基于贝叶斯优化极端梯度提升树的电缆状态分类研究

佘　维[1,2,3]，王　欣[1,3]，陈　斌[4]，吕钟毓[1,3]，张海丽[1,3]，田　钊[1,3]

(1. 郑州大学 网络空间安全学院　河南 郑州 450002；2. 嵩山实验室　河南 郑州 450046；
3. 郑州市区块链与数据智能重点实验室　河南 郑州 450002；4. 许继电气股份有限公司　河南 许昌 461099)

摘要：针对多分类问题中样本类间不均衡引起的电缆状态分类准确性不高的问题，提出一种基于贝叶斯优化极端梯度提升树的电缆状态分类方法。首先，利用贝叶斯优化对极端梯度提升树算法里面的超参数进行训练，获取最优超参数配置。其次，将最优超参数配置应用于极端梯度提升树算法中，得到 Bo-XGBoost 分类模型。最后，通过实例验证该分类方法相较于 SVM、TabNet、LightGBM 等方法有更高的准确性，可为电缆状态分类提供一种新方向。

关键词：贝叶斯优化；极端梯度提升树；电缆状态分类；超参数优化

中图分类号：TP183　　**文献标志码**：A　　**文章编号**：1671-6841(2025)06-0001-07

DOI：10.13705/j.issn.1671-6841.2024079

Research on Cable Condition Classification Based on Bayesian Optimization of Extreme Gradient Boosting Tree

SHE Wei[1,2,3], WANG Xin[1,3], CHEN Bin[4], LYU Zhongyu[1,3], ZHANG Haili[1,3], TIAN Zhao[1,3]

(1. *School of Cyber Science and Engineering*, *Zhengzhou University*, *Zhengzhou* 450002, *China*;
2. *Songshan Laboratory*, *Zhengzhou* 450046, *China*; 3. *Zhengzhou Key Laboratory of Blockchain and Data Intelligence*, *Zhengzhou* 450002, *China*; 4. *XJ Electric Co. Ltd.*, *Xuchang* 461099, *China*)

Abstract: Addressing the issue of low accuracy in cable condition classification due to imbalanced sample classes in multiclass classification problems, a cable condition classification method based on Bayesian-optimized extreme gradient boosting was proposed. Firstly, Bayesian optimization was employed to train the hyperparameters within the XGBoost algorithm, with the aim of acquiring the optimal hyperparameter configuration. Then, this optimal hyperparameter configuration was applied to the XGBoost algorithm, which resulted in the Bo-XGBoost classification model. Finally, the verification through case studies demonstrated that this classification method achieved higher accuracy compared to methods such as SVM, TabNet, and LightGBM, thereby providing a new direction for cable condition classification.

Key words: Bayesian optimization; extreme gradient boosting tree; cable condition classification; hyperparameter optimization

0　引言

电缆的数量在电力现代化进程中迅速增长，形成了一个庞大且复杂的网络体系。但由于复杂的操作环境、工况变化以及外部因素的影响，容易出现缺陷或故障，给电力系统带来潜在威胁，及时准确地掌握电缆的运行状态成为当前发展的必然需求[1]。随着电缆的规模不断扩大，检测手段也日益多样化，为准确评估电缆状态提供了丰富的数据支持[2]。因此，对电缆运行状态各项参数进行综合分析，提取核心指标，构建多维度风险评估模型，对电缆状态分

收稿日期：2024-04-19
基金项目：嵩山实验室预研项目(YYYY022022003)；河南省重点研发与推广专项(科技攻关)(212102310039)
第一作者：佘维(1977—　)，男，教授，主要从事复杂系统建模与仿真、机器学习、区块链与数据智能研究，E-mail：wshe@zzu.edu.cn。
通信作者：田钊(1985—　)，男，副教授，主要从事群体智能、机器学习、区块链与数据智能研究，E-mail：tianzhao@zzu.edu.cn。

