

采用预测模型与模糊理论的风电机组状态参数异常辨识方法

孙 鹏¹, 李 剑², 寇晓适¹, 吕中宾¹, 姚德贵¹, 王 吉¹, 王磊磊¹, 滕卫军¹

(1. 国网河南省电力公司电力科学研究院, 河南 郑州 450000;

2. 重庆大学 输配电装备及系统安全与新技术国家重点实验室, 重庆 400030)

摘要: 为提高风电机组的停运预警能力, 基于风电场数据采集与监控(SCADA)系统数据提出了一种风电机组状态参数的异常辨识方法。对参数进行划分, 针对与环境因素密切相关的状态参数, 采用神经网络建立了状态参数预测模型。采用本机组近期 SCADA 样本、本机组历史样本和其他机组近期样本分别作为预测模型的训练数据, 对比分析了基于 3 类样本建立的模型的预测精度。采用平均绝对误差对基于本机组历史样本和其他机组近期样本建立的预测模型进行选择。定义了异常程度指标量化预测残差的异常程度。为了提高异常辨识的精度, 采用模糊综合评判对筛选出的预测模型的异常辨识结果进行融合。最后, 以国内某风场的 1.5 MW 风电机组为例进行了异常分析, 并传统的风电机组状态参数异常检测方法进行了对比, 实例分析结果表明所提出的异常辨识方法具有更高的准确性。

关键词: 风电机组; 风电场数据采集与监控系统; 预测模型; 模糊综合评判; 异常辨识

中图分类号: TM 315

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2017.08.012

0 引言

随着全球风电产业的迅猛发展, 风电场的运行维护问题受到了广泛关注。风电场数据采集与监控 SCADA(Supervisory Control And Data Acquisition)系统作为风电机组状态监测系统的重要组成部分, 能够为监测风电机组及组件的运行状态提供数据支持。通过对风电机组上各类传感器信号的采集, SCADA 系统可提供自然环境参数、设备运行参数与电网参数等多种状态信息。充分挖掘 SCADA 数据中所包含的故障信息对提高风电机组的停运预警能力和可靠性具有显著意义^[1]。

风电场 SCADA 系统中的状态参数如设备温度参数、机组输出功率、转子转速等受自然环境与运行工况的影响显著, 这些参数难以直接用于风电机组的状态监测与故障诊断。通过建立 SCADA 状态监测参数的预测模型, 可实现状态参数的准确预测, 降低外部环境和工况因素对风电机组实时状态参数的影响^[2]。较多研究者利用状态参数预测模型辨识机组故障前状态参数的异常状态, 进而实现机组的早期故障预警^[3-5]。基于统计的异常辨识理论通常认为正常状态处于随机模型的高概率区间, 而异常状态处于低概率区间^[6], 因此预测模型的预测误差可用来检测状态参数与机组的异常状态。采用上述异常

辨识方法时需注意 2 点: 一点是异常辨识的灵敏度受预测模型预测精度的影响^[7]; 另一点是为保证异常辨识的有效性, 预测模型的训练数据来自机组处于健康状态下的数据^[8]。为提高状态参数的预测精度, 神经网络^[9-10]、支持向量机^[11]、自适应模糊推理系统^[12]、非线性估计技术^[13]等多种数据挖掘算法均被用于建立风电机组状态参数的预测模型。此外, 还有学者针对预测模型的输入参数选择方法开展研究^[14-15]。

然而, 鲜有研究考虑了训练数据对基于预测模型的异常辨识方法有效性的影响。目前预测模型训练数据的选择缺乏通用的规则。文献^[16]指出自投运起风电机组各部件始终处于磨损与劣化的过程中, 因此预测模型训练数据的时间尺度应该尽量短。已有研究的训练数据通常来自近期数周或数月内的 SCADA 数据, 然而在故障停机前难以判定机组是否处于正常状态, 当训练样本中含有机组异常时的数据时异常辨识的精度与有效性将降低。当训练数据选自历史 SCADA 数据时, 模型预测精度将受机组组件磨损与劣化的影响。同时, 由于同一风电场内不同机组的状态参数取值分布具有较大差异, 选取其他机组数据作为训练数据的预测模型的预测精度需要进一步的研究。

针对以上问题, 本文基于预测模型与模糊综合评判提出了风电机组状态参数的异常辨识方法。采用神经网络建立了机组状态参数预测模型, 对比了不同类型训练样本建立的模型的预测精度。定义了异常程度指标量化状态参数预测残差的异常程度, 采用模糊理论对筛选出的预测模型的异常辨识结果进行融

收稿日期: 2017-02-27; 修回日期: 2017-07-04

基金项目: 国家电网公司重大科技专项(智能变电站母线及智能组件可靠性研究)

Project supported by Major Project of State Grid Corporation of China(Research on Reliability of Buses and Components in Smart Substation)

合。最后,通过对某风电场的 1.5 MW 机组进行实例分析,验证了本文提出的异常辨识方法的有效性和准确性。

1 风电机组状态参数与数据选择

表 1 为采集自某风电场 SCADA 系统的风电机组状态参数,相应的传感器位置如图 1 所示。

表 1 风电机组状态参数
Table 1 Status parameters of WT

编号	风电机组状态参数	单位	类别
1	齿轮箱输入轴温度	°C	1
2	齿轮箱输出轴温度	°C	1
3	齿轮箱油温	°C	1
4	齿轮箱冷却水温度	°C	1
5	主轴温度(转子侧)	°C	1
6	主轴温度(齿轮箱侧)	°C	1
7	发电机绕组温度(U 相)	°C	1
8	发电机绕组温度(V 相)	°C	1
9	发电机绕组温度(W 相)	°C	1
10	发电机轴承 a(前)温度	°C	1
11	发电机轴承 b(后)温度	°C	1
12	发电机冷却风温度	°C	1
13	控制柜温度	°C	1
14	变频控制器温度	°C	1
15	输出功率	kW	1
16	转子转速	r/min	1
17	偏航位置	(°)	2
18	对风偏差	(°)	2
19	主液压	bar	2
20	转子刹车系统液压	bar	2

