而抬高理论线损率基准。 |
| 网架结构 | 影响等效阻抗与分流比。网架结构影响等效电阻，优化网架可以缩短电流传输的平均电气距离，在同等负荷下降低损耗占比(减少分子)，可以实现理论线损率的降低。 |
| 用户总数 | 反映负荷分散度与接点损耗。用户数增加意味着接点电阻损耗增多且负荷分布变广，在增加物理损耗(分子)的同时，若供电密度未显著提升，将导致理论线损率呈结构性上升。 |

1.2 原始数据采集及处理

1.1 节选取的关联特征，数据基础源于用户用电信息采集系统提供的 15 min 级量测数据，涵盖配电变压器关口及用户侧的三相电压(u_A, u_B, u_C)、电流(i_A, i_B, i_C)、有功功率(p_A, p_B, p_C)及无功功率($q_A,$

q_B, q_C), 旨在构建动态运行变量, 同期线损管理系统则提供日级供电量数据。

为提升数据质量, 本文参照文献[14]对原始数据进行了清洗与修复。针对采集系统(15 min 级)与线损管理系统(日级)间的数据粒度差异, 将 15 min 级电气参数按日聚合为平均值, 以匹配日频分析需求。聚合公式为

$$H_{av,t} = \frac{1}{3} \sum_{F \in \{A,B,C\}} \left(\frac{1}{96} \sum_{j=1}^{96} h_{F,j,t} \right) \quad (1)$$

式中: $H_{av,t}$ 为第 t 日的平均物理量(电流 $I_{av,t}$ 、电压 $U_{av,t}$ 、有功功率 $P_{av,t}$ 、无功功率 $Q_{av,t}$); $h_{F,j,t}$ 为 F 相在第 t 日第 j 个采样点的量测值。

1.3 关联特征量化计算

本节基于 1.2 节预处理后的数据集, 对负载率、三相不平衡度、负荷形状系数、功率因数及首末端压降进行量化计算。静态物理属性与供电量数据分别取自电网台账与同期线损系统。

1) 负载率

负载率 η_t 是评估变压器经济运行状态的核心特征, 计算公式为

$$\eta_t = \frac{\sqrt{P_{av,t}^2 + Q_{av,t}^2}}{S_N} \quad (2)$$

式中: S_N 为配电变压器的额定容量。

2) 三相不平衡度

采用国标定义的电流不平衡度 ε_t 进行量化。

$$\varepsilon_t = \frac{\max(I_{A,t}, I_{B,t}, I_{C,t}) - \min(I_{A,t}, I_{B,t}, I_{C,t})}{\max(I_{A,t}, I_{B,t}, I_{C,t})} \quad (3)$$

式中: $I_{A,t}$ 、 $I_{B,t}$ 、 $I_{C,t}$ 分别为第 t 日 A、B、C 三相的日平均电流。

3) 负荷形状系数

负荷形状系数 K_t 刻画了日负荷曲线的波动形态, 定义为

$$K_t = \frac{\sqrt{\frac{1}{3} \sum_{F \in \{A,B,C\}} \left(\frac{1}{96} \sum_{j=1}^{96} i_{F,j,t}^2 \right)}}}{I_{av,t}} = \frac{I_{rms,t}}{I_{av,t}} \quad (4)$$

式中: $i_{F,j,t}$ 为 F 相在第 t 日第 j 个采样点的电流瞬时值; $I_{rms,t}$ 为三相电流的日均方根值。

4) 功率因数

日平均功率因数 $\cos \theta_t$ 计算公式为

$$\cos \theta_t = \frac{P_{av,t}}{\sqrt{P_{av,t}^2 + Q_{av,t}^2}} \quad (5)$$

此外, 工程中常采用日有功用电量 $W_{P,t}$ 与无功

用电量 $W_{Q,t}$ 计算日平均功率因数。

$$\cos \theta_t = \frac{W_{P,t}}{\sqrt{W_{P,t}^2 + W_{Q,t}^2}} \quad (6)$$

结果显示两者 $\cos \theta_t$ 的相关系数为 0.989, 对模型性能无实质性影响, 两种方式的功率因数输入模型的输出对比详见附录 A。

5) 首末端压降

首末端压降 ΔU_t 反映了台区供电范围内的电压质量。计算公式为

$$\Delta U_t = U_{source,t} - U_{end,t} \quad (7)$$

式中: $U_{source,t}$ 为变压器二次侧首端日平均电压;

$U_{end,t}$ 为典型末端用户日平均电压, 满足式(8)。

$$U_{end,t} = \min\{U_{av,t,1}, U_{av,t,2}, \dots, U_{av,t,S}\} \quad (8)$$

式中: $U_{av,t,s}$ 为台区内第 s 个末端用户的日平均电压; S 为该台区的用户总数。

2 数据驱动的理论线损率计算模型

为精确捕捉理论线损率与电气特征间的复杂非线性关系, 本文提出一种基于 IGWO 优化 LightGBM 超参数的计算模型。该模型利用 IGWO 的自适应寻优能力提升 LightGBM 的计算精度与泛化性能。本节将依次阐述 LightGBM 原理、IGWO 改进策略及 IGWO-LightGBM 模型的构建流程。

2.1 基于 LightGBM 的理论线损率计算基础模型

本文选取 LightGBM 作为计算理论线损率的基础框架。作为一种基于梯度提升决策树(gradient boosting decision tree, GBDT)的高性能机器学习算法^[32], LightGBM 的核心优势在于能高效处理高维度、多特征的电网数据, 并准确捕捉其间的非线性特征。模型性能的提升主要依赖两项技术。

1) 直方图

LightGBM 采用直方图算法(如图 1 所示)将连续特征离散化为有限区间, 将最优分裂点的搜寻转化为对直方图单元的遍历。该方法在保留关键信息的同时降低了计算复杂度, 能够满足海量连续型电气特征数据的实时处理需求。

2) 带深度限制的按叶子生长策略

区别于传统的按层生长策略, LightGBM 采用带深度限制的按叶子生长策略(如图 2 所示)。该机制优先扩展当前增益最大的节点, 而非均匀分裂同层节点。鉴于电气特征对理论线损率影响的非均衡性, 该策略能更精准地挖掘高增益分裂路径, 有效捕捉关键特征及其物理交互逻辑。

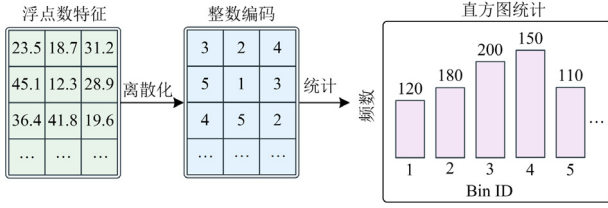


图1 直方图工作原理流程图

Fig. 1 Flowchart of the working principle of a Histogram.

