

DOI:10.3969/j.issn.1001-4551.2022.04.019

基于极端梯度提升回归模型的电梯 钢丝绳磨损预测方法^{*}

陈向俊^{1,2}, 傅军平^{1,2}, 陈栋栋^{1,2}, 李 科^{1,2}, 李黎苹^{1,2}, 吕林锋^{1,2}

(1. 浙江省特种设备科学研究院, 浙江 杭州 310020; 2. 浙江省特种设备检验技术研究重点实验室, 浙江 杭州 310053)

摘要: 目前对电梯钢丝绳磨损量预测的研究还存在不足, 针对这一问题, 采用基于极端梯度提升(XGBoost)算法的机器学习方法, 对电梯钢丝绳磨损率的预测进行了理论分析、数据采集和实验测试研究。首先, 在目标损失函数中增加了额外的正则化项, 并使用了贝叶斯超参数优化, 提出了优化后的算法 BO-XGBoost; 然后, 用自制的电梯钢丝绳疲劳试验机对曳引轮直径、载荷力、频率和包角这4个因素进行了试验, 得到了用于预测钢丝绳磨损率的数据; 最后, 用 BO-XGBoost 算法对钢丝绳磨损率进行了预测分析, 同时与多元线性回归(MLR)、随机森林(RF)以及支持向量机(SVM)等机器学习算法进行了比较。研究表明: BO-XGBoost 算法的建模效果和回归效果最好, 其泛化能力也最高, 能适应不同工况的实验, 在预测钢丝绳磨损率方面优于其他几个模型, 预测值与试验值达到了99.1%的准确率, 证明了该方法的有效性。

关键词: 电梯; 钢丝绳; 磨损预测; 极端梯度提升; 回归模型

中图分类号: TH236; TU857

文献标识码: A

文章编号: 1001-4551(2022)04-0554-07

Elevator wire rope wear prediction method based on extreme gradient lifting regression model

CHEN Xiang-jun^{1,2}, FU Jun-ping^{1,2}, CHEN Dong-dong^{1,2},
LI Ke^{1,2}, LI Li-ping^{1,2}, LV Lin-feng^{1,2}

(1. Zhejiang Academy of Special Equipment Science, Hangzhou 310020, China; 2. Zhejiang Provincial Key Laboratory of Special Equipment Inspection Technology Research, Hangzhou 310053, China)

Abstract: Aiming at the problem of the limitations and deficiencies in the current research on the prediction of elevator wire rope wear, based on the machine learning method of the extreme gradient boosting (XGBoost) algorithm, theoretical analysis, data collection and experimental testing were carried out on the prediction of wire rope wear rate. First, an additional regularization term was added to the objective loss function and Bayesian hyperparameter optimization was used, and the optimized algorithm BO-XGBoost (Bayesian optimization-eXtreme gradient boosting) was proposed; then a self-made elevator wire rope fatigue testing machine was used to determine the diameter, load force, frequency and wrap angle of the researched traction sheave, and the data used to predict the wear rate of wire rope were obtained; finally, the BO-XGBoost algorithm was used to predict and analyze the wear rate of wire rope, and at the same time, it was compared with the other machine learning algorithms, such as multiple linear regression (mixed logistic regression, MLR), random forest (RF), and support vector machine (SVM). The research results show that the BO-XGBoost algorithm has the best modeling effect and regression effect, and its generalization ability is also the highest. It can adapt to experiments under different working conditions and is better than other models in predicting wire rope wear rate. The predicted value and the experimental value have reached an accuracy of 99.1%, which proves the effectiveness of the method.

