

基于 BO-LightGBM 算法的 XLPE 配电电缆绝缘状态评估

罗正均¹, 叶刚¹, 周箩鱼¹, 李涛², 陈楠¹, 张志熙¹

(1. 长江大学 电工电子国家级实验教学示范中心, 湖北 荆州 434023;

2. 湖北师范大学 电气工程与自动化学院, 湖北 黄石 435002)

摘要:为提升电缆绝缘状态评估的精度,本文提出了一种基于贝叶斯优化(BO)算法与轻量级梯度提升机(LightGBM)算法的电缆绝缘状态评估方法。首先将数据集中所有特征进行组合,形成不同的特征子集,通过遍历所有的特征子集,找到五折交叉验证的准确率最高所对应的特征组合,完成对输入特征的筛选。然后使用 BO 算法对 LightGBM 中的 7 个超参数进行寻优。最后利用本文所提出的 BO-LightGBM 算法完成对电缆绝缘状态的评估。结果表明:本文提出的特征子集法与主成分分析法和互信息筛选法相比能更好地提升模型表现;经过 BO 算法优化后,LightGBM 模型的精度能得到进一步的提升,与粒子群优化算法(PSO)和遗传算法优化(GA)相比,BO 算法的计算效率能在几乎相同的精度下分别提升约 80% 和 86.9%;与其他常用机器学习算法进行对比,本文模型的相关性能指标均为最优。

关键词: XLPE 电缆; 状态评估; 机器学习; 贝叶斯优化算法; 轻量级梯度提升机算法

Insulation status assessment of XLPE distribution cable based on BO-LightGBM algorithm

LUO Zhengjun¹, YE Gang¹, ZHOU Luoyu¹, LI Tao², CHEN Nan¹, ZHANG Zhixi¹

(1. National Demonstration Center for Experimental Electrical and Electronic Education,

Yangtze University, Jingzhou 434023, China;

2. School of Electrical Engineering and Automation, Hubei Normal University, Huangshi 435002, China)

Abstract: To improve the accuracy of cable insulation status assessment, this paper proposed an assessment model of insulation condition based on Bayesian optimization (BO) algorithm and light gradient boosting machine (LightGBM) algorithm. First, all the features in the dataset were combined to form different feature subsets. By traversing all the feature subsets, the optimal feature combination corresponding to the highest accuracy from five-fold cross-validation was identified to complete the input feature selection. Then, the BO algorithm was used to optimize seven hyperparameters in LightGBM. Finally, the proposed BO-LightGBM algorithm was used to assess the cable insulation status. The results show that the feature subset method proposed in this paper can better improve model performance compared with principal component analysis (PCA) and mutual information-based feature selection methods. After optimization by the BO algorithm, the accuracy of the LightGBM model is further enhanced. Compared with particle swarm optimization (PSO) algorithm and genetic optimization (GA) algorithm, the computational efficiency of BO algorithm increases by approximately 80% and 86.9% at the same accuracy level, respectively. Furthermore, compared with other commonly used machine learning algorithms, the performance metrics of the proposed model are optimal.

Key words: XLPE cable; state assessment; machine learning; Bayesian optimization (BO) algorithm; light gradient boosting machine (LightGBM) algorithm

0 引言

交联聚乙烯(XLPE)电缆由于其优异的力学性能、耐热性能和绝缘性能,已成为电力系统的重要输电设备^[1]。XLPE 电缆敷设线路较长且运行环境复杂,随着运行年限增加,电缆的绝缘将逐渐老

化,剩余寿命逐渐降低。因此,开展电缆绝缘状态评估对保障供电安全具有重要意义。

现阶段针对电缆绝缘状态评估的研究方法主要分为两种^[2]。第一种是对电缆绝缘参数开展现场试验或者实验室测量,获取电缆老化特征参数进行分析评估^[3-7]。例如有学者对不同水树老化程度电缆开展超低频介质损耗测试,利用获取的电压-电

基金项目:国家自然科学基金资助项目(62473135)。

流($U-I$)滞回曲线的偏转角变化率反映电缆的老化程度^[8]。也有研究发现介质损耗因数曲线随着电缆老化程度的增加而不断上移,可适用于判断电缆绝缘老化状态^[9]。上述方法虽然可以获得特征参数与绝缘老化之间的关系,但是获得的电缆特征数据较少,评价指标单一,对电缆的绝缘状态往往只能做到相对表征。第二种方法主要是通过数学工具去分析过去电缆失效行为的趋势^[10-12]。例如采用模糊层次分析法对电缆的特征参数进行赋权,通过建立绝缘参数相对劣化度的隶属函数判断电缆绝缘老化状态等级^[13]。这类方法的核心是选取合适的电缆绝缘特征参数,对特征参数采取不同方式的赋权处理,最后按评级需求使用合适的隶属度函数,按最大隶属度原则判断电缆的老化状态等级。该方法对于特征参数的选取与赋权具有很强的主观判断性,且对评估后的电缆老化状态等级难以进行客观验证。

除了上述传统研究方法外,近年来,基于人工智能和机器学习的方法也被广泛地应用在电缆绝缘状态评估中。R SAHOO等^[14]利用支持向量机(SVM)、 K 近邻(KNN)、人工神经网络和朴素贝叶斯等机器学习算法分析了电缆绝缘的健康状况。YUN F等^[15]利用SVM算法,通过已公开的电缆数据集对其进行训练,完成了对电缆健康等级的划分。SVM是目前最常用的机器学习线性分类算法模型,但对于非线性问题分类效果表现一般,需要通过升维来解决。PAN Q S等^[16]利用极端梯度提升(XGBoost)算法对电缆健康状态进行评估,结果表明该方法拥有比SVM更好的性能表现,但XGBoost模型在训练时非常耗费计算资源,占用内存大,计算速度慢,在小数据集上训练容易出现过拟合的问题,导致训练集准确率下降。