类和评估具有深远的研究和应用价值[3]。

近年来,基于机器学习的电缆状态监测方法得到更多应用。文献[4]提出了一个基于大数据技术的电力数据质量评估框架,该框架能够整合实时和历史数据,为电力大数据分类提供一体化的计算环境,并实现对不同类型数据的存储支持。文献[5]提出了一种基于概率神经网络算法的电缆状态评估方法,运用因子分析法对不完备信息下的数据库进行预处理,提出基于概率神经网络的多源信息融合算法,在不完备信息条件下对电缆整体状态进行综合评估。文献[6]通过对11种典型电缆缺陷进行模拟试验,提出了基于多状态量特征及变化规律的高压电缆状态综合评价方法,并通过实例验证了其有效性。文献[7]在极化-去极化电流(PDC)检测诊断方法的基础上,对PDC法进行改进,将提取的单一状态分类特征量改良为具有不同权重的多个状态分类特征量,提高了电缆状态分类的准确性。文献[8]提出了一种基于长短时记忆网络的配网电缆状态评估方法,该方法利用加权Bagging算法对所述历史特征参数进行融合得到加权后的融合特征参数,然后构建基于长短时记忆网络的配网电缆绝缘状态评估模型,实现电缆状态在线监测。

上述文献虽然涉及对电缆状态特征量的分析评估,但对于因多分类问题中样本类间不平衡而导致电缆状态分类准确性不高的问题未进行深入具体探讨。为弥补这一不足,本文提出了一种基于贝叶斯优化极端梯度提升树的电缆状态分类方法,缓解了传统分类模型可能面临的过拟合问题,提高了状态分类的准确性。

1 相关知识

1.1 极端梯度提升树

极端梯度提升(extreme gradient boosting,XGBoost)树[9]是建立在梯度提升决策树(gradient boosting decision tree,GBDT)基础上的机器学习模型,其目标是将速度和效率发挥到极致。极端梯度提升树的一般过程如下。

1) 初始化:在算法开始时,XGBoost会构建一个简单的基模型作为起点,对数据进行初步预测。

2) 迭代训练:通过多次迭代,每次基于前次预测值与真实值间的误差(残差)训练新决策树,旨在纠正前次错误。

3) 加权融合:每次迭代后,新的决策树模型会被添加到模型中,并与之前的模型进行加权融合。这个加权融合的过程是根据每个模型的预测性能来决定的,预测性能越好的模型会得到更大的权重。

4) 损失函数优化:每一步迭代均计算损失函数,该函数综合了预测误差与模型复杂度,XGBoost通过最小化此函数提升模型性能。

5) 添加正则化项:为防止过拟合,损失函数中融入正则项,平衡模型拟合能力与泛化能力。

6) 终止条件:当达到预设迭代次数或模型在验证集上性能下降时,迭代停止,输出最终优化模型。

XGBoost通过梯度提升策略持续增强模型,每轮迭代精准拟合残差。其利用二阶梯度直方图加速训练,并有效防控过拟合,这使得XGBoost在处理大规模数据集和高维特征时具有出色的性能。

1.2 交叉熵损失函数

交叉熵损失函数[10]是一种在分类问题中常用的损失函数,它用于衡量模型的预测概率分布与真实数据分布之间的差异,通常在监督学习的分类任务中用于优化模型的参数,计算公式为

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)), \tag{1}$$

其中:p代表真实的概率分布,在多分类问题中,p表示样本的真实类别;q代表模型预测的概率分布,其中的每个元素表示模型预测样本属于相应类别的概率。交叉熵具有非负性和凸函数性质。交叉熵损失函数在极端梯度提升树中能够更好地满足电缆状态分类的需求,优化特定性能指标,处理不平衡数据,并生成合理的概率分布。

1.3 贝叶斯优化

贝叶斯优化[11]是一种自动化调优机器学习模型超参数的智能方法,它基于先验知识不断更新对目标函数(通常是模型性能指标)的估计,智能地选择下一个超参数配置,以优化模型性能。且对噪声和异常值具有强鲁棒性,且能动态调整策略,高效利用计算资源加速优化流程[12]。

