

基于 Relief 算法的风电机组故障特征参数提取方法^{*}

王正宇^{1 2} 张扬帆^{1 2} 段向阳³ 宋 鹏^{1 2} 杨伟新^{1 2}

(1. 国网冀北电力有限公司电力科学研究院(华北电力科学研究院有限责任公司) 北京 100045;

2. 国家电网公司风光储联合发电运行技术实验室 北京 100045;

3. 河北大唐国际新能源有限公司 河北承德 067000)

摘 要: 随着风电技术的快速发展,风电装机容量快速增长,风机 SCADA 系统监测的数据也越来越全面。风电机组部件故障与运行环境密切相关,可通过重点关注风电场中工况恶劣的风机来提高整体可用率。文章提出一种基于 Relief 算法从风电机组 SCADA 数据中提取故障特征参数的方法,根据 SCADA 历史数据和风机故障记录,利用 Relief 算法筛选出与风机部件故障相关度高的故障特征参数,作为判断风机运行工况和故障率的特征量。采用实际风电场的 SCADA 数据,以变桨故障和偏航故障为例进行计算,验证了本方法的有效性。

关键词: 风电机组; Relief 算法; SCADA; 故障特征参数

中图分类号: TM315 文献标识码: A DOI: 10.16308/j.cnki.issn1003-9171.2017.10.010

Selection Method of Fault Characteristic Parameters for Wind Turbine Based on Relief Algorithm

Wang Zhengyu^{1 2} Zhang Yangfan^{1 2} Duan Xiangyang³ Song Peng^{1 2} Yang Weixin^{1 2}

(1. State Grid Jibei Electric Power Co. Ltd. Research Institute ,

North China Electric Power Research Institute Co. Ltd. ,Beijing 100045 ,China;

2. State Grid Wind-Solar-Energy Storage Generation Laboratory ,Beijing 100045 ,China;

3. Hebei Datang International Renewable Energy Co. Ltd. ,Chengde 067000 ,China)

Abstract: With the rapid development of wind power technology ,the capacity of wind power is growing rapidly and monitoring data of SCADA system is also becoming more comprehensive. Faults of wind turbine components are closely related to work condition ,so the overall availability of whole wind farm could be improved by focusing on the wind turbines operating in harsh condition. In this paper ,a method based on SCADA data to judge harsh condition is proposed. According to historical data of SCADA and fault record ,the fault characteristic parameters are selected by Relief algorithm as the characteristic quantity of failure rate of wind turbine. Taking pitch fault and yaw fault as example ,the SCADA data of the actual wind farm are simulated and the effectiveness of the method is verified.

Key words: wind turbine ,Relief algorithm ,SCADA ,fault characteristic parameters

0 引言

近年来,随着生态环境的日益恶化和能源危机的迫在眉睫,风力发电越来越受到世界各国的重视,全球风机装机容量迅猛增长^[1]。同时,相比于传统发电,风力发电技术尚未成熟,风电机组

的故障率较高^[2]。目前风电机组没有成熟的预防性和诊断性试验体系,振动等在线监测技术也不成熟^[3]。

风机故障与运行环境尤其是恶劣工况密切相关。近年来,风机 SCADA 系统快速更新换代^[4],其监测的风机运行状态参数愈发全面,并

^{*} 国家科技支撑计划(2015BAA06B03)

且已能实现秒级数据的采集和存储。目前,基于 SCADA 数据的风电机组故障诊断技术研究已取得不少成果,但如何根据 SCADA 数据对故障进行长期预测还有待研究^[5-7]。SCADA 系统监测的状态参数众多,首先需要确定与恶劣工况和部件故障相关性强的故障特征参数,才能进一步进行故障预测。**Relief 算法**是一种经典的特征评估算法,在各行各业得到了广泛的应用,为基于 SCADA 数据的风电机组故障特征评估提供了一种解决思路^[8-10]。

本文提出一种风电机组故障特征参数选择方法,基于风电机组 SCADA 系统监测及存储的运行参数及故障信息,构建变桨系统正常运行数据集和故障运行数据集,采用 Relief 算法筛选与变桨系统故障相关性较大的 SCADA 参数,再进行相关性分析,最终确定故障特征参数,作为估计风电机组故障未来概率大小的特征量。

1 故障特征参数提取算法

1.1 Relief 算法简介

大数据是最近几年提出的热点概念。所谓大数据,一方面指数据样本数量巨大,另一方面指数据样本的维数高。大数据处理问题的不断出现,使许多算法的性能受到不相关或冗余特征的负面影响。