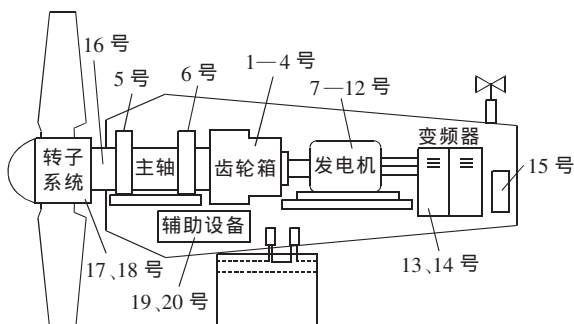


图 1 风电机组的主要组件与传感器位置

Fig.1 Main components of WT and its sensor positions

根据各参数与自然环境的相关性可将状态参数分为 2 类。

a. 第 1 类状态参数包含设备温度、转子转速与机组输出功率。根据风电机组设备的热量传导过程,设备温度与风速和环境温度密切相关。图 2 为齿轮箱输入轴温度与风速及环境温度的关系。

b. 第 2 类状态参数包含偏航位置、对风偏差、主液压与转子刹车系统液压,该类参数的取值与自然环境关系较弱。

由于仅通过设定阈值即可实现第 2 类状态参

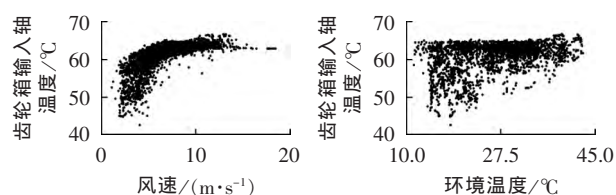


图 2 齿轮箱输入轴温度与风速及环境温度的关系

Fig.2 relationship between gearbox input shaft temperature and wind speed or environmental temperature

数的异常辨识,故本文仅探讨第 1 类状态参数的异常辨识方法。

本文所采用的 SCADA 数据来自华北地区某陆上风电场。该风电场包含 34 台 1.5 MW 双馈式风电机组,所有机组均为同一型号。SCADA 数据自 2011 年 2 月 15 日开始采集,采样率为每秒 1 个采样点。

2 风电机组状态参数预测模型

建立第 1 类状态参数的预测模型时需考虑输入参数、样本数据、训练算法 3 点因素。本文从双馈式风电机组的工作原理出发,考虑各状态参数与自然环境、机组工作特征的关系,选择各状态参数预测模型的输入参数。选用神经网络算法建立环境敏感类状态参数(第 1 类参数)的预测模型。最后,对比了采用不同样本建立的预测模型的预测精度。

2.1 输入参数选择

输入参数选择是简化预测模型结构与确保预测模型预测精度的关键,根据变速恒频风电机组的运行原理与内部发热及散热机制^[17],选取状态参数预测模型的输入参数如表 2 所示。

表 2 第 1 类状态参数预测模型的输入参数

Table 2 Input parameters of prediction models based on data type 1

目标状态参数	预测模型输入参数
t 时刻输出功率	t 时刻风速、 t 时刻对风偏差、 t 时刻桨距角
t 时刻转子转速	t 时刻风速、 t 时刻对风偏差、 t 时刻桨距角
t 时刻设备温度	t 时刻风速、 t 时刻输出功率、 t 时刻环境温度、 $t-1$ 时刻设备温度

2.2 训练样本选择与数据预处理

选取以下 3 类 SCADA 数据作为状态参数预测模型的训练样本。

a. 本机组近期 SCADA 数据:本机组近期 SCADA 数据提取自目标机组近期 60 d 内的 SCADA 数据(不包含当天数据)。

b. 本机组历史 SCADA 数据:本机组历史 SCADA 数据提取自上一年目标机组与近期数据相同时间段的 SCADA 数据(不包含当天数据)。

c. 其他机组近期 SCADA 数据:其他机组近期

SCADA 数据提取自风电场内其他机组近期 60 d 内的 SCADA 数据(不包含当天数据)。

3 类样本均提取自机组处于正常运行时段的数据,训练、测试样本比例设置为 6:4。

在特定情况下风电机组通过桨距角调节对其输出功率进行限制,图 3 为 2 个时段内机组桨距角、输出功率与风速的关系,由图可见,5 月 1 日至 5 月 30 日风电机组的功率曲线正常,而 6 月 1 日至 6 月 5 日风电机组的输出功率由于桨距角增大而削减。因此,在特定风速下输出功率的削减是由变桨系统控制造成的,与风电机组的异常状态无关,因此样本数据中需删除限功率阶段的 SCADA 数据。在进行模型训练前还需对训练样本进行平均处理、归一化处理与无效数据剔除。对于设备温度参数采用 10 min 平均处理,对于变化较快的输出功率与转子转速采用 1 min 平均处理。

