XGBoost 算法在风机主轴承故障预测中的应用

王桂兰,赵洪山,米增强
(华北电力大学 电子与电气工程学院,河北 保定 071003)

摘要:针对传统机器学习算法处理海量风机数据采集与监视控制(SCADA)监测数据效率低和准确度差的问题,提出利用极端梯度提升(XGBoost)算法预测风机主轴承故障。首先,对风机主轴承 SCADA 数据开展特征分析,挖掘和发现特征与故障之间的关联关系,并评估各特征的重要性;然后利用 XGBoost 算法构建主轴承故障预测模型,进行模型评估;最后,依据 SCADA 系统收集的实测数据对模型进行训练和测试,并调整 XGBoost 模型的主要参数,提高预测准确率。通过与经典梯度提升决策树(GBDT)算法诊断结果相对比,结果表明 XGBoost 在风机主轴承故障预测的效率和准确度方面均优于 GBDT 算法,是处理 SCADA 大规模数据集的有效工具。

关键词:XGBoost;数据采集与监视控制;主轴承;故障预测;大数据
中图分类号:TM 315 **文献标识码**:A DOI:10.16081/j.issn.1006-6047.2019.01.011

0 引言

近年来,随着国家清洁能源战略的实施和发展,建设和运营中的风电场数量和规模都快速、大幅增加。为了提高风电场的运行和监控水平,多数风电场安装了数据采集与监视控制 SCADA(Supervisory Control And Data Acquisition)系统。随着对大规模风电场运行状态的长期监测,海量的 SCADA 数据逐渐被收集。如国内某著名风电企业截至 2016 年在国内有约 1 万台风电设备,每台风电机组每秒产生 1 000 个点的量测数据,每秒大约有 10^7 个点的数据被采集和集中处理[1]。如何利用海量的 SCADA 数据,基于机器学习算法挖掘数据内部规模、提高风机安全运行水平成为目前迫切需要解决的重要问题[2]。文献[3]采用 SCADA 运行数据,结合风电机组的运行原理,建立了发电性能模型,将实时运行数据作为模型输入,通过分析模型预测残差实时监测风电机组发电性能的异常变化。文献[4]通过对风电机组 SCADA 振动信号数据的建模和分析,实现了对塔架、传动链、叶轮等机组重要部件的状态监测。然而,目前 SCADA 数据的整体利用水平仍偏低,且针对风机特定部件的故障诊断研究较少。

风机主轴轴承是风机传动系统的核心部件,主要承受径向力,其性能的好坏不仅影响传递效率,而且也决定了主传动链的维护成本,所以风机主轴轴承应具有良好的调心性能、抗振性能和运转平稳性。有较多的文献采用传统机器学习算法或信号处理技术研究风机主轴承的故障检测与分析,却较少考虑利用已收集的海量 SCADA 数据来实现。文献[5]提出一种基于盲源分离和流形学习算法的风电机组轴承故障特征提取方法,实现了风电机组轴承故障特征的准确提取。文献[6]基于振动信号,利用经验模态分解(EMD)降噪和谱峭度法实现了滚动轴承的早期故障诊断。文献[7]基于振动信号时域、频域和包络谱等多源特征,采用决策融合方法构建了直驱风力发电机组轴承故障诊断模型。文献[8]研究了基于集合经验模态分解(EEMD)和模糊 C 均值聚类的风电机组齿轮箱故障诊断。

然而,以上基于传统机器学习算法的研究仅适应于有限数据样本空间[9-10],针对大样本数据集时效率较低,很难满足工程实际需求。如针对某风机主轴承采集的 3 个月的 SCADA 监测数据,数据条目约为 73.6 万条,数据规模约为 120 MByte,运用经典机器学习算法支持向量机 SVM(Support Vector Machine)进行建模,模型构建时间超过 24 h,逻辑回归 LR(Logistic Regression)等算法也表现出相似的特性。另一方面,当前热门的深度学习算法能够实现数据特征的自动识别与提取[2],常用于图像处理、语音识别、文本识别等领域,但不适合处理特征明显的 SCADA 数据等结构化数据。因此需要寻找适应大规模结构化数据集的高效机器学习算法来处理 SCADA 数据。极端梯度提升 XGBoost(eXtreme Gradient Boosting)算法[11-12]是适应大规模数据集的并行集成学习算法,所谓集成学习算法是指利用多个弱分类器构成 1 个强分类器。XGBoost 算法能充分利用 CPU 多线程并行计算,提高运算效率,同时利用多种优化策略,使算法精度获得大幅提升。目前,XGBoost 算法已应用于电力短期负荷预测[13]、滚动轴承故障检测[14]、台区合理线损率估计[15]等多个行业领域。