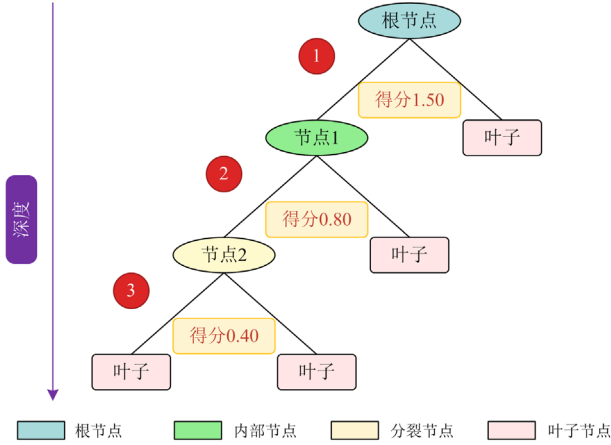


图2 按叶子生长策略分裂原理图

Fig. 2 Diagram of splitting principle based on leaf growth strategy

2.2 IGWO 超参数寻优算法

为提升 LightGBM 在理论线损率计算中的泛化性能, 需对其超参数空间进行高效搜索。本文选用灰狼优化算法 (grey wolf optimizer, GWO) 作为基础寻优框架^[33], 具备良好的局部搜索能力与收敛速度。GWO 通过模拟灰狼群体的社会等级与狩猎行为实现寻优, 其核心数学模型为

$$D = |C \cdot X_p(g) - X(g)| \quad (9)$$

$$X(g+1) = X_p(g) - A \cdot D \quad (10)$$

式中: $X(g)$ 为灰狼个体第 g 次迭代时的位置向量; $X_p(g)$ 为猎物 (由 α 、 β 、 δ 狼引导的最优解) 第 g 次迭代时的位置向量; D 为灰狼与猎物间的距离向量; A 和 C 为协同系数向量。针对标准 GWO 在处理复杂高维参数空间时易陷入局部最优及收敛失衡的问题, 本文从以下 3 个维度进行改进。

1) 基于 Tent 混沌映射的种群初始化

标准 GWO 采用的随机初始化策略难以保证初始种群 (超参数组合 $X(0)$) 在搜索空间中均匀分布。为此, 采用遍历性更优的 Tent 混沌映射生成初始种群, 增强搜索初期的多样性与覆盖率, 公式为

$$x_{g+1} = \begin{cases} 2x_g, & 0 \leq x_g < 0.5 \\ 2(1-x_g), & 0.5 \leq x_g \leq 1 \end{cases} \quad (11)$$

式中: x_g 为第 g 次迭代生成的混沌序列值。

2) 非线性收敛因子

系数向量 A 和 C 的动态变化对平衡算法的探索与开发至关重要, 其计算公式为

$$A = 2a \cdot r_1 - a \cdot E \quad (12)$$

$$C = 2 \cdot r_2 \quad (13)$$

式中: r_1 、 r_2 为 $[0,1]$ 范围内的随机向量; E 为与 r_1 同维度的全 1 向量。标准 GWO 采用线性收敛因子 a , 易导致全局探索不足。为此, 引入基于 Sigmoid 函数的非线性收敛因子 $a(g)$, 表示为

$$a(g) = a_s - \frac{a_s - a_e}{1 + \exp\left[-10\left(\frac{g}{G_{\max}} - 0.5\right)\right]} \quad (14)$$

式中: G_{\max} 为最大迭代次数; a_s 、 a_e 分别为收敛因子的起始与终止值, $a_s = 2$, $a_e = 0$ 。该策略使算法在前期维持较强的全局搜索能力, 后期则平滑过渡至精细化局部搜索。

3) 多样性保持与早熟停滞规避策略

为缓解后期种群多样性丧失的问题, 引入双重判据触发扰动机制: ① 个体聚集度判据, 即种群平均欧氏距离 c 低于阈值 σ (式(15)); ② 寻优停滞判据, 即全局最优解连续多代未更新。

$$c = \frac{1}{H} \sum_{u=1}^H \sqrt{\sum_{k=1}^h (p_{uk} - p_{vk})^2} < \sigma \quad (15)$$

式中: H 为种群规模; h 为参数空间维度; p_{uk} 、 p_{vk} 分别为第 u 、 v 个个体的第 k 维分量。一旦触发判据, 则对种群施加随机扰动以恢复多样性, 确保超参数全局最优解搜索更为充分。

2.3 IGWO-LightGBM 的理论线损率计算模型

IGWO-LightGBM 理论线损率计算模型的工作流程如图 3 所示。该模型架构以 IGWO 算法作为上层优化器, 并选取均方根误差 (root mean square error, RMSE) 作为自适应寻优的适应度函数值, 原因在于 RMSE 对计算偏差具有较高的敏感性, 能够引导 IGWO 在复杂的高维超参数空间中高效搜索, 从而找到使计算精度最优的参数组合。均方根误差 E_{RMSE} 计算公式为

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N_{\text{train}}} \sum_{i_{\text{train}}=1}^{N_{\text{train}}} (y_{i_{\text{train}}} - \hat{y}_{i_{\text{train}}})^2} \quad (16)$$

式中: N_{train} 为模型训练的样本总数; $\hat{y}_{i_{\text{train}}}$ 为模型对第 i_{train} 个训练样本计算出的理论线损率; $y_{i_{\text{train}}}$ 为对应训练样本的理论线损率目标值 (选取方法详见 4.1 节)。

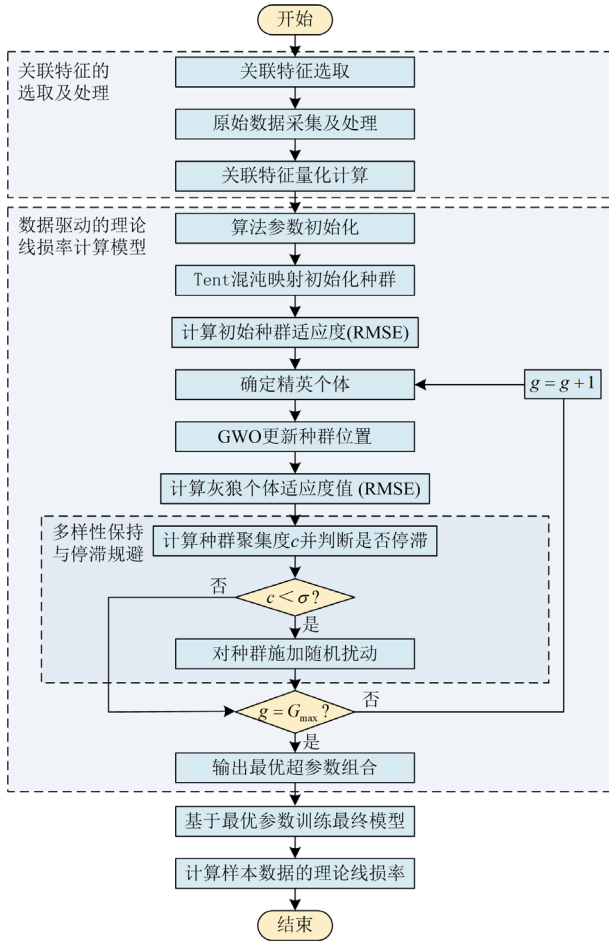


图3 IGWO-LightGBM 理论线损率计算流程图

Fig. 3 Flowchart of theoretical line loss rate calculation for the IGWO-LightGBM

3 基于 SHAP 的可解释机制

为解析 IGWO-LightGBM 模型的内部决策逻辑, 本文引入源于博弈论的 SHAP(shapley additive explanations)方法^[34]。该方法将模型输入特征类比为“博弈玩家”, 将计算过程视为“合作博弈”, 旨在公平量化各特征对计算值 $f(x)$ 与平均计算值 $E[f(x)]$ 之差(即博弈收益)的贡献程度。SHAP 通过计算特征在所有可能子集中的加权平均边际贡献, 有效消除了特征输入顺序的影响, 确保了贡献度评估的唯一性与公平性。