Key words: elevator; wire rope; wear prediction; extreme gradient lifting; regression model

收稿日期: 2021-09-27

基金项目: 国家自然科学基金资助项目 (U1709210, 51805382)

作者简介: 陈向俊(1977-), 男, 浙江东阳人, 高级工程师, 主要从事特种设备检验检测技术方面的研究。E-mail: 1085469675@qq.com

0 引言

截至2020年12月,我国国内电梯总量将达到了850万台,中国的电梯产销量和电梯数总量在世界上依旧保持第一。但随着电梯市场和需求的日益扩大,电梯的安全问题越来越严峻^[1]。

电梯钢丝绳是关系到电梯正常运行的重要部件之一,易发生腐蚀磨损和疲劳断裂。2019年7月14日,上海市嘉定区某电梯维护作业人员徐某在进行电梯钢丝绳更换作业时,在刚更换了一根钢丝绳后,其余所有的钢丝绳(6根)全部断裂,徐某被上方脱落的链条打中。后经调查发现,该电梯的钢丝绳已经严重磨损,处于一碰即断的临界点,从而引发了此次安全事故。

随着电梯曳引钢丝绳的广泛使用,钢丝绳的安全防护问题非常值得关注,除了要经常对其进行润滑之外,对曳引轮槽表面磨损情况、轮槽尺寸都需要进行定期检查,以保证电梯的安全运行^[2]。因此,钢丝绳磨损量的预测对于提高钢丝绳的使用寿命和电梯安全性能是非常重要的。

通过文献检索发现,国内现有针对电梯钢丝绳磨损的研究并不多见。陈海龙^[3]在对起重机钢丝绳磨损进行了研究,结果表明,直径的减少是起重机钢丝绳磨损报废的直接原因,并提出了一种通过测量钢丝绳直径的减少来值判断钢丝绳磨损量的方法;但他的研究中,实验的样本量太少,且其研究仅针对钢丝绳直径对磨损量的影响,不够全面。王平等人^[4]应用绳索振动理论结合BP(back propagation)神经网络的方法,对钢丝绳磨损与张力偏差之间的关系进行了研究,并设计了实验对该方法的实用性进行了验证;但是其研究也仅针对张力与磨损量的关系,同样不够全面。王晔^[5]研究了不同因素对钢丝绳磨损量的影响,但是他仅从理论方面进行了阐述,并未用实验对此进行验证。任洁^[6]分析了钢丝绳实际磨损的图片,并得到结论,即多数钢丝绳磨损为人为安装因素引起,不涉及到钢丝绳自身的因素。

因此,目前对电梯钢丝绳磨损量预测的研究存在明显的局限和不足。

为了解决这个问题,笔者以梯度提升回归(gradient boosting regression,GBR)算法为基础,围绕影响钢丝绳磨损的4个主要因素^[7]采集实验数据,利用经过贝叶斯优化的极端梯度提升模型(BO-XGBoost)进行训练和验证,预测钢丝绳在不同情况下的磨损量,并将其与实际磨损量进行比较。

1 理论模型

1.1 梯度提升回归(GBR)

梯度提升算法本质上是一种监督学习方法。对于许多复杂的数据集而言,它是一种非常可靠的方法^[8,9]。GBR就是一种弱学习器的回归算法,最后得到强学习器,它可以用于回归和分类的问题。在梯度提升过程中,先找到一个预测准确性比较弱的模型 $F_m(x)$,得到其预测值与实际 y 值之间的差异,即残差 $R(x)$ 。

GBR不会改变 $F_m(x)$,而是通过增加估计器 $h(x)$ 构建新的模型,逐步减小残差。当满足下式时,效果最好:

$$h(x) = y - F_m(x) \quad (1)$$

1.2 极端梯度提升(XGBoost)

极端梯度提升算法是以梯度提升算法为基础的一种优化算法^[10,11],沿用了梯度提升的框架,因此同样是一种监督算法。XGBoost在目标函数的损失函数中新加入一个正则化项,它能过渡最终学习到的权重比,防止过拟合现象。

接下来笔者介绍XGBoost的理论框架。

梯度提升树模型的估计输出 $\hat{y}_i$ 可以表示为所有树的预测分数 $f_k(x_i)$ 的总和:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \Gamma \quad (2)$$

式中: Γ —回归树空间; K —回归树的数量; x_i —样本 i 对应的特征。

目标函数是机器学习问题中最基本的表达方式^[12],其提升的过程会一直进行,直到目标函数变得最简且最优。笔者将模型中采用的函数集做近似化处理,并定义正则化的目标函数为:

$$\Phi = \sum_{i=1}^n l(y_i, \hat{y}_i) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (3)$$