数据样本的维数过大经常会引起维数灾难(Curse of Dimensionality)等问题。因此,对大规模数据进行处理前,需要通过特征选择对数据样本进行特征选择以降低样本维数。特征选择是指从输入数据集中选择出与某种评估标准相关性较大的特征数据组成特征子集。在利用算法处理大数据之前,首先选择出关键数据,有助于精简算法模型,提高工作效率。因此,特征选择在数据挖掘和机器学习中扮演着一个极其重要的角色,处理风机 SCADA 数据前应当对其进行数据选择。

最早由 Kira 等于 1992 年提出的 Relief 算法是一种经典的特征评价算法,其借鉴了最近相邻学习算法的思路,得到了广泛的应用^[11]。Relief 算法选择特征参数的目标是:能够根据特征参数区分同类的样本和不同类的样本。**其核心思想**是对样本的每一个待选参数赋予权值来表征其与样本类别的相关程度,利用假设间隔(Hypothe-

ses-Margin)对权值进行迭代计算,通过样本训练,选择权值较大的待选参数作为特征参数,组成特征子集,忽略权值较小的特征参数^[12]。

Relief 算法是一种**有监督的特征评估方法**,它要求的样本集是带有分类标签的,也就是说样本类别是已知的。Relief 算法适合处理样本数量大、维数大,同时分类较少的特征选择问题。随着风机 SCADA 系统的逐渐发展完善,其监测的运行数据类别越来越多,采集周期也越来越短。对于一个大型风电场,年度 SCADA 数据数量大,维数大,可认为属于大数据范畴,可采用 Relief 算法进行处理。对于风机部件故障,应先对 SCADA 数据进行分类,根据故障与否将其 SCADA 数据分为正常数据集和故障数据集 2 部分,再采用 Relief 算法进行样本训练,提取故障特征参数。

1.2 风电机组故障特征参数提取方法

本文采用 Relief 算法对风机 SCADA 数据进行处理,步骤如下。

假设有 n 个样本(包括故障数据集和正常数据集),每个样本有 m 个待选参数(与故障相关的 SCADA 运行参数),第 i 个样本的第 j 个待选参数赋予权值为 w_{ij} 。设 w_{ij}^0 初值为零,对 $m \times n$ 个待选参数依次进行样本训练:

$$w_{i+1,j} = w_{i,j} + \frac{\text{diff}(S, x_i, M(x_i))}{n} - \frac{\text{diff}(Y, x_i, H(x_i))}{n}$$

$$\text{diff}(x, x') = \begin{cases} |x - x'| & x \neq x' \\ 0 & x = x' \end{cases}$$

式中, Y 代表与抽取的样本 x_i 同类别的样本集合, S 代表与 x_i 不同类的样本集合, $H(x_i)$ 与 $M(x_i)$ 分别代表与 x_i 同类和不同类的最近邻点;

$$\text{diff}(x, x') = \begin{cases} |x - x'| & x \neq x' \\ 0 & x = x' \end{cases}$$

对样本分类影响较大的待选参数,与同类样本的距离较近,与异类样本的距离较远,故根据公式计算出来的权值较大。因此,根据样本训练得到的权值能够反映特征参数与样本类别的相关程度,可根据权值大小选择出特征参数组成特征子集。

由于部分 SCADA 参数之间相关性较强,本文在采用 Relief 算法得到与风机故障相关性较大(权值较大)的参数类别后,需要进一步去除冗余参数以确定最终的特征参数子集。本文的处理方法是对权值较大的参数进行两两相关性分析,

计算皮尔逊相关系数(Pearson Correlation Coefficient)判断相关性大小,若一对参数相关性强,则保留权重较大的参数,舍弃权重较小的参数。

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

上式为皮尔逊相关系数计算公式, $\text{cov}(X,Y)$ 表示变量 X,Y 的协方差, $\sigma_X \sigma_Y$ 表示变量 X,Y 标准差之积。

