

改进的 XGBoost 杂散电流预测及可解释模型

孟昭亮^{1,2,3}, 张泽涛^{1*}, 杨媛², 李国锋³, 陶崇勃³, 牛一疆³

¹西安工程大学电子信息学院, 陕西 西安 710600;

²西安理工大学国际工学院, 陕西 西安 710048;

³中车永济电机有限公司电力电子事业部, 陕西 西安 710000

摘要 为了解决影响地铁轨道的杂散电流特征众多, 常规特征选择方法影响模型预测精度及模型结果可解释性差的问题, 提出基于最优特征改进极端梯度提升(XGBoost)的杂散电流预测模型。利用遗传算法的灵活性和较强的搜索能力, 在包含原始 V 个特征的集合中逐代寻找使目标函数均方误差(MSE)最小的前 M 个特征, 建立最优特征选取方法下的杂散电流预测模型(OFS-XGBoost)。同时为了解决 OFS-XGBoost 模型预测结果较好, 但是黑箱模型对预测结果解释性不足的问题, 提出基于 SHAP 理论的归因分析框架, 根据杂散电流特征样本的边际贡献, 以易于理解的方式显示特征集合对模型预测结果的影响, 提高模型可解释性。结果表明: 所提模型的预测误差仅为 1.684%, 低于相同优化策略下的随机森林、反向传播(BP)神经网络等预测模型; 基于 SHAP 值的归因分析方法可以从全局与个体角度解释输入特征对杂散电流预测结果的影响, 在提高模型可解释性的基础上辅助地铁智能化健康管理。

关键词 极端梯度提升; 特征选择; 遗传算法; 杂散电流; SHAP 分析

中图分类号 TG172.84

文献标志码 A

DOI: 10.3788/LOP202259.1215011

Improved XGBoost Stray Current Prediction and Interpretable Model

Meng Zhaoliang^{1,2,3}, Zhang Zetao^{1*}, Yang Yuan², Li Guofeng³, Tao Chongbo³, Niu Yijiang³

¹College of Electronic Information, Xi'an Polytechnic University, Xi'an 710600, Shaanxi, China;

²International Engineering College, Xi'an University of Technology, Xi'an 710048, Shaanxi, China;

³Power Electronics Division of CRRC Yongji Motor Co., Ltd., Xi'an 710000, Shaanxi, China

Abstract To address the issue that there are many characteristics influencing the stray current of a subway track, the conventional feature selection method affects the prediction accuracy of the model, and the interpretability of the model results is poor, this paper proposes a stray current prediction model based on optimal feature improved eXtreme Gradient Boosting (XGBoost). Using the flexibility and the strong searchability of the genetic algorithm, we found the first M features that minimizing the mean square error (MSE) of the objective function generation by generation in the set containing the original V features. Simultaneously, the stray current prediction model under the optimal feature selection method (OFS-XGBoost) is established. To address the issue that the prediction results of the OFS-XGBoost are good, however, the machine learning black-box model has an insufficient explanatory ability for the prediction results, an attribution analysis framework based on SHAP theory is proposed to show the influence of feature set on the prediction results of the model in an understandable way based on the marginal contribution of stray current feature samples to improve the inference accuracy. The results show that the prediction error of the proposed model is only 1.684%, which is lower than the prediction models such as random forest and back propagation (BP) neural network under the same optimization strategy. The attribution analysis method based on

收稿日期: 2021-07-16; 修回日期: 2021-08-10; 录用日期: 2021-08-17

基金项目: 西安市科技计划(2020KJRC0029)

通信作者: *zz667892021@163.com

SHAP value explains the impact of input characteristics on stray current prediction results from a global and individual perspective, helping intelligent subway health management based on improving model interpretability.

Key words extreme gradient boosting; feature selection; genetic algorithm; stray current; SHAP analysis