2 Bo-XGBoost模型

本节提出的贝叶斯优化极端梯度提升树(Bo-XGBoost)模型具体流程见图1。

1) 初始化Bo-XGBoost模型中的目标函数,

$$J(f_t) = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) + C, \tag{2}$$

其中:y_i表示样本i的实际值;$\hat{y}_i^{t-1}$表示前$t-1$棵决策树对样本的预测值(将前$t-1$棵决策树中的每棵

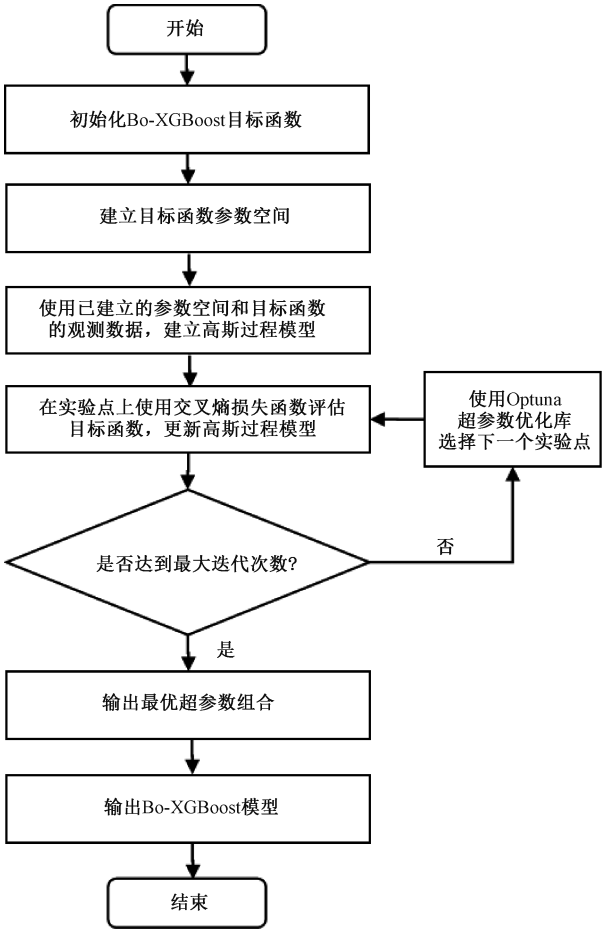

图 1　Bo-XGBoost 算法流程图

Figure 1　The flow chart of Bo-XGBoost algorithm

决策树的预测值相加所得)；$L(y_i,\hat{y}_i^{t-1})$ 表示与 y_i、$\hat{y}_i^{t-1}$ 有关的损失函数；$f_t(x_i)$ 表示第 t 棵决策树对样本 i 的预测值；$\Omega(f_t)$ 表示第 t 棵树的模型复杂度。决策树复杂度计算公式为

$$\Omega(f_t)=\gamma T_t+\lambda\frac{1}{2}\sum_{j=1}^{T}w_j^2, \tag{3}$$

其中：T 表示叶子结点数量；w_j 表示该树中第 j 个叶子结点的值。

2) 建立参数空间。选择 *learning_rate*、*max_depth*、*gamma*、*subsample* 作为 Bo-XGBoost 模型的可选参数,构建参数空间 χ,

$$\forall x_i\in\chi。 \tag{4}$$

参数空间 χ 为 Bo-XGBoost 模型提供了一个明确、可量化的搜索范围,可以更加高效地找到能使目标函数 $J(f_t)$ 达到最优的参数值。

3) 建立高斯过程模型。高斯过程(Gaussian process, GP)模型是一种用于建模连续函数的概率方法,是 Bo-XGBoost 模型中目标函数 $J(f_t)$ 的先验概率模型。它假设任意一组输入变量对应的输出都服从一个联合高斯分布,可以通过已知的观测点推断出未知点的函数值及其不确定性。使用参数空间 χ 和目标函数 $J(f_t)$ 的观测数据作为输入,建立一个高斯过程模型,

$$GP=\{\chi,J(f_t)\}。 \tag{5}$$

高斯过程是多元高斯概率分布的范化,它由一个均值函数 $m:X\to R$ 和一个半正定的协方差函数 $k:X\times X\to R$ 构成[13],

$$f(x)\sim GP(m(x),k(x,x')), \tag{6}$$

其中:均值函数 $m(x)=E[f(x)]$ 是用来描述函数的平均值随输入变化的函数;协方差函数 $k(x,x')=E[(f(x)-m(x))(f(x')-m(x'))]$ 用于描述函数值在不同输入点之间的相关性,通常设置均值函数 $m(x)=0$。