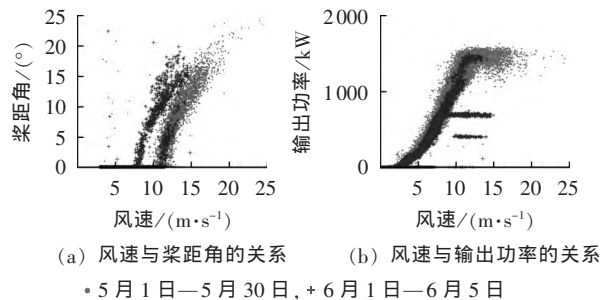


图 3 机组桨距角、输出功率与风速的关系
Fig.3 Relationship between wind turbine pitch angle or output power and wind speed

2.3 训练算法

采用反向传播神经网络 BPNN(Back-Propagation Neural Network)训练预测模型。BPNN 模型包含 1 个输入层、1 个隐层和 1 个输出层,隐层和输出层的传递函数分别采用双极性 S 型函数和线性函数。隐节点个数取为 2~10 个,在随机初始权重下对每个 BPNN 模型训练测试 10 次并获得其平均绝对误差(MAE)的平均值,确定 BPNN 模型的最优隐节点个数。图 4 为 BPNN 模型预测精度与隐节点个数的关系,以图 4 为例,4 个状态参数 BPNN 模型的最优隐节点个数分别为 6、10、7、6。

3 训练样本数据类型对预测模型预测精度的影响规律

3.1 近期数据模型精度分析

采用 MAE、均方根误差(MSE)与平均相对误差(MAPE) 3 类指标评估预测模型的预测精度。表 3 为采用 9 号风电机组近期样本数据建立的 14 个设备温度参数预测模型的预测精度,可见对于不同温度参数,预测精度存在较大差异,但是 MAE 的最大

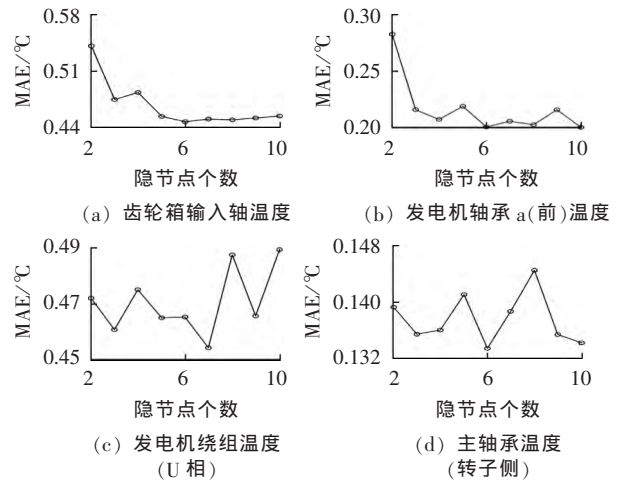


图 4 BPNN 模型预测精度与隐节点个数的关系
Fig.4 Relationship between BPNN prediction accuracy and hidden node quantity

表 3 风机温度参数的预测精度
Table 3 Prediction accuracies of WT temperature parameters

状态参数	评估指标		
	MSE/°C	MAE/°C	MAPE/%
齿轮箱输入轴温度	0.55	0.38	1.03
齿轮箱输出轴温度	0.23	0.17	0.35
齿轮箱油温	1.06	0.78	1.55
齿轮箱冷却水温度	1.34	0.85	1.71
主轴承温度(转子侧)	0.16	0.13	0.27
主轴承温度(齿轮箱侧)	0.16	0.14	0.29
发电机绕组温度(U 相)	1.93	0.72	1.23
发电机绕组温度(V 相)	1.92	0.71	1.22
发电机绕组温度(W 相)	1.94	0.72	1.25
发电机轴承 a(前)温度	0.41	0.17	0.43
发电机轴承 b(后)温度	0.45	0.21	0.45
发电机冷却风温度	1.52	0.88	1.92
控制柜温度	0.15	0.12	0.46
变频控制器温度	0.71	0.42	1.15

值小于 1 °C。表 4 为采用 9 号风电机组近期样本数据建立的输出功率与转子转速的预测模型的预测精度,可见本文建立的预测模型具有更高的精度。

表 4 输出功率与转子转速的预测精度
Table 4 Prediction accuracies of output power and rotor speed

状态参数	评估指标		
	MSE	MAE	MAPE/%
输出功率	68.62 kW	39.12 kW	11.52
转子转速	0.37 r/min	0.24 r/min	1.96

3.2 样本数据类型对预测模型预测精度的影响

以 17 号风电机组发电机轴承 a(前)温度的预测结果为例对比分析采用 2.2 节中 3 类训练样本建立的预测模型的预测精度。采用盒须图对比 3 类训练样本的数据分布,图 5 为 17 号风电机组近期样本(样本 1)与历史样本(样本 2)中风速、环境温度和发电机组轴承 a(前)温度的盒须图,由图可见,2 类样