1 引言

地铁机车的供电系统通常采用直流牵引电,由于在实际状况中,钢轨对地面的绝缘并非理想状况无穷大,因此存在部分电流经钢轨泄漏至大地,形成杂散电流^[1]。杂散电流对城市建设有着较大影响,长时间泄漏的杂散电流会危害地铁轨道、道床、埋地钢筋以及腐蚀附近金属通信设备,对地铁安全运行埋下隐患^[2]。Zaboli等^[3]在MATLAB软件中对一列装有牵引变电站(TPS)、走行轨和第三轨的电动列车进行了仿真。在模拟模型中,通过二维有限元分析,比较了不同土壤类型、有无集电垫时的杂散电流泄漏情况。Ma等^[4]在分析自适应神经模糊推理系统原理和结构的基础上,采用减法聚类算法得到模糊推理规则数,确定网络结构,预测杂散电流。王承涛^[5]使用基于卷积神经网络和反向传播(BP)神经网络的地铁杂散电流泄漏等级预测方法。刘江等^[6]运用极端梯度提升(XGBoost)算法预测地铁列车故障类型,挖掘特征量与目标变量之间的关联,结果表明迭代收敛速度和预测准确率都优于随机森林同类树模型。Li等^[7]使用极端学习机算法预测杂散电流,与多元回归算法相比,误差较小。然而,还需要对神经网络模型中的激活函数和隐藏层节点的数量进行仔细的选择和测试。上述研究对杂散电流的预测多采用神经网络及相关优化算法,预测效果优于传统回归模型,但仍存在收敛速度较慢、最终预测模型对样本依赖过大的问题,同时激活函数和隐藏层数量等关键参数的选择会导致网络结构的改变,影响模型性能。由于地铁杂散电流影响因素复杂,现有的预测方法对特征输入缺乏量化与筛选,会降低模型泛化性。

为改善上述问题,增强算法泛化性和增加收敛速度,在XGBoost预测算法的基础上进行改进,考虑到特征维数高,在XGBoost中加入正则项控制模型复杂度,防止过拟合,并且通过调整树的深度增强模型的收敛能力。同时利用遗传算法全局隐含并行性和易跳出局部极值的优点,筛选最优特征组合,将优化后的特征组合输入预测模型,得到接近于真实值的杂散电流预测值。最后利用shaply value(SHAP)可解释性分析框架,将机器学习黑箱

模型转换为容易理解的分析图表,帮助地铁运维人员精确管理杂散电流影响源。

2 XGBoost模型

eXtreme Gradient Boosting(XGBoost),即极端梯度提升,是一种提升学习算法。XGBoost算法在预测过程中的核心步骤是通过不断形成新的决策树来拟合上一步预测的残差,逐步缩减真实值与预测值之间的残差,从而有效地提高预测精度。在求解最优预测值的过程中,XGBoost突破传统GBDT方法只利用一阶导数的局限,对损失函数进行二阶泰勒展开的同时添加了正则项,因此在减小过拟合风险的同时提高了预测速度。

给定地铁轨道杂散电流数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_f, y_f)\}$, 其中 $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(V)})$ 为输入样本实例, V 为特征个数, $i = 1, 2, \dots, f, f$ 为样本个数, 预测模型采用第 K 次迭代的预测值作为输出结果。对第 i 个轨道杂散电流样本 X_i 的预测值为 \hat{y}_i , 表达式为

$$\hat{y}_i = \Phi(x_i) = \sum_{k=1}^K f_k(x_i). \quad (1)$$

得到杂散电流预测模型训练过程中损失函数的表达式为

$$O_{bj} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad (2)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\theta\|^2, \quad (3)$$

式中: $\Phi(\cdot)$ 为每个样本弱学习器的累加; O_{bj} 为目标函数, 目的是使预测值尽量接近真实值而且有一定的泛化能力; $\sum_i l(y_i, \hat{y}_i)$ 代表损失函数; $\sum_k \Omega(f_k)$ 代表正则化项; y_i 是轨道杂散电流的真实值, \hat{y}_i 为杂散电流预测输出; T 为叶子节点数; θ 为树中叶子节点预测值构成的向量; γ 为叶子惩罚正则项, 在预测过程中起到剪枝作用; λ 为叶子权重惩罚正则项。

在模型训练过程中, 采用梯度提升策略, 即保留经过训练的模型, 每一次迭代后添加一个新的回归树到模型中。假设第 i 个杂散电流样本在第 k 轮迭代的预测结果为 $\hat{y}_i^{(k)}$, $f_i(x_i)$ 为加入的新的回归树, 其推导过程为

$$\hat{y}_i^{(0)} = 0, \quad (4)$$

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i), \quad (5)$$

$$\hat{y}_i^{(2)} = f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i), \quad (6)$$

⋮

$$\hat{y}_i^{(k)} = \sum_{k=1}^K f_k(x_i) = \hat{y}_i^{(k-1)} + f_k(x_i). \quad (7)$$