4) 选择下一个实验点。通过高斯过程模型预测目标函数的均值和不确定性后,Bo-XGBoost 模型使用 Optuna 超参数优化库智能地选择下一个实验点 x_i(一组新的参数),从而高效地探索搜索空间,

$$x_i=\mathrm{argmax}_x uncertainty(GP)。 \tag{7}$$

5) 在实验点上评估目标函数。在选定的实验点 x_i 上运行目标函数 $J(f_x)$,获取相应的目标函数值 y_i 和交叉熵损失函数评估数值。这样就可以得到实验点的实际效果,用于更新高斯过程模型,

$$y_i=J(f_{x_i})。 \tag{8}$$

6) 更新高斯过程模型。将新的观测结果 (x_i,y_i) 加入高斯过程模型中,更新模型的参数和不确定性,以此不断提升代理模型的准确性和预测能力,

$$GP'=update(GP,x_i,y_i)。 \tag{9}$$

7) 重复迭代。重复以上步骤,每次都选择新的实验点 x_i,评估目标函数 $J(f_x)$,更新代理模型,直到达到预定的迭代次数或者目标函数收敛到一个预定的阈值,最终输出 Bo-XGBoost 模型,具体算法流程伪代码如算法 1 所示。

算法 1　Bo-XGBoost 模型流程伪代码

输入:树的深度(*max_depth*),学习率(*learning_rate*),子采样率(*subsample*),控制树生长参数(*gamma*)的取值范围,高斯过程(*GP*),交叉熵损失函数($f(\cdot)$),最大优化轮数(N),验证数据集(NP)。

输出:Bo-XGBoost 模型。

1) $\chi\leftarrow initialHyperParams()$

2) $(max\_depth,learning\_rate,subsample,gamma)\leftarrow paramsRange(minval,maxval)$

3) $GP\leftarrow GaussianProcessModel()$,$f(\cdot)\leftarrow Cross\text{-}Entropy\ Loss()$,$BO\leftarrow initialBO(GP,f)$

4) $N\leftarrow max\_epoch$

5) WHILE $i<N$

6) $x_i \leftarrow BO.findnextpoint()$

7) $xgboost\_model \leftarrow xgboost.train(params = x_i)$

8) $fitness \leftarrow mean\_e(NP\_labels, xgboost\_model.predict(NP\_features))$

9) $BO.update(x_i, fitness)$

10) END WHILE

11) $best\_params \leftarrow BO.get\_best()$

12) $BO\text{-}xgboost\_model \leftarrow xgboost.train(params = best\_params)$

在上述代码中,Bo-XGBoost 模型优化器使用高斯过程模型来建立超参数与验证误差之间的关系,并通过 Optuna 超参数优化库来选择下一个要评估的超参数组合。然后,使用选定的超参数训练 Bo-XGBoost 模型,并在验证集上通过交叉熵损失函数来评估其性能。最后,将评估结果反馈给 Bo-XGBoost 模型优化器,以更新其内部的高斯过程模型。在达到最大优化轮数后,算法返回在优化过程中找到的最优超参数组合,最终输出 Bo-XGBoost 算法模型。

3 基于 Bo-XGBoost 模型的电缆状态分类方法

本文在电缆状态分类方面提出基于 Bo-XGBoost 模型的分类方法,减轻了传统分类模型面临的过拟合问题,具体过程见图 2。

先对收集的原始电缆数据集 D 进行清洗,去除重复、错误或缺失严重的样本。异常数据不仅会降低模型预测的成功率,而且会出现预测结果与真实值偏差较大的现象。然后进行数据无量纲归一化处理,将原始指标数据通过简单的数学变换来消除各指标量纲的影响,把所有的指标值规范化到同一个范围内进行比较,从而使所有指标数据之和等于 1,解决各指标数值不能直接进行合成、无法综合分析与比较的问题。将获得的电缆状态数据进行数据无量纲归一化处理,对原始数据进行线性变换,将其映射到[0,1]之间,计算为

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{10}$$

式中:min 是样本的最小值;max 是样本的最大值。

处理后的数据集为 D',将 D' 按照一定比例划分为训练集和测试集,选择大部分数据作为训练集,用于训练 Bo-XGBoost 模型,剩余部分作为测试集,用于评估模型的性能。

