

基于三重伯努利随机森林的充电桩故障诊断方法

吴玘琪¹, 李林¹, 赵品彰¹, 鲍进²

(1. 江苏省计量科学研究所 (江苏省能源计量数据中心), 南京 210023;

2. 国网江苏省电力有限公司营销服务中心, 南京 210019)

摘要: 针对直流充电桩多类别故障数据不均衡、运行数据噪声高以及响应慢的问题, 采用一种引入元代价敏感学习机制的三重伯努利改进随机森林方法; 该方法在特征子空间选择、特征内软采样及分裂准则切换 3 个阶段嵌入受控随机性, 并在特征选择与节点分裂中通过权重矩阵引入类别代价信息, 显著增强对不平衡小样本类别的识别能力, 同时保持模型多样性与泛化性能; 在包含 19 类故障、9 项运行特征的真实充电桩数据集上进行试验测试, 在测试集准确率 (96.76%) 和 F_1 分数 (0.948) 方面均优于 C4.5-RF (66.01%、0.605) 和 CART-RF (86.33%、0.834), 训练时间仅 54.93 s, 较 C4.5-RF 缩短约 92.8%; 经消融试验证实, 三重伯努利机制与代价敏感信息的结合实现了充电桩故障诊断准确率的提高和计算成本的降低。

关键词: 直流充电桩; 故障诊断; 随机森林; 伯努利随机过程; 动态权重函数; 元代价敏感学习

A Fault Diagnosis Method for Charging Piles Based on Triple Bernoulli RF

WU Jiqi¹, LI Lin¹, ZHAO Pinzhang¹, BAO Jin²

(1. Jiangsu Institute of Metrology (Jiangsu Energy Measurement Data Center), Nanjing 210023, China;

2. Marketing Service Center of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210019, China)

Abstract: To address the challenges of imbalanced multi-class fault data, high noise in operational data, and slow diagnostic response in DC charging stations, this paper proposes an improved random forest (RF) method, termed Sandwich-RF, which integrates a triple Bernoulli mechanism with a meta cost-sensitive learning strategy. This method embeds controlled randomness into three stages: feature subspace selection, soft sampling within features, and splitting criterion switching. Furthermore, the category cost information is incorporated into both the feature selection and the node splitting through a dynamic weight matrix, significantly enhancing the recognition capability for imbalanced small sample categories while maintaining model diversity and generalization performance. Experiments are conducted on a real-world dataset containing 19 fault categories and 9 operational features, which shows that Sandwich-RF achieves a test accuracy of 96.76% and an F_1 score of 0.948, outperforming the C4.5-RF (66.01%, 0.605) and the CART-RF (86.33%, 0.834), with a training time of only 54.93 s, which is approximately 92.8% faster than that of the C4.5-RF. Ablation experiments confirm that the combination of the triple Bernoulli mechanism and cost-sensitive learning information effectively improves the fault diagnosis accuracy of charging piles while reducing computational costs.

Keywords: DC charging station; fault diagnosis; RF; Bernoulli stochastic process; dynamic weighting function; meta cost-sensitive learning

0 引言

当前充电桩的故障诊断方法主要可分为两大类: 基于规则的专家知识方法与基于数据驱动的机器学习方法^[1-4]。

在基于规则的诊断方法中, 研究者通常通过构建专

家系统、故障树模型、模糊逻辑推理或知识图谱进行推断和识别^[5]。这类方法具有可解释性强、推理路径明确的优点^[6]。文献 [7] 中设计了基于故障树的直流充电桩故障诊断专家系统, 通过专家知识库和推理引擎实现对典型故障的识别。文献 [8] 提出了一种基于知识图

收稿日期: 2025-08-26; 修回日期: 2026-04-10。

基金项目: 国家电网有限公司科技项目 (5700-202318272A-1-1-ZN)。

作者简介: 吴玘琪 (1989-), 男, 硕士, 工程师。

引用格式: 吴玘琪, 李林, 赵品彰, 等. 基于三重伯努利随机森林的充电桩故障诊断方法[J]. 计算机测量与控制, 2026, 34(5): 129-136.

谱的故障识别方法，构建了面向充电桩领域的语义图谱并结合改进 K-BERT 模型，实现了高效的故障分类。此类方法依赖先验规则的完备与维护成本，面对新型故障工况和复杂电气系统迁移适应性不足。

基于机器学习的数据驱动方法则凭借其建模能力强、无需显式规则构建、适应性强等优点获得广泛关注^[9]。在 HVAC 系统中引入随机森林 (RF, random forest) 与主成分分析法，可以实现对早期故障的精确识别^[10]。将数字孪生技术与机器学习模型结合，实现空调系统部件的状态预测与预警^[11]。文献 [12] 系统研究了深度神经网络如卷积神经网络 (CNN, convolutional neural network)、循环神经网络 (RNN, Recurrent Neural Network) 和自编码器等在异常检测中的性能，为大规模工业故障诊断提供了理论依据。然而，这类方法也存在训练样本依赖性强、容易过拟合、模型可解释性弱、特征选择效率低等问题。