收稿日期:2018-01-06;**修回日期**:2018-11-08
基金项目:国家科技支撑计划资助项目(2015BAA06B03);国家自然科学基金资助项目(51677072)
Project supported by the National Key Technology Support Program(2015BAA06B03) and the National Natural Science Foundation of China(51677072)

风机主轴承 SCADA 数据具有变量多、数据量大等特点。XGBoost 算法是高效并行机器学习算法，适应多数据类型和多变量的情况；与传统故障诊断方法相比，其能克服传统方法在数据规模较大情况下效率不够高的问题，有效利用大规模数据提升故障预测或诊断精度。本文基于风电机组主轴承 SCADA 数据，利用 XGBoost 算法研究风机主轴承故障预测，挖掘各个特征相关度并评估其重要性，实现了主轴承故障预测，为后续提升 SCADA 数据的利用水平、开展基于大数据的故障预测和诊断提供一种思路。

1 XGBoost 算法原理及模型介绍

XGBoost 算法是针对梯度提升机 GBM（Gradient Boosting Machine）的一个 C++实现，主要用于处理工业界规模问题。XGBoost 模型属于基于树集成的学习模型，它的基本思想是把多棵分类准确率较低的树模型组合起来，构建 1 个准确率相对更高的模型。该模型会不断地迭代提升，每次迭代生成 1 棵新树拟合前一棵树的残差，这个过程被称为梯度提升。简单而言，提升算法是首先使用简单的模型拟合数据，得到一个比较一般的结果，然后不断向模型中添加简单模型，提升模型的准确率。虽然多数情况下，XGBoost 模型中为层数较浅的决策树，但随着树的增多，集成模型的复杂度逐渐变高，直到接近数据本身的复杂度时，训练达到最佳水平。

XGBoost 模型由多棵分类回归树 CART（Classification And Regression Trees）集成，下面通过一个简单的例子说明 XGBoost 算法的工作原理，如附录中的图 A1 所示。

图 A1 为 2 棵 CART 集成的 XGBoost 模型，2 棵树分别根据不同的特征构建，用于预测某人对电脑游戏的喜好程度，其叶子节点用分值表示喜好程度。对每棵树的预测分数进行求和得到最终得分。如针对小男孩，第 1 棵树的预测分值为+2，第 2 棵树的预测分值为+0.9，则总分值为 2.9；而对于老年人预测分值和为-1.9，说明小男孩对游戏的喜好程度明显强于老年人。通过这个例子可以看出，模型中 2 棵树试图互补，从而得到最佳结果。

根据上文的描述，可以得到 XGBoost 的模型表示如式（1）所示。

$$\hat{y}_i=\sum_{k=1}^{K}f_k(x_i)\quad f_k\in F \tag{1}$$

其中，K 为树的数目；f_k 为函数空间 F 中的 1 个函数；$\hat{y}_i$ 为预测值；x_i 为输入的第 i 个样本；F 为所有可能的 CART 的集合。

在树集成学习算法中很难一次列出所有可能的 CART；相反地，XGBoost 算法使用梯度提升策略，不是一次得到所有的树，而是每次添加 1 棵新树，不断修补之前的测试结果。假设在第 t 步的预测值为 $\hat{y}_i^{(t)}$，则可得到如下的推导过程：

$$\begin{cases}\hat{y}_i^{(0)}=0\\ \hat{y}_i^{(1)}=f_1(x_i)=\hat{y}_i^{(0)}+f_1(x_i)\\ \hat{y}_i^{(2)}=f_1(x_i)+f_2(x_i)=\hat{y}_i^{(1)}+f_2(x_i)\\ \vdots\\ \hat{y}_i^{(t)}=\sum_{k=1}^{t}f_k(x_i)=\hat{y}_i^{(t-1)}+f_t(x_i)\end{cases} \tag{2}$$

因此，得到 XGBoost 算法的目标优化函数，如式（3）所示。

$$\begin{aligned}f_{\text{obj}}^{(t)}=&\sum_{i=1}^{n}l(y_i,\hat{y}_i^{(t)})+\sum_{i=1}^{t}\Omega(f_i)=\\&\sum_{i=1}^{n}l(y_i,\hat{y}_i^{(t-1)}+f_t(x_i))+\\&\Omega(f_t)+\text{cons tan }t\end{aligned} \tag{3}$$