SHAP 框架通过构建一个线性的加性特征归因模型, 将单样本计算值 $f(x)$ 分解为各特征 SHAP 值的线性叠加。

$$f(x) = \phi_0 + \sum_{m=1}^M \phi_m \quad (17)$$

式中: $f(x)$ 为模型对样本 x 的最终计算值; M 为特征总数; ϕ_0 为模型的基准值, 代表全体样本的平均

计算水平; ϕ_m 为第 m 个特征对本次计算贡献的 SHAP 值, 量化了第 m 个特征将结果从基准值 ϕ_0 推向最终值 $f(x)$ 的贡献量。若 $\phi_m > 0$, 表明该特征对计算结果具有正向抬升作用, 反之则具有抑制作用; 绝对值 $|\phi_m|$ 衡量了贡献的大小, 该过程如图 4 所示。

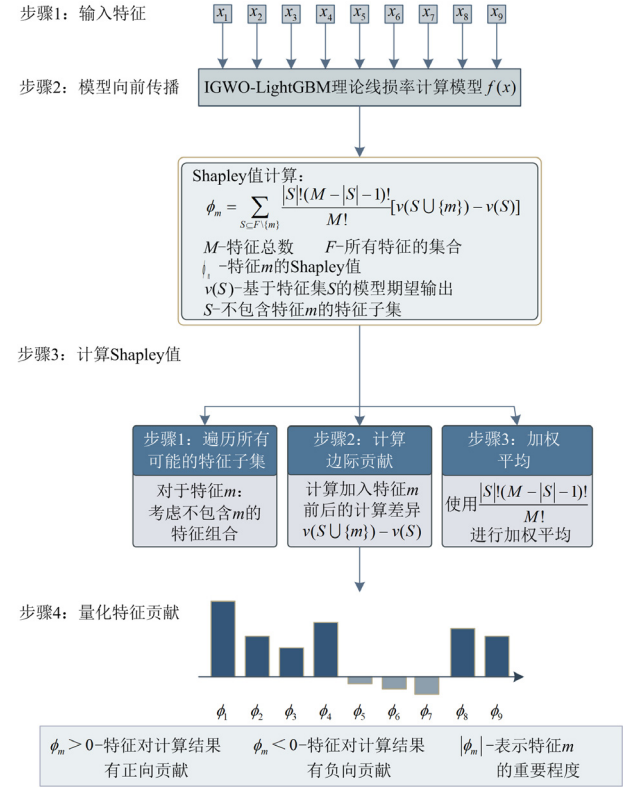


图4 SHAP 值计算步骤图

Fig. 4 Calculation steps of SHAP values

针对集成树模型计算复杂度随特征数量呈指数级增长的问题, 本文采用高效的 TreeSHAP 算法。该算法通过利用 LightGBM 的树结构特性, 将计算复杂度从指数级 $O(T \cdot 2^M)$ 降低至多项式级 $O(T \cdot L \cdot D^2)$ 。其中, T 、 L 、 D 分别对应树的数量、最大叶子节点数及深度。TreeSHAP 获取的 ϕ_m 值为模型决策提供了量化的可追溯解释路径。

4 算例分析

4.1 数据集

本文选取某市电网 8000 个低压配电台区的真实运行数据作为算例样本。数据集时间跨度为 2021 年 4 月 1 日至 15 日, 包含 15 min 采样频率的动态电气参数(三相电压、电流、有功及无功功率)及日级供电量数据。静态物理属性(供电半径、网架结构、用户总数)直接调取自电网台账, 相关原始数据采集

与特征量化流程详见本文 1.2 节和 1.3 节。

为构建数据驱动模型的训练“真值”并实现理论线损率的高效逼近, 本文严格依据《配电网理论线损计算导则》(GB/T 44768-2024)^[2]规定的低压交流配电网理论线损计算方法, 计算得到各台区每日的理论线损电量, 进而获得理论线损率。实验数据集按时间先后顺序进行划分: 前 10 天的数据及其导则计算结果作为训练集, 用于驱动模型学习; 后 5 天的数据作为测试集, 利用对应的导则计算结果作为精度评估的参考基准, 验证所提模型的计算性能。

4.2 性能指标

为评估模型及对比算法的计算精度与拟合优度, 本文选取 RMSE、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^2)作为性能评价指标, 表示为

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i_{\text{test}}=1}^{N_{\text{test}}} (y_{i_{\text{test}}} - \hat{y}_{i_{\text{test}}})^2} \quad (18)$$

$$E_{\text{MAE}} = \frac{1}{N_{\text{test}}} \sum_{i_{\text{test}}=1}^{N_{\text{test}}} |y_{i_{\text{test}}} - \hat{y}_{i_{\text{test}}}| \quad (19)$$

$$E_{\text{MAPE}} = \frac{1}{N_{\text{test}}} \sum_{i_{\text{test}}=1}^{N_{\text{test}}} \frac{|y_{i_{\text{test}}} - \hat{y}_{i_{\text{test}}}|}{y_{i_{\text{test}}}} \times 100\% \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i_{\text{test}}=1}^{N_{\text{test}}} (y_{i_{\text{test}}} - \hat{y}_{i_{\text{test}}})^2}{\sum_{i_{\text{test}}=1}^{N_{\text{test}}} (y_{i_{\text{test}}} - \bar{y})^2} \quad (21)$$

式中: N_{test} 为测试集的样本总数; $\hat{y}_{i_{\text{test}}}$ 为模型对第 i_{test} 个测试样本的理论线损率计算值; $y_{i_{\text{test}}}$ 为对应测试样本的理论线损率参考值; \bar{y} 为所有样本理论线损率参考值的平均值。在理论线损率计算中, RMSE 对较大误差具有高敏感性, 用以识别极端工况下的计算偏差; MAE 反映计算值与参考值之间的平均绝对偏差; MAPE 从相对误差维度衡量模型在不同线损水平台区的普适性; R^2 表征模型对理论线损率波动的整体拟合优度。前 3 项指标值越小, 且 R^2 越接近 1, 表明模型的计算精度与综合性能越优。