针对上述机器学习在电缆绝缘状态评估领域所存在的问题,本文提出利用轻量级梯度提升机(LightGBM)算法实现对电缆绝缘状态的分类。LightGBM算法是由微软研究院KE G L等^[17]提出的一种新的基于梯度提升决策树(GBDT)算法的新型集成算法。目前LightGBM算法已在气象目标分类^[18]、风险状态预测等领域获得了广泛的应用,但是在绝缘评估领域中相关研究还较少。电气设备绝缘状态评估涉及多种数据类型和大量特征参量,且绝缘状态的变化通常与环境因素和电气参数之间

存在复杂的非线性关系。相比传统算法,LightGBM算法在处理高维数据和非线性关系方面表现出色,同时在效率和性能上也有显著提升。在机器学习中,超参数调优对于提高模型性能和泛化能力至关重要。LightGBM算法无论是在回归还是分类任务中,均涉及大量的超参数需要调整。贝叶斯优化(BO)则是一种为各分类器超参数进行调优的算法^[19]。该方法以目标函数的概率统计模型为基础,通过采集函数,确认迭代观测点,实现高效的全局搜索,避免了传统优化算法容易陷入局部最优的问题。

本文将BO与LightGBM算法进行结合,得到BO-LightGBM算法模型,利用公开数据集对模型进行训练,同时采用特征子集法完成对输入特征的筛选。通过BO算法找出LightGBM的最佳超参数,解决LightGBM调优困难的问题。通过特征子集法将特征量的全部特征进行组合,然后遍历来找到模型最佳特征输入,以提高模型精度,达到对电缆绝缘状态的准确评估。

1 算法与模型

1.1 轻量级梯度提升机(LightGBM)算法

GBDT算法在机器学习中拥有广泛的应用,如XGBoost和并行梯度提升回归树(PGBRT)均是基于其实现的,但是在特征维数高、数据量大的情况下,GBDT的效率和扩展性表现往往不尽人意^[8]。LightGBM基于梯度单侧采样与独家特征捆绑技术,可在达到几乎相同精度的情况下,将传统GBDT的训练过程加快,并减少内存占用。LightGBM算法具体实现如下。

假设模型输入数据集为 $D=\{(x_{ij}, C_i)\}$,其中 $i=1,2,3,\dots,n_1; j=1,2,3,\dots,n_2; n_1$ 为数据集的样本总数; n_2 为数据集中每个样本的总特征数; C_i 表示第 i 个样本目标类别标签; x_{ij} 表示第 i 个样本的第 j 个特征参数。LightGBM算法首先会初始化特征参数 x_{ij} 在第 k 类的预测状态 $F_{k,0}(x_{ij})$,如式(1)所示。

$$F_{k,0}(x_{ij}) = 0 \quad (1)$$

初始化后按照式(2)开始迭代计算特征参数 x_{ij} 在每个类别上的概率($p_{k,m-1}(x_{ij})$),迭代次数为 m ,具体分类问题中总的类别数为 K 。

$$p_{k,m-1}(x_{ij}) = \frac{e^{F_{k,m-1}(x_{ij})}}{\sum_{k=1}^K e^{F_{k,m-1}(x_{ij})}} \quad (2)$$

随后算法会通过计算当前预测值和真实值之间的差异得到特征参数 x_{ij} 的负梯度 $\tilde{y}_{i,k}$ ，如式(3)所示，式中 $y_{i,k}$ 为特征参数 x_{ij} 的真实概率。负梯度是每次迭代中新的树需要拟合的目标。

$$\tilde{y}_{i,k} = y_{i,k} - p_{k,m-1}(x_{ij}) \quad (3)$$

叶子节点上的样本集合是 $R_{h,k,m}$ ，叶子节点数为 h ，叶子节点分裂后的叶子节点值 $\gamma_{h,k,m}$ 如式(4)所示。分裂值决定了样本如何被划分到不同的叶子节点中，直接影响树模型的性能。

$$\gamma_{h,k,m} = \frac{K-1}{K} \frac{\sum_{x_{ij} \in R_{h,k,m}} \tilde{y}_{i,k}}{\sum_{x_{ij} \in R_{h,k,m}} |\tilde{y}_{i,k}|(1 - |\tilde{y}_{i,k}|)} \quad (4)$$

得到分裂后的叶子节点值后，算法会将模型按照式(5)进行刷新优化。模型的学习率设置为 η ，用于控制每棵树对最终模型的贡献。 I 为某一叶子节点的样本集合的索引，用于表示分裂后样本的归属。 H 为决策树的叶子节点数，用于控制树的复杂度。

$$F_{k,m}(x_{ij}) = F_{k,m-1}(x_{ij}) + \eta \sum_{h=1}^H \gamma_{h,k,m} I \quad (5)$$

最终得到的模型如式(6)所示， M 为最大迭代次数，即构建的决策树总数。

$$F_k(x) = \eta \sum_{m=1}^M \sum_{h=1}^H \gamma_{h,k,m} I \quad (6)$$

1.2 贝叶斯优化(BO)算法

贝叶斯优化是一种用于全局优化的算法，特别适用于计算代价高且无梯度信息的黑盒函数。它在机器学习中常用于超参数调优。贝叶斯优化的核心思想是使用一个代理模型(通常是高斯过程)来近似目标函数，并通过优化采集函数来选择下一个评估点^[9]，使得在少量采样下可以快速逼近全局最优解。通过这种迭代更新的方式，贝叶斯优化算法能实现对昂贵目标函数的高效优化。

贝叶斯优化算法的主要步骤为：①选择初始观测点评估目标函数，并完成初始化。②使用高斯过程等代理模型对目标函数进行建模。③定义并优化采集函数，利用代理模型的输出，通过采集函数结合预测均值和不确定性计算并选择下一个评估点。④计算该点的目标函数值并进行记录，使用新数据点更新代理模型。⑤重复上述步骤，直到满足停止条件。