为不失一般性，XGBoost 算法的目标函数由损失值和正则项 $\Omega(f_t)$ 两部分组成。在一般情况下，将损失函数的泰勒级数推广到 2 阶，并移出常数项，则第 t 步的目标函数变为：

$$f_{\text{obj}}^{(t)}=\sum_{i=1}^{n}\left(g_if_t(x_i)+\frac{1}{2}h_if_t^2(x_i)\right)+\Omega(f_t) \tag{4}$$

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

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

其中，g_i、h_i 分别为 1 阶和 2 阶导数。

对于树，定义其正则项为：

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{5}$$

其中，w_j 为树 f 中第 j 个叶子节点上的分值；T 为树 f 中叶子节点的总数目；γ 和 λ 为 XGBoost 自定义参数，γ 为 L1 正则的惩罚项，λ 为 L2 正则的惩罚项。从式（5）可以看出，XGBoost 算法对树的复杂度项增加了一个 L2 正则化项，即针对每个叶结点的得分增加 L2 平滑，其目的是避免过拟合。

最终得到目标函数为：

$$f_{\text{obj}}=-\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda}+\gamma T \tag{6}$$

$$G_j=\sum_{i\in I_j}g_i,\ H_j=\sum_{i\in I_j}h_i$$

其中，I 定义为每片叶子上的样本集合，$I_j=\{i\mid q(x_i)=j\}$，I_j 为第 j 片叶子上的样本集合，结构函数 q 表示输入样本到叶子索引号之间的映射关系，$q(x_i)=j$ 表示样本 x_i 在对应树中叶子节点的索引号为 j。

由式（6）可知，目标值越小，整个树的结构越优。

2 风机主轴承 SCADA 数据特征分析

特征分析的目标是最大限度地从原始数据中提

取特征以供算法和模型使用,其有效性很大程度上决定了模型的准确度。特征处理是特征工程的核心部分,包括数据预处理、特征选择、降维等操作。风机主轴承 SCADA 数据主要包括风速、温度等 10 个特征值,详细说明如表 1 所示。

表 1　风机主轴承 SCADA 数据特征

Table 1　SCADA data feature of wind turbine main bearing

序号	名称	中文含义
1	wind_speed	风速
2	gearbox_temperature_input_shaft1	齿轮箱输入轴 1 温度
3	gearbox_temperature_output_shaft2	齿轮箱输入轴 2 温度
4	gearbox_oil_temperature_gearbox	齿轮箱油温
5	generator_bearing_temperature_b	发电机轴承 b 温度
6	nacelle_temperature	机舱内部温度
7	grid_power	并网功率
8	main_bearing_rotor_side_temperature	主轴承转子侧温度
9	main_bearing_gearbox_side_temperature	齿轮箱侧主轴承温度
10	converter_motor_speed	减速比

XGBoost 算法采用 CART 模型,对数据类型具有较强的适应能力,可以处理离散、连续等多种类型的数据,因此无需对 SCADA 数据进行标准化、归一化、特征离散化等处理。对于缺失值,由于 XGBoost 算法自身能够处理数据缺失的情况,因此也不需要对缺失数据进行任何处理。但针对错误数据,如不符合物理定律或自然常识的数据需要进行人工排查并修正或删除。运用 XGBoost 算法的特征选择工具可以筛选出对分类和预测更有效的特征,挖掘特征与结果之间的关联关系,主要特征依据相关度排序如表 2 所示。

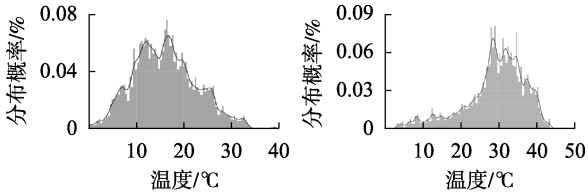
表 2　特征重要性排序

Table 2　Sorting of feature importance

排名	特征名称	分值
1	nacelle_temperature	0.190 271 817
2	main_bearing_gearbox_side_temperature	0.157 367 668
3	main_bearing_rotor_side_temperature	0.138 769 671
4	generator_bearing_temperature_b	0.123 032 904
5	gearbox_oil_temperature_gearbox	0.095 851 216
6	gearbox_temperature_output_shaft2	0.077 253 219
7	gearbox_temperature_input_shaft1	0.070 100 143
8	converter_motor_speed	0.054 363 376
9	grid_power	0.040 057 225
10	wind_speed	0.028 612 303