充电桩故障诊断当前面临三方面难点：1) 故障数据通常具有严重的类别不平衡性，稀有故障样本数量有限，导致训练过程偏倚^[13]；2) 运行数据具有高维度特性，包含大量冗余和噪声信息，影响模型判别性能^[14]；3) 传统方法在泛化能力与实时响应之间存在矛盾^[15]。随机森林及其改进变体因鲁棒性、可解释的特征重要度与低部署成本在工业场景中广泛采用^[16]。Deep Balanced Cascade Forest 利用级联森林与重采样策略提升少数类召回，但训练推理链路较深，在线部署时延与内存开销偏大^[17]。在变压器故障识别上，SMOTE+RF 用 SMOTE 平衡 DGA 样本后准确率超过 96%，验证了传统机器学习加重采样在电力工业诊断的有效性^[18]。现有工作大多采用策略固定，对多源少数类与多标签场景覆盖和适应性均不足^[19]。此外传统 RF 在构造树结构过程中仍存在样本重叠率高、特征使用率偏低的问题，难以进一步提升模型性能。

基于上述挑战，提出一种引入元代价敏感学习机制的三重伯努利改进随机森林故障诊断方法。该方法在构建决策树过程中，分别在特征子空间划分、特征选择概率控制与节点划分准则选择 3 个阶段依次引入典型伯努利采样机制和非典型的广义伯努利机制，以增强模型结构多样性和特征利用效率，由此命名为 Sandwich-RF；同时在特征选择与节点分裂中通过权重矩阵嵌入类别代价信息，提升不平衡小样本类别的识别能力并保持较高的泛化性能。在真实工业运行数据上的试验结果表明，所提出的 Sandwich-RF 在准确率、 F_1 分数、训练时间等指标上均优于多种传统基线模型。

1 改进随机森林算法

1.1 随机森林分类算法

随机森林算法构建过程如图 1 所示。决策树构建中

的分裂指标通常有两种：Gini 系数^[20]和信息增益率^[21]。

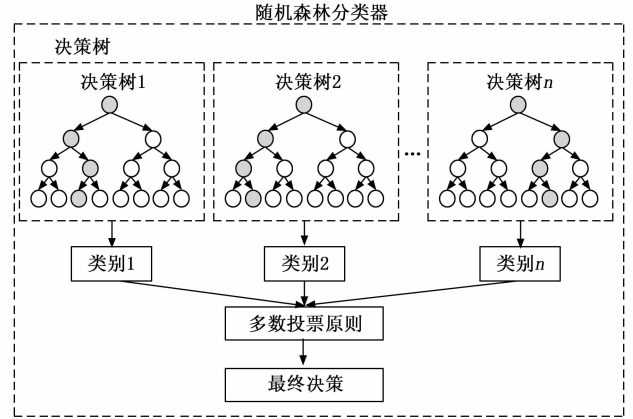


图 1 随机森林构建过程

Gini 系数的计算首先计算每个标签数据占总体的概率 p_i ，求和后再用 1 减去该值：

$$Gini = 1 - \sum_{i=1}^k p_i^2$$

信息增益率 Gain Ratio (A) 需要计算数据集的熵 $H(D)$ ，按照某特征 A 分裂后的加权熵 $IG(A)$ ，分裂信息 Split Info (A)：

$$Gain\ Ratio = \frac{IG(A)}{Split\ Info(A)}$$

完成决策树构建后，随机森林算法采用多数投票原则得到最终结果。

1.2 伯努利随机过程

温度调节的广义伯努利采样机制，是一种“非二元随机机制”，实质是一种带有温度调控的随机选择过程^[22]。 λ_{temp} 不再决定一个固定的二元事件是否发生，而是作为温度参数，用于调整多个候选划分指标。数学上可以借用 softmax 函数的思想来描述这一过程^[23]。设某节点划分准则候选集为 $\{g_1, g_2, \dots, g_k\}$ ，其对应的启发度指标为 $\{s_1, s_2, \dots, s_k\}$ 。则通过一个温度调节的概率分布计算公式：

$$P(g_i) = \frac{\exp(s_i/\lambda_{temp})}{\sum_{j=1}^k \exp(s_j/\lambda_{temp})}$$

其中： $\lambda_{temp} > 0$ 控制选择的集中程度，当 λ_{temp} 趋于 0 时，分布趋于贪婪选择，只选择最大的 s_i ， λ_{temp} 趋于 ∞ 时，分布趋于平均，即随机选择。

1.3 元代价敏感学习机制

元代价敏感学习机制 (Meta Cost-Sensitive Learning) 是通过实时性能反馈动态调整损失函数中各类别的权重。权重函数定义为：

$$w_c^{(t)} = \underbrace{\sqrt{\frac{N_{max}}{N_c}}}_{\text{静态权重}} \times \frac{\overset{\text{动态调节项}}{1}}{1 + \exp[-k(R_c^{(t)} - \mu)]}$$

其中: N_c 是类别 c 的样本量; $R_c^{(t)}$ 是类别 c 在 t 轮时的召回率; μ 为设定的召回率阈值; k 为调节斜率。该函数包含两个关键分量静态基数项和动态反馈项, 前者补偿样本不平衡, 样本量 N_c 越小则权重越大, 后者使用 Sigmoid 函数将召回率 $R_c^{(t)}$ 映射为权重增益因子。

进一步定义类别敏感权重矩阵:

$$W(t) = \begin{bmatrix} w_1(t) & 0 & \cdots \\ 0 & 0 & 0 \\ \cdots & 0 & w_c(t) \end{bmatrix}$$