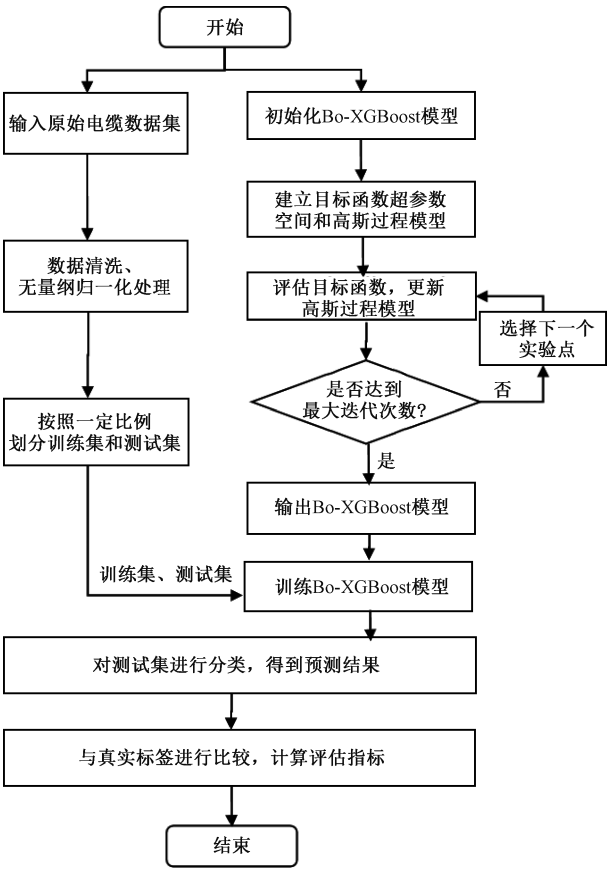

图 2 基于 Bo-XGBoost 算法的电缆状态分类流程图

Figure 2 The flow chart of cable condition classification based on Bo-XGBoost algorithm

接着采用本文所提到的基于 Bo-XGBoost 模型的电缆状态分类方法对已标记的数据集 D_{train} 进行训练。通过这一过程,Bo-XGBoost 模型将学习并建立起电缆状态与各种特征之间的关联关系。

1) 初始化 Bo-XGBoost 模型目标函数,确定目标函数超参数空间 χ,并为每一个参数设置可取值范围:$max\_depth = [1,15]$,$learning\_rate = [0.000\,5, 0.1]$,$subsample = [0.5, 1]$,$gamma = [0.000\,01, 5]$。

2) 建立高斯过程模型。高斯过程模型 $M(x)$ 将根据输入的已知超参数配置和相应性能指标值,建立一个概率模型来预测未知超参数配置的性能,即

$$M(x) = GP[m(x), k(x, x')], \tag{11}$$

其中:$m(x)$ 是均值函数,它定义了 x 处随机变量的期望;$k(x, x')$ 是协方差函数,它定义了 x 和 x' 处随机变量之间的协方差,协方差函数是高斯过程的核心,决定了随机变量的空间相关性。

3) 评估目标函数。在选定的超参数配置下,采用式(1)计算交叉熵损失。这个损失值就是目标函数的输出,用于衡量当前超参数配置下的模型性能,并被用于更新 Bo-XGBoost 模型优化中的模型参数

和后验分布。

4) Bo-XGBoost 模型优化。Bo-XGBoost 模型优化具体过程包括:选择优化算法、寻找最佳超参数配置、更新高斯过程模型。本文选用 Optuna 超参数优化库迭代寻找最佳的超参数配置。在每次迭代中,优化算法会根据高斯过程模型的预测选择下一个超参数配置,并在目标函数上评估其性能。随着迭代的进行,高斯过程模型将不断更新,以更好地捕捉超参数与性能指标之间的关系,这有助于优化算法更智能地选择下一个配置。继续迭代到一定的迭代次数时终止,最终将得到 Bo-XGBoost 模型 f。

5) 通过模型 f 对测试数据 D_{test} 进行预测,得到预测结果 y', 并将预测结果与真实标签 y 进行比较,计算评估指标。这个评估阶段的目的是了解模型的泛化性能,即模型在未见过的数据上的表现。通过这个过程,可以对模型的准确度和可靠性进行全面评估。

最终,根据评估指标和业务需求,对模型的输出进行解释和分析,得出对电缆状态的分类结论。

4　仿真实验及分析

4.1　数据集的选取

为了验证 Bo-XGBoost 模型在电缆状态分类中的有效性,本文收集 1 000 条电缆数据并对其进行数据预处理,确保数据的准确性和完整性。将处理后的数据统一按照 8:2比例划分训练集和测试集,测试集与训练集在时间和空间上保持独立,以确保样本的代表性和合理性。