将式(7)结果代入式(2),得到

$$O_{bj}^{(k)} = \sum_{k=1}^K l[y_i, \hat{y}_i^{(k-1)} + f_k(x_i)] + \Omega(f_k) + C_{\text{constant}}, \quad (8)$$

对目标函数进行二阶泰勒展开,且引入正则项,得到

$$O_{bj}^{(k)} = \sum_{j=1}^T [G_j \theta_j + \frac{1}{2} (H_j + \lambda) \theta_j^2] + \gamma T, \quad (9)$$

式中: $g_i = \partial_{\hat{y}_i^{(k-1)}} l(y_i, \hat{y}_i^{(k-1)})$, $h_i = \partial_{\hat{y}_i^{(k-1)}}^2 l(y_i, \hat{y}_i^{(k-1)})$, $G_i = \sum_{i \in I_j} g_i$, $H_i = \sum_{i \in I_j} h_i$ 。叶子节点权重值 θ_j 是一个不

确定的值,所以目标函数 $O_{bj}^{(k)}$ 对 θ_j 求一阶导数,可以解出叶子节点 j 的最优值 θ_j^* ,将 θ_j^* 代入目标函数, $O_{bj}^{(k)}$ 取得最小值:

$$O_{bj}^{(k)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j}{H_j + \lambda} + \gamma T. \quad (10)$$

3 最优特征构建与特征分析

3.1 最优特征构建

最优特征构建就是筛选原始特征,去除对预测结果没有帮助甚至影响结果的特征,达到降低学习任务的复杂度,提高预测模型准确性的目的^[8]。在采集杂散电流输入特征时,因影响因素众多,前期采集尽可能多的特征数据,但是如果对特征集合不加筛选,直接输入到预测模型,可能会导致预测模型体积大、训练时间过长的问题^[9]。因此,需要对杂散电流原始输入特征进行选择,将相关度最高的特征集合用于预测系统,提高模型精度。根据给定的杂散电流原始数据集,从数量为 D 的原始特征中挑选出最有效的 d ($d < D$) 个特征^[10],得到保留原有特征集合的绝大部分信息的最优特征集合 O 。

3.2 基于遗传算法的最优特征选取

遗传算法在选择特征时具有易跳出局部极值和全局隐含并行性的特点,因此适用于特征选择^[11]。遗传算法把原始特征集合编码为染色体,由 0 和 1 字符组成,0 代表不选择此特征,1 代表选中该特征^[12]。目标是求得仅有 d 个 1 的染色体,以原始特征集合作为输入的预测值为 Y_o ,真实负荷为 Y ,则最优特征组合寻优过程可以如下表示。

1) 初始化群体规模 D ,代表生成特征子集的个数,则把所有特征描述为 D 个 0 和 1 字符组成的二进制字符串,字符串代表遗传算法中的染色体,“0”

和“1”的个数和位置都是随机的。

2) 目标选出一条有且仅有 d 个 1 的染色体,适应度函数采用均方误差(MSE),通过适应度函数获得迭代之后第 k 代群体 $\text{pop}(k)$ 的个体适应度。适应度函数为

$$E_{\text{MS}} = \frac{1}{f} \sum_{i=1}^f (y_i - \hat{y}_i)^2, \quad (11)$$

式中: y_i 和 \hat{y}_i 为杂散电流的真实值与预测值。

3) 使用轮盘选择法作为算子,得到杂散电流特征群体 $\text{newpop}(k)$ 。

4) 基于特征群体 $\text{newpop}(k)$,通过多点交叉算子,设置交叉率为 0.6,产生新的群体 $\text{crosspop}(k)$,此处采用精英保留策略,最优个体不参加交叉操作。

5) 根据群体进化情况,判断是否停止搜索。当 $k < K$ 时,则 $k = k + 1$,那么新的群体 $\text{pop}(k) = \text{crosspop}(k - 1)$,返回到步骤 4),否则停止进化,进入步骤 6)。

6) 验证测试特征子集的预测性能,验证所选特征子集的质量。

3.3 基于 SHAP 归因分析的杂散电流预测模型可解释性分析框架

相对于其他线性模型,XGBoost 模型有着较好的准确率与预测速度,但是模型的可解释性较差,所以 XGBoost 通常被认为是黑箱模型,无法了解样本的特征值是如何影响最终的预测结果的。但在具体任务中,地铁运营人员还是想要了解样本特征与杂散电流预测结果的关系,所以从特征与结果的关联度进行分析,便于提高模型精度与分析特征重要度^[13]。