2 融合代价敏感机制的 Sandwich-RF 故障诊断模型

2.1 三次伯努利随机决策树构建

充电桩故障数据具备“类别多且不均衡、特征相关性强、噪声与工况漂移并存”的典型特征, 考虑三重伯努利可提供多视角决策边界, 代价敏感在节点级别提升稀有和关键故障的区分度, 因此本文在决策树构建过程中引入 3 次伯努利机制, 如图 2 所示。三重伯努利机制优化的随机森林算法与传统随机森林仅在子空间与候选阈值附近随机不同, 在子空间选择降低个体树对冗余和共线特征的绑定, 从源头减小树间相关性; 利用温度软化采样在“探索-利用”间调节, 缓解互信息估计噪声导致的贪婪偏置; 其准则随机切换打破单一准则在不均衡与多值属性下的系统性偏置并扩大样本分类边际。

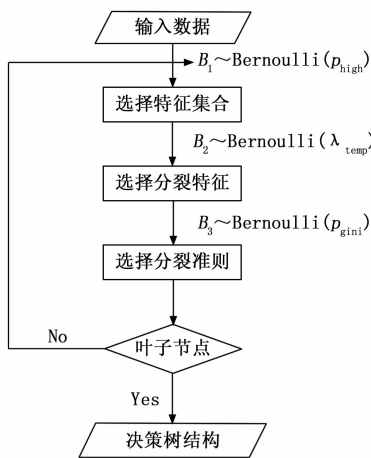


图 2 三次伯努利随机决策树构建过程

第一次伯努利机制用于选择特征集, 每棵树可以从高相关特征集和低相关特征集中的特征空间中进行选择。

互信息 (Mutual Information) 用于衡量某个特征和标签之间的信息共享程度, 反映了知道特征 A 能减少多少对 Y 的不确定性, 通常用于评估某特征分类能力^[24]。

若对于特征 A 有取值集合 $X = \{x_1, x_2, x_3, \dots, x_k\}$, 对于标签 Y 有取值集合 $Y = \{y_1, y_2, y_3, \dots, y_m\}$ 则互信息的具体计算公式可以写为:

$$MI(A; Y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \log_2 \left(\frac{P(x, y)}{P(x)P(y)} \right)$$

其中: $P(x)$ 指特征值为 x 的边缘概率, $P(y)$ 指标签为 c 的边缘概率, $P(x, y)$ 指特征值为 x 标签为 y 的联合概率。

其中边缘与联合概率通过估算获得, 兼顾不均衡与实时性需求。在本文中可以使用加权频数加 Dirichlet 平滑的方法估算得到。设加权总量 $N_w = \sum_i w_{y_i}$, 平滑 $\alpha = 1$, 则有:

$$\hat{p}(x, y) = \frac{n(x, y; w) + \alpha}{N_w + \alpha |X||Y|}, n(x, y; w) = \sum_{x_i = x, y_i = y} w_{y_i}$$

则边缘概率为:

$$\hat{p}(x) = \sum_y \hat{p}(x, y), \hat{p}(y) = \sum_x \hat{p}(x, y)$$

高相关特征集 F_{high} 和低相关特征集 F_{low} 由特征与标签之间的互信息来决定。对于给定特征集合 $F = \{f_1, f_2, f_3, \dots, f_k\}$, 第一次伯努利选择的过程可以表示为 $S = Bernoulli(p_{high})$, 其中 P_{high} 表示是选择高相关特征集的概率。若 $S=1$, 表示选择高相关特征集, 否则选择低相关特征集。对于每棵树, 使用伯努利分布对其进行特征子空选择:

$$F_{noise} = \begin{cases} F_{high}, & \text{if } B_1 \sim Bernoulli(p_{high}) = 1 \\ F_{low}, & \text{otherwise} \end{cases}$$

第二次伯努利机制用于选出当前节点的最佳划分特征。确定从某一相关特征集中选择, 即确定子空间 F_{chosen} 后, 引入基于互信息的 Softmax 权重机制, 定义第 j 个特征被选中的概率为:

$$P(f_j) = \frac{\exp(I(f_j, y) / \lambda_{temp})}{\sum_{k \in F_{chosen}} \exp(I(f_k, y) / \lambda_{temp})}$$

其中: $I(f_j, y)$ 为特征与标签之间的互信息, λ_{temp} 为温度系数, 控制选择的“软性程度”。当 λ_{temp} 增大时, 所有特征的选择概率会趋于均等, 反之, 当 λ_{temp} 较小, 选择的特征会更偏向于那些具有较高互信息的特征。 λ_{temp} 用于控制第二次伯努利分布选择特征时的权重分配。较低的 λ_{temp} 会让算法更加偏向于选择高信息量的特征, 因此特征选择的过程更“冷”, 即选特征时更加“确定”。较高的温度系数会增加选择特征的随机性和多样性。通过这种方式引导模型偏向信息量大的特征, 同时保留一定的随机性。

第三次伯努利机制用于随机选取当前节点的划分准则, 使用 Gini 指数或信息增益率 (GainRatio)。设定伯努利分布概率 P_{high} , 则划分准则 C 的选取为:

$$C = \begin{cases} Gini, & \text{if } B_3 \sim Bernoulli(p_{gini}) = 1 \\ GainRatio, & \text{otherwise} \end{cases}$$