式中: n —给定的数据样本; λ —正则化超参数; γ —引入额外叶子的复杂度成本; j —叶子节点; ω_j^2 —叶子节点 j 权重的L2范数; $\sum_{i=1}^n l(y_i, \hat{y}_i)$ —描述模型与训练数据拟合程度的训练损失函数; $\gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2$ —正则化惩罚项,目的是减小模型复杂度,降低过拟合的风险。

对于给定的数据集,每个叶子节点 j 都存在一个预测分数 $f_k(x_i)$,即叶子权重。叶子权重 ω_j 是这棵树中在叶节点 j 处所有样本的回归值。其中, $j \in \{1, 2, \dots$

T_i 是这棵树的叶子数。所有的树都是在加法学习过程中顺序构建的,每棵树都从它之前的树中学习,并更新预测值中的残差。因此 $\hat{y}_i^{(k-1)}$ 已经包含了所有树的迭代结果。

对于第 k 次迭代, $\hat{y}_i^{(k)}$ 可以表示为 $\hat{y}_i^{(k-1)} + f_k(x_i)$, 且目标函数 $\Phi_{(k)}$ 的表达式为:

$$\Phi_{(k)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(k-1)} + f_k(x_i)) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (4)$$

由于需要在损失函数中优化目标,笔者在此处将式(4)按二阶泰勒展开做近似处理,即:

$$\begin{aligned} \Phi_{(k)} \simeq & \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(k-1)}) + g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \\ & \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \end{aligned} \quad (5)$$

式中: g_i —损失函数的一阶梯度统计量; h_i —损失函数的二阶梯度统计量。

g_i 的具体表达式如下式所示:

$$g_i = \partial_{\hat{y}_i^{(k-1)}} l(y_i, \hat{y}_i^{(k-1)}) \quad (6)$$

h_i 的具体表达式如下式所示:

$$h_i = \partial_{\hat{y}_i^{(k-1)}}^2 l(y_i, \hat{y}_i^{(k-1)}) \quad (7)$$

在迭代过程步骤 k 中去除掉常数项,得到以下近似目标值:

$$\Phi_{(k)} \simeq \sum_{i=1}^n \left[g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (8)$$

用叶子的分数向量和样本映射到叶子 j 的索引映射函数来定义树,这个过程可以表示为:

$$\sum_{i=1}^n f_k(x) = \sum_{j=1}^T \omega_j \quad (9)$$

同时式(8)可以改写为:

$$\Phi_{(k)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (10)$$

给定一个固定的树结构,每个叶子节点上的最优叶子权重值 ω_j^* 和极值 $\Phi_{(k)}^*$ 通过简单的二次规划求解,具体如下式:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (11)$$

$$\Phi_{(k)}^* = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (12)$$

式(12)可以看作是一个评估函数,用于衡量给定的叶子评分向量的合适程度,其值越小越好。为了避免陷入无尽的树结构,在实际应用中,笔者采用贪心算法来寻找最优的树结构。

1.3 基于贝叶斯优化的 XGBoost 回归模型

超参数是指在机器学习过程中需要提前设置值的参数,超参数对模型的预测准确性有显著影响。在机器学习实践中调整超参数通常采用网格搜索(grid search, GS)和随机搜索(random search, RS)方法。由于 GS 容易受到维数约束的影响,维数越多则花费时间指数增长,这已被证明对于 XGBoost 是不可行的方法^[13,14]。RS 通过尝试超参数的随机组合来选择最佳超参数配置的方法,但 RS 的缺点在于不必要的高方差。

贝叶斯优化(Bayesian optimization, BO)通过目标以前的评估结果重新构建概率模型^[15,16],寻得最小化目标函数的值,被广泛应用于机器学习超参数调优。贝叶斯超参数优化是一种为序列模式模型提供的求全局优化的方法^[17],序列化模型全局优化(sequential model-based optimization, SMBO)是把贝叶斯优化的一个形式化的定义。