SHAP 是在合作博弈论的启发下构建的一个机器学习结果解释模型,所有的特征都视为“贡献者”。在 SHAP 分析框架中,每个预测样本对应一个预测值,SHAP 值为某条预测样本中各个特征所获得的数值^[14]。假设第 i 个样本为 x_i ,第 i 个样本的第 v 个特征为 x_{i-v} ,模型对该样本的预测值为 y_i ,样本目标变量的平均值为 y_{base} ,那么得到 SHAP 值的公式为

$$y_i = y_{\text{base}} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{iv}). \quad (12)$$

从式(12)可以看到, $f(x_{i1})$ 为第 i 个样本中第 1 个特征对最终预测值 y_i 的贡献值,当 $f(x_{i1}) > 0$,表明此特征促进预测值增大,是正向影响作用;反之,表明在该特征影响下会减小最终预测值,是负向作用。

4 实例分析

4.1 模型数据采集与评价体系

为了验证所提方法的有效性,采用地铁轨道杂

散电流分析常用的四层动态杂散电流结构^[15]作为算例,仿真平台为 MATLAB 与 Python。轨道设置 20 个监测点位,可以在整条线路的不同位置采集相应的杂散电流数据。采集的原始参数有轨道电位 U_{rail} 、地铁排流网电位 U_{drain} 、地铁轨道-排流网电位 U_{rd} 、地铁轨道排流网-埋地金属电位 $U_{\text{rail-metal}}$ 、地铁埋地金属电位 U_{metal} 、地铁轨道电流 I_{rail} 、地铁排流网电流 I_{drain} 、地铁排流网-埋地金属电流 $I_{\text{drain-metal}}$ 、埋地金属电流 I_{metal} 、列车距变电所距离 L 、轨地过渡电阻 R_{trans} 、纵向电阻 R_{long} 、埋地金属电阻 R_{metal} 、排流网电阻 R_{trans} 。采集数据集共 500 组,选取其中 350 组作为 OFS-XGBoost 预测模型训练集,余下 150 组数据作为模型测试集,以检测模型正确性。

基于本文的杂散电流预测方向,评价指标采用 MSE、均方根误差(RMSE)、平均绝对误差(MAE)、决定系数(R^2)。其中 MSE、RMSE、MAE 三个评价指标越小,代表误差越小; R^2 也称拟合优度,越接近于 1,代表目标变量对输入特征的解释程度越高,拟合效果越好。评价指标的具体公式:

$$E_{\text{MS}} = \frac{1}{f} \sum_{i=1}^f (y_i - \hat{y}_i)^2, \quad (13)$$

$$E_{\text{RMS}} = \sqrt{\frac{1}{f} \sum_{i=1}^f (y_i - \hat{y}_i)^2}, \quad (14)$$