4.3 模型参数设置

为保证实验的公平性与可复现性, 本节详述 IGWO-LightGBM 模型的参数配置, 包括 IGWO 算法控制参数与 LightGBM 超参数搜索空间两部分。IGWO 的控制参数基于多次预实验, 以算法收敛行为和初步解质量为评估依据, 在综合平衡收敛速度与全局搜索能力的前提下确定具体参数数值, 详见

表 2。LightGBM 的 9 个超参数则通过 IGWO 在其预设范围内进行自适应寻优, 各参数的具体搜索空间分布及最终寻优结果详见表 3。

表 2 IGWO 算法参数

Table 2 IGWO algorithm parameters

| 特征 | 灰色关联度平均值 |
|---------|----------|
| 种群规模 | 20 |
| 迭代次数最大值 | 50 |
| 混沌次数最大值 | 5 |
| 收敛因子初始值 | 2 |
| 收敛因子结束值 | 0 |

表 3 LightGBM 模型参数表

Table 3 LightGBM model parameters

| 参数 | 取值范围 | 寻优结果 |
|-------------------|------------|-------|
| num_leaves | [20,100] | 40 |
| max_depth | [5,15] | 12 |
| learning_rate | [0.01,0.1] | 0.064 |
| n_estimators | [50,500] | 495 |
| min_child_samples | [20,100] | 21 |
| subsample | [0.6,1.0] | 0.84 |
| colsample_bytree | [0.6,1.0] | 0.73 |
| reg_alpha | [0,1.0] | 0 |
| reg_lambda | [0,1.0] | 0 |

4.4 计算结果分析

4.4.1 优化算法对比

为检验本文所提 IGWO 算法在超参数寻优中的性能, 将其与 GWO、贝叶斯优化算法(Bayesian optimization, BO)及 CatCMA 优化算法^[35]进行对比。为确保实验公平, 各算法迭代次数统一设为 50 次(与表 2 中一致), 余项参数遵循算法推荐配置。此外, 为检验算法在不同梯度提升决策树模型上的泛化能力, 将 4 种优化算法分别应用于 LightGBM、XGBoost(eXtreme gradient boosting)模型的理论线损率计算任务中。如图 5 所示, 在不同模型中, 经 IGWO 优化的模型组合在不同框架下均取得最优性能。其中, IGWO-LightGBM 的 RMSE、MAE 和 MAPE 指标均为最低, 且 R^2 高达 0.9950; 而在 XGBoost 上, IGWO-XGBoost 的表现同样优于其他对比组合。结果表明, IGWO 凭借其改进的搜索机制, 能够在高维参数空间中稳定定位全局更优解, 在不同梯度提升模型上均展现出较强的泛化能力。

4.4.2 计算模型对比

为验证所提 IGWO-LightGBM 模型在理论线损率计算任务上的精度, 选取标准 LightGBM、CatBoost (Categorical Boost)、支持向量回归(support vector regression, SVR)、随机森林(random forest, RF)、XGBoost、反向传播神经网络(back propagation neural

network, BP)及线性回归(linear regression, LR)7 种机器学习模型作为对比基准。如图 6 所示, 本文模型在各项评价指标上均取得最优性能, 其 RMSE、MAE 和 MAPE 均为最低, R^2 最接近于 1, 表明拟合精度最优。其中, 集成学习模型普遍优于 LR 等简单模型, 侧面印证了理论线损率与其物理影响因素间存在复杂的非线性映射关系, 需采用非线性模型进行拟合。

为深入剖析本文模型的性能, 图 7—图 9 从拟合优度、分布一致性和误差稳定性 3 个维度对计算结果进行可视化检验。

图 7 通过散点图展示了本文模型的计算值与理

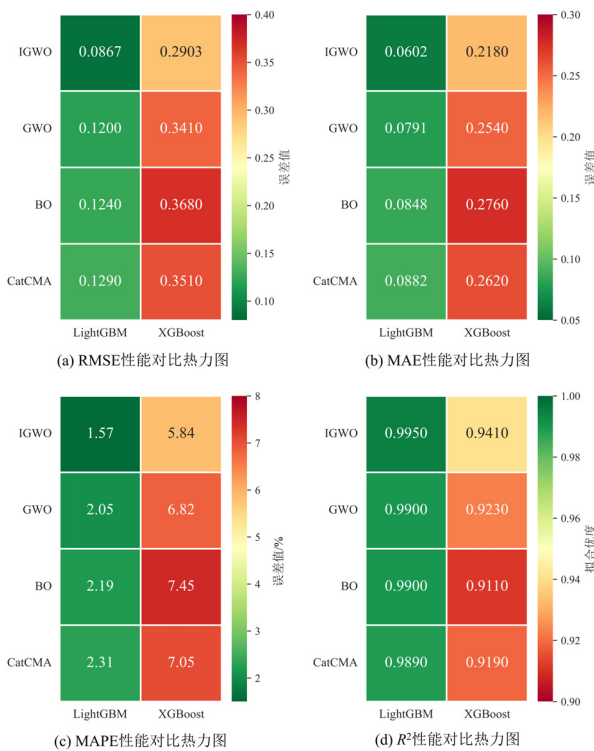


图 5 不同优化算法下的模型性能对比图

Fig. 5 Comparison of model performance under different optimization algorithms

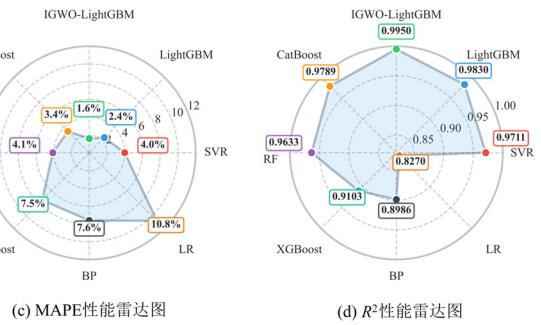
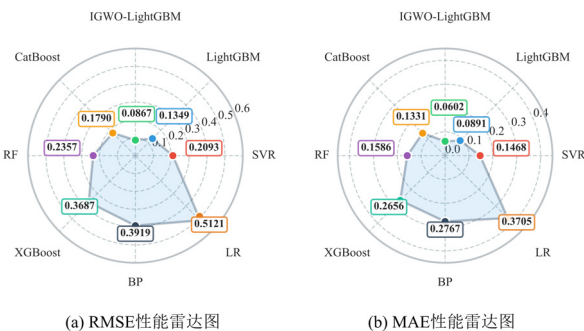


图 6 不同理论线损率计算模型性能对比图

Fig. 6 Comparison of calculation models' performance of different theoretical line loss rate

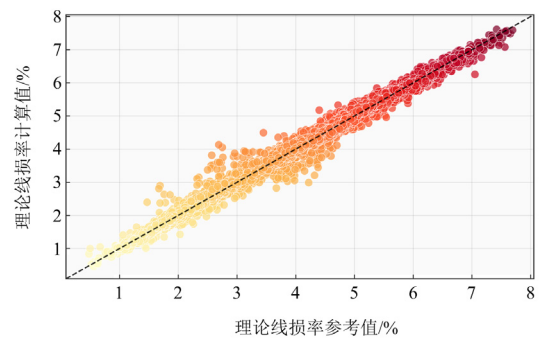


图 7 理论线损率计算值与参考值的散点分布图

Fig. 7 Scatter distribution diagram of calculated versus reference theoretical line loss rates