SMBO 伪代码流程如表 1 所示。

表 1 SMBO 伪代码

步骤数	操作
1	定义 $S_0; H = \phi$
2	for $t = 1$ to T do
3	$\lambda^* = \operatorname{argmin}_{S_{t-1}} (L)$
4	$c = L(\lambda^*)$
5	$H = H \cup (\lambda^*, c)$
6	根据更新的 S_t , 拟合新模型 H
7	End for
8	在 λ 中返回具有最小 c 的 H

λ —超参数; L —损失函数; H —记录了所有超参数的历史集合; S_t —从超参数 λ — L 的映射概率模型

整个循环一共需要进行 T 次。首先每次先根据已有的历史试验集合找出一个更可能拥有更小损失函数的超参数集合 λ , 之后计算实际的损失函数 L 的值, 加入历史集合后更新 S 。

由于贝叶斯优化是优化目标函数的首选,笔者提出的基于贝叶斯超参数优化的极端梯度提升算法(BO-XGBoost)对应的流程如图 1 所示。

疲劳试验机分为机械部分和电控部分:(1)机械部分。由伺服电机带动试验轮正反转运行,以此来拖动钢丝绳进行疲劳寿命试验。实验轮和导向轮为可拆卸结构,用来测试不同直径的钢丝绳,载荷力通过加载轮来调节其大小;(2)电控部分。主要为人机界面,在设定试验频率、试验次数以及包角等参数后,即开始进行疲劳试验。

整个试验在曳引比 1:1 的情况下进行。当 PLC 通讯正常时,系统会跳转到参数设置界面来设置相关实验参数。参数设置界面如图 3 所示。

2.2 数据采集

目前一般以钢丝绳直径的磨损率来考察钢丝绳的失效程度。磨损率涉及到的因素较多,钢丝绳的磨损率

主要受载荷力、曳引轮直径、运动频率、包角的影响较大。笔者将上述 4 个因素作为实验考察的因素,将数据实验测得的磨损率记录下来,用于创建机器学习模型。

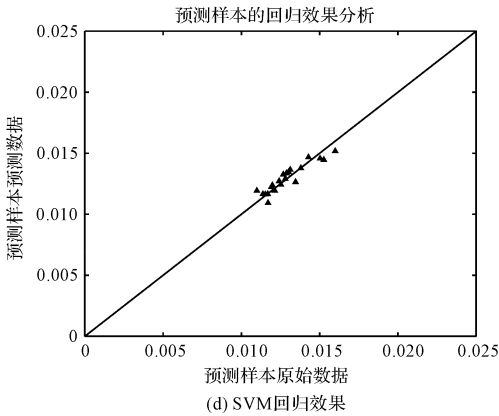
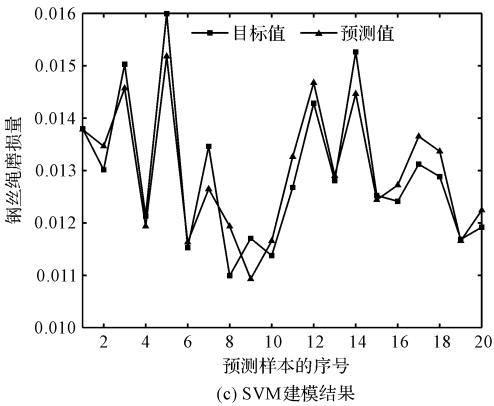
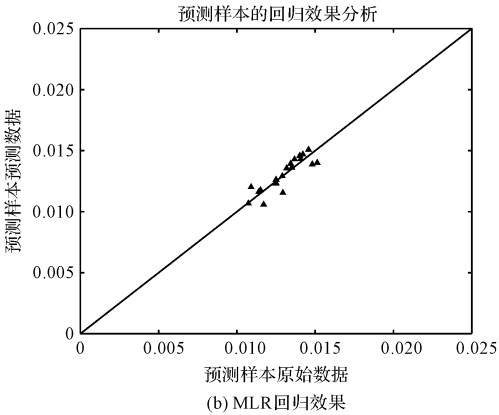
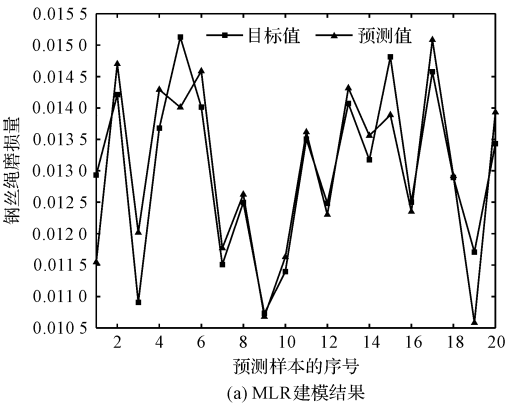
由于钢丝绳疲劳实验流程很长,笔者采用 130 组数据作为训练数据;且模型训练完毕后,再用 20 组新数据作为测试之用,故总共采集 150 组数据。测试模型的数据如表 2 所示。