由表 2 可知,风机主轴承转子侧温度和机舱内部温度是 2 个重要的特征指标,其与预测准确性具有较高相关度。在工程实践中,这 3 个参数也是风机监测的重要参考指标。图 1 展示了机舱内部温度和主轴承转子侧温度数据分布情况,从图 1 可见它们基本符合正态或偏正态分布。

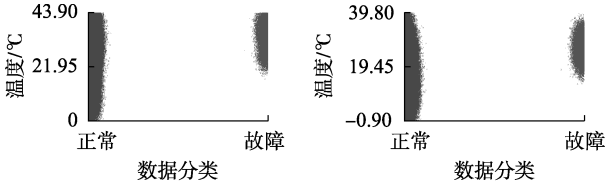
分别绘制训练数据集中机舱内部温度和风机主轴承转子侧温度与数据分类的关系,如图 2 所示。从图中可以看出,正常数据分布在有限范围内,而故障数据的分布范围更广,说明温度特征可以作为衡量故障的重要指标。



(a) 机舱内部温度数据分布　(b) 风机主轴承转子侧数据分布

图 1　机舱内侧温度和转子侧温度数据分布图

Fig.1　Data distribution of cabin inner temperature and main bearing rotor side temperature



(a) 机舱内部温度数据分布与数据分类的关系　(b) 风机主轴数据分类承转子侧温度数据分布与数据分类的关系

图 2　机舱内部温度和主轴承转子侧温度与数据分类的关系

Fig.2　Relationship between cabin inner temperature, main bearing rotor side temperature and data classification

3　测试结果

3.1　实验测试环境

针对某风场收集的 2 组不同风机主轴承 SCADA 故障和正常数据进行测试,其中一组作为训练数据,另一组作为测试数据。将风机主轴承故障修复后的数据作为正常数据,修复前的数据作为故障数据。每组数据包括约 3 个月的数据,SCADA 系统的采集频率为每秒 6 次,数据规模约为 120 MByte。测试环境采用 Python 语言的 Scikit-Learn 模块包,其具有操作简单、效率高、无访问限制、BSD 开源协议等特征,可调用封装好的 XGBoost 算法及相关模块。

XGBoost 算法最终可生成多棵 CART,附录中的图 A2 展示了一棵高度为 2 的基于风机主轴承 SCADA 数据的 CART。可根据某属性值分类,判断值是否在某一个范围内,最终得到权重值,用于分类预测。

3.2　参数优化

XGBoost 算法使用了几十个参数,为了提高模型的性能,参数的调整十分必要。XGBoost 算法主要包括通用类型参数、Booster 参数和学习任务参数 3 种类型的参数,每类参数包括多个具体的参数值,其中对算法性能影响较大的是 Booster 参数。XGBoost 算法的详细信息如表 3 所示。

在固定其他参数的基础上,改变树的高度和树的个数,比较了训练数据和测试数据的 Log 损失和分类错误率。损失函数定义为:

$$f_{\text{loss}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log_2 p_i + (1-y_i)\log_2(1-p_i)\right] \quad (7)$$

表3 XGBoost 算法的主要参数及说明

Table 3 Main parameters and description of XGBoost algorithm

序号	名称	中文含义	默认值
1	eta	步长	[0.01,0.2]
2	max_depth	树的最大高度	[3,10]
3	max_leaf_nodes	最大叶子的数量	2×max_depth
4	gamma	最小损失函数下降值	0
5	max_delta_step	树权重改变最大步长	0
6	colsample_bytree	随机列数占比	[0.5,1]
7	lambda	权重的 L2 正则化项	1
8	alpha	权重的 L1 正则化项	1
9	min_child_weight	最小叶子节点权重和	1
10	subsample	随机采样比例	[0.5,1]