论线损率参考值之间的拟合关系。图中每个散点代表一个样本, 若参考值与模型计算值一致, 散点应严格分布于图中对角线 $y = x$ 上。图 7 中绝大多数散点紧密沿对角线分布, 表明模型具备优良的拟合能力, 计算结果与参考值高度吻合。

图 8 从统计分布的角度检验模型计算值与参考值总体分布的一致性。图中两条曲线分别为全部样本的模型理论线损率计算值与参考值的概率密度曲线。纵坐标“概率密度”通过核密度估计方法生成, 该方法与传统的直方图相比, 避免了因分箱位置和宽度选择不当引入的主观偏差, 能真实、稳定地揭示数据背后潜在分布形态。由图 8 可知, 计算值与参考值的两条概率密度曲线高度重合。为量化评估两者的分布一致性, 采用 Kolmogorov-Smirnov(K-S) 检验进行统计推断。该检验以“两组数据分布无显著差异”为原假设, 通过计算累积分布函数间的最大距离来判定其是否同源。K-S 检验的 P 值高达 0.9972, 远大于常规阈值 0.05, 表明统计学上无法拒绝原假设。证实了该模型不仅复现了数据的中心趋势, 还准确再现了理论线损率的整体波动规律(分布形状)。

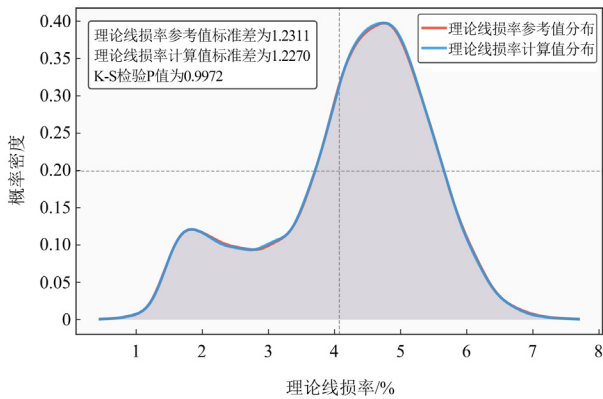


图8 理论线损率计算值与参考值的概率密度分布对比

Fig. 8 Comparison of probability density distributions between calculated and reference theoretical line loss rates

另外，模型的误差稳定性也得到了验证。图9绘制了模型绝对误差与模型理论线损率计算值的关系散点图。结果显示，误差在整个计算区间内均被控制在较低水平，未出现随计算值增大而发散或收敛的趋势(即无明显异方差现象)，验证了模型在不同区间均具备稳健的计算能力。

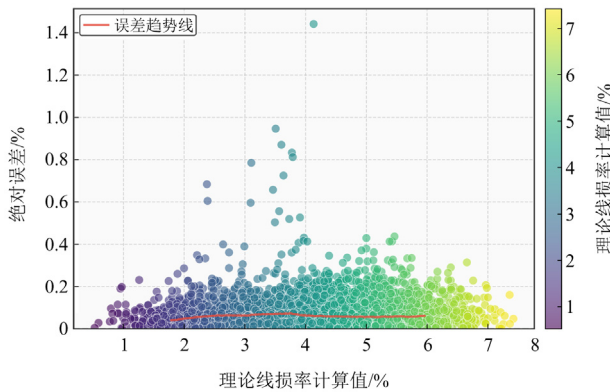


图9 绝对误差与理论线损率计算值关系图

Fig. 9 Distribution of model absolute error across calculated values

此外，为验证模型在大规模实时计算中的可行性，对比基于导则的传统计算方法与本文方法在测试集上的计算时间。如表4所示，本文模型处理40 000个样本的时间为0.079 s，低于传统方法，展现了其满足大规模台区在线实时计算的高效性能。

表4 不同方法的理论线损率计算时间对比

Table 4 Comparison of computational time for theoretical line loss rate using different methods

| 方法 | 计算时间/s |
|------|-----------|
| 传统方法 | 21 869.16 |
| 本文模型 | 0.079 |

为评估模型推算理论线损电量的工程实用性，将模型在测试集输出的理论线损率乘以当日供电量，得出各台区的理论线损电量，并与基于导则计算的同期参考值进行对比。虽然理论线损率与理论线损电量的相对误差在数学上等效，但电量绝对指标的准确量化对电费核算与节能改造具有更直接的指导意义。如图10所示，本文模型在理论线损计算中精度最高，相对误差在[0,1%]内的样本占比为59.8%，优于次优模型的50.5%。此外，随机选取100个台区在2021年4月11日至15日期间的数据绘制拟合曲线(图11)，结果显示计算值与参考值的变化趋势高度一致。同时，8000个台区的理论线损散点图(图12)表明，样本点紧密分布于对角线附近。

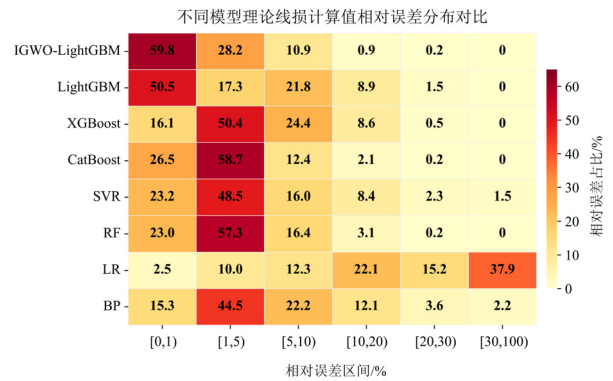


图10 不同模型理论线损计算值相对误差分布对比图

Fig. 10 Comparison of relative error distributions in theoretical line loss calculations across different models

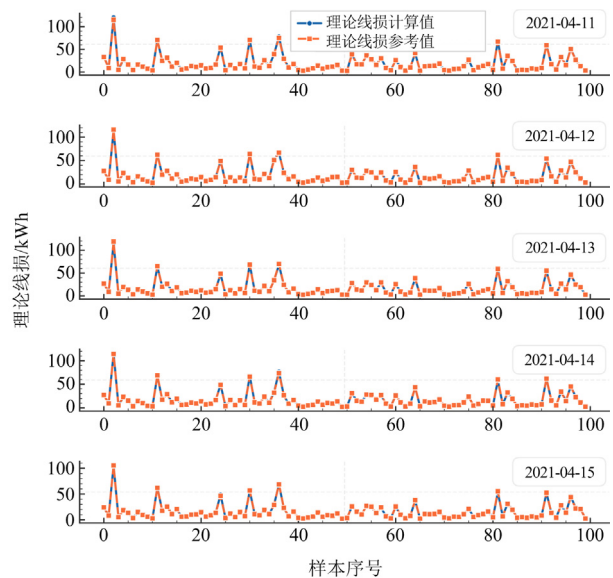


图11 随机100个台区理论线损值拟合曲线

Fig. 11 Fitting curves of theoretical line loss values for 100 randomly selected transformer districts