风电机组故障特征参数选择流程如图 1 所示。

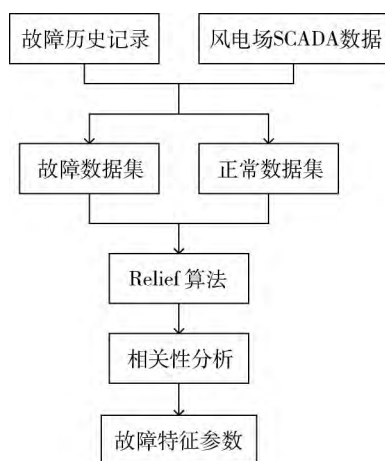


图 1 风电机组故障特征参数选择流程图

2 算例

本文以河北地区某风电场作为研究对象。该风电场工程总装机容量为 199.5 MW, 安装 1.5 MW 风机 133 台, 其中 A 型号风机 40 台, B 型号风机 93 台。

风机的变桨、偏航系统运行状态直接和运行环境尤其是风况直接相关, 故障率较高。本文以变桨、偏航故障为例, 分析风电场整年的变桨和偏航故障记录, 选择故障特征参数, 验证本文算法的有效性。

2.1 A 型号风机

风电场提供了 A 型号风机一整年的 SCADA 数据, 数据分辨率 10 min。数据表中共包含 38 个数据参数, 其中温度参数 18 个, 其他如变桨电机扭矩、振动幅值等数据参数 7 个, 有效数据包含有功功率、无功功率、发电参数、叶片桨距角信号 2 个、风速、发电机转速信号 6 个。

选取每台风机与变桨/偏航故障相关的 36 个 SCADA 数据参数(其余参数置零), 计算 1 年内的平均值、变化幅度以及方差, 作为待选故障

特征参数。

(1) 变桨故障

根据运行日志, A 型号总共 40 台风机在统计的 1 年时间内有 19 台风机变桨系统发生故障, 累计故障 28 次。以故障风机的待选特征参数构建故障数据集, 正常风机的待选特征参数构建正常数据集, 输入 Relief 算法, 计算各个状态参数对风机变桨故障的影响权重。

各状态参数权重分布如图 2 所示, 根据权重柱图突跳点位置, 将权重阈值设为 500。进一步对待选特征参数进行两两冗余分析, 去除冗余特征参数(相关系数阈值 0.8)。最终得到 A 型号风机变桨系统故障特征参数的选择结果如表 1 所示。

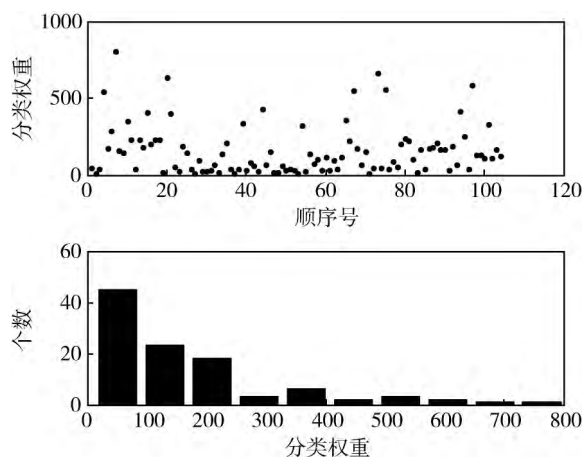


图 2 A 型号风机状态参数权重(变桨故障)

表 1 A 型号风机变桨系统故障特征参数

故障特征参数	权重
机侧电感温度(平均值)	803.10
风速(平均值)	664.72
驱动链摆动幅度(变化幅度)	632.28
控制用发电机最小转速(变化幅度)	587.06
风速(方差)	559.08
机舱电池温度(方差)	550.29
NCC320 温度(平均值)	544.67

根据故障历史记录, 表中所列特征参数与 A 型号风机变桨系统的运行工况和故障概率相关性较强, 未来可根据 SCADA 数据计算特征参数值来确定需要重点关注的 A 型号风机。

(2) 偏航故障

根据运行日志, A 型号总共 40 台风机在统计

的 1 年时间内有 10 台风机偏航系统发生故障,累计故障 15 次。以故障风机的待选特征参数构建故障数据集,正常风机的待选特征参数构建正常数据集,输入 Relief 算法,计算各个状态参数对风机变桨故障的影响权值。

各状态参数权重分布如图 3 所示,根据权重柱图突跳点位置,将权重阈值设为 400。进一步对待选特征参数进行两两冗余分析,去除冗余特征参数(相关系数阈值 0.8)。最终得到 A 型号风机偏航系统故障特征参数的选择结果如表 2 所示。