2.2 三重伯努利 RF 与元代价敏感学习耦合机制

为提升不平衡数据中少数类与高代价类别的识别能

力, 在 Sandwich-RF 的第二次伯努利 (特征软采样) 与第三次伯努利 (节点分裂准则选择) 两个关键阶段注入动态代价信息。动态代价由类别频次与阶段性召回率驱动, 随训练过程自适应更新; 在第二次伯努利中用于调整候选特征被采样的概率, 使与高代价/小样本类别相关的特征更易被选取; 在第三次伯努利中用于替换无权的纯度度量, 使分裂更倾向于产生能够区分高代价类别的划分。该融合在保持三重伯努利带来的结构多样性的同时, 显著提高了对弱势类别的关注度, 从而兼顾整体精度与宏平均指标。三重伯努利 RF 与代价敏感学习的融合伪代码如下。

```

Input: 训练集 D、子空间参数、温度系数  $\lambda_{temp}$ ,
第一次/第三次伯努利概率 ( $P_{high}, P_{gin}$ ) 等,
初始类权重向量  $W_0$ ; 评估方式验证集
Output: 训练好的融合模型
1:  $W \leftarrow W_0$ 
2: for  $t = 1 \dots T$  do
3:   由 D 自助采样得到训练子集  $D_t$ 
4:   执行第一次伯努利: 选择本树特征子空间  $F'$  (受  $P_{high}$  控制)
5:   从根节点开始递归建树:
6:     若满足停止条件 (纯度/深度/样本数), 标记为叶结点并返回
7:   第二次伯努利 (特征软采样, 注入代价):
8:     依据互信息与当前权重  $W$  计算候选特征采样概率 (受  $\lambda_{temp}$  调节)
9:     采样候选集并选定分裂候选特征  $F''$ 
10:    第三次伯努利 (分裂准则选择, 注入代价):
11:      以概率  $P_{gin}$  采用“代价加权 Gini”, 否则采用“代价加权增益率”
12:      在候选阈值上搜索最优划分, 生成子结点
13:      递归处理子结点
14:    使用验证集评估各类别阶段性召回率  $R^{(c)}$ 
15:    依据  $R^{(c)}$  与频次对  $W$  做小步动态更新
16: end for
17: 返回决策树结构
    
```

2.3 直流充电桩故障诊断模型

在完成训练后, 模型被应用于直流充电桩的故障诊断任务, 如图 3 所示。

故障诊断系统主要由 4 个功能模块构成: 数据采集模块、数据预处理模块、故障诊断模块与预警响应模块。

首先, 数据采集模块获取充电桩运行过程中的关键参数, 具体包括: 输出电压、电流, 电池 SOC、电池温度、充电桩内部温度、充电枪温度、单次订单的充电量、车辆需求电压与电流等。随后, 数据预处理模块对采集到的数据进行清洗, 包括缺失值补全、异常值修正等处理过程。同时, 基于互信息理论划分高相关特征集与低相关特征集。

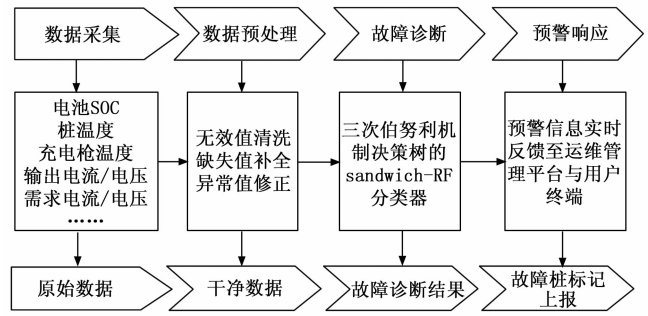


图 3 基于 Sandwich-RF 的直流充电桩故障诊断模型框架

核心的故障诊断模块构建于本文提出的 Sandwich-RF 算法之上。该模块由随机森林决策树构成。每棵树在构建过程中均经历三阶段采样, 融合元代价敏感学习机制, 通过大多数投票机制对直流充电桩的状态进行判别, 输出具体故障类型。最后, 预警响应模块接收诊断模块的输出结果, 将预警信息实时反馈至运维管理平台与用户终端, 实现高效及时的故障告警与响应, 降低人工成本, 提升充电基础设施的智能管理水平。

3 试验设计与结果分析

3.1 数据集介绍与预处理

目前实际使用的直流充电桩中, 常见的结构如图 4 所示。

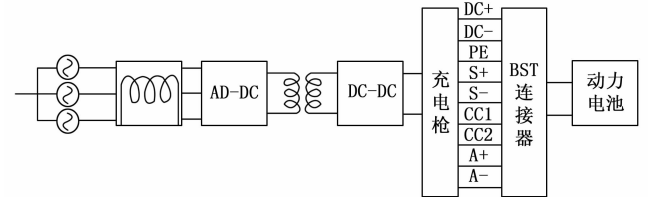


图 4 直流充电桩结构图

其中, 直流充电桩 BST 连接器中 BST-BMS 元件最容易出现故障。BST-BMS 元件指 DC-DC 模块输出端器件或者隔离电路中连接车端的接口。用于 BMS 通信、电压采样控制等。这部分过温通常由 3 个原因引起: 1) 接触点氧化或连接器接触不良导致接触电阻过大; 2) 充电桩长时间工作, 散热不足; 3) 外部环境如夏季阳光直射。与 BST-BMS 元件低电流的电气属性不同, BST 输出连接器通常工作在大电流下为整车供电。因此这部分过温除了接触电阻过大的原因, 还有可能存在大电流工作下发热的情况。此外还存在常见的故障如: 等待接收车端 BCL 报文超时, 主回路继电器异常和交流接地故障等。