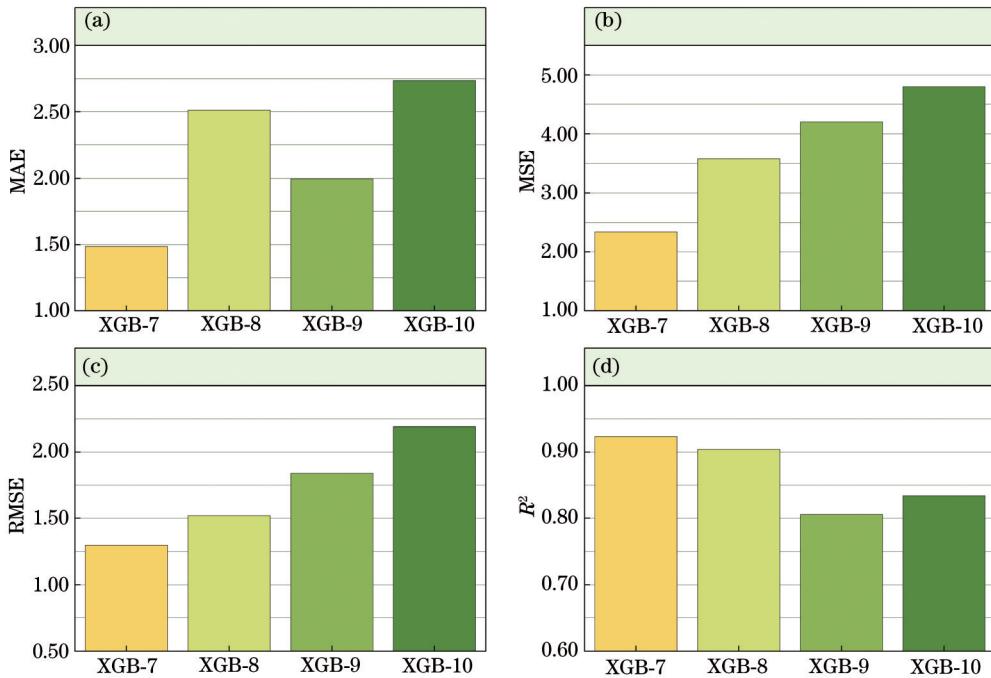


图 2 特征组合误差评价指标直方图。(a)MAE;(b)MSE;(c)RMSE;(d) R^2

Fig. 2 Feature combination error contrast histograms. (a) MAE; (b) MSE; (c) RMSE; (d) R^2

$$E_{\text{MA}} = \frac{1}{f} \sum_{i=1}^f |y_i - \hat{y}_i|, \quad (15)$$

$$R^2 = \frac{\sum_{i=1}^f (y_i - \hat{y}_i)^2}{\sum_{i=1}^f (y_i - \bar{y}_i)^2}. \quad (16)$$

4.2 最优特征选择

预测模型在每一代包含不同特征个数 M 时选出的最优特征集合预测误差如图 1 所示。可以看

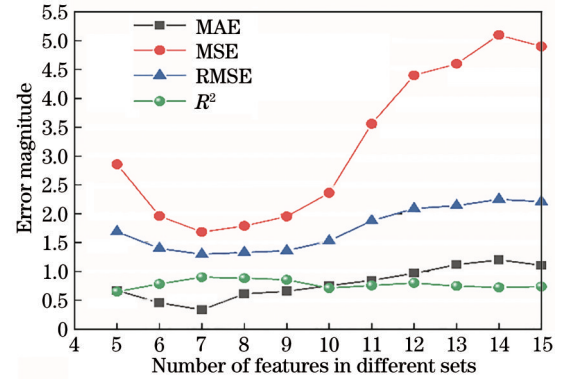


图 1 不同特征数组组合下误差对比

Fig. 1 Error comparison under different characteristic number combinations

出:不同的特征组合会导致预测模型不同的偏差,当特征组合在 7 至 10 的区间范围内时,四项误差评价指标均优于其他特征组合。

以直方图的形式放大显示特征数量在 7 至 10 区间内的误差评价指标。从图 2 可以直观看出,

当特征个数为 7 时各项评价指标优于其他特征集合, 当将这个特征组合输入至 XGBoost 预测模型时, 误差较小, 可以取得良好的预测效果。

图 3 为特征组数为 7 时的迭代过程, 初始迭代过程的 MSE 为 8.471, 迭代结束后 MSE 为 1.684。迭代过程中, U_{rail} 、 U_{drain} 、 I_{rail} 、 R_{trans} 四个特征量对最终预测结果作用较大, 从第 99 次迭代开始保持不变; 当迭代到 300 次附近时最优特征集合为 U_{rail} 、 U_{drain} 、 I_{rail} 、 R_{trans} 、 L 、 R_{long} 、 I_{metal} , 此时 MSE 下降至 2.464, 误差较初始下降 70.91%, 特征组合已能达到较好的预测效果; 当迭代次数在 400 次附近时, MSE 降至最小且保持不变, 此时得到最优特征组合 $M=7$, 即 U_{rail} 、 U_{drain} 、 I_{rail} 、 U_{rd} 、 I_{drain} 、 L 、 R_{trans} 。

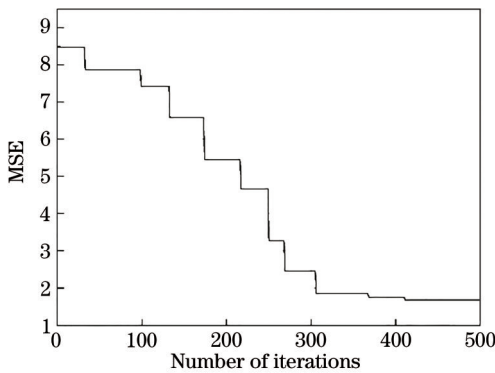


图 3 最优特征迭代过程

Fig. 3 Optimal characteristic iterative procession

4.3 预测模型性能对比

从表 1 可以看出: 与直接输入特征的 XGBoost 预测模型相比, 最优特征组合方法的 MSE 下降 80.1%, 且 MSE、MAE、 R^2 这三个误差评估指标都优于对比模型, 表明使用的最优特征选择方法与 XGBoost 模型融合较好, 取得良好的预测结果。