1.3 BO-LightGBM 模型

在机器学习算法中，不能随着训练而自动调

整，需要人为设置的参数被称之为超参数^[20]。例如决策树算法中的树深、叶子节点样本数等。超参数的选择对模型性能有着关键性的影响。使用调优后的超参数可以帮助模型在不同的数据集上达到更好的准确率，同时也能防止过拟合，确保模型对新数据具有良好的预测能力。本研究采用将BO算法与LightGBM算法相结合的方式来完成对LightGBM的超参数设置。LightGBM虽然具有众多超参数，但并非所有的都对最终模型的性能都有显著影响。因此本文选取了其中7个超参数，涵盖了树结构、学习率、正则化等方面，既保证模型的调优空间充足，又避免过高的参数维度导致优化效率下降。本研究所采用的超参数取值范围如表1所示，并据此构建BO算法中的7维参数搜索空间。

表1 LightGBM待调优超参数

Table 1 Hyperparameters to be tuned of LightGBM

| 超参数名称 | 取值范围 | 作用 |
|------------------|------------|------------|
| max_depth | [3,30] | 控制最大树深 |
| num_leaves | [10,100] | 控制叶子节点数 |
| n_estimators | [100,1000] | 控制迭代次数 |
| colsample_bytree | [0.8,1] | 设置数据采样比例 |
| learning_rate | [0.01,0.3] | 设置学习率 |
| reg_alpha | [0,1] | 设置L1正则化项系数 |
| reg_lambda | [0,1] | 设置L2正则化项系数 |

为提高模型训练的稳定性与泛用性，采用五折交叉验证准确率作为BO算法中的目标函数。一般来说，电缆的老化是由多种条件导致的。选取特定的特征组合而不是用全部的特征量进行训练，这样可以降低特征维度，减小特征冗余，提高模型性能。针对数据集中的全部特征量，建立所有特征子集，给每个特征子集创建原生LightGBM模型，输出每个模型的五折交叉验证准确率，将准确率最高的特征子集作为LightGBM的输入特征。使用BO优化算法对该输入特征下的LightGBM算法进行超参数调优，得到BO-LightGBM模型，以此完成对电缆绝缘的老化评估，算法评估的具体流程如图1所示。

2 BO-LightGBM 模型用于电缆绝缘状态评估

2.1 电缆老化状态等级划分

对电缆绝缘状态进行定量判断，有助于电力人员进行维护与检修。有学者对电缆的绝缘状态划分为优、良、差3个等级^[21]，也有划分为4个等级

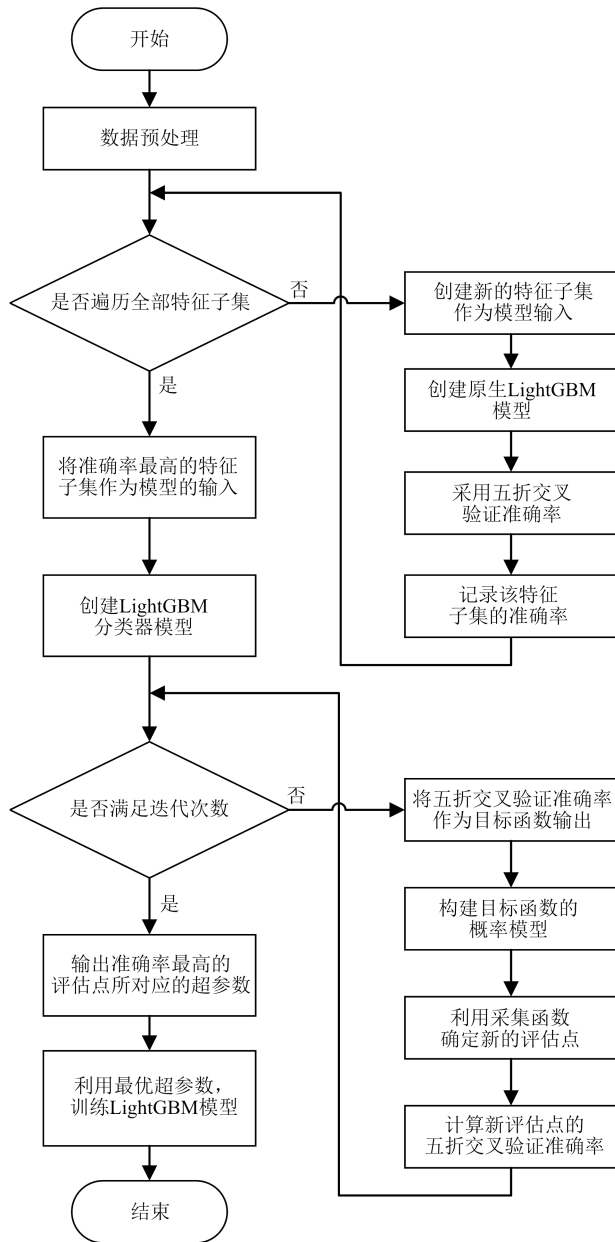


图1 BO-LightGBM 模型流程图
Fig.1 BO-LightGBM model flowchart

的^[10]。等级划分越精确,对电缆处理维护的措施也更细致。国际能源研究机构对美国、加拿大、澳大利亚和以色列的21家电力公司进行了调查,总结了典型电力资产不同健康指数水平的管理行为^[21]。Kinetics Lab 和 Hydro One 将电力类别资产健康指数分为5个等级,健康指数越大,说明电力设备的运行状态越好,健康指数与实际运行状态的关系和具体管理措施如表2所示。

2.2 数据预处理

数据预处理在机器学习中至关重要,在实际应用中一些非数值型特征数据是不能直接被机器学

表2 电缆健康指数
Table 2 Cable health index

| 健康指数 | 定义 | 管理措施 |
|------|-----------------|--------|
| 5 | 几乎全新 | 按计划期管理 |
| 4 | 有老化痕迹 | 正常维护 |
| 3 | 存在问题,不干预会加速老化 | 增加维护频率 |
| 2 | 存在严重问题,不干预会导致失效 | 置换或修复 |
| 1 | 完全恶化,无法继续使用 | 立即更换 |