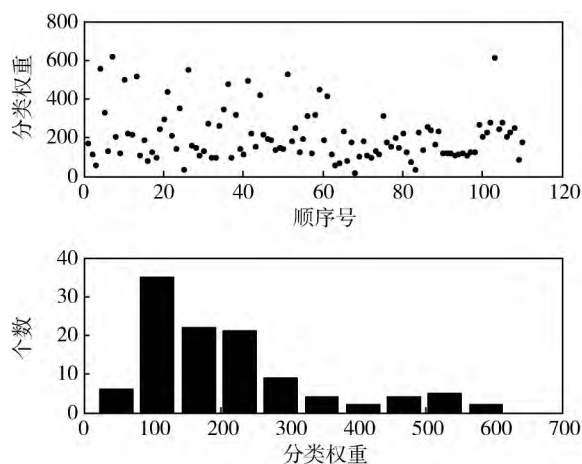


图 3 A 型号风机状态参数权重(偏航故障)

表 2 A 型号风机偏航系统故障特征参数

故障特征参数	权重
机侧电感温度(平均值)	618.89
控制用发电机最小转速(变化幅度)	613.97
NCC320 温度(平均值)	558.38
滤波板温度(平均值)	550.76
变桨电机 3 温度(变化幅度)	531.14
机侧半导体温度(平均值)	517.96
网侧电感温度(平均值)	499.52
最大发电机绕组温度(平均值)	496.66
驱动方向发电机轴承温度(变化幅度)	480.86
变桨电机 3 扭矩(平均值)	453.02
驱动链摆动幅度(方差)	441.77
变桨电机 1 温度(平均值)	424.62
变桨电机 3 扭矩(方差)	416.40

根据故障历史记录,表中所列特征参数与 A 型号风机偏航系统运行工况和故障概率相关性较强,未来可根据 SCADA 数据计算特征参数值

来确定需要重点关注的 A 型号风机。

2.2 B 型号风机

风电场共安装 B 型号风机 93 台,风电场提供了一整年的 SCADA 数据,数据表包含 106 个数据变量,与 A 型号机组 SCADA 数据相比增加了一些中间变量。

选取每台风机与变桨/偏航故障相关的 40 个 SCADA 数据参数(其余参数置零),计算 1 年内的平均值、变化幅度以及方差,作为故障特征参数。

(1) 变桨故障

根据运行日志, B 型号 93 台风机在统计的 1 年时间内共有 62 台风机变桨系统发生故障,累计故障 261 次。以故障风机的待选特征参数构建故障数据集,正常风机的待选特征参数构建正常数据集,输入 Relief 算法,计算各个状态参数对风机变桨故障的影响权值。

各状态参数权重分布如图 4 所示,根据权重柱图突跳点位置,将权重阈值设为 1 500。进一步对待选特征参数进行两两冗余分析,去除冗余特征参数(相关系数阈值 0.8)。最终得到 B 型号风机变桨系统故障特征参数的选择结果如表 3 所示。

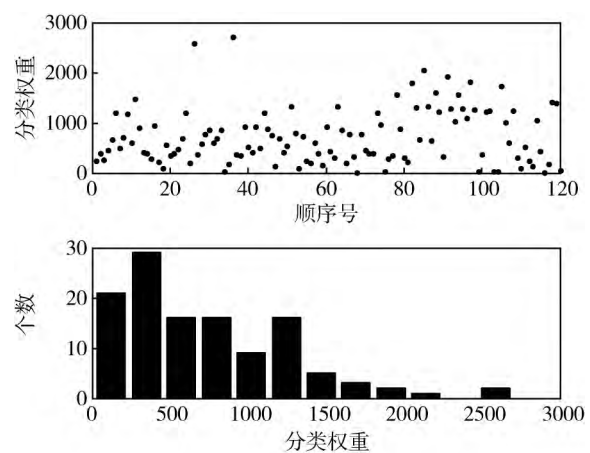


图 4 B 型号风机状态参数权重(变桨故障)

表 3 B 型号风机变桨系统故障特征参数

故障特征参数	权重
10 s 内风向偏差(变化幅度)	2 702.49
轴 1 控制柜温度(变化幅度)	2 063.28
轴 2 控制柜温度(变化幅度)	1 935.00
顶部控制柜温度(平均值)	1 571.53
轴 2 电池柜温度(变化范围)	1 557.23

根据故障历史记录,表中所列特征参数与 B 型号风机变桨系统的运行工况和故障概率相关性较强,未来可根据 SCADA 数据计算特征参数值来确定需要重点关注的 B 型号风机。