本的取值分布相似。图 6 为 14—23 号风电机组发电机轴承 a(前)温度的盒须图,由图可见,10 台风电机组的样本分布存在很大差异。

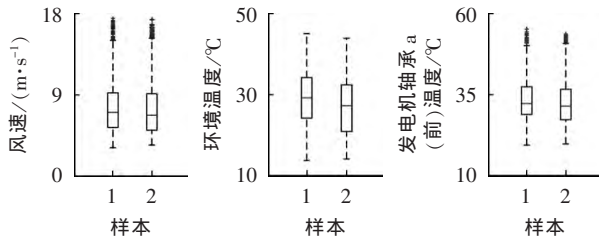


图 5 17 号风电机组近期样本与历史样本的盒须图

Fig.5 Boxplots of resent and historical data samples of WT 17

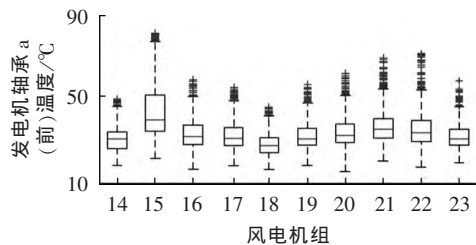


图 6 10 台风电机组发电机轴承 a(前)温度的盒须图

Fig.6 Boxplot of generator bearing a(front) temperatures of ten WTs

采用 3 类样本分别建立 17 号风电机组发电机轴承 a(前)温度的预测模型,预测精度如表 5 所示。由表 5 可见:基于本机组近期样本建立的预测模型具有最高的预测精度;基于本机组历史数据建立的预测模型的预测精度明显低于前者,原因可能是随着机组设备的老化,状态参数间的关联关系发生了微小变化,从而降低了历史数据模型的精度;以其他风电机组(15、16、18、19 号风电机组)近期数据作为训练样本建模,模型的预测精度存在较大的分散性;由于 15 号与 18 号风电机组的发电机轴承 a(前)温度分布与 17 号风电机组差异过大(如图 6 所示),因此以这 2 台风电机组近期数据为训练样本的模型预测精度较差。

表 5 基于 3 类样本建立的预测模型预测发电机轴承 a(前)温度的预测精度

Table 5 Prediction accuracies of generator bearing a (front) temperature, based on three sample types

类别	训练样本	评估指标		
		MSE/°C	MAE/°C	MAPE/%
1	17 号机组近期数据	0.36	0.21	0.61
2	17 号机组历史数据	0.60	0.45	1.23
3	15 号机组近期数据	2.51	2.22	5.76
	16 号机组近期数据	0.45	0.34	0.95
	18 号机组近期数据	2.66	1.26	3.22
	19 号机组近期数据	0.43	0.31	0.88

为进一步分析目标参数分布对预测模型预测精度的影响,本文定义样本相似度指标量化 2 台风电机组的目标参数分布的相似程度。获取训练样本对应

时段内 2 台风电机组的目标参数时间序列,则样本相似度的计算方法为:

$$\lambda_{SSD} = \frac{\sum_{n=1}^N h_1(n)h_2(n)}{\sqrt{\sum_{n=1}^N h_1^2(n) \sum_{n=1}^N h_2^2(n)}} \quad (1)$$

其中, λ_{SSD} 为样本相似度; h_1 和 h_2 为 2 台风电机组目标参数分布的直方图; N 为直方图的区间个数。 λ_{SSD} 的取值在 0~1 范围内, λ_{SSD} 越大表明 2 台风电机组的目标参数分布的相似度越高。

选取 21—23 号风电机组为目标机组,采用 12—20 号风电机组近期数据分别建立目标机组发电机轴承 a(前)温度的预测模型,计算目标机组与 12—20 号风电机组发电机轴承 a(前)温度的 λ_{SSD} 指标。图 7 为以 12—20 号风电机组近期数据为训练样本建立的 21 号风电机组发电机轴承 a(前)温度模型的预测误差直方图,由图可见目标参数分布的 λ_{SSD} 与模型的预测精度密切相关。图 8 为 λ_{SSD} 与基于 12—20 号风电机组近期样本建立的预测模型的 MAE 的关系,可见随着 λ_{SSD} 的升高,预测模型的 MAE 呈减小趋势。

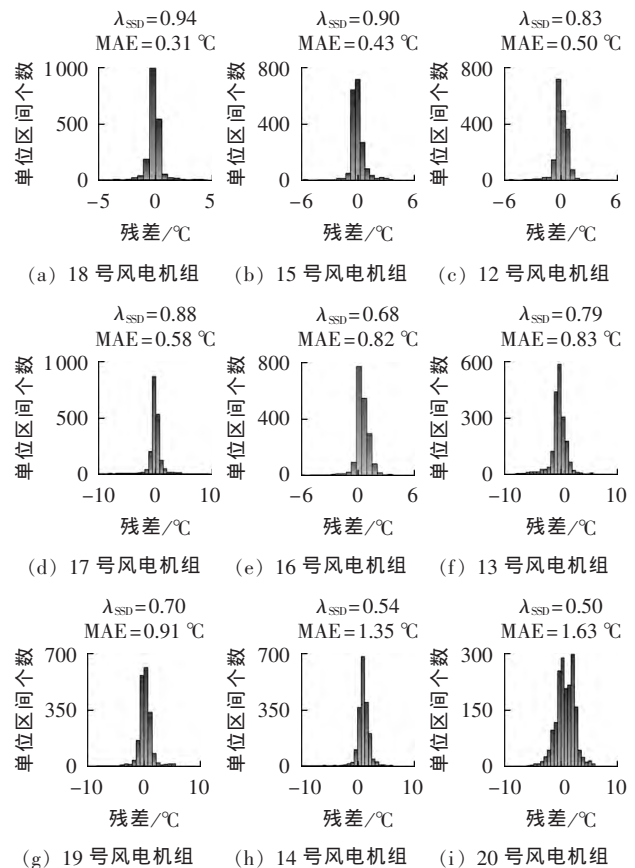


图 7 基于其他机组近期数据建立的预测模型预测 21 号风电机组发电机轴承 a(前)温度模型的误差分布

Fig.7 Error distribution of prediction model for generator bearing a(front) temperature of WT 21, based on resent data of other WTs