表 1 原模型与优化模型的误差对比

Table 1 Error comparison between original model and optimized model

Model	MSE	RMSE	MAE	R^2
XGBoost	8.471	2.910	3.491	0.787
OFS-XGBoost	1.684	1.297	0.334	0.923

将 OFS-XGBoost 地铁轨道杂散电流预测模型与常用预测模型进行对比, 结果如表 2 所示。相对于传统的支持回归机 (SVR) 预测模型, OFS-XGBoost 的 MSE 大幅度下降, MAE 降低了 95.2%, RMSE 降低了 82.0%, R^2 提高了 34.5%。SVR 预测模型收敛性较差, 根本原因在于 SVR 模

表 2 最优特征组合方法下各模型预测性能指标

Table 2 Prediction performance index of different models based on optimal feature combination method

Model	MSE	RMSE	MAE	R^2	Prediction time / s
SVR	51.908	7.201	6.888	0.686	4.3
BP	7.391	2.773	2.355	0.775	2.4
RF	2.945	1.715	2.383	0.846	1.8
OFS-XGBoost	1.684	1.297	0.334	0.923	1.1

型预测结果受核函数影响, 极易陷入局部最优, 同时预测用时最多, 为 4.3 s。与 BP 神经网络预测模型相比, OFS-XGBoost 的 MSE 下降了 77.2%, RMSE 下降了 53.2%, MAE 降低了 85.8%, R^2 提升了 19.1%。与同为集成树类型的随机森林 (RF) 算法相比, OFS-XGBoost 的 RMSE 相差不大, 有着更好的拟合程度。综合来看, OFS-XGBoost 模型的 R^2 值较其他模型最大, 说明模型拟合能力相对最好, 预测迭代用时相比三种算法最短, 为 1.1 s。

不同模型的预测曲线与真实曲线的对比如图 4 所示, 可以直观地看到各个预测模型对真实杂散电流曲线的拟合情况。图 4 显示 20 组杂散电流的预测曲线, 可以相对清晰地反映出各模型的杂散电流预测情况。

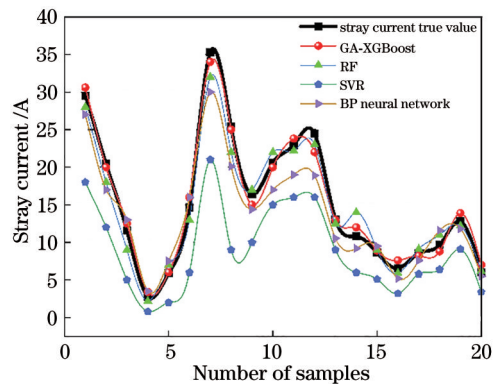


图 4 地铁轨道杂散电流预测各模型对比图

Fig. 4 Comparison of different models for stray current prediction

4.4 SHAP 预测结果解释性分析

在 SHAP 框架下, XGBoost 模型变量重要度排名及影响结果的正负效应如图 5 SHAP 摘要图所示。从图 5 可以看到, y 轴按照输入特征的重要度, 由大到小排名分别是轨地电压、轨道-排流网电压、轨道电流、列车与变电所距离、排流网电压、排流网电流和轨道过渡电阻, x 轴表示特征对应的 SHAP 值。在图中每个特征点代表相应数据集中的特

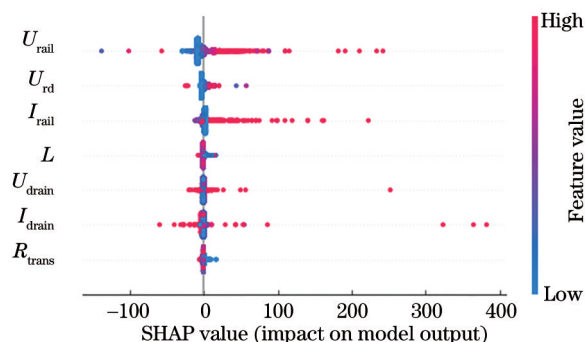


图 5 SHAP 特征分析概要图

Fig. 5 SHAP feature analysis sketch chart

征样本,颜色从Low到High的变化表示样本特征数值从小到大,正负的SHAP值表示特征与预测结果的正负相关性。就特征重要度第一的轨道电压 U_{rail} 而言,数值较大的样本 SHAP 值为正,表示对结果起着抬升的作用,会增加杂散电流最终的预测值。颜色靠近变量条下半部分的特征点主要集中于 SHAP 值为负的区域,意味着当轨道电压越小,杂散电流预测值也会越小。