习模型使用的。因此,需要对数据集中不同的数据类型进行预处理。根据式(7),可以将有序分类条件数据转换为数值数据。

$$x = \frac{i - 0.5}{N}, i = 1, 2 \dots N \quad (7)$$

式(7)中: N 为评分的总数; i 为评分的顺序。例如,有的数据集中某种特征被划分为优、良、差3种等级,则可将差评转换为0.17,良评级可转换为0.5,优评级可转换为0.83。

除此之外,不同类型的数值型数据具有不同的单位量纲,特征之间可能存在较大差异。归一化可以降低不同特征之间由量纲差异而造成的对诊断精度的影响,特征归一化的方法有很多,本研究选择 Min-Max 归一化,如式(8)所示。

$$Y_{\text{norm}} = \frac{Y - Y_{\text{min}}}{Y_{\text{max}} - Y_{\text{min}}} \quad (8)$$

式(8)中: Y 为原始特征值; Y_{max} 在特征列中观察到的最大值; Y_{min} 是在特征列中观察到的最小值。

2.3 模型性能指标

在机器学习中,准确率(accuracy)、精确率(precision)、召回率(recall)、F1分数(F1-Score)是评估分类模型性能的重要指标。

准确率是指模型正确预测的样本数占总样本数的比例,公式为式(9)。

$$A = \frac{T_p + T_N}{T_p + F_p + T_N + F_N} \quad (9)$$

式(9)中: T_p 是真正例; T_N 是真负例; F_p 是假正例; F_N 是假负例。

精确率表示在所有被模型预测为正的样本中,实际为正的样本的比例,计算公式为式(10)。

$$P = \frac{T_p}{T_p + F_p} \quad (10)$$

召回率是指模型正确预测为正类的样本数占实际正类样本总数的比例,公式为式(11)。

$$R = \frac{T_p}{T_p + F_N} \quad (11)$$

F1-Score (F_1)是精确率和召回率的调和平均数,用于两者之间达到平衡,计算公式为式(12)。

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (12)$$

通过上述4个指标可以全面准确衡量模型的性能,对应指标越高,说明模型的性能越好。

2.4 电缆老化状态评估流程

本文从模型特征选择和超参数优化两个方面出发,提出了有效的电缆绝缘状态评估方法,主要有以下3个步骤:

(1)对数据进行预处理,将电缆各项特征参数进行归一化或数值化。构建所有特征量的全部特征子集,利用原生LightGBM模型,输出所有特征子集的五折交叉验证准确率。找到准确率最高的特征子集,确定BO-LightGBM模型的输入。

(2)利用BO算法对LightGBM的超参数进行寻优。将LightGBM算法的五折交叉验证准确率作为BO算法中的目标函数。使用高斯过程作为目标函数的概率模型,利用采集函数对概率模型进行判断,确定评估点。将评估点带入目标函数,求得准确率,以此不断迭代,找到准确率最高对应的评估点。

(3)根据最佳特征子集与最优评估点所对应的特征组合和超参数取值,构建LightGBM模型,使用测试集进行绝缘状态评估判断,最后输出各项性能指标结果。

3 算例与分析

3.1 数据集描述

本文数据集取自加拿大西部某大型公用事业公司公开在Kaggle网站的XLPE电缆数据^[21-22],其中包含了2018年2 500根15 kV XLPE电缆的数据。数据集中给出了每根电缆的运行年限、局部放电、视觉条件、中性点腐蚀、负载和健康指数等信息。2 500条电缆健康指数的分布情况如图2所示,其中1~5依次对应上述5个健康指数。

5个特征量中运行年限、局部放电、中性点腐蚀和负载属于数值数据,视觉条件属于有序分类数据。运行年限是指电缆投产正常工作的使用时间。局部放电是评判XLPE电缆运行状态的重要判据,当XLPE电缆开始老化时,局部放电的强度和数量也会相应增加^[23]。视觉条件由电缆检查人员检查电缆是否变色,是否存在表面裂纹和污染来判定,分

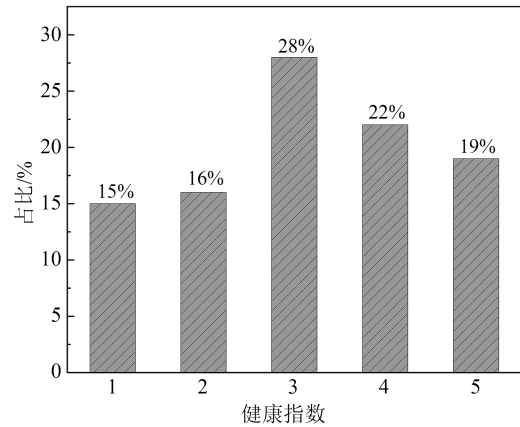


图2 数据集电缆健康指数分布

Fig. 2 Distribution of cable health index in the dataset

为优、良、差3个等级。中性点腐蚀是地下电力电缆重大失效过程的开始,表示电缆与土壤、水等环境发生电化学反应对电缆中性点的破坏程度^[24]。负载反映了电缆的峰值电流,电缆负载过大可能导致老化速度加快。

3.2 不同特征选择方法对LightGBM的影响

针对本文所使用15 kV电缆数据集中的5个特征,可以建立全部的31种特征子集,具体情况如表3所示。

表3 不同特征组合输入

Table 3 Different feature combination inputs

| 特征子集编号 | 输入特征 |
|--------|-------------------------|
| 1 | 运行年限 |
| 2 | 局部放电 |
| 3 | 视觉条件 |
| 4 | 中性点腐蚀 |
| 5 | 负载 |
| 6 | 运行年限、局部放电 |
| | |
| 26 | 运行年限、局部放电、视觉条件、中性点腐蚀 |
| | |
| 30 | 局部放电、视觉条件、中性点腐蚀、负载 |
| 31 | 运行年限、局部放电、视觉条件、中性点腐蚀、负载 |