对电缆状态的分类主要基于四个关键指标:运行时间、接地故障、局部放电以及介质损耗,根据国家电网江苏省电力科学研究院专家评定[14],将电缆状态(class)分为轻度老化、中度老化和严重老化三大类,表 1 所示为电缆的部分数据。

表 1　电缆运行部分参数数据

Table 1　Some parameters of cable operation

序号	运行时间/月	接地故障	局部放电/pC	介质损耗	class
1	180	0	7.0	0.036	1
2	216	1	10.6	0.026	3
3	156	0	9.0	0.030	1
4	188	1	6.0	0.035	2
5	156	0	20.0	0.109	3
…	…	…	…	…	…

注:接地故障中的 0 表示未发生,1 表示发生过故障;class 中 1 表示轻度老化,2 表示中度老化,3 表示重度老化。

4.2　Bo-XGBoost 优化结果分析

本文采用 Bo-XGBoost 模型对电缆样本数据进行状态分类,设置迭代次数为 30 轮,其结果如图 3 所示。

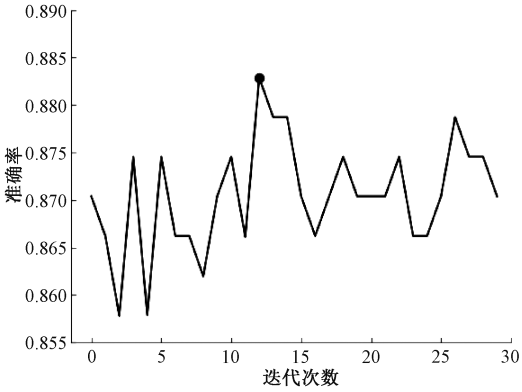

图 3　Bo-XGBoost 优化结果图

Figure 3　Bo-XGBoost optimization result graph

通过观察图 3,可以发现在不同迭代次数下,准确率一直在变化。经过 30 次迭代后,得出 Bo-XGBoost 模型最佳的迭代次数为 13 次,可达到最高的分类准确率 0.883,输出最佳超参数配置为:$max\_depth = 8$;$learning\_rate = 0.043\ 25$;$subsample = 0.712\ 78$;$gamma = 0.001\ 65$。对于电缆状态数据而言,Bo-XGBoost 模型能够在较短时间内找到超参数的最优配置从而提高分类的准确性。

4.3　对比分析

为了更全面地展示本文模型性能,选择经典分类算法(SVM、Random Forest)、深度学习算法(GoogleNet[15]、TabNet[16])、集成算法(GBDT[17]、LightGBM[18]、AdaBoost[19])以及优化前的 XGBoost 算法作为对比实验的基准算法,在准确率、召回率和处理时间方面进行对比分析。

4.3.1　*准确率和召回率对比*　在电缆数据多分类问题中,分类模型的准确率是指该模型在所有电缆数据类别上正确分类的样本数占总样本数的比例,计算公式为

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%,$$

其中:TP 是真正例;FP 是假正例;TN 是真反例;FN 是假反例。准确率的值越高表示模型性能越好。

召回率衡量了模型识别的正样本数量与实际正样本数量之间的比例。召回率表示模型在所有实际为正样本的样本中,正确识别为正样本的比例,计算公式为

$$recall = \frac{TP}{TP + FN} \times 100\%。$$

召回率越高,表示所有真正为正例的样本中,被正确预测的比例越高,即漏报率越低。

本文采用准确率和召回率作为各算法模型对电缆状态进行分类的指标,具体结果见表2。

表2 各分类算法模型准确率和召回率对比

Table 2 Comparison of accuracy and recall of various classification algorithm models

算法名称	准确率/%	召回率/%
SVM	65.00	69.70
Random Forest	75.00	82.60
GoogleNet	82.66	81.27
TabNet	83.33	83.33
GBDT	78.33	79.93
LightGBM	80.00	85.62
AdaBoost	73.33	80.94
XGBoost	81.67	82.27
Bo-XGBoost	88.29	93.98