3 实验及结果分析

笔者在 MATLAB 上编写 4 种模型的程序,并导入表 2 中的数据,分别运行后得到模型的建模结果分析图和回归效果分析图。各模型的预测结果如图 4 所示。

表 2 测试模型试验数据集

数据集	曳引轮直径/mm	频率/(r·min ⁻¹)	载荷力/N	角度/(°)	磨损率
训练集	430	85	5 800	120	0.016 651
	450	75	5 400	130	0.009 096
	490	60	6 000	120	0.017 429
	400	65	4 200	140	0.014 536
	460	80	4 000	160	0.012 359
	450	70	5 800	120	0.019 450
	...	...	...	...	...
测试集	430	80	4 000	150	0.019 777
	460	75	4 000	120	0.017 990
	490	85	4 400	140	0.010 469
	480	80	6 000	160	0.009 314
	450	75	4 600	140	0.011 860
	...	...	...	...	...



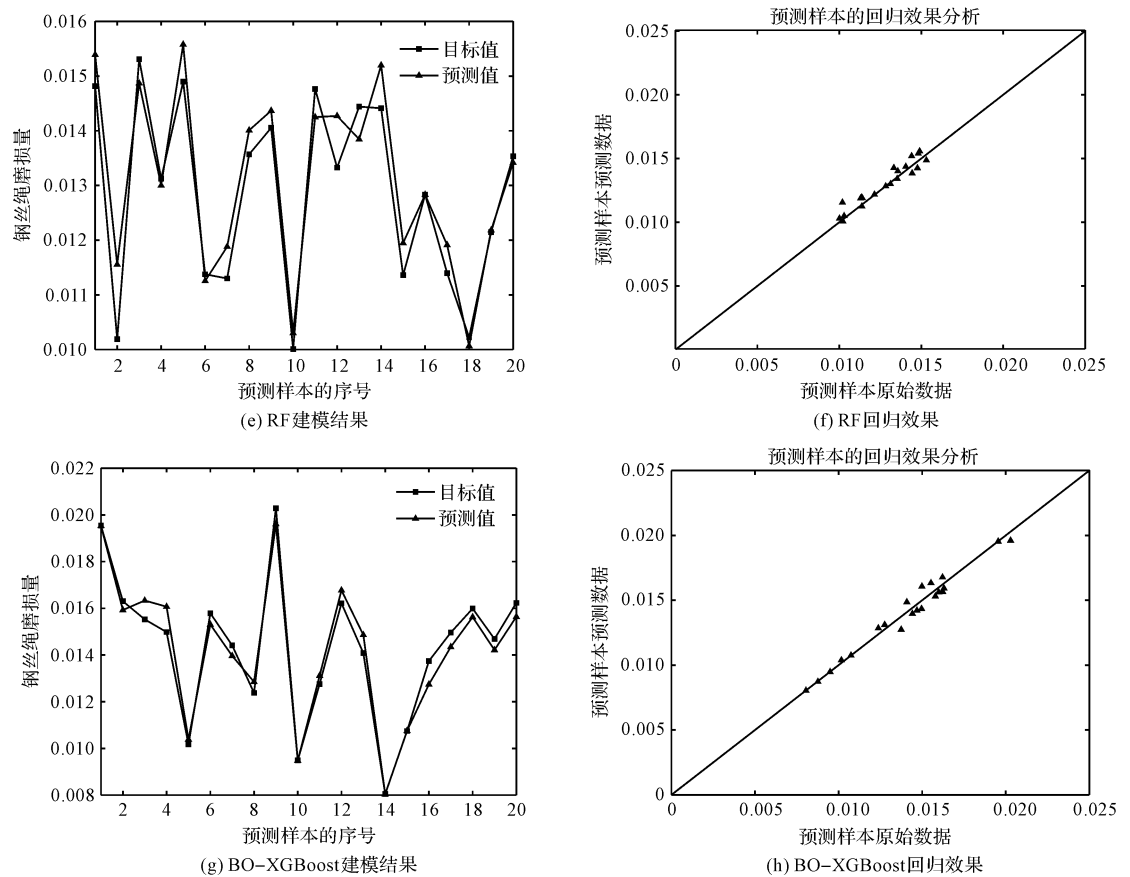


图4 不同模型预测结果曲线

由图4(a~h)可以发现,BO-XGBoost模型的预测值与目标值更接近。因此,不难看出BO-XGBoost模型的模拟效果最好,RF和SVM模型其次,MLR模型最差。

在算法原理上,RF对异常值不敏感,且对训练集一视同仁,且每棵树分裂特征随机;采用MLR算法处理复杂问题时,存在较大的局限性;SVM虽然有一定的泛化能力,但核函数的选取非常困难,同时对缺失数据也很敏感;BO-XGBoost对代价函数进行了二阶泰勒展开,同时考虑一阶与二阶导数,支持自定义代价函数(二阶可导),在超参数问题上加以贝叶斯优化后,其优势很大。

为了评估预测模型的可靠度,笔者计算每个模型结果的平均绝对误差(MAE)、均方根误差(RMSE)和决定系数(R^2)。MAE是评估回归模型预测误差的最常用方法之一,它是方程中给出的预测误差的平均绝对值,如下式所示:

$$MAE = \frac{1}{N} \sum_{i=1}^N |PM - AM| \tag{13}$$