其中，y_i 为输入实例对应的真实类别；p_i 为模型预测的第 i 个样本为第 1 类的概率。

分类错误率采用 0-1 损失 MCE(Mean Consequential Error)表示，其定义公式为：

$$r_{\mathrm{mce}} = \frac{1}{N}\sum_{\hat{y}_i \neq y_i} 1 \tag{8}$$

通过调整参数，依次改变树的高度为 2、4、6、8，在不同树数目的情况下，针对训练和测试数据集分别比较 Log 损失和分类的错误率，其结果如图 3 所示。

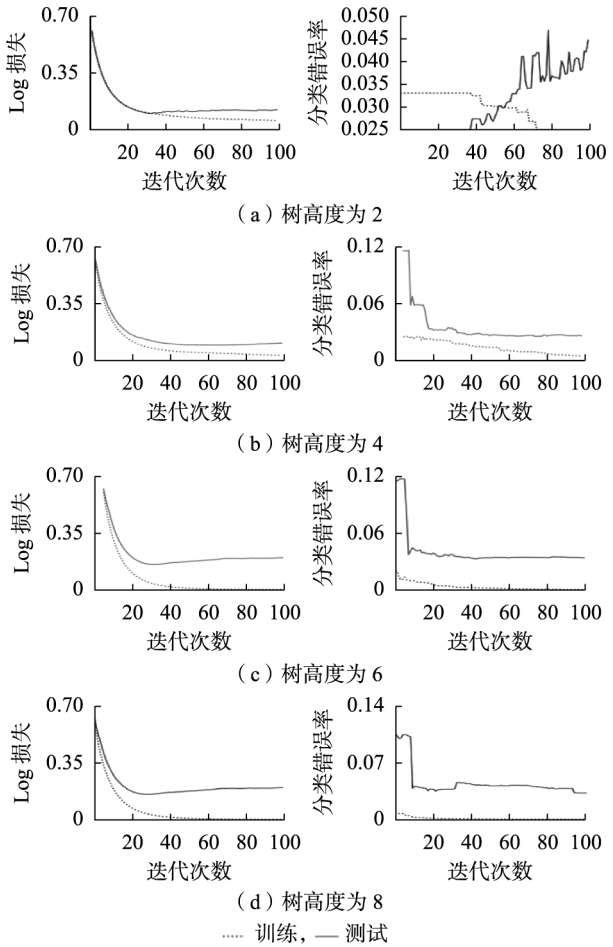


图3 不同树高度时 Log 损失和分类错误率对比

Fig.3 Comparison of Log loss and classification error rate among different tree heights

通过对比调整参数后的 Log 损失和错误率发现：当树的高度为 4、6 时，XGBoost 模型的损失函数和分类错误率都比较合理，XGBoost 模型在训练数据和测试数据中表现出相似的特性；当树的高度为 2 时，由于 XGBoost 模型过于简单，在训练数据中虽然较好，但在测试数据时错误率很高，说明其拟合情况有待加强；当树的高度为 8 时，树的复杂度很高，XGBoost 模型在训练和测试数据集上的效果出现较大差异。

3.3 实验结果

为测试 XGBoost 算法的性能和准确率，将其和传统梯度提升决策树 GBDT(Gradient Boosting Decision Tree)[16]进行比较。本文没有将 XGBoost 算法与支持向量机、逻辑回归等算法进行比较，是因为这些算法在大样本数据集上的效率偏低。根据参数优化的结果，分别测试 2 种算法在树的高度为 4 和 6 时的效率和错误率，结果如图 4 所示。从图 4 可见：随着迭代次数的增多，XGBoost 算法的效率相对 GBDT 算法有明显的提升，基本呈线性增长；多数情况下 XGBoost 算法的准确率优于 GDBT 算法，通过合理设置参数，XGBoost 的错误率比 GDBT 算法的错误率降低了 0.15%，同时在测试数据集上，XGBoost 算法的预测准确率达到 97.7%。

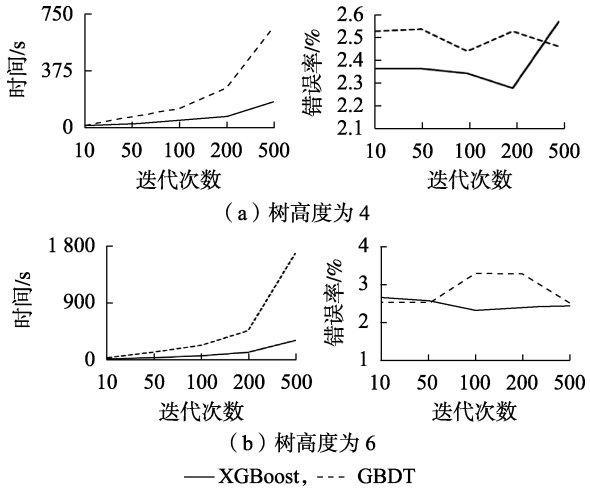


图4 不同树高度下 2 种算法的效率和错误率对比

Fig.4 Comparison of efficiency and error rate among two algorithm with different tree heights

4 结论

XGBoost 算法具有速度快、效果好、支持自定义损失函数等特点，是处理大规模数据集的有效解决方案。本文利用 XGBoost 算法对实测 SCADA 风机数据进行建模，分析风机主轴承状态，并预测其故障。通过实验对比，证明了该方法的有效性。下一步将针对更多、更丰富的 SCADA 实测数据集进行挖掘并识别故障准确类别，进一步验证算法的工程实践能力。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

[1] 汪喆. 基于大数据分析的风电机组运行状态监测评估[D]. 北京:华北电力大学,2016.
WANG Zhe. Research on wind turbine condition monitoring based on big data analysis[D]. Beijing:North China Electric Power University,2016.