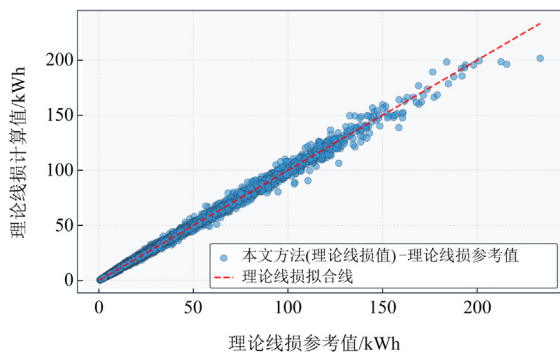


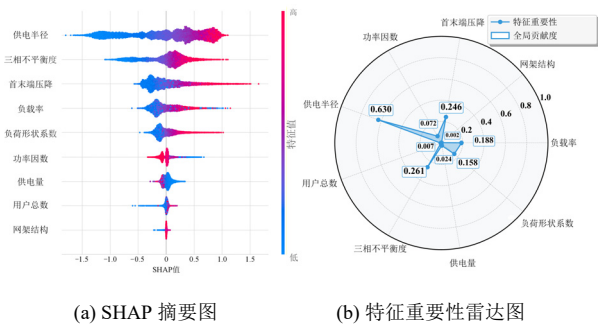
图 12 理论线损计算值与参考值散点分布

Fig. 12 Scatter distribution of theoretical line loss calculated values vs. reference values

进一步证实, 模型不仅能精准拟合理论线损率, 在计算理论线损电量时也具备良好的泛化稳定性。

4.5 可解释性分析

为验证模型的决策逻辑与物理机制的一致性, 采用 TreeSHAP 方法进行全局可解释性分析。图 13 分别展示了各特征对理论线损率的全局影响分布及其按 SHAP 绝对平均值排序的重要性权重。结果显示, 供电半径、三相不平衡度、首末端压降、负载率及负荷形状系数是影响计算结果的前 5 大核心特征。值得注意的是, 供电量的 SHAP 贡献度较低, 这具有深层的物理内涵: 理论线损率本质上是损耗占比的相对指标, 在模型充分捕获阻抗与电流分布等关键特征的前提下, 供电量的绝对数值对该比例的直接影响的有限。这证明模型并不简单依赖总量数据, 而是有效识别了决定线损水平的关键驱动因子。



(a) SHAP 摘要图 (b) 特征重要性雷达图

图 13 关联特征的全局重要性及影响分布

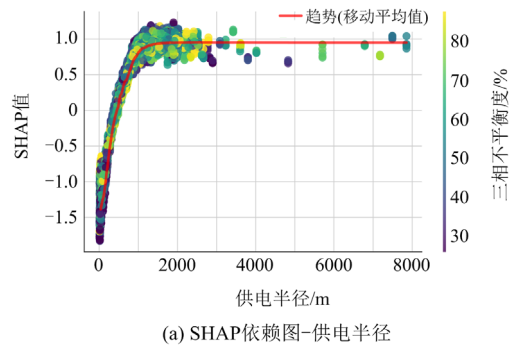
Fig. 13 Global importance and impact distribution of associated features

图 13(a)中高特征值(红点)普遍对应正向 SHAP 值, 即特征值增大会推高理论线损率, 这与基本物理规律高度契合。从物理机制看, 供电半径增加意味着线路电阻增大, 压降升高则反映了电流或阻抗的提升。依据焦耳定律, 这些因素将促使线损电量

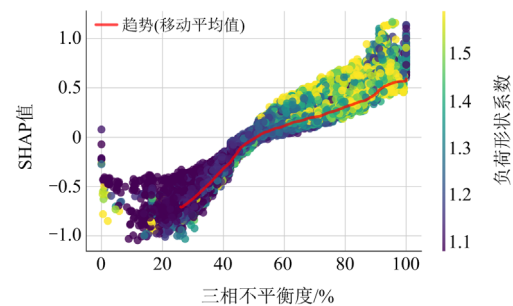
呈非线性增长, 进而在供电量变化相对平缓时抬升理论线损率水平。模型准确反映了功率因数对理论线损率的抑制作用: 功率因数越高(趋近于 1), 其 SHAP 值往往为负。这是由于高功率因数减小了传输等额有功功率所需的总电流, 从而有效降低了物理损耗及对应的理论线损率。

除单特征分析外, 本文进一步采用 SHAP 依赖图探究特征间的非线性交互效应。该方法以目标特征值为横轴, 对应的 SHAP 值为纵轴, 并引入颜色映射来表征第二特征的数值分布, 从而直观量化特征间的协同作用。若散点颜色呈现出规律性的色阶渐变或分层, 则表明两特征间存在显著的交互耦合关系。

图 14(c)展示了首末端压降与供电半径的协同效应: 当压降较大且供电半径较长时, 二者对理论线损率的正向贡献显著增强。这一现象符合焦耳定律中电流平方项与电阻项的乘积关系, 体现出“重载”与“长线路”叠加时损耗加剧的物理效应, 使得理论线损率在同等供电水平下呈现大于单一因素作用之和的增长幅度。图 14(b)与图 14(d)揭示了三相不平衡度与负荷形状系数之间的交互关系: 当系统同时存在较高不平衡度(三相电流不均)和高形状系数(负荷峰谷差大)时, 其对理论线损率的提升作用强于单独影响之和, 其物理本质在于: 空间与时间维度的负荷特性叠加加剧了电流有效值的偏离, 导致均方根电流显著大于平均电流, 促使线损电量突破线性增长边界, 呈现加速上升态势。



(a) SHAP 依赖图-供电半径



(b) SHAP 依赖图-三相不平衡度

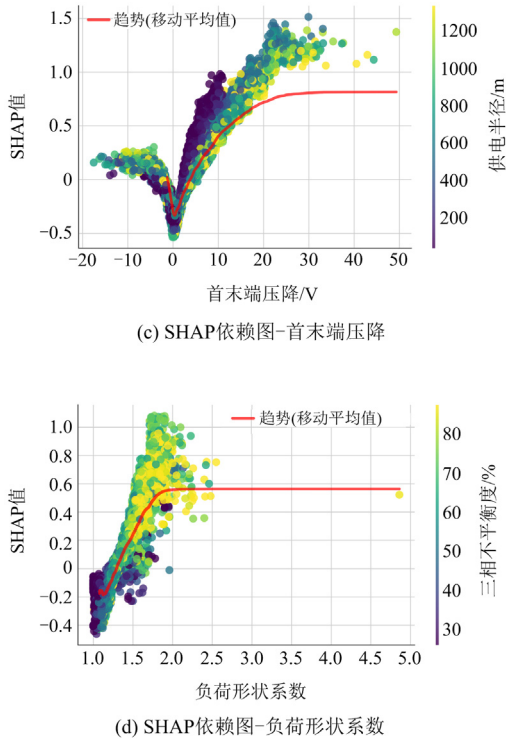


图 14 关键特征间的 SHAP 依赖图

Fig. 14 SHAP dependency graph between key features

模型还捕捉到实际电网运行中的边际效应递减现象。如图 14(a)与图 14(d)所示, 当供电半径或负荷形状系数超过特定阈值后, 其 SHAP 贡献趋于平缓, 这是由于在电网实际运行中, 极长线路或极端负荷峰值常受限于电压质量约束、保护定值或无功补偿能力, 使得损耗的实际增量趋于饱和。模型能够学习到这一非线性饱和特性, 说明其能隐式拟合电网运行的物理边界。