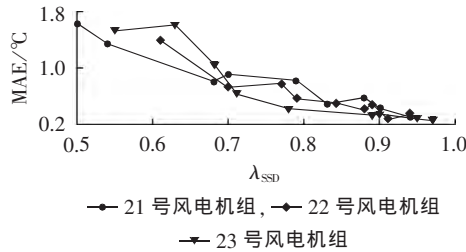


图 8 SSD 指标与预测模型 MAE 误差的关系
Fig.8 Relationship between SSD and MAE of prediction model

4 预测模型的选择方法

由 3.2 节的分析结果可知,由于设备的老化和风电机组的差异性,基于本机组历史数据和其他机组近期数据建立的预测模型的精度低于以本机组近期数据为训练样本建立的模型。预测模型对机组正常情况下状态参数的准确预测是状态参数异常辨识的前提,因此,需对以本机组历史数据和其他机组近期数据为训练样本建立的模型进行选择。采用 MAE 对以历史数据为样本建立的模型进行筛选,如果模型的 MAE 符合式(2)所示准则,则选择相应的模型。

$$E_{MAE_i} \leq w E_{MAE}^{ref} \quad (2)$$

其中, E_{MAE_i} 为正常模型 i 对测试样本的 MAE; E_{MAE}^{ref} 为以本机组近期样本数据建立的预测模型对测试样本的 MAE; w 为权重系数,本文选取为 2。

为提高选择效率,本文提出了以其他机组近期数据为训练样本建立的预测模型的快速选择方法。图 9 为预测模型选择流程,主要包含风电机组分类、目标参数分布相似度计算与预测模型选择。在风电机组分类中,提取训练样本所属时段内所有机组的状态参数时间序列,计算各风电机组状态参数的均值和方差,采用 k -means 算法对风电机组进行分类。图 10 为基于发电机轴承 a(前)温度的风电机组聚类结果,可见在同分类下状态参数的取值分布相似。

5 状态参数异常程度量化

以表 5 中基于本机组近期数据、本机组历史数据与其他机组(16、19 号风电机组)近期数据建立的预测模型(分别简称为模型 1、2、3、4)的测试结果为例,基于核密度估计法得到预测残差的概率密度函数^[18],如图 11 所示。根据基于统计的异常辨识理论^[6],对预测残差的取值范围进行区间划分,以分位数 0.025、0.25、0.75、0.975 对应的残差值,将预测残差划分为 3 个区间,如图 12 所示。

采取通过滑动残差窗口采集 24 h 内的状态参数预测残差序列(滑动步长为 2 h),统计预测残差落入 3 个区间内的个数,定义状态参数的异常程度指标表征残差的异常程度,计算公式为:

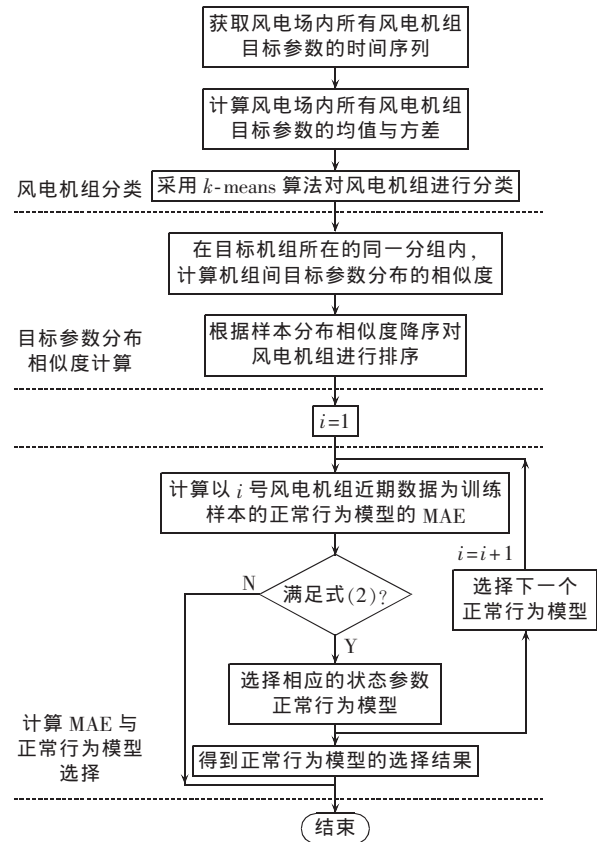


图 9 基于其他机组近期样本的预测模型的选择流程

Fig.9 Flowchart of selecting prediction models based on recent samples of other WTs

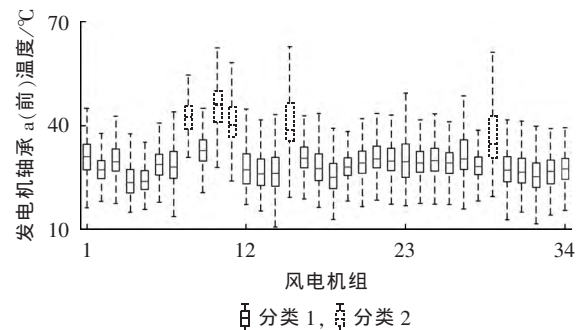


图 10 基于发电机轴承 a(前)温度参数的风电机组聚类结果

Fig.10 Results of WT clustering based on generator bearing a(front) temperature

$$\lambda_{ALL} = 1 - \frac{N_1 C_1}{\sum_{k=1}^3 N_k C_k} \quad (3)$$

其中, λ_{ALL} 为状态参数预测残差的异常程度指标; N_k 为残差序列中取值处于区间 k 内的数目; C_k 为残差序列中取值处于区间 k 内的惩罚因子, C_1 、 C_2 、 C_3 分别设为 1、3、5。 λ_{ALL} 越大,表明状态参数的异常程度越高。