建立原生LightGBM模型,将31种特征子集依次输入,得到对应的五折交叉验证的准确率,最终输出结果如图3所示。计算结果表明,特征子集10(局部放电与视觉条件)所对应的准确率最高,达到0.968 8。将全部特征进行输入的特征子集31的准确率为0.966 4,与最佳特征子集相比,准确率略有下降。说明本文的方法可以做到使用较少特征量的情况下,达到最佳训练效果。值得注意的是,特

征子集5(负载)所对应的准确率最低,仅有0.232 8,说明电缆负载情况不能很好地表征其绝缘状况。

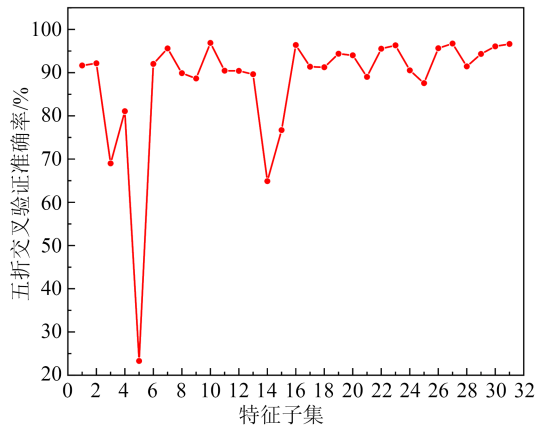


图3 不同特征子集准确率计算结果

Fig. 3 Calculation results of accuracy rate for different feature combinations

除了本文所提出的特征选择方法以外,机器学习中有其他可以进行特征选择的方法,如降维算法中常用的主成分分析法(PCA),过滤法中的互信息检验等。主成分分析法并不是简单地将某一个特征进行舍弃来实现降维,而是将特征量之间的相关信息去除,保留为特定维度的主成分。主成分维度的选取一般是通过计算不同个数主成分的方差解释率,再设定好一个阈值(一般为0.95),选择刚好大于该阈值的主成分数。本文计算不同主成分数的累计方差解释率碎石图如图4所示。

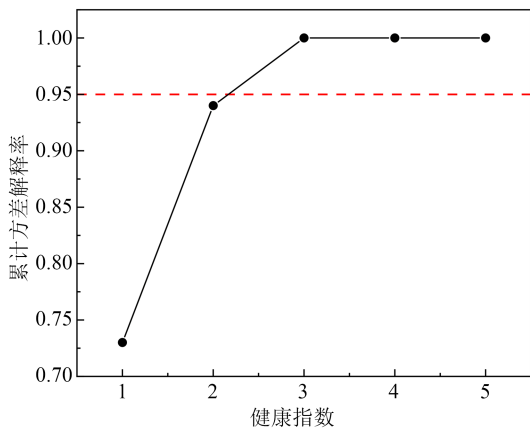


图4 主成分分析法碎石图

Fig. 4 The scree plot of PCA

从图4可以看出,在大于0.95阈值情况下,可以将5个特征量降维成3个主成分。使用3个主成分降维后,输入到原生LightGBM算法,计算五折交叉准确率为0.959 2。与直接将5个特征进行输入的

0.966 4相比,准确率反而有所下降。其原因可能在于使用主成分分析法进行降维会天然丢掉一部分信息,而这部分信息恰恰又可以使样本间存在区分。在特征数量较少时,采用降维算法效果可能会有较大的信息损失。

互信息是信息论中用来衡量两个随机变量相互依存程度的概念。在特征选择过程中,可以用其来评估特征量与目标变量的相关性。互信息值较高意味着特征量与目标变量的相关性强,更适合模型训练。互信息值为零则表示特征与目标变量无关。5个特征与标签健康指数之间的互信息值计算结果如表4所示。

表4 互信息法计算结果

Table 4 Calculation results of mutual information method

| 特征量 | 互信息值 |
|-------|-------|
| 运行年限 | 1.392 |
| 局部放电 | 1.418 |
| 视觉条件 | 0.950 |
| 中性点腐蚀 | 1.184 |
| 负载 | 0.015 |

从表4可以看出,负载的互信息值较低,接近于0,说明负载与标签健康指数之间无明显关系,其结果符合上述关于特征子集5的分析。将负载作为无效特征,仅使用运行年限、局部放电、视觉条件、中性点腐蚀作为特征输入,计算原生LightGBM算法的五折交叉验证准确率为0.956 4,与特征子集26的计算结果一致。与全部特征输入的准确率0.966 4相比,去除无关特征后准确率反而下降,其原因可能是因为使用过多的强相关特征导致模型发生过拟合,而加入一个弱相关特征反而可以提高模型整体性能。同时,强相关特征之间可能存在较高的冗余信息,加入一个弱相关特征可能会引入新的信息,从而提升模型的性能,这也可以说明为什么局部放电与视觉条件的特征组合表现最佳。

综上所述,使用主成分分析法和互信息检验法,均不能很好地提升LightGBM算法的交叉验证准确率。而在数据集特征数不多的情况下,利用本文所提出的筛选特征子集的方法,可以快速、简单地找到最佳特征输入,提高模型准确率。

3.3 不同调优算法对LightGBM的影响

得到最佳特征输入后,使用BO算法对LightGBM算法中的7个超参数进行寻优,迭代次数设置为

50次,以五折交叉验证的准确率作为目标函数调优值。最终超参数计算结果表5所示,调优曲线如图5所示。

表5 LightGBM最优超参数
Table 5 LightGBM optimal hyperparameters

| 超参数名称 | 取值 |
|------------------|-------|
| max_depth | 14 |
| num_leaves | 37 |
| n_estimators | 279 |
| colsample_bytree | 0.941 |
| learning_rate | 0.011 |
| reg_alpha | 0.518 |
| reg_lambda | 0.215 |