上述分析表明, 模型有效提取了特征间的非线性交互关系, 所构建的数据映射关系反映了底层的物理损耗机理。特征归因结果与客观物理规律的契合, 证明模型的高计算精度源自对台区电气运行特性的合理表征, 而非对训练样本的过拟合。

5 结论

本文针对低压台区理论线损率计算中物理模型效率低、数据驱动方法可解释性不足及在工程应用中受限的问题, 提出一种融合可解释性技术的数据驱动理论线损率计算方法。主要研究工作与结论如下。

1) 筛选出影响理论线损率水平的关联电气特征, 通过融合台区静态拓扑属性与动态运行变量, 为模型构建提供特征基础。

2) 提出一种 IGWO 与 LightGBM 融合的理论线损率计算模型。通过引入 Tent 混沌映射、非线性收敛因子及多样性保持策略, 提升了 IGWO 在 LightGBM 超参数空间中的寻优能力。算例分析表明, 本文模型的计算精度优于其他对比模型。

3) 引入 SHAP 方法实现了对模型决策过程的量化解释。分析结果表明, 模型的决策逻辑与基于线损物理成因驱动理论线损率变化的客观规律一致。

后续研究可集中于融合更多维度的异构数据以进一步提升模型性能, 并将此可解释计算方法拓展至异常诊断与降损策略评估等应用场景。

附录 A

模型性能指标如表 A1 所示, 两种功率因数输入方式下模型均保持优异精度, R^2 均高于 0.99, MAPE 均低于 1.8%, 各项指标高度接近。

表 A1 不同功率因数计算方式下模型的性能对比

Table A1 Performance comparison of the model under different

| power factor calculation methods | | | | |
|---|--------|--------|--------|--------|
| 功率因数公式 | RMSE | MAE | MAPE/% | R^2 |
| $\cos \theta_l = \frac{W_{P,d}}{\sqrt{W_{P,d}^2 + W_{Q,d}^2}}$ | 0.0952 | 0.0673 | 1.75 | 0.9939 |
| $\cos \theta_l = \frac{P_{av,d}}{\sqrt{P_{av,d}^2 + Q_{av,d}^2}}$ | 0.0867 | 0.0602 | 1.57 | 0.9950 |

参考文献

- [1] 谢荣斌, 杜帆, 程湘, 等. 三相不平衡及谐波对三相四线低压配电网线损的影响[J]. 电力系统保护与控制, 2020, 48(21): 22-30.
XIE Rongbin, DU Fan, CHENG Xiang, et al. Influence of three-phase imbalance and harmonic on line loss of three-phase four-wire low-voltage distribution network[J]. Power System Protection and Control, 2020, 48(21): 22-30.
- [2] 中国电力企业联合会. 配电网线损理论计算导则: GB/T 44768—2024[S]. 北京: 中国标准出版社, 2024.
- [3] 张明慧, 张逸, 刘必杰, 等. 考虑供用电双方无功补偿潜力的省级技术降损互动管控平台[J]. 电力系统保护与控制, 2023, 51(23): 160-169.
ZHANG Minghui, ZHANG Yi, LIU Bijie, et al. Provincial technology loss reduction interactive management and control platform considering reactive power compensation potential of power supply and consumption sides[J]. Power System Protection and Control, 2023, 51(23): 160-169.
- [4] 李端超, 王松, 黄太贵, 等. 基于大数据平台的电网线损与窃电预警分析关键技术[J]. 电力系统保护与控制, 2018, 46(5): 143-151.

- LI Duanchao, WANG Song, HUANG Taigui, et al. Key technologies of line loss and stealing electricity prediction analysis based on big data platform[J]. *Power System Protection and Control*, 2018, 46(5): 143-151.
- [5] 汪颖, 李龙, 陈韵竹, 等. 基于核心关联指标和虚拟等值电阻的台区理论线损计算[J]. *电力系统自动化*, 2024, 48(23): 121-135.
- WANG Ying, LI Long, CHEN Yunzhu, et al. Theoretical calculation of line loss in distribution station areas based on core correlation indicators and virtual equivalent resistance[J]. *Automation of Electric Power Systems*, 2024, 48(23): 121-135.
- [6] 刘庭磊, 王韶, 张知, 等. 采用负荷电量计算低压配电台区理论线损的牛拉法[J]. *电力系统保护与控制*, 2015, 43(19): 143-148.
- LIU Tinglei, WANG Shao, ZHANG Zhi, et al. Newton-Raphson method for theoretical line loss calculation of low-voltage distribution transformer district by using the load electrical energy[J]. *Power System Protection and Control*, 2015, 43(19): 143-148.
- [7] 张恺凯, 杨秀媛, 卜从容, 等. 基于负荷实测的配电网理论线损分析及降损对策[J]. *中国电机工程学报*, 2013, 33(增刊 1): 92-97.
- ZHANG Kaikai, YANG Xiuyuan, BU Congrong, et al. Theoretical line loss analysis and loss reduction countermeasures of distribution network based on load measurement[J]. *Proceedings of the CSEE*, 2013, 33(S1): 92-97.
- [8] 郝思鹏, 蔡欣灵, 张仰飞, 等. 三相不平衡与线损的量化分析[J]. *电网技术*, 2021, 45(4): 1547-1552.
- HAO Sipeng, CAI Xinling, ZHANG Yangfei, et al. Quantitative analysis between three-phase unbalance and line losses[J]. *Power System Technology*, 2021, 45(4): 1547-1552.
- [9] 乔辉, 韩璟琳, 李征, 等. 采用日累计发电量重构光伏出力曲线的线损精确计算方法[J]. *电力系统保护与控制*, 2024, 52(24): 161-168.
- QIAO Hui, HAN Jinglin, LI Zheng, et al. An accurate calculation method of line loss based on reconstructing a photovoltaic output curve by daily cumulative power generation[J]. *Power System Protection and Control*, 2024, 52(24): 161-168.
- [10] ZHU M, XU C, DONG S, et al. An integrated multi-energy flow calculation method for electricity-gas-thermal integrated energy systems[J]. *Protection and Control of Modern Power Systems*, 2021, 6(1): 1-12.
- [11] GORBACHEV S, GUO J, MANI A, et al. MPC-based LFC for interconnected power systems with PVA and ESS under model uncertainty and communication delay[J]. *Protection and Control of Modern Power Systems*, 2023, 8(4): 1-17.
- [12] 荣雪琴, 丁莹莹, 刘勇. 基于数据驱动与相关性的电能误差分析方法研究[J]. *电测与仪表*, 2025, 62(1): 101-109.
- RONG Xueqin, DING Yingying, LIU Yong. Research on data-driven and correlation-based electric energy error analysis method[J]. *Electrical Measurement & Instrumentation*, 2025, 62(1): 101-109.
- [13] 周王峰, 李勇, 郭钊秀, 等. 基于 DAE-LSTM 神经网络的配电网日线损率预测[J]. *电力系统保护与控制*, 2021, 49(17): 48-56.
- ZHOU Wangfeng, LI Yong, GUO Yixiu, et al. Daily line loss rate forecasting of a distribution network based on DAE-LSTM[J]. *Power System Protection and Control*, 2021, 49(17): 48-56.
- [14] HU W, GUO Q, WANG W, et al. Loss reduction strategy and evaluation system based on reasonable line loss interval of transformer area[J]. *Applied Energy*, 2022, 306.
- [15] AHSAN F, DANA N H, SARKER S K, et al. Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review[J]. *Protection and Control of Modern Power Systems*, 2023, 8(3): 1-42.
- [16] JAZEBI S, MOGHIMI HAJI M, NAGHIZADEH R A. Distribution network reconfiguration in the presence of harmonic loads: optimization techniques and analysis[J]. *IEEE Transactions on Smart Grid*, 2014, 5(4): 1929-1937.
- [17] 陈光宇, 张盛杰, 杨里, 等. 基于多维场景划分的台区线损率异常研判及关联用户精准追踪方法[J]. *电力系统保护与控制*, 2024, 52(16): 162-177.
- CHEN Guangyu, ZHANG Shengjie, YANG Li, et al. Station line loss rate anomaly identification and accurate tracking method of associated users based on multi-dimensional scene division[J]. *Power System Protection and Control*, 2024, 52(16): 162-177.
- [18] ZHOU K, WANG T, CHEN X, et al. A fault diagnosis method for power systems based on temporal tissue-like P systems[J]. *Protection and Control of Modern Power Systems*, 2024, 9(2): 101-114.
- [19] QUEIROZ L M O, ROSELLI M A, CAVELLUCCI C, et al. Energy losses estimation in power distribution systems[J]. *IEEE Transactions on Power Systems*, 2012, 27(4): 1879-1887.
- [20] 乐健, 王靖, 廖小兵, 等. 基于随机场景的配电网源网储协同规划方法[J]. *电力系统保护与控制*, 2025, 53(4): 132-147.
- LE Jian, WANG Jing, LIAO Xiaobing, et al. A source-n