通过表2对比,可以看出经过贝叶斯参数优化后的极端梯度提升树算法在准确率和召回率上相较于优化前有明显提升。在准确率方面,相较于其他分类算法,Bo-XGBoost以88.29%领先。从召回率的结果来看,与其他分类算法相比,Bo-XGBoost相对较好,召回率可以达到93.98%,具有较强的正例识别能力。

ROC(receiver operating characteristic)曲线通过绘制真正例率(true positive rate,TPR)和假正例率(false positive rate,FPR)之间的关系图,为模型性能评估提供了客观视角。高性能模型对应的*ROC*曲线靠近左上角,表明在保持低误报率的同时有效识别正例。反之,性能较差的模型*ROC*曲线接近对角线或更低位置,分类效果不佳。*AUC*(area under the curve)值作为*ROC*曲线下面积,量化分类器性能,范围为0.5~1,值越大表示分类器综合性能越强。本文比较了Bo-XGBoost模型在三类状态数据样本上的表现,绘制了相应的*ROC*曲线并计算了*AUC*值。经过2 000轮迭代后结果如图4所示,证明了本文方法在解决多分类样本预测问题上的可行性。

4.3.2 处理时间对比 处理时间是指系统或算法完成特定数据处理任务所需的总时间长度。这个时间从数据处理任务开始执行算起,直到任务完成并输出结果为止。本文采用处理时间作为各算法模型对电缆状态进行分类的另一个评估标准,以衡量不同模型在处理电缆状态分类任务时的速度表现。具体结果如表3所示,通过对比不同模型的处理时间指标,可以更加全面地了解各模型在性能上的差异,为实际应用中的模型选择和优化提供参考依据。

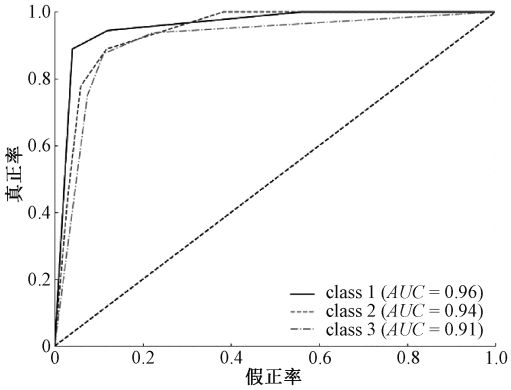

图4 *ROC*曲线图

Figure 4 *ROC* curve graph

表3 各分类算法模型处理时间对比

Table 3 Comparison of processing time for various classification algorithm models

算法名称	处理时间/s
SVM	0.928
Random Forest	0.186
GoogleNet	5.653
TabNet	6.452
GBDT	0.178
LightGBM	0.151
AdaBoost	0.133
XGBoost	0.111
Bo-XGBoost	0.082

从表3结果看出,Bo-XGBoost展现出了相对较高的数据处理能力。与传统的XGBoost算法相比,时间从0.111 s缩短至0.082 s,有较为显著的提升。与深度学习算法和集成算法相比,本文模型在处理电缆数据集上展现出更好的实时性能。

综上所述,本文模型在电缆状态分类中表现出较好的效果。这一优势主要体现在经贝叶斯优化后的极端梯度提升树模型在处理电缆样本数据时更为鲁棒,可以有效缓解模型过拟合问题。

5 结语

本文针对电缆状态分类问题,提出了利用贝叶斯超参数优化对极端梯度提升树算法进行优化的方法。通过实例验证,相比传统的电缆状态分类方法,本文所提出的模型在准确性和有效性方面得到提升,为电缆状态监测领域提供了一种相对较优的分类方法。

参考文献:

[1] 江秀臣, 盛戈皞. 电力设备状态大数据分析的研究和应用[J]. 高电压技术, 2018, 44(4): 1041-1050.
JIANG X C, SHENG G H. Research and application of big data analysis of power equipment condition[J]. High voltage engineering, 2018, 44(4): 1041-1050.

[2] 江秀臣, 罗林根, 余钟民, 等. 区块链在电力设备泛在物联网应用的关键技术及方案[J]. 高电压技术, 2019, 45(11): 3393-3400.
JIANG X C, LUO L G, YU Z M, et al. Technologies and solutions of blockchain application in power equipment ubiquitous Internet of Things[J]. High voltage engineering, 2019, 45(11): 3393-3400.