[2] 刘辉海,赵星宇,赵洪山,等. 基于高度自编码网络模型的风电机组齿轮箱故障检测[J]. 电工技术学报,2017,32(17):156-163.
LIU Huihai,ZHAO Xingyu,ZHAO Hongshan,et al. Fault detection of wind turbine gearbox based on deep autoencoder network[J]. Transactions of China Electrotechnical Society,2017,32(17):156-163.

[3] 郭鹏,姜漫利,李航涛. 基于运行数据和高斯过程回归的风电机组发电性能分析与监测[J]. 电力自动化设备,2016,36(8):10-15.
GUO Peng,JIANG Manli,LI Hangtao. Performance analysis and monitoring based on SCADA data and Gaussian process regression for wind turbine power generation[J]. Electric Power Automation Equipment,2016,36(8):10-15.

[4] 郭鹏,徐明,白楠,等. 基于 SCADA 运行数据的风电机组塔架振动建模与监测[J]. 中国电机工程学报,2013,33(5):128-135.
GUO Peng,XU Ming,BAI Nan,et al. Wind turbine tower vibration modeling and monitoring driven by SCADA data[J]. Proceedings of the CSEE,2013,33(5):128-135.

[5] 赵洪山,李浪,王颖. 一种基于盲源分离和流形学习的风电机组轴承故障特征提取方法[J]. 太阳能学报,2016,37(2):269-275.
ZHAO Hongshan,LI Lang,WANG Ying. Fault feature extraction method of wind turbine bearing based on blind source separation and manifold learning[J]. Acta Energiae Solaris Sinica,2016,37(2):269-275.

[6] 苏文胜,王奉涛,张志新,等. EMD 降噪和谱峭度法在滚动轴承早期故障诊断中的应用[J]. 振动与冲击,2010,29(3):18-21.
SU Wensheng,WANG Fengtao,ZHANG Zhixin,et al. The EMD noise reduction and spectral kurtosis method for early fault diagnosis of roller bearing[J]. Journal of Vibration and Shock,2010,29(3):18-21.

[7] 安学利,蒋东翔,李少华. 基于决策融合的直驱风力发电机组轴承故障诊断[J]. 电网技术,2011,35(7):36-41.
AN Xueli,JIANG Dongxiang,LI Shaohua. Fault diagnosis of spherical roller bearing of direct-drive wind turbine based on decision fusion[J]. Power System Technology,2011,35(7):36-41.

[8] 王军辉,贾嵘,谭泊. 基于 EEMD 和模糊 C 均值聚类的风电机组齿轮箱故障诊断[J]. 太阳能学报,2015,36(2):319-324.
WANG Junhui,JIA Rong,TAN Bo. Fault diagnosis of wind turbine's gearbox based on EEMD and Fuzzy C means clustering[J]. Acta Energiae Solaris Sinica,2015,36(2):319-324.

[9] 戴志辉,李芷筠,焦彦军,等. 基于 BP 神经网络的小样本失效数据下继电保护可靠性评估[J]. 电力自动化设备,2014,34(11):129-134.
DAI Zhihui,LI Zhijun,JIAO Yanjun,et al. Reliability assessment based on BP neural network for relay protection system with a few failure data samples[J]. Electric Power Automation Equipment,2014,34(11):129-134.

[10] 肖怀硕,李清泉,施亚林,等. 灰色理论-变分模态分解和 NSGA-Ⅱ优化的支持向量机在变压器油中气体预测中的应用[J]. 中国电机工程学报,2017,37(12):3643-3653.
XIAO Huaishuo,LI Qingquan,SHI Yalin,et al. Prediction of dissolved gases in oil for transformer based on grey theory-variational mode decomposition and support vector machine improved by NSGA-Ⅱ[J]. Proceedings of the CSEE,2017,37(12):3643-3653.

[11] CHEN T,GUESTRIN C. XGBoost:a scalable tree boosting system[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco,USA:ACM,2016:785-794.

[12] TORLAY L,PERRONEBERTOLOTTI M,THOMAS E,et al. Machine learning-XGBoost analysis of language networks to classify patients with epilepsy[J]. Brain Informatics,2017(11):1-11.