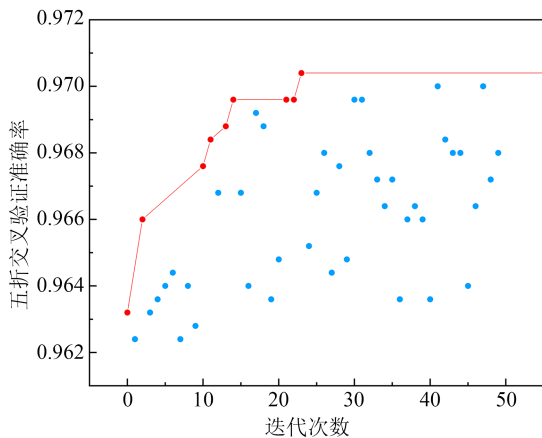


图5 BO调优曲线
Fig.5 BO tuning curve

除了BO算法以外,本文还采用了粒子群优化算法(PSO)、遗传优化(GA)调优算法与其进行对比分析。PSO通过模拟鸟群觅食行为,利用个体间的协作和信息共享来寻找全局最优解,适合连续变量的优化问题,但在维度过高时可能陷入局部最优或搜索效率降低。GA算法通过模拟自然选择和遗传过程,利用选择、交叉和变异操作来优化问题的解,适合多目标优化问题,能有效挑出局部最优解,但代价是收敛速度较慢。两种调优算法的迭代次数均设置为50,五折交叉验证的准确率作为调优函数目标值。两种调优曲线如图6所示,调优计算结果对比如表6所示。

从图6可以看出,PSO算法与GA算法都是最优迭代方式,调优曲线随迭代次数增加而上升,直到收敛。两种算法中PSO迭代收敛速度最快,在第13代时即寻得全局最优,目标函数值为0.97。GA算法则是在第33代时完成收敛,目标函数值为

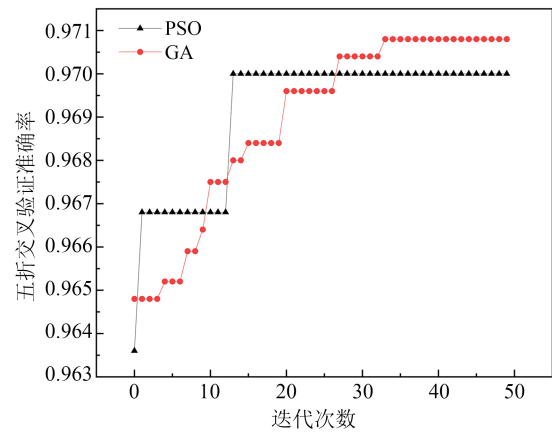


图6 PSO与GA调优曲线
Fig.6 PSO and GA tuning curves

表6 不同调优算法计算结果
Table 6 Calculation results of different optimization algorithms

| 调优算法 | 调优后五折交叉验证准确率 | 训练时间/s |
|------|--------------|---------|
| BO | 0.970 4 | 44.198 |
| PSO | 0.970 0 | 221.252 |
| GA | 0.970 8 | 336.036 |

0.970 8。BO算法由于每次迭代都是根据采样函数计算概率模型中的最大值所对应的观测点来确定的,所以在迭代过程中会出现一定的随机性,导致准确率产生波动,在图5中,每次迭代出现最优位置后用红线标出。从图5可以看到,在50次迭代次数中,第23代为最佳观测点,对应的五折交叉验证准确率为0.970 4,之后的观测点均未超过第23代的结果。由表6可知,3种调优算法的性能表现几乎一致,且均较调优前的0.968 8有所提升。但是三者的调优效率具有较大的差别,BO算法兼具效率与精度。在几乎一致的准确率的情况下,BO算法的运行时间大幅降低,较GA算法效率提升约86.9%,较PSO算法效率提升约为80%。

综合以上分析,对比三种算法,BO算法在速度和精度的综合考量上较PSO与GA算法更具优势,BO的迭代策略非常适合高维空间的参数优化问题。

3.4 实验结果与分析

利用最佳特征输入与最优超参数训练LightGBM算法模型。将数据集2 500条数据按照7:3的比例来构建训练集与测试集,random_state随机种子设置为18。

最终模型测试集输出的准确率达0.985 3,精确

率为0.983 9,召回率为0.986 9,F1-Score为0.985 3。
对应混淆矩阵如图7所示。

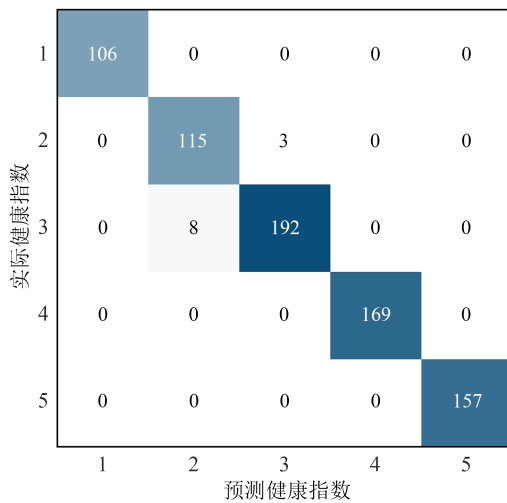


图7 测试集混淆矩阵

Fig.7 Confusion matrix of test set

从图7可以看出,在测试集750条数据里健康指数为1、4、5的样本评估效果很好,未出现错误。而健康指数为2、3的样本评估结果稍有偏差,分别有3条、8条样本预测错误。同时训练集模型准确率为0.964 0,小于测试集结果,说明利用BO调优后的超参数没有使模型发生过拟合。实验结果表明利用局部放电和视觉条件特征组合以及超参数寻优后所建立的BO-LightGBM模型可以很好地评估电缆的老化状态水平。