6 采用模糊综合评判的状态参数异常辨识

由于样本数据的差异性,基于不同样本数据的

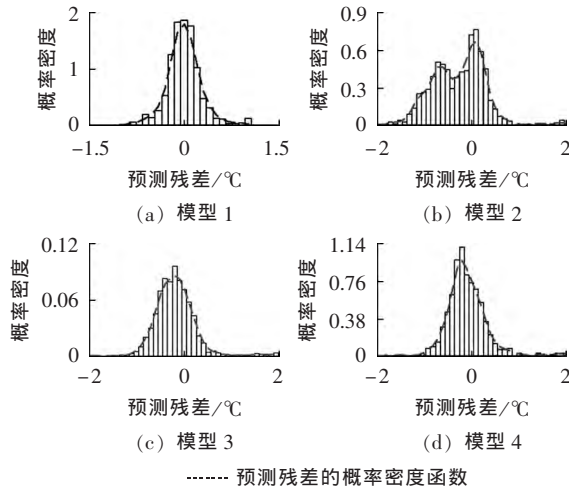


图 11 表 5 中预测模型预测残差的概率密度函数

Fig.11 PDFs of residual error of prediction models in Table 5

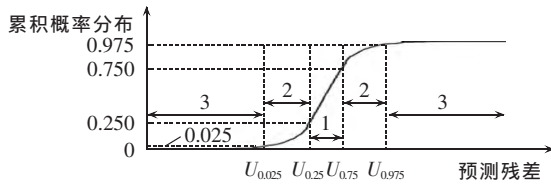


图 12 状态参数预测残差的区间划分

Fig.12 Divisions of residual error of status parameter prediction

预测模型的异常辨识能力存在很大不同,因此根据各预测模型异常辨识的精度,采用模糊综合评判方法融合各预测模型的异常辨识结果。

6.1 模糊矩阵与隶属函数

模糊矩阵作为模糊综合评判的重要组成部分,用于表征评估指标与评估结果间的相互联系。本文选取各预测模型的异常程度指标作为指标层的评价指标,将机组状态参数异常状态的评估结果划分为 3 个等级: $L = \{l_1, l_2, l_3\} = \{\text{正常}, \text{注意}, \text{异常}\}$ 。根据状态参数预测模型的选择结果,将被选择的预测模型预测残差的异常程度指标代入其模糊隶属函数即可得到模糊评判矩阵 V :

$$V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} & v_{n3} \end{bmatrix} \quad (4)$$

其中, $v_{ij} (j=1, 2, 3)$ 为第 i 个预测模型异常程度指标的隶属度。

本文采用三角形和半梯形组合的分布隶属度函数,以状态参数预测残差的异常程度指标为基础,隶属于机组 3 个异常状态评估等级的隶属度可分别按式(5)、(6)、(7)计算。

$$v_{i1} = \begin{cases} 1 & x_i < 0.4 \\ 3.67 - 6.67x_i & 0.4 \leq x_i < 0.55 \\ 0 & x_i \geq 0.55 \end{cases} \quad (5)$$

$$v_{i2} = \begin{cases} 0 & x_i < 0.4 \text{ 或 } x_i \geq 0.7 \\ 6.67x_i - 2.67 & 0.4 \leq x_i < 0.55 \\ 4.67 - 6.67x_i & 0.55 \leq x_i < 0.7 \end{cases} \quad (6)$$

$$v_{i3} = \begin{cases} 0 & x_i < 0.55 \\ 6.67x_i - 3.67 & 0.55 \leq x_i < 0.7 \\ 0 & x_i \geq 0.7 \end{cases} \quad (7)$$

其中, x_i 为第 i 个预测模型预测残差的异常程度指标。

状态参数的预测精度直接影响异常辨识的准确性,因此以各状态参数预测模型对测试样本的预测精度计算各指标的权重。

$$\omega_i = \left(1 - r_i / \sum_{i=1}^M r_i \right) / \sum_{i=1}^M \left(1 - r_i / \sum_{i=1}^M r_i \right) = \left(1 - r_i / \sum_{i=1}^M r_i \right) / (M-1) \quad (8)$$

其中, r_i 为第 i 个预测模型的 MAE; M 为被选择的预测模型的个数。

6.2 异常辨识流程

图 13 为采用模糊综合评价的风电机组状态参数异常辨识流程。

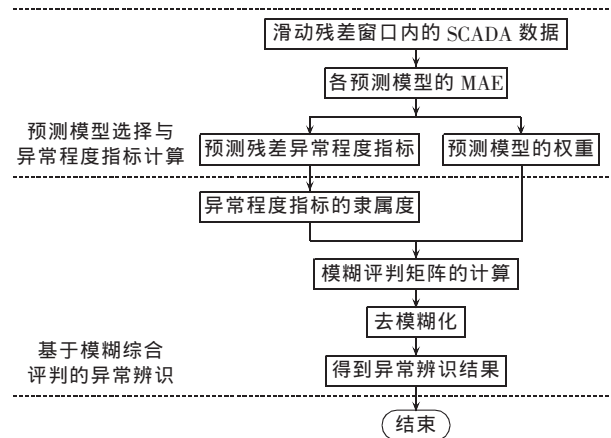


图 13 风电机组状态参数异常辨识流程

Fig.13 Flow chart of wind turbine status parameter anomaly detection

风电机组状态参数异常辨识流程如下。

a. 预测模型建立与异常程度指标计算: 分别选择本机组近期数据、本机组历史数据和其他机组近期数据为 BPNN 模型的训练样本,建立状态参数的预测模型;基于测试数据获得各预测模型的 MAE,根据 4 节内容对预测模型进行选择;根据式(8)获得各预测模型的权重;根据式(3)计算各预测模型预测残差的异常程度指标。