(2) 偏航故障

根据运行日志, B 型号 93 台风机在统计的 1 年时间内共有 19 台风机偏航系统发生故障,累计故障 33 次。以故障风机的待选特征参数构建故障数据集,正常风机的待选特征参数构建正常数据集,输入 Relief 算法,计算各个状态参数对风机变桨故障的影响权重。

各状态参数权重分布如图 5 所示,根据权重柱图突跳点位置,将权重阈值设为 1 600。进一步对待选特征参数进行两两冗余分析,去除冗余特征参数(相关系数阈值 0.8)。最终得到 B 型号风机偏航系统故障特征参数的选择结果如表 4 所示。

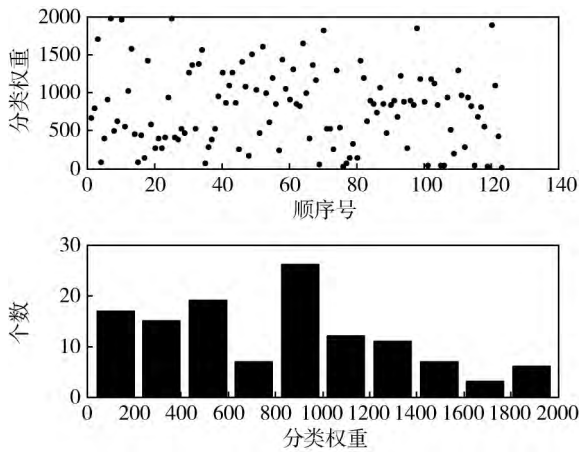


图 5 B 型号风机状态参数权重(偏航故障)

表 4 B 型号风机偏航系统故障特征参数

故障特征参数	权重
PLC 发电机转速(方差)	1 983.48
轴 1 温度(平均值)	1 897.88
轴 3 控制柜温度(平均值)	1 849.94
主开关吸合次数(平均值)	1 817.15
功率(方差)	1 702.70
齿轮箱轴承温度(平均值)	1 653.14
主轴承温度(变化范围)	1 609.27

根据故障历史记录,表中所列特征参数与 B 型号风机偏航系统的运行工况和故障概率相关性较强,未来可根据 SCADA 数据计算特征参数值来确定需要重点关注的 B 型号风机。

从以上算例结果可以看出,风机不同类型故障对应的特征参数具有各自的特点。

(1) 变桨系统

(a) 风机运行风速(平均值、方差)、风向偏差(变化幅度)、控制柜/电池柜温度(平均值)能够反映风机变桨系统运行工况的恶劣程度以及故障概率;

(b) 变桨系统不同故障特征参数的权重相差较大,几个权重较大的故障特征参数即可反映风机的运行工况恶劣程度以及变桨系统故障概率。

(2) 偏航系统

(a) 风机控制用发电机转速(方差)、变桨电机/控制柜温度(平均值)、主轴承温度(平均值、变化范围)等故障特征参数能够比较有效地反映风机偏航系统运行工况的恶劣程度以及故障概率;

(b) 偏航系统不同待选故障特征参数的权重相差较小,需要选择较多的故障特征参数才能有效反映风机的运行工况恶劣程度以及偏航系统故障概率;

(c) 偏航系统故障特征参数选择结果较差可能是由于风电场故障风机偏航故障次数较少,故障数据集样本不足,导致 Relief 算法无法有效计算特征参数权重。

(3) 不同型号风机的 SCADA 系统记录数据种类不同,因此计算得到的故障特征参数也不同: B 型号风机 SCADA 系统记录数据种类较多(待选特征参数较多),最终的故障特征参数选择结果较好(特征参数权重大、种类少),这从侧面说明较全面的 SCADA 测点有助于选择故障特征参数反映风机的运行工况和故障概率。

3 小结

本文提出一种基于 Relief 算法的风电机组故障特征参数提取方法。首先结合风电机组 SCADA 数据和故障历史记录建立样本集合,再通过样本训练和相关性分析选取故障特征参数。本文以某实际风电场的变桨偏航故障为例,提取 2 种型号风机的变桨偏航故障特征参数,计算结果验证了本文方法的有效性。选取风电机组故障特征参数有助于风电机组的状态监测、状态评估及异常识别,对降低机组故障率,提高运行可靠性具有重要意义和工程价值。