4 讨论

为进一步验证本文模型性能的优越性,本节将BO-LightGBM算法与原生LightGBM算法和其他机器学习分类算法SVM、XGBoost、KNN、随机森林(RF)进行比较。所对比的算法均按照7:3的比例划分训练集和测试集,算法特征输入为默认全部特征,超参数也选择为默认值。各测试集的性能指标比较结果如表7所示。

从表7可以看出,本文提出的电缆绝缘状态评估方法,无论是准确率、精确率、召回率还是F1-Score均优于其他机器学习算法。BO-LightGBM算法准确率较原生LightGBM算法提升了1.87%,精确率提升了1.6%,召回率提升了1.84%,F1-Score提升了1.71%。所有算法中SVM表现最差,BO-LightGBM算法准确率较其提升了4.93%,精确率提升了4.85%,召回率提升了4.52%,F1-Score提升了

表7 不同算法性能对比

Table 7 Comparison on performance of different algorithms

| 算法模型 | 性能指标 | | | |
|-------------|---------|---------|---------|----------|
| | 准确率 | 精确率 | 召回率 | F1-Score |
| BO-LightGBM | 0.985 3 | 0.983 9 | 0.986 9 | 0.985 3 |
| LightGBM | 0.966 6 | 0.967 9 | 0.968 5 | 0.968 2 |
| XGBoost | 0.964 0 | 0.965 8 | 0.965 3 | 0.965 5 |
| SVM | 0.936 0 | 0.935 4 | 0.941 7 | 0.937 5 |
| RF | 0.957 3 | 0.962 1 | 0.965 8 | 0.963 7 |
| KNN | 0.950 6 | 0.950 8 | 0.956 2 | 0.953 1 |

4.78%。这说明SVM作为一种线性分类器,不能较好地处理非线性问题。同时LightGBM较XGBoost性能略有提升,这主要是因为算法的树枝生长策略不同。XGBoost每遍历一次数据都会同时将同一层的叶子进行分裂,便于进行多线程优化,但也会产生许多不必要的额外开销。LightGBM叶子生长策略是直接从当前所有叶子中找到分裂增益最大的一个叶子,然后再进行分裂。

最后,为验证本文所提方法的普适性。本研究重新选取一个20 kV电缆数据集进行测试,同样按照7:3的比例划分训练集与测试集。数据集中包含了2017年总共3 943条电缆的数据,特征量与15 kV数据集一致,同样为运行年限、局部放电、视觉条件、中性点腐蚀和负载,每条电缆的健康指数作为标签已经给出。最终本研究所提模型在测试集中准确率为0.962 7、精确率为0.961 6、召回率为0.960 2、F1-Score为0.960 9,同样具有优异的表现。

综上所述,利用本文提出的特征选择方法和BO算法对LightGBM算法进行优化后,可以较好地提升模型的性能,综合表现较目前常用机器学习算法相比,均为最优,体现了该模型的优越性。在未来的研究中,可以进一步优化调优算法,减少不必要的迭代次数,从而提升训练效率。除了对电缆绝缘状态进行评估外,还能扩展算法在其他电力设备健康诊断领域的应用,如变压器和断路器,形成智能化电力资产管理体系。

5 结论

(1)利用本文所提出的遍历特征子集的方法可以简单、有效地对数据集的特征进行筛选,找到最佳特征输入。对于本文所使用的数据集,主成分分析法和互信息检验法均不能很好地对特征进行筛选,甚至会导致模型性能进一步下降。

(2)BO、PSO和GA调优算法对于模型的性能提升表现几乎一致,但是3种算法的计算效率差别较大。BO算法兼顾了计算的精度与效率,运行时间较GA算法提升约86.9%,较PSO算法提升约为80%,非常适合高维的超参数优化。

(3)针对加拿大西部资产公开的15 kV电缆数据集,BO-LightGBM相比其他机器学习算法,表现出最优的模型性能。本文提出的基于BO-LightGBM算法的多分类电缆老化状态评估模型,精度高、性能好,各项参数指标均达到98%以上,能够有效地评估电缆状态。