在 SHAP 概要图中,首先看到特征值大小对预测的影响,但是想要了解特征与预测结果的确切形式时,必须查看 SHAP 依赖图。SHAP 依赖图可以更为详细地解释特征样本是如何影响最终预测结果的。图 6 轨道电压 SHAP 分析依赖图中的每一个点代表该特征的样本值,横坐标轴是特征值大小,纵坐标轴是相应特征的 SHAP 值。轨道电压的 SHAP 值基本呈现一个线性增加的趋势,表征在预测模型中,轨道电压越大,相应的杂散电流值也会越高。

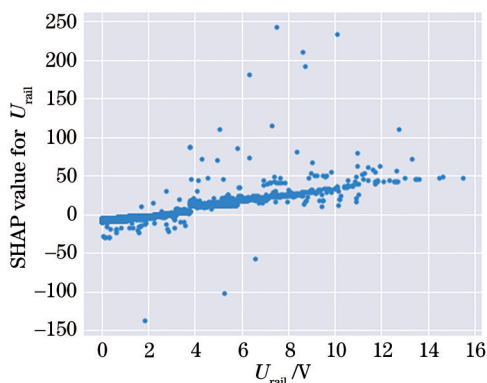


图 6 轨道电压 SHAP 分析依赖图

Fig. 6 Dependence diagram of rail voltage SHAP analysis

从图 7 过渡电阻 R_{trans} 的 SHAP 依赖图来看,当地铁过渡电阻为 5000, 10000, 15000 $m\Omega \cdot km$ 左右

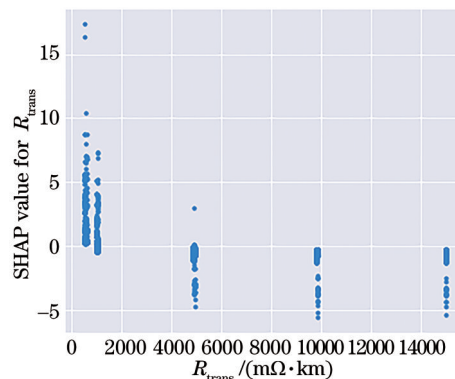


图 7 过渡电阻 SHAP 分析依赖图

Fig. 7 Dependence diagram of transition resistance SHAP analysis

时,整体的 SHAP 值皆为负,此时特征 R_{trans} 的作用为减小杂散电流的预测值。而当地铁轨道过渡电阻降低为 500, 1000 $m\Omega \cdot km$ 左右时,此时 SHAP 值转变为正,过渡电阻越小,SHAP 值越大,此时相应的杂散电流也会越高。

5 结 论

地铁杂散电流影响因素较多,不同特征组合的输入对预测模型的精度与模型泛化能力影响较大,通过对输入特征的量化选择,在不损失数据信息量的前提下获得最优的特征组合,会提高预测模型的精度与泛化性。与随机森林、BP 神经网络等方法相比, XGBoost 模型在训练之前对输入特征数据进行排序,存储为 block 结构,在之后的预测过程中这个结构被反复使用,很大程度上减少了计算量,可以在预测中实现并行计算,因此具有更快的预测速度。并且在处理有缺失值的特征时, XGBoost 内部的稀疏感知算法能够自动分析分裂方向,提高模型可靠性。

使用 SHAP 分析框架对 XGBoost 杂散电流预测模型进行分析与解释,分别从全局和个体角度揭示特征在预测过程中的重要度和特征样本数值大小与杂散电流预测值的正负相关性。通过 SHAP 值解释了预测模型对杂散电流的预测原因,增加了预测结果的可信度,同时为地铁运营人员制定预防或控制杂散电流策略提供分析依据。

参 考 文 献

[1] 陈超, 缪希仁. 轨交杂散电流监测与腐蚀预测[J]. 电器与能效管理技术, 2017(20): 67-74.
Chen C, Miao X R. Subway stray current monitoring