试验数据集使用充电桩头部企业星星充电提供的充电桩运行故障数据, 有 19 种故障和 9 个输入特征, 在特征数据的收集方面考虑到了来自车端与桩端的两端特征, 具体故障种类见表 1, 具体特征类别见表 2。

表 1 故障类型统计表

故障类型	故障可能原因与判定机制
BST 充电连接器故障	充电连接器(锁止、温度、辅助触点/检测)任一必需量缺失/越界,或插拔寿命计数告警
BST 输出连接器过温故障	外部母线端子温度或接触电阻或压降超过阈值
K1/K2 继电器外侧电压过大	K1K2 继电器触点焊死、泄放回路失效,电流电压异常
BMS 报文中单体电压过高	单体电池电压超过阈值,热失控前兆
主回路继电器故障	继电器拒动、误动、粘连,触点或线圈故障、驱动欠压,电流电压异常
BST 电流过大	BST 输出短路/负载突变、BMS 限流失效、环路不稳
BST 检测点 2 电压故障	采样偏移、母线电容/均压异常、接触不良,检测点 2 电压偏离期望值
充电枪过温保护	触点电阻大、线缆发热、环境高温,充电枪温度超过阈值
BMS 报文中 SOC 过高	BMS 报文中 SOC 过高,SOC 值超出 100%或变化速率异常
枪未正确连接 / 连接断开	检测点电流电压、报文异常
BST 其他异常	BST 电源模块上报非分类故障位(过压/欠压/风扇故障等)
桩离线	接收报文异常,通信板掉电
充电桩转空闲	本地策略异常、接触器抖动、看门狗复位
电表数据异常	电流电压表异常、计量误差超出阈值
订单充电量超过最大允许值	单笔订单累计充电量超过最大允许值
订单持续超过最大允许时长停止	单笔订单累计充电时长超过最大允许值
交流接地故障	对地短路、潮湿、绝缘老化、线缆破损,电阻低于阈值
BST-BMS 元件输出连接器过温	元件温度超过阈值
等待接收车端 BCL 报文超时	BCL 帧计数、丢包率、握手响应时间超过阈值

数据采集覆盖完整季节周期,在 2024 年度的春(4 月)夏(8 月)秋(10 月)冬(12 月)季开展,共获得 160 000 条有效记录,其中 4 月与 10 月各 30 000 条,8 月与 12 月各 50 000 条。该配置旨在纳入不同季节环境条件对充电设备与车辆侧行为的影响。研究对象为直流充电桩,为避免设备偏置,采用随机型号采样策略(不对品牌/型号设定先验约束),以增强数据在不同设备配置下的外部有效性。全部数据均来自实际运营条件下的常规运行日志与测量记录(非干预式采集),在不影响正常服务的前提下完成。采集过程中未对具体工况进行刻意控制,仅依据自然到站与会话过程记录形成样本。剔除异常值和缺失值后数据共 146 714 条,按 70%和 30%划分训练和测试集。

如表 2 所示,数据集中故障类型涵盖了通信、连接、BMS、充电枪和主回路等多个模块,覆盖了充电桩从交流输入到直流输出、从控制通信到执行硬件的全链路问题,其中通信类和温度类问题占比高,集中在车桩交互环节和高功率部件。低占比的高风险类故障(如电流过大、电压过高)虽然较为少见,但对系统稳定性和安全性影响大,也需要对其进行重点关注预防。

表 2 数据集输入特征

序号	特征	来源	均值	标准差	取值范围
1	输出电压/V	桩端	357.44	159.90	[0,713.1]
2	输出电流/A	桩端	41.06	52.66	[0,329]
3	SOC	车端	30.56	35.13	[0,100]
4	车辆电池温度/℃	车端	37.57	20.27	[-40,215]
5	桩温度/℃	桩端	57.73	25.84	[-40,96]
6	充电枪温度/℃	桩端	52.73	26.57	[-50,255]
7	订单充电量/kWh	桩端	76.94	4 485.07	[0,348.91]
8	车需电压/V	车端	127.64	218.65	[0,732.6]
9	车需电流/A	车端	34.89	72.91	[0,400]

3.2 对比试验及分析

试验 1: 网格搜索最优三重伯努利 RF 参数组合

为验证原始三重伯努利 RF 在直流充电桩故障诊断任务中的有效性,对不同的参数配置进行试验,分析不同配置下的训练时间、准确率和 F_1 分数。

经网格搜索,得到树棵数($n_{estimators}$)、树深度(max_{depth})、选择高相关特征集概率(P_{high})、选择 Gini 分裂指标概率(P_{gini})和温度系数(λ_{temp})最佳组合如表 3 所示。其训练时间为 54.931 0 s,预测时间为 0.306 9 s,测试集准确率达 95.67%, F_1 分数为 0.927 5。这表明在特征选择的过程中,使用高相关特征集的概率为 0.8 能够在保持较低计算开销的同时,获得较高的准确性和鲁棒性。

表 3 三重伯努利 RF 参数组合

参数	数值
$n_{estimators}$	20
max_{depth}	9
P_{high}	0.8
P_{gini}	0.2
λ_{temp}	2

该最佳参数组合下的分类混淆矩阵如表 4 所示。由混淆矩阵可知,类别 6、8、11、12、14、15、16、19 的预测精度为 1.00 或接近 1.00。总体而言,所提出的算法在大多数充电桩故障类型的诊断中仍表现出较高的准确性与鲁棒性,但某些类别如 14、16、19 存在混淆。这些故障分别是:BST 电流过大、BMS 报文中单体动力电池电压过高、BST 充电连接器故障。