[13] 李广野,李伟,田小蕾,等. 基于 XGBoost 算法的短期电力负荷预报[J]. 智能电网,2017,7(4):274-285.
LI Guangye,LI Wei,TIAN Xiaolei,et al. Short-term electricity load forecasting based on the XGBoost algorithm[J]. Smart Grid,2017,7(4):274-285.

[14] 张钰,陈珺,王晓峰,等. Xgboost 在滚动轴承故障诊断中的应用[J]. 噪声与振动控制,2017,37(4):166-170.
ZHANG Yu,CHEN Jun,WANG Xiaofeng,et al. Application of Xgboost to fault diagnosis of rolling bearings[J]. Noise and Vibration Control,2017,37(4):166-170.

[15] 王守相,周凯,苏运. 基于随机森林算法的台区合理线损率估计方法[J]. 电力自动化设备,2017,37(11):39-45.
WANG Shouxiang,ZHOU Kai,SU Yun. Line loss rate estimation method of transformer district based on random forest algorithm[J]. Electric Power Automation Equipment,2017,37(11):39-45.

[16] LIU J,WU C. A gradient-boosting decision-tree approach for firm failure prediction:an empirical model evaluation of Chinese listed companies[J]. Journal of Risk Model Validation,2017,11(2):43-64.

作者简介:

王桂兰

王桂兰(1979—),女,河北保定人,讲师,博士研究生,主要研究方向为风机故障预测与诊断、大数据分析技术(**E-mail**:wang.guilan@163.com);

赵洪山(1965—),男,河北保定人,教授,博士,主要研究方向为电力系统混杂建模、动态分析与控制以及电力系统优化计算等(**E-mail**:zhaohshcn@126.com);

米增强(1960—),男,河北保定人,教授,博士,主要研究方向为电力系统自动化与混杂电力系统动态分析与控制(**E-mail**:mizengqiang@sina.com)。

(下转第 83 页 continued on page 83)

[12] TIAN Pauchang. Estimation of wind energy potential using different probability density functions[J]. Applied Energy, 2011, 88(5): 1848-1856.

[13] 鲍海波,韦化,郭小璇,等. 考虑风电不确定性的概率区间潮流模型与算法[J]. 中国电机工程学报,2017,37(19):5633-5642.
BAO Haibo, WEI Hua, GUO Xiaoxuan, et al. The model and algorithm of probabilistic interval power flow considering wind power uncertainty[J]. Proceedings of the CSEE, 2017, 37(19): 5633-5642.

[14] 吴京秋,孙奇,杨伟,等. 基于 D-S 证据理论的短期负荷预测模型融合[J]. 电力自动化设备,2009,29(4):66-70.
WU Jingqiu, SUN Qi, YANG Wei, et al. Short-term load forecast model fusion based on D-S evidence theory[J]. Electric Power Automation Equipment, 2009, 29(4): 66-70.

[15] SHAFER G A. Mathematical theory of evidence[M]. Princeton, USA: Princeton University Press, 1976.

[16] YAGER R R, KACPRZYK J, FEDRIZZI M. Advances in the Dempster-Shafer theory of evidence[M]. New York, USA: John Wiley & Sons, 1994.

作者简介:

郭小璇

郭小璇(1986—),女,辽宁鞍山人,工程师,博士,主要研究方向为电力系统运行与分析(**E-mail**:guo_xiaoxuan@live.cn);

龚仁喜(1962—),男,广西桂林人,教授,博士研究生导师,博士,通信作者,主要研究方向为电力电子及新能源发电技术(**E-mail**:rxgong@ gxu.edu.cn);

鲍海波(1988—),男,安徽合肥人,高级工程师,博士,主要研究方向为电力系统最优运行与规划。

Modeling of wind speed uncertainty based on evidence theory

GUO Xiaoxuan[1,2], GONG Renxi[1], BAO Haibo[3]

(1. School of Electrical Engineering, Guangxi University, Nanning 530004, China;
2. Electric Power Research Institute of Guangxi Power Grid Corporation, Nanning 530023, China;
3. Nanning Power Supply Bureau of Guangxi Power Grid Corporation, Nanning 530031, China)

Abstract: Wind speed determines the output power of a wind farm, and its model is an important basis for researching the operation and planning of power system with wind power. A modeling method of wind speed uncertainty based on the evidence theory is proposed. The concept of BPA (Basic Probability Assignment) in the evidence theory is used to describe wind speed. The implementation method of determining the focal points and trust functions of BPA based on the measured historical data is proposed, and two modeling methods of equal probability interval and equal value interval are designed. Taking the measured wind speed of a wind farm as an example, the simulative results of the proposed model are compared with those of the probability distribution and interval distribution model, and the results show that the proposed model can determine the cumulative plausibility probability distribution and the cumulative belief probability distribution of wind speed, and can describe and deal with the uncertainty information of wind speed more effectively.