b. 模糊综合评判: 根据式(5)—(7)求解各异常程度指标隶属于不同异常状态等级的隶属度;根据式(9)获得输出层的评估值。

$$B = [b_1 \ b_2 \ b_3] = [\omega_1 \ \omega_2 \ \cdots \ \omega_n] \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} & v_{n3} \end{bmatrix} \quad (9)$$

根据最大隶属度原则,将最大评估值 $b_{\max} = \max(b_i)$ ($i=1, 2, 3$) 相对应的状态参数异常状态等级 l_i 作为异常辨识广义模型的最终辨识结果。

7 实例分析

本节通过 2 个实例验证提出的异常辨识方法的有效性。实例 1 中,将本文提出的异常辨识方法与传统的阈值诊断方法比较,通常阈值的大小取决于预测误差的置信区间,本文选取 97.5% 与 99% 置信区间对应的预测残差值作为阈值。实例 2 中,与基于单一模型的异常辨识方法对比。

7.1 实例 1

24 号风电机组由于齿轮箱输出轴过热故障于 2011 年 6 月 25 日停机。根据风电场的维修记录,该机组在此前的 2 个月内均处于运行状态。

本实例主要阐述齿轮箱输出轴温度的异常辨识过程。由于该风场 SCADA 系统的投运时间为 2011 年 2 月 15 日,因此不考虑基于本机历史数据的预测模型。根据 4 节中的模型选择方法,仅有基于本机近期数据的预测模型可用于状态参数的异常辨识。根据模型测试结果,预测误差 97% 置信区间的边界值为 $\pm 0.6^\circ\text{C}$,预测误差 99% 置信区间的边界值为 $\pm 1^\circ\text{C}$ 。

图 14 为 2011 年 6 月 18 日至 6 月 21 日间齿轮箱输出轴温度的预测结果。计算得到齿轮箱输出轴温度在 2011 年 6 月 21 日 00:00—24:00 间的 λ_{AI} 为 0.91,因此该状态参数处于异常状态。

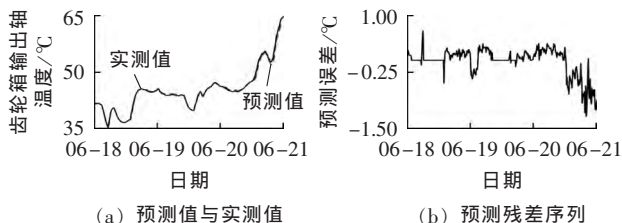


图 14 2011 年 6 月 18 日至 6 月 21 日 24 号机组齿轮箱输出轴温度的预测结果

Fig.14 Results of gearbox output shaft temperature prediction, from June 18 to June 21, 2011

由图 14(b)可知采用传统的阈值诊断方法也能辨识出该状态参数的异常状态,但是可能会导致机组状态的误诊断。图 15(a)为 2011 年 5 月 25 日至 6 月 25 日 24 号机组齿轮箱输出轴温度的预测残差序列,由图可见在 6 月 4 日预测残差出现了显著增大。图 15(b)为 2011 年 6 月 4 日的风速时间序列,

可见风速在短小时内迅速升高,在 05:30—06:00 之间风速的变化为 12.3 m/s 。图 15(d)为 2011 年 6 月 4 日 03:00—09:00 间齿轮箱输出轴温度的预测残差序列,预测结果表明风速在短小时内的快速升高是造成预测残差显著增大的根本原因。

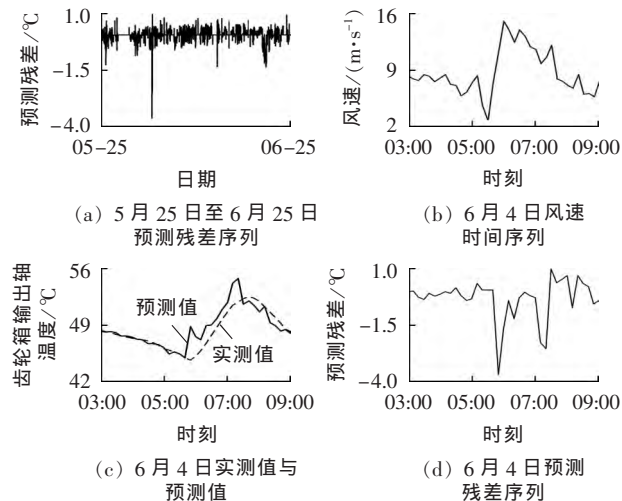


图 15 短时风速波动对状态参数预测残差的影响

Fig.15 Impact of short-term wind-speed fluctuation on residual error of status parameter prediction