- and corrosion prediction[J]. *Electrical & Energy Management Technology*, 2017(20): 67-74.
- [2] 林晓鸿. 地铁多区间杂散电流计算方法及特性分析[D]. 成都: 西南交通大学, 2019.
Lin X H. Study on calculation method and characteristic analysis of metro stray current in multiple power supply sections[D]. Chengdu: Southwest Jiaotong University, 2019.
- [3] Zaboli A, Vahidi B, Yousefi S, et al. Evaluation and control of stray current in DC-electrified railway systems[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(2): 974-980.
- [4] Ma C Y, Zhang D L, Wang Z, et al. Study on ANFIS application in coal mining stray current security prediction[J]. *Key Engineering Materials*, 2010, 426/427: 216-219.
- [5] 王承涛. 一种基于卷积神经网络及BP神经网络的地铁杂散电流泄漏等级预测方法: CN107991873A[P]. 2018-05-04.
Wang C T. A prediction method of metro stray current leakage level based on convolution neural network and BP neural network: CN107991873A[P]. 2018-05-04.
- [6] 刘江, 许康智, 蔡伯根, 等. 基于XGBoost的列车车载设备故障预测方法[J]. *北京交通大学学报*, 2021, 45(4): 95-106.
Liu J, Xu K Z, Cai B G, et al. XGBoost-based fault prediction method for on-board train control equipment[J]. *Journal of Beijing Jiaotong University*, 2021, 45(4): 95-106.
- [7] Li J S, Liu Z B, Yi H, et al. Stray current prediction model for buried gas pipelines based on multiple regression models and extreme learning machine[J]. *International Journal of Electrochemical Science*, 2021, 16: 210253.
- [8] 张爱武, 董喆, 康孝岩. 基于XGBoost的机载激光雷达与高光谱影像结合的特征选择算法[J]. *中国激光*, 2019, 46(4): 0404003.
Zhang A W, Dong Z, Kang X Y. Feature selection algorithms of airborne LiDAR combined with hyperspectral images based on XGBoost[J]. *Chinese Journal of Lasers*, 2019, 46(4): 0404003.
- [9] 张俐, 陈小波. 动态加权条件互信息的特征选择算法[J]. *电子与信息学报*, 2021, 43(10): 3028-3034.
Zhang L, Chen X B. Feature selection algorithm for dynamically weighted conditional mutual information[J]. *Journal of Electronics & Information Technology*, 2021, 43(10): 3028-3034.
- [10] 李雷, 苏明昕, 苏波, 等. 基于适应性函数的自适应性遗传算法及其实验验证[J]. *自动化技术与应用*, 2020, 39(8): 11-13, 19.
Li L, Su M X, Su B, et al. Adaptive genetic algorithm and its experimental verification based on adaptive function[J]. *Techniques of Automation and Applications*, 2020, 39(8): 11-13, 19.
- [11] 庞伟帆, 傅戈雁, 王明雨, 等. 基于响应面法和遗传神经网络模型的高沉积率激光熔覆参数优化[J]. *中国激光*, 2021, 48(6): 0602112.
Pang Y F, Fu G Y, Wang M Y, et al. Parameter optimization of high deposition rate laser cladding based on the response surface method and genetic neural network model[J]. *Chinese Journal of Lasers*, 2021, 48(6): 0602112.
- [12] Wang Y, Sherry Ni X. A XGBoost risk model via feature selection and Bayesian hyper-parameter optimization[J]. *International Journal of Database Management Systems*, 2019, 11(1): 1-17.
- [13] 张秋霞, 何留杰, 张来顺. 基于联盟博弈的云任务调度及Shapley值法的收益分配模型[J]. *计算机应用与软件*, 2020, 37(5): 275-280, 320.
Zhang Q X, He L J, Zhang L S. Cloud tasks scheduling based on coalition game and profit allocation model with Shapley value method[J]. *Computer Applications and Software*, 2020, 37(5): 275-280, 320.
- [14] Someh N G, Pishvae M S, Sadjadi S J, et al. A decision-making model for performance evaluation and profit sharing in a diagnostic laboratory network[J]. *Journal of Evaluation in Clinical Practice*, 2020, 26(5): 1498-1503.
- [15] 李威. 地铁杂散电流腐蚀监测及防护技术[M]. 徐州: 中国矿业大学出版社, 2004: 6-8.
Li W. Metro stray current corrosion monitoring and protection technology[M]. Xuzhou: China University of Mining & Technology Press, 2004: 6-8.