导致上述类别间出现混淆的主要原因在于其特征模

式存在较高相似性，从而降低了模型的判别能力。具体而言，“BST 电流过大”（第 14 类）故障主要表现为输出电流显著升高，并伴随桩温度与充电枪温度上升，同时 SOC 增加速率加快；“BMS 报文中单体动力蓄电池电压过高”（第 16 类）故障的直接特征是车需电压与输出电压升高，但在高压条件下，电流亦可能偏高，导致其特征分布与第 13 类部分重叠；“BST 充电连接器故障”（第 19 类）通常引发电流和电压的波动异常，同时产生局部发热（桩温度与枪温度升高），其数据模式亦与第 13 类存在相似性。由此，电流、电压、温度等核心特征在不同故障类型中呈现出相近的异常变化，使得模型在特征空间中难以形成明确的分类边界。从硬件层面来看，这种混淆可能还与传感器测量的耦合效应有关。由于输出电压、输出电流与车需电压、车需电流在物理上存在较强相关性，单一故障往往会同时影响多个传感器测量值，从而进一步降低不同类别在特征维度上的区分度。

表 4 分类混淆矩阵表

序号	故障类别
1	等待接收车端 BCL 报文超时
2	BST-BMS 元件输出连接器过温
3	交流接地故障
4	订单持续超过最大允许时长停止
5	订单充电量超过最大允许值
6	电表数据异常
7	充电桩转空闲
8	桩离线
9	BST 其他异常
10	枪未正确连接/连接断开
11	BMS 报文中 SOC 过高
12	充电枪过温保护
13	BST 检测点 2 电压故障
14	BST 电流过大
15	主回路继电器故障
16	BMS 报文中单体动力蓄电池电压过高
17	K1/K2 继电器外侧电压过大
18	BST 输出连接器过温故障
19	BST 充电连接器故障

这些类别普遍存在的样本数量不足或特征分布与邻近类别高度重叠的问题，使得模型在训练过程中对其边界样本的识别能力有限，进而增加了漏检与错检的风险。

试验 2：融合元代价敏感学习机制有效性验证

为进一步处理上述混淆类别与小样本类别的故障数据，验证本文提出的融合元代价敏感学习机制的有效性，试验分别使用原始三重伯努利 RF、融合元代价敏感机制改进的完全 Sandwich-RF 模型、以及融合 SMOTE-ENN 的三重伯努利 RF 模型进行实验。SMOTE-ENN 是一种在类别不平衡问题中常用且有效的重采样技术，结合了

合成少数类过采样技术和编辑最近邻，主要流程是先使用 SMOTE 增加少数类，再使用 ENN 清理所有数据。

试验中保持随机森林模型的其他参数设置一致，仅在改进模型中引入动态类别权重，以在特征软采样与分裂准则选择阶段增加对高代价类别的关注度。结果见表 5。

表 5 不同代价敏感策略的效果

方法	类别 13 召回率/%	类别 17 召回率/%	类别 18 召回率/%	整体准确率/%
三重伯努利 RF	80	85	40	95.67
+元代价敏感学习	86	88	52	96.76
+SMOTE-ENN	82	78	49	94.95

结果表明，融合元代价敏感机制后，易混淆类别的平均召回率显著提高，部分类别的召回率提升幅度超过 23%，整体平均召回率与 F_1 分数均有所增加。这表明该机制能够在保持总体性能稳定的同时，有效改善少数类与高代价类别的识别能力，从而增强诊断结果的可信度与工程应用的可靠性。

为进一步阐明元代价敏感学习机制提升少数类识别性能的内在逻辑，结合动态权重矩阵的作用机制与模型训练过程分析：元代价敏感学习的核心在于通过动态权重矩阵实现定向关注少数类与保持泛化性的平衡。根据权重函数 $w_c^{(i)}$ 的定义，其中静态基数项 $\frac{N_{\max}}{N_c}$ 优先补偿样本量差异，对应试验中数据集少数类类别 13（BST 电流过大）、17（BMS 单体电压过高）、18（BST 充电连接器故障），其静态权重为多数类的 3~5 倍，初始即赋予更高错分代价；动态反馈项则实时修正偏差，当三类召回率低于阈值时，Sigmoid 函数进一步提升权重，强制模型聚焦其分类误差。该权重矩阵通过两大环节作用于模型训练：1) 在第二次伯努利特征软采样中，特征采样概率被动态权重优化，避免少数类关键特征被多数类特征压制；2) 在第三次伯努利分裂准则选择中，权重嵌入纯度度量（如代价加权 Gini 系数），使节点分裂优先减少少数类错分，例如在 13 类与 17 类特征重叠区，模型会优先选择降低 13 类错分的分裂点，最终实现三类召回率提升。

对比而言，SMOTE-ENN 性能略低于元代价敏感学习机制源于三方面适配性不足：1) 其线性插值合成少数类样本时，易混入重叠特征区，导致伪样本干扰；2) ENN 清理误删少数类边界样本，加剧有效信息损失；3) 离线预处理无法应对训练中少数类召回率动态下降问题，而元代价机制通过每轮权重更新可主动抵消该偏倚，故更适配充电桩多类别不平衡、特征重叠度高的故障数据特性。