式中:PM—预测的磨损量;AM—实际的磨损量;
对于预测误差,笔者利用RMSE来验证预测的有效性,即:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (PM - AM)^2} \tag{14}$$

评估参数 R^2 被广泛称为确定系数,用于确定预测值与实际值的接近程度,即:

$$R^2 = \left(\frac{N(\sum PM \times AM) - (\sum PM)(\sum AM)}{\sqrt{[N\sum PM^2 - (\sum PM)^2][N\sum AM^2 - (\sum AM)^2]}} \right) \tag{15}$$

如前所述,MAE和RMSE的值越低表示该模型的效率越高, R^2 值在0~1之间,且越接近1表示模型的预测精度越好。

4种模型的MAE、RMSE、 R^2 值如表3所示。

表3 钢丝绳磨损结果预测			
算法模型	平均绝对误差(MAE)	均方根误差(RMSE)	决定系数(R^2)
MLR	0.006 91	0.006 7	0.827
SVM	0.006 35	0.005 9	0.844
RF	0.004 74	0.005 2	0.885
BO-XGBoost	0.002 86	0.003 6	0.991

通过表3,并结合前文的图4可以看出:在此次试验中,BO-XGBoost模型比其他3个模型效果都要好;具体表现为BO-XGBoost的MAE与RMSE均要小于

MLR、SVM 和 RF,且 BO-XGBoost 的 R^2 值较高,达到了 0.991,与图 4 反映的结果类似,即 MLR 显示的预测性能最低,RF 和 SVM 其次,BO-XGBoost 的预测效果最优。

4 结束语

针对目前对电梯钢丝绳磨损量预测的不足,笔者采用基于极端梯度提升(XGBoost)算法的机器学习方法,对电梯钢丝绳磨损率的预测进行了研究。笔者基于自制的钢丝绳疲劳试验机,对钢丝绳磨损的 4 个主要因素的数据进行采集,并用基于贝叶斯超参数优化的极端梯度提升回归算法 BO-XGBoost 来预测钢丝绳疲劳磨损率;同时将 BO-XGBoost 与多元线性回归(MLR)、随机森林(RF)以及支持向量机(SVM)算法进行了比较。

研究结果表明:

(1)BO-XGBoost 模型对多因素影响下钢丝绳磨损率的预测效果最佳,模型的泛化能力最好,其预测可靠度优于 MLR、RF、SVM 等模型;

(2)对于电梯钢丝绳磨损,采用基于极端梯度提升回归模型算法,其预测精度达到了 99% 以上,预测值基本等于实际磨损值;将其用于钢丝绳磨损程度的参考,在不同工况下选择合适的曳引轮直径、载荷力、频率和包角,有利于降低钢丝绳的磨损率,提高特种设备的安全。

在后续的工作中,笔者将研究钢丝绳的磨损量与其使用寿命之间的规律,以进一步提高钢丝绳使用的安全性。

参考文献(References):

- [1] 王英儒. 曳引电梯钢丝绳失效分析[J]. 中国电梯,2019,10(12):61-63.
- [2] 王彦辉. 浅谈曳引钢丝绳原因分析[J]. 建筑工程技术与设计,2020,5(1):4109.
- [3] 陈海龙. 起重机钢丝绳磨损报废测量探析[J]. 中国设备工程,2021,12(5):179-181.
- [4] 王 平,张岳明,陆向军,等. 基于电梯钢丝绳磨损与张力

偏差的研究[J]. 工业仪表与自动化装置,2019,11(2):61-62.

- [5] 王 晔. 基于高速电梯轿厢气压变化调节技术[J]. 特种设备安全技术,2018,10(5):85-87.
- [6] 任 洁. 电梯曳引钢丝绳磨损因素分析[J]. 机电工程,2019,8(3):40-42.
- [7] PONRAJO A S, VIGNESWARAN T. Daily evapotranspiration prediction using gradient boost regression model for irrigation planning[J]. *The Journal of Supercomputing*, 2020,76(8):157-161.
- [8] 尤光辉,林 正,陈向俊,等. 基于正交设计的电梯曳引钢丝绳磨损研究[J]. 起重运输机械,2016,6(8):104-108.
- [9] 陈 达. 基于极端梯度提升模型的钢材料性能预测研究[D]. 秦皇岛:燕山大学计算机学院,2020.
- [10] 孟 杭,黄细霞,涂修建. 基于时间序列和 Xgboost 的钢卷仓储吞吐量预测[J]. 计算机应用,2019(12):24-28.
- [11] ZHANG Wen-gang, WU Chong-zhi, ZHONG Hai-yi. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization[J]. *Geoscience Frontiers*, 2021,48(12):264-267.
- [12] LYU Fei, FAN Xin-yu, DING Fa-xing, et al. Prediction of the axial compressive strength of circular concrete-filled steel tube columns using sine cosine algorithm-support vector regression[J]. *Composite Structures*, 2021,273(10):64-69.
- [13] 曹 斐,周 彧,王春晓,等. 一种改进的支持向量回归的混凝土强度预测方法[J]. 硅酸盐通报,2021,4(3):36-38.
- [14] DU Wen-ke, CHEN Rui-han, CONG Zheng-long. Application of support vector regression in prediction model using genetic algorithm optimized[J]. *Journal of Physics: Conference Series*, 2021,99(8):28-30.
- [15] 崔佳旭,杨 博. 贝叶斯优化方法和应用综述[J]. 软件学报,2018,10(12):3068-3090.
- [16] 王 雪,韩 韬. 基于贝叶斯优化随机森林的变压器故障诊断[J]. 电测与仪表,2021,13(6):167-173.
- [17] RATIMA P, TOSHNIWAL D. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance[J]. *Journal of Cleaner Production*, 2021,279(1):120-124.

[编辑:李 辉]

本文引用格式:

陈向俊,傅军平,陈栋栋,等. 基于极端梯度提升回归模型的电梯钢丝绳磨损预测方法[J]. 机电工程,2022,39(4):554-560.

CHEN Xiang-jun, FU Jun-ping, CHEN Dong-dong, et al. Elevator wire rope wear prediction method based on extreme gradient lifting regression model[J]. *Journal of Mechanical & Electrical Engineering*, 2022,39(4):554-560.

《机电工程》杂志: <http://www.meem.com.cn>