[3] 鞠平, 周孝信, 陈维江, 等. "智能电网+"研究综述[J]. 电力自动化设备, 2018, 38(5):2-11.
JU P, ZHOU X X, CHEN W J, et al. "Smart Grid Plus" research overview[J]. Electric power automation equipment, 2018, 38(5): 2-11.

[4] LIU H, HUANG F P, LI H, et al. A big data framework for electric power data quality assessment[C]//The 14th Web Information Systems and Applications Conference. Piscataway: IEEE Press, 2017: 289-292.

[5] 李玉凌, 李二霞, 樊勇华, 等. 一种基于概率神经网络算法的电缆状态评估方法: CN106251059-B[P]. 2016-12-21.
LI Y L, LI E X, FAN Y H, et al. A cable condition evaluation method based on probabilistic neural network algorithm:CN106251059-B[P]. 2016-12-21.

[6] 周路遥, 曹俊平, 王少华, 等. 基于多状态量特征及变化规律的高压电缆状态综合评估[J]. 高电压技术, 2019, 45(12): 3954-3963.
ZHOU L Y, CAO J P, WANG S H, et al. Comprehensive state evaluation of high voltage cable based on multi-state variables characteristics and variation law[J]. High voltage engineering, 2019, 45(12): 3954-3963.

[7] 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 X, MA T, WANG X F, et al. Evaluation method of cable insulation ageing state based on PDC multi feature weight[J]. Insulating materials, 2022, 55(9): 80-88.

[8] 董芝春, 冯宝, 周慧彬, 等. 一种基于长短时记忆网络的配网电缆状态评估方法及系统: CN117171642-A[P]. 2023-12-05.
DONG Z C, FENG B, ZHOU H L, et al. A distribution network cable condition evaluation method and system based on long short-term memory network:CN117171642-A[P]. 2023-12-05.

[9] 李占山, 刘兆赓. 基于 XGBoost 的特征选择算法[J]. 通信学报, 2019, 40(10): 101-108.
LI Z S, LIU Z G. Feature selection algorithm based on XGBoost[J]. Journal on communications, 2019, 40(10): 101-108.

[10] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. Annals of operations research, 2005, 134(1): 19-67.

[11] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. 软件学报, 2018, 29(10): 3068-3090.
CUI J X, YANG B. Survey on Bayesian optimization methodology and applications[J]. Journal of software, 2018, 29(10): 3068-3090.

[12] 李亚茹, 张宇来, 王佳晨. 面向超参数估计的贝叶斯优化方法综述[J]. 计算机科学, 2022, 49(S1):86-92.
LI Y R, ZHANG Y L, WANG J C. Overview of Bayesian optimization methods for hyperparameter estimation[J]. Computer science, 2022, 49(S1):86-92.

[13] 王方成, 刘玉敏, 崔庆安. 基于高斯过程回归的混合型参数建模及优化[J]. 统计与决策, 2023, 39(1): 34-39.
WANG F C, LIU Y M, CUI Q A. Hybrid parameter modeling and optimization based on Gaussian process regression[J]. Statistics & decision, 2023, 39(1), 34-39.

[14] 谭笑, 陈杰, 刘洋, 等. 基于朴素贝叶斯的电缆老化状态评估方法和装置: CN111610407B[P]. 2022-09-01.
TAN X, CHEN J, LIU Y, et al. Cable aging state assessment method and apparatus based on naive Bayes: CN111610407-A[P]. 2020-09-01.

[15] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1-9.

[16] ARIK S Ö, PFISTER T. TabNet: attentive interpretable tabular learning[J]. Proceedings of the AAAI conference on artificial intelligence, 2021, 35(8): 6679-6687.

[17] TIAN Z H, ZHANG R, HOU X Y, et al. FederBoost: private federated learning for GBDT[J]. IEEE transactions on dependable and secure computing, 2024, 21(3): 1274-1285.

[18] ZHAO G S, WANG Y, WANG J. Intrusion detection model of Internet of Things based on LightGBM[J]. IEICE transactions on communications, 2023, 106(8): 622-634.

[19] 张淑芬, 董燕灵, 徐精诚, 等. 基于目标扰动的 AdaBoost 算法[J]. 通信学报, 2023, 44(2): 198-209.
ZHANG S F, DONG Y L, XU J C, et al. AdaBoost algorithm based on target perturbation[J]. Journal on Communications, 2023, 44(2): 198-209.