试验 3：三重伯努利采样机制有效性验证

为验证本文提出的 Sandwich-RF 算法中伯努利机制

的有效性, 设计消融试验, 结果见表 6。

表 6 试验结果统计表

P_{high}	P_{gini}	λ_{temp}	训练时间/s	测试集准确率/%
0.8	0.2	2	54.931 0	96.76
1	0.2	2	65.798 4	94.47
0	0.2	2	51.262 4	42.54
0.8	0	2	57.564 2	94.95
0.8	1	2	52.125 0	90.34
0.8	0.2	1	49.102 7	93.01
0.8	0.2	10	72.942 8	92.32

由表 6 可知, 高 P_{high} 值 (接近 1) 表示所有分裂在高相关特征集合中, 这带来较高的分裂精度和效率, 但是缺少随机性, 所有树的特征集相似。低 P_{high} 值 (接近 0) 表示所有分裂在低相关特征集合中, 信息含量少, 模型学习不到数据的特征, 分类精度很低。高 P_{gini} 值 (接近 1) 会使得算法更多地依赖 Gini 系数进行特征选择。Gini 系数适用于分类任务且计算效率较高, 因此训练速度较快, 但可能在某些复杂的高维数据集上不如信息增益率灵活。低 P_{gini} 值 (接近 0) 会使得算法更多地依赖信息增益率。信息增益率能够更好地处理特征之间的冗余, 但计算过程可能较为复杂, 分裂信息可能导致计算开销增加, 容易出现过拟合。较低的 λ_{temp} ($=1$) 会使得模型的决策树更加简单, 特征选择更加偏向高信息量的特征, 可能导致过拟合。较高的 λ_{temp} ($=10$) 会使得模型更加复杂, 增加决策树的多样性, 适合高维数据集, 但也可能带来过高的计算复杂度。

为明确各机制在解决数据不均衡问题中的核心作用以及其贡献度, 设计试验, 结果见表 7。

表 7 试验结果统计表

P_{high}	P_{gini}	λ_{temp}	测试集准确率/%	测试集 F_1 分数
0.8	0.2	2	96.76	0.948
0.8	0	2	94.95	0.930
0.8	1	2	90.34	0.884
/	0.2	/	89.96	0.871

当仅关闭第三次采样机制时, P_{gini} 取 0 或 1。第一、二次采样机制用于实现特征子空间与特征内软采样, 一同打开或关闭。与仅保留第三次采样机制的基线相比, sandwich-RF 在测试集 F_1 上提升 +0.077 (第一、二次采样机制贡献度占 76.6%, 第三次采样机制贡献度占 23.4%), 测试集准确率提升 +6.80% (第一、二次采样机制贡献度占 73.4%, 第三次采样机制贡献度占 26.6%)。

这表明特征子空间选择与特征内软采样是缓解长尾不均衡、提升少数类识别的核心, 而分裂准则切换主要提升分裂稳健性与宏平均表现。固定单一分裂准则会明

显劣化性能 (信息增益率优于 Gini), 进一步验证了节点级自适应切换的必要性。

试验 4: Sandwich-RF 与其他分类算法对比

为验证 Sandwich-RF 的优越性, 将其与传统随机森林分类算法以及一些工程中常用的机器学习分类算法进行对比试验。

将 Sandwich-RF 与传统的 C4.5 随机森林和 CART 随机森林算法进行对比, 传统随机森林的参数设置与 Sandwich-RF 保持一致。试验结果如表 7 所示, C4.5 随机森林训练时间为 767.160 s, 训练集准确率高达 99.52%, 但测试集准确率仅为 66.01%, F_1 分数为 0.601。这表明, C4.5 随机森林在该情景下存在严重问题, 出现了过拟合现象, 无法有效处理测试集中的复杂模式。CART 随机森林的表现略好, 但与 Sandwich-RF 相比, 其性能仍存在一定差距, 具体表现在准确率较低, 训练时间较长。

此外本研究还对比了其他常用的机器学习分类算法 SVM 和 XGBoost, 其中 SVM 参数设置采用径向基函数 (RBF) 核, 惩罚系数 $C=0.1$, 并使用一对多 (One-vs-Rest) 策略; XGBoost 参数设置 boost round=20, max depth=9。由表 8 可知, SVM 在训练时间和准确率上表现均不及 Sandwich-RF。XGBoost 在训练时间上大幅领先其他模型, 但是准确率略有欠缺。

表 8 Sandwich-RF 与其他常用分类模型对比结果

方法	训练时间/s	训练集准确率/%	测试集准确率/%	训练集 F_1 分数	测试集 F_1 分数
Sandwich-RF	54.93	96.26	96.76	0.932	0.948
C4.5	767.16	99.52	66.01	0.994	0.605
CART	260.05	86.89	86.33	0.839	0.834
SVM	94.76	72.80	73.88	0.732	0.744
XGBoost	3.54	93.59	93.68	0.936	0.924

与传统随机森林和其他机器学习中常用的分类算法相比, Sandwich-RF 在测试集上的表现更好, 具有较低的训练时间和较高的准确率。这表明 Sandwich-RF 能够更有效地应对充电桩故障诊断任务中的复杂性与不确定性。

综上, 本文提出的包含元代价敏感学习机制的三重伯努利改进随机森林方法, 即 Sandwich-RF, 增加了决策树的多样性, 提升了直流充电桩诊断准确率的同时降低了计算复杂度, 减少了训练时间, 具有较高的工程应用价值。