- network-storage collaborative planning method for distribution networks based on random scenarios[J]. Power System Protection and Control, 2025, 53(4): 132-147.
- [21] 谭煌, 于海波, 刘婧, 等. 基于三相不平衡网络能量流的高比例新型源荷配电台区碳排放流计算[J]. 供用电, 2024, 41(7): 55-62, 83.
TAN Huang, YU Haibo, LIU Jing, et al. Carbon emission flow calculation for high proportion new source-load distribution station area based on three-phase unbalanced network energy flow[J]. Distribution & Utilization, 2024, 41(7): 55-62, 83.
- [22] 王二威, 秦亮, 皇甫成, 等. 基于空间分布和时序特征的配电网台区三相不平衡调整方法[J]. 电力系统自动化, 2023, 47(19): 97-105.
WANG Erwei, QIN Liang, HUANGFU Cheng, et al. Three-phase unbalance adjustment method for distribution station areas based on spatial distribution and time-series characteristics[J]. Automation of Electric Power Systems, 2023, 47(19): 97-105.
- [23] 王晨欣, 夏嘉航, 赵成勇, 等. 降低直流线路损耗的并联电容换相换流器单位功率因数控制策略[J]. 电力系统自动化, 2022, 46(5): 141-150.
WANG Chenxin, XIA Jiahang, ZHAO Chengyong, et al. Unity power factor control strategy for DC line loss reduction in shunt capacitor commutated converter[J]. Automation of Electric Power Systems, 2022, 46(5): 141-150.
- [24] DEGHANPOUR K, YUAN Y, WANG Z, et al. A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation[J]. IEEE Transactions on Smart Grid, 2019, 10(6): 5942-5951.
- [25] LIN C, CHEN C, CHUANG H, et al. An expert system for three-phase balancing of distribution feeders[J]. IEEE Transactions on Power Systems, 2008, 23(3): 1488-1496.
- [26] HENRIQUE L F, SILVA W N, SILVA C C A, et al. Optimal siting and sizing of distributed energy resources in a smart campus[J]. Electric Power Systems Research, 2023, 217.
- [27] 余传祥, 潘傲然, 毛文鹏, 等. 基于 ICSO 的智能电表误差和线损率联合评估模型[J]. 电力系统保护与控制, 2025, 53(13): 117-127.
YU Chuanxiang, PAN Aoran, MAO Wenpeng, et al. Joint evaluation model of smart meter error and line loss rate based on ICSO[J]. Power System Protection and Control, 2025, 53(13): 117-127.
- [28] ABEYSINGHE S, ABEYSEKERA M, WU J, et al. Electrical properties of medium voltage electricity distribution networks[J]. CSEE Journal of Power and Energy Systems, 2020, 7(3): 497-509.
- [29] CARPANETO E, CHICCO G, AKILIMALI J S. Loss partitioning and loss allocation in three-phase radial distribution systems with distributed generation[J]. IEEE Transactions on Power Systems, 2008, 23(3): 1039-1049.
- [30] FANG L, MA K, LI R, et al. A statistical approach to estimate imbalance-induced energy losses for data-scarce low voltage networks[J]. IEEE Transactions on Power Systems, 2019, 34(4): 2825-2835.
- [31] XIANG Y, RIBEIRO P F, COBBEN J F G. Optimization of state-estimator-based operation framework including measurement placement for medium voltage distribution grid[J]. IEEE Transactions on Smart Grid, 2014, 5(6): 2929-2937.
- [32] KE G, MENG Q, Finley T, et al. Lightgbm: a highly efficient gradient boosting decision tree[J]. Advances in neural information processing systems, 2017, 30.
- [33] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in engineering software, 2014, 69: 46-61.
- [34] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [35] HAMANO R, SAITO S, NOMURA M, et al. CatCMA: stochastic optimization for mixed-category problems[C]// Proceedings of the Genetic and Evolutionary Computation Conference, 2024: 656-664.

收稿日期: 2025-08-27; 修回日期: 2026-01-29

作者简介:

蔡延华(1976—), 男, 硕士研究生, 高级经济师、工程师, 研究方向为大数据分析、配用电与计量技术; E-mail: 526241527@qq.com

张浩博(2001—), 男, 通信作者, 硕士研究生, 研究方向为人工智能在电力系统中的应用; E-mail: 13192980119@163.com

彭显刚(1964—), 男, 教授, 硕士生导师, 研究方向为电力系统优化运行与控制、人工智能在电力系统中的应用。E-mail: epxg@gdut.edu.cn

(编辑 许威)