参考文献

- [1] Global Wind Report 2015: Annual Market Update [R]. Global Wind Energy Council 2016.
- [2] Tavner P. Offshore wind turbines: reliability, availability and maintenance [M]. London: The Institution of Engineering and Technology 2012.
- [3] 金晓航, 孙毅, 单继宏, 等. 风力发电机组故障诊断与预测技术研究综述 [J]. 仪器仪表学报, 2017, 38(5): 1041-1053.
- [4] 梁颖, 方瑞明. 基于 SCADA 和支持向量回归的风电机组状态在线评估方法 [J]. 电力系统自动化, 2013, 37(14): 7-12.
- [5] Rodríguez-López M A, López-González L M, López-Ochoa L M, et al. Development of indicators for the detection of equipment malfunctions and degradation estimation based on digital signals (alarms and events) from operation SCADA [J]. Renewable Energy 2016, 99(12): 224-236.
- [6] Sun P, Li J, Wang C, et al. A generalized model for wind turbine anomaly identification based on SCADA data [J]. Applied Energy 2016, 168: 550-567.
- [7] 李学伟. 基于数据挖掘的风电机组状态预测及变桨系统异常识别 [D]. 重庆: 重庆大学 2012.
- [8] 赖冰凌, 王新宇. Relief 算法在关门车故障自动识别中的应用 [J]. 铁路计算机应用, 2007, 16(1): 21-23.
- [9] 李颖新, 李建更, 阮晓钢. 肿瘤基因表达谱分类特征基因选取问题及分析方法研究 [J]. 计算机学报, 2006, 29(2): 324-330.
- [10] 范文兵, 王全全, 雷天友, 等. 基于 Relief 算法的故障图像识别与匹配方法 [J]. 兵工自动化, 2010, 29(10): 60-63.
- [11] Robnik-Sikonja M, Kononenko I. Comprehensive Interpretation of Relief's Estimates [C]//Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 2001: 433-440.
- [12] Kononenko I. Estimating attributes: Analysis and extensions of RELIEF [C]//European Conference on Machine Learning. Springer-Verlag New York, Inc. 1994: 171-182.
- 收稿日期: 2017-09-28
- 作者简介: 王正宇(1990—), 男, 硕士, 助理工程师, 研究方向为新能源并网技术。
- (本文编辑 徐伟婷)
- *****
- (上接第 49 页)
- [2] Rogers J, Fink S, Porter K. Examples of wind energy curtailment practices [R]. Columbia: National Renewable Energy Laboratory 2010.
- [3] Fink S, Mudd C, Porter K, et al. Wind energy curtailment case studies May 2008-May 2009 [R]. Columbia: National Renewable Energy Laboratory 2009.
- [4] 王乾坤. 国内外风电弃风现状及经验分析 [J]. 华东电力, 2012, 40(3): 378-381.
- [5] 朱向东. 目前中国风电弃风现状及对策 [J]. 能源与节能, 2012(10): 30-30.
- [6] 梅华威, 米增强, 白俊良. 风电场及机组出力损失计算模型与方法 [J]. 电力系统自动化, 2014, 38(2): 12-16.
- [7] 姜文玲, 冯双磊, 孙勇, 等. 基于机舱风速数据的风电场弃风电量计算方法研究 [J]. 电网技术, 2014, 38(3): 647-652.
- [8] 甘槐樟. 利用损失电量评价指导风电场的生产管理 [J]. 湖南电力, 2011, 31(4): 44-46.
- [9] 丁坤, 吕清泉, 蔡旭, 等. 采用样板机法计算风电场弃风电量的实测数据统计分析 [J]. 可再生能源, 2016, 34(1): 56-63.
- [10] 卢斯煜, 陈雷, 赵睿, 等. 电网弃风电量计算方法及其在电网规划中的应用 [J]. 广东电力, 2016, 29(10): 31-36.
- [11] 任华, 姚秀萍, 张新燕, 等. 风电场弃风电量统计方法研究 [J]. 华东电力, 2013, 41(10): 2148-2152.
- [12] 宋颖巍, 刘岩, 韩震焘, 等. 基于修正风速的风电并网运行弃风电量评估 [J]. 电气应用, 2014(23): 42-45.
- [13] 白恺, 宋鹏, 刁嘉, 等. 基于可靠性和发电性能的新能源发电设备状态评估方法 [J]. 中国电力, 2016, 49(7): 140-144.
- 收稿日期: 2017-09-28
- 作者简介: 张海军(1986—), 男, 本科, 助理工程师, 研究方向为风电机组运行检修。
- (本文编辑 石磊)