参考文献 References

- [1] 赵艾萱,刘健,徐龙,等.基于PDC法的在运XLPE电缆绝缘状态评估[J].高电压技术,2019,45(5):1542-1550.
ZHAO Aixuan, LIU Jian, XU Long, et al. Insulation status evaluation of in-service cables based on polarization and depolarization current[J]. High Voltage Engineering,2019,45(5):1542-1550.
- [2] CHMURA L, JIN H F, CICHECKI P, et al. Use of dissipation factor for life consumption assessment and future life modelling of oil-filled high-voltage power cables[J]. IEEE Electrical Insulation Magazine,2012,28(1):27-37.
- [3] 赵世林,周凯,何珉,等.冲击电压下电缆的介电响应特性与绝缘状态评估[J].高电压技术,2019,45(4):1297-1304.
ZHAO Shilin, ZHOU Kai, HE Min, et al. Dielectric response characteristics and insulation condition evaluation under impulse voltage for cables[J]. High Voltage Engineering,2019,45(4):1297-1304.
- [4] 王尉军,陈旻,殷慧,等.基于多维介电参数雷达谱图的XLPE电缆老化状态综合评估[J].绝缘材料,2024,57(2):72-79.
WANG Weijun, CHEN Min, YIN Hui, et al. Comprehensive evaluation for ageing condition of XLPE cable based on radar chart of multidimensional dielectric parameters[J]. Insulating Materials, 2024,57(2):72-79.
- [5] 王昊月,孙茂伦,赵凯杰,等.电缆绝缘老化的高频频域介电谱诊断评估方法[J].中国电机工程学报,2023,43(9):3630-3642.
WANG Haoyue, SUN Maolun, ZHAO Kaijie, et al. High voltage FDS diagnosis and evaluation method for cable insulation aging [J]. Proceedings of the CSEE,2023,43(9):3630-3642.
- [6] 宋微,贾步超,张鹏,等.基于IRC法和局部放电法的车载高压电缆绝缘状态评估[J].高压电器,2024,60(1):118-127,135.
SONG Wei, JIA Buchao, ZHANG Peng, et al. Insulation condition assessment of vehicle high voltage cable based on IRC method and partial discharge method[J]. High Voltage Apparatus,2024, 60(1):118-127,135.
- [7] 王天,白银浩,吕中宾,等.交联聚乙烯电缆绝缘老化试验及其检测技术[J].绝缘材料,2022,55(6):6-15.
WANG Tian, BAI Yin hao, LÜ Zhongbin, et al. Progress in insulation ageing test and detecting technology of XLPE cable[J]. Insulating Materials,2022,55(6):6-15.
- [8] 王璨,赵学童,张杰,等.基于超低频介损和U-I滞回曲线的XLPE电缆水树老化状态评估[J].中国电机工程学报,2023,43(13):5261-5271.
WANG Can, ZHAO Xuetong, ZHANG Jie, et al. Condition assessment of water-tree aged XLPE cables via very-low frequency dielectric loss and U-I hysteresis curve[J]. Proceedings of the CSEE,2023,43(13):5261-5271.
- [9] 叶刚,姚星辰,李涛,等.XLPE电缆绝缘老化的时频域介电特性[J].高电压技术,2018,44(11):3713-3719.
YE Gang, YAO Xingchen, LI Tao, et al. Time/frequency domain dielectric characteristics of XLPE cable insulation aging[J]. High Voltage Engineering,2018,44(11):3713-3719.
- [10] 马骁,马弢,王晓峰,等.基于PDC多特征量权重的电缆绝缘老化状态评估方法研究[J].绝缘材料,2022,55(9):80-88.
MA Xiao, MA Tao, WANG Xiaofeng, et al. Evaluation method of cable insulation ageing state based on PDC multi feature weight[J]. Insulating Materials,2022,55(9):80-88.
- [11] 刘青,张浩然,尚英强,等.基于主成分分析法的高压服役电缆主绝缘老化状态评估[J].绝缘材料,2022,55(11):56-62.
LIU Qing, ZHANG Haoran, SHANG Yingqiang, et al. Assessment on ageing condition of served HV cable main insulation based on principal component analysis[J]. Insulating Materials, 2022,55(11):56-62.
- [12] 李登淑,王昕,赵明,等.基于集对分析动态加权法的XLPE绝缘老化状态评估[J].水电能源科学,2020,38(7):194-197,172.
LI Dengshu, WANG Xin, ZHAO Ming, et al. Evaluation of XLPE insulation aging status based on set pair analysis and dynamic weighting method[J]. Water Resources and Power, 2020, 38(7):194-197,172.
- [13] 张琦,夏向阳,杨志峰,等.基于模糊理论的XLPE电缆绝缘状态评估研究[J].绝缘材料,2014,47(2):84-88.
ZHANG Qi, XIA Xiangyang, YANG Zhifeng, et al. Insulation condition assessment study for XLPE cable based on fuzzy theory[J]. Insulating Materials,2014,47(2):84-88.
- [14] SAHOO R, KARMAKAR S, PANIGRAHY S. Health index analysis of XLPE cable insulation using machine learning technique[C]//2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering. Prayagraj, India: IEEE,2020.
- [15] YUN F, DONG H, LIANG C, et al. Feature selection of XLPE cable condition diagnosis based on PSO-SVM[J]. Arabian Journal for Science and Engineering,2023,48(5):5953-5963.
- [16] PAN Q S, ZHANG C, WEI X Y, et al. Assessment of MV XLPE cable aging state based on PSO-XGBoost algorithm[J]. Electric Power Systems Research,2023,221:1-8.
- [17] KE G L, MENG Q, FINLEY T, et al. Lightgbm: a highly efficient gradient boosting decision tree[J]. Advances in Neural Information Processing Systems,2017,30:1-9
- [18] 欧阳彤,汪玲,朱岱寅,等.基于LightGBM的气象目标分类技术[J].雷达科学与技术,2023,21(6):621-629.
OUYANG Tong, WANG Ling, ZHU Daiyin, et al. Meteorologi-

- cal target classification technology based on LightGBM[J]. Radar Science and Technology,2023,21(6):621-629.
- [19] SHAHRIARI B, SWERSKY K, WANG Z, et al. Taking the human out of the loop: a review of bayesian optimization[J]. Proceedings of the IEEE,2015,104(1):148-175.
- [20] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:6-9.
ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press,2016:6-9.
- [21] DONG M, LI W Y, NASSIF A B. Long-term health index prediction for power asset classes based on sequence learning[J]. IEEE Transactions on Power Delivery,2021,37(1):197-207.
- [22] 15kV XLPE underground cable inspection dataset[DS/OL]. [2024-04-20]. <https://www.kaggle.com/datasets/utilityanalytics/utility-underground-cable-dataset1>.
- [23] CHANG C K, BOYANAPALLI B K. Assessment of the insulation status aging in power cable joints using support vector machine[J]. IEEE Transactions on Dielectrics and Electrical Insulation,2021,28(6):2170-2177.
- [24] LEE C K, KWAK K S, PARK J B, et al. Concentric neutrals corrosion localization and its impedance analysis in the underground power cable system based on the reflectometry[C]// 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings. Graz, Austria: IEEE,2012.

收稿日期:2024-09-26;修回日期:2024-11-28。

作者简介:

罗正均(2001-),男(汉族),四川成都人,硕士生,主要从事电力设备绝缘状态检测与诊断的研究;

通信作者:叶刚(1980-),男(汉族),湖北荆州人,副教授,博士,主要从事高电压与绝缘技术、电力设备绝缘状态检测的研究。