4 结束语

本文围绕直流充电桩多类别且不均衡的故障识别需求, 构建了融合元代价敏感学习的三重伯努利随机森林模型 Sandwich-RF。该方法在特征子空间、特征内软采样与分裂准则三层引入受控随机性, 并在特征选择与节

点分裂环节注入类别代价信息,实现对高代价、小样本类别的定向关注与模型多样性的平衡,有效提升了直流充电桩故障诊断系统的准确性与泛化能力。基于直流充电桩实际运营的故障数据开展了多组对比试验,结果表明该方法在整体准确率上较传统树模型有显著提升,少数类召回率改善尤为明显,同时保持了较低的训练与推理开销;和其他机器学习中常用的分类算法相比,也表现出更好的性能。

尽管试验结果证明所提出的分类模型效果取得了显著提升,但仍存在一定的研究局限性。例如当前试验数据来源仍以单一充电桩企业的数据为主,可能存在对市面上其他品牌直流充电桩的普适性问题。

参考文献:

- [1] 姚望. 基于直流充电桩输出电压的故障诊断研究 [D]. 贵州: 贵州大学, 2023.
- [2] GAO X M, YUAN G T, ZHANG M J. Fault detection of electric vehicle charging piles based on extreme learning machine algorithm [C] //Fourth International Conference on Computing Methodologies and Communication, 2020: 849-852.
- [3] 郑经豪, 刘洪儒, 郑焜, 等. 基于改进决策树的直流充电桩充电模块故障诊断方法 [J]. 电器工业, 2025 (1): 54-58.
- [4] JANG K, HONG S, KIM M, et al. Adversarial autoencoder based feature learning for fault detection in industrial processes [J]. IEEE Transactions on Industrial Informatics, 2022 (18): 827-834.
- [5] 宋龙飞, 陈玉清, 金振俊. 基于故障树和产生式规则的故障诊断专家系统设计 [J]. 中国舰船研究, 2024 (19): 84-92.
- [6] 郑聪, 彭庆忠, 周海峰, 等. 基于人工智能专家系统的船舶电力系统故障诊断研究 [J]. 广州航海学院学报, 2021 (29): 5-8.
- [7] 杨莎莎. 基于故障树的直流充电桩故障诊断专家系统研究 [D]. 北京: 北京交通大学, 2019.
- [8] 田文海. 基于知识图谱的充电桩故障识别方法研究 [D]. 北京: 北京交通大学, 2023.
- [9] 黄友锐, 沈玉轩, 徐善永. 融合机器学习的永磁同步电机数字孪生故障诊断技术研究 [J]. 电测与仪表, 2025 (62): 181-187.
- [10] GOURABPASIA H, NIK B M. Knowledge discovery by analyzing the state of the art of data-driven fault detection and diagnostics of building HVAC [J]. CivilEng, 2021 (2): 986-1008.
- [11] HOSAMOH H, SVENNEVIGP R, SVIDT K, et al. A digital twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics [J]. Energy and Buildings, 2022 (261): 111988.
- [12] NASEERS, SALEEMY, KHALIDS, et al. Enhanced network anomaly detection based on deep neural networks [J]. IEEE Access, 2018 (6): 48231-48246.
- [13] SHEN W, FAN W, CHEN C. An electric vehicle charging pile fault diagnosis system using Borderline-SMOTE and LightGBM [C] //Tenth International Symposium on Precision Mechanical Measurements, SPIE, 2021.
- [14] YINC, TANGZ L, WENGX F, et al. A diagnostic method for open-circuit faults in DC charging stations based on improved s-transform and lightGBM [J]. Energies, 2024, 17 (2): 404.
- [15] 计喆, 金月华. 基于智能技术的充电桩网络运维与故障诊断方法分析 [J]. 集成电路应用, 2024 (41): 378-379.
- [16] WANG Q, TAAL C, FINK O. Integrating expert knowledge with domain adaptation for unsupervised fault diagnosis [J]. IEEE Transactions on Instrumentation and Measurement, 2022 (71): 1-12.
- [17] CHENH, LIC S, YANGW X, et al. Deep balanced cascade forest: a novel fault diagnosis method for data imbalance [J]. ISA Transactions, 2022 (126): 428-439.
- [18] RAHMAN A P, MUHAMMAD A A, PUTRAE, et al. Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique [J]. Electric Power Systems Research, 2023 (220): 109361.
- [19] CHENW X, YANGK X, YU Z W, et al. A survey on imbalanced learning: latest research, applications and future directions [J]. Artificial Intelligence Review, 2024 (57): 137.
- [20] 李蔚, 吴懿范, 毛静宇, 等. 基于改进随机森林算法的汽轮机振动故障诊断研究 [J]. 浙江电力, 2024 (9): 107-116.
- [21] SALZBERG S L. C4.5: programs for machine learning by J. Ross Quinlan [J]. Machine Learning, 1994 (16): 235-240.
- [22] TALAGRAN D M. Bernoulli processes [M]. Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems, Springer International Publishing, 2021: 175.
- [23] AHMAD S S, EHSAN R, STEFAN S. Evaluation of the generalized bernoulli trial-transient adaptive subcell (GBT-TAS) collision scheme in treating rarefied gas flows [J]. Computers & Fluids, 2020, 213: 104740.
- [24] BROWN G, POCOCK A, ZHAO M J, et al. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection [J]. The Journal of Machine Learning Research, 2012, 13 (1): 27-66.