Key words: wind speed; wind power; evidence theory; basic probability assignment; cumulative plausibility probability distribution; cumulative belief probability distribution

(上接第 77 页 continued from page 77)

Application of XGBoost algorithm in prediction of wind motor main bearing fault

WANG Guilan, ZHAO Hongshan, MI Zengqiang

(School of Electrical & Electronic Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: To improve efficiency and accuracy of the processing of scale wind turbine SCADA data, XGBoost algorithm is used to predict the main bearing fault of the wind turbine instead of traditional machine learning algorithm. Firstly, the feature analysis is carried out on the SCADA data of main bearing, the relationship between the feature and fault is excavated and discovered, and the importance of each feature is evaluated. Then the XGBoost algorithm is used to construct the main bearing fault prediction model, and the model is evaluated. Finally, the fault prediction model is trained and tested according to the measured data collected by SCADA system, and the main parameters of the XGBoost model are adjusted to improve the prediction accuracy. Compared with the diagnosis results of classical GBDT (Gradient Boosting Decision Tree) algorithm, the results show that XGBoost is superior to GBDT algorithm in terms of efficiency and accuracy on wind turbine main bearing fault prediction, and it is an effective tool to deal with large scale data sets of SCADA.

Key words: XGBoost; SCADA; main bearing; fault prediction; big data

附录

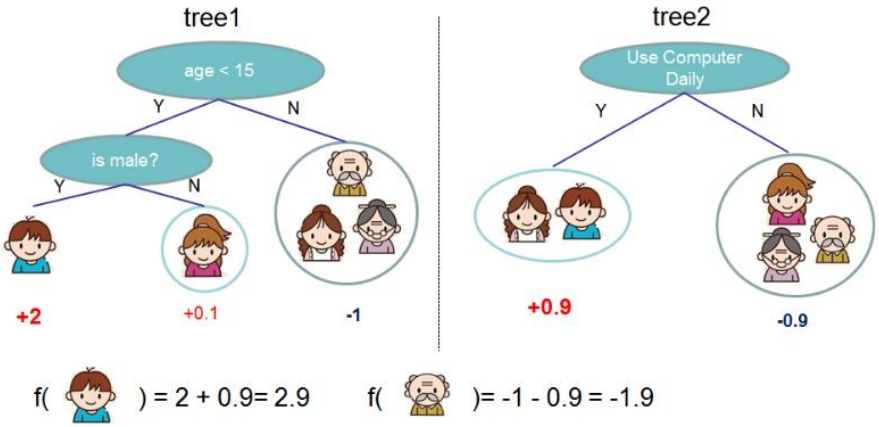


图 A1 棵分类回归树的集成

Fig.A1 Integration of two CARTs

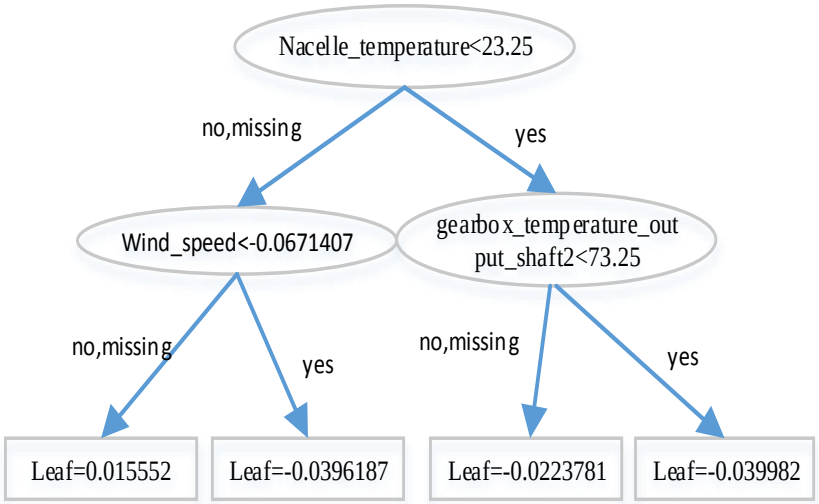


图 A2 高度为 2 的分类回归树

Fig.A2 CART with height 2