图 16 为在预测模型训练样本所属时段风速在 30 min 内的变化情况,由图可见 30 min 内风速的波动值大多分布在 $-5 \sim 5 \text{ m/s}$ 范围内,由于缺乏训练数据,预测模型无法实现风速短时快速变化下的准确预测。由上述分析可知,预测残差的突然增大不能反映齿轮箱输出轴温度的异常状态,传统的阈值诊断方法产生了误判。根据本文提出的异常辨识方法,计算得到 2011 年 5 月 25 日至 2011 年 6 月 25 日异常程度指标的时间序列,如图 17 所示,由图可见在 6 月 4 日异常程度指标幅值并未出现明显的升高。2011 年 6 月 4 日 01:00 至 2011 年 6 月 5 日 01:00 间齿轮箱输出轴温度的异常程度指标为 0.42,参数状

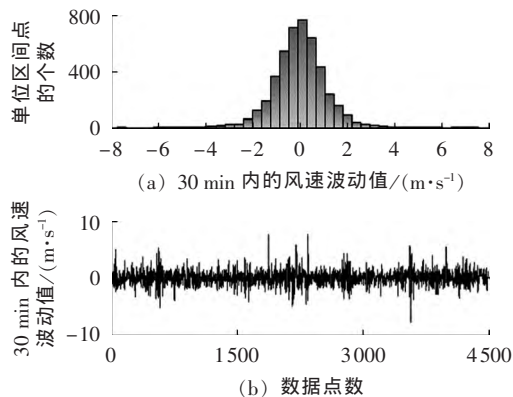


图 16 训练数据所在时段风速在 30 min 内的变化情况

Fig.16 Wind-speed variation for 30 min within training data period

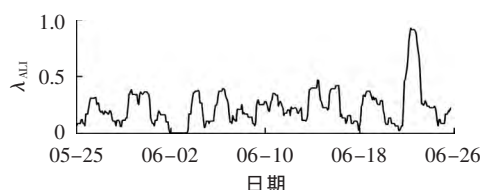


图 17 2011 年 5 月 25 日至 2011 年 6 月 25 日间 24 号
风电机组齿轮箱输出轴温度预测残差的 ALI 指标
Fig.17 ALI values of residual error of gearbox
output shaft temperature prediction,
from May 25 to June 25, 2011

态判断为正常。因此,相比于传统的阈值诊断方法,本文提出的异常辨识方法具有更高的准确度。

7.2 实例 2

22 号风电机组由于发电机碳刷氧化故障于 2012 年 6 月 11 日停机,停机时间为 72 h。本实例主要阐述发电机轴承 a(前)温度的异常辨识过程。根据预测模型的选择结果,以本机组近期数据、本机组历史数据、20 号机组近期数据和 25 号机组近期数据为训练样本的模型可用于状态参数的异常辨识,4 个模型的误差如表 6 所示。

表 6 基于 3 类样本建立的预测模型预测
发电机轴承 a(前)温度的预测精度

Table 6 Prediction accuracies of generator bearing a(front)
temperature, based on three sample types

模型	训练样本	评估指标		
		MSE/°C	MAE/°C	MAPE/%
a	本机组近期数据	0.41	0.17	0.43
b	本机组历史数据	0.41	0.18	0.47
c	20 号风电机组近期数据	0.50	0.29	0.71
d	25 号风电机组近期数据	0.53	0.33	0.82

图 18 为 2012 年 6 月 7 日至 2012 年 6 月 10 日间发电机轴承 a(前)温度的预测残差序列,可见 4 个模型的预测残差均出现异常。但以本机组近期数据

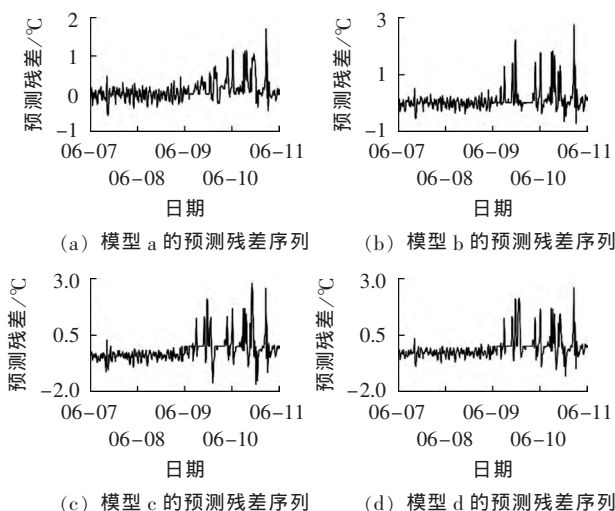


图 18 22 号风电机组发电机轴承 a(前)
温度的预测残差序列

Fig.18 Sequences of residual error of gearbox output
shaft temperature prediction of WT 22

为训练样本的模型的预测残差的幅值变化幅度明显小于其他 3 个模型。可以推断,在本机组近期数据模型的训练样本中,可能已经掺杂了机组异常状态的数据,导致状态参数的异常辨识灵敏度下降。

获得 4 类预测模型预测残差的累积分布函数,根据第 5 节方法将预测残差划分为 3 个离散区间。根据式(3)得出各模型预测残差的 λ_{ALI} , 计算结果如表 7 所示。基于式(5)~(7)计算各预测模型预测残差隶属于 3 种机组异常状态的隶属度,则机组状态参数异常的模糊评价矩阵为:

$$V = \begin{bmatrix} 0 & 0.87 & 0.13 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

表 7 各预测模型的异常程度指标

Table 7 ALI values of different prediction models

预测残差区间	异常辨识指标			
	模型 a	模型 b	模型 c	模型 d
区间 1	110	80	25	23
区间 2	12	31	76	81
区间 3	22	33	43	40
λ_{ALI}	0.57	0.76	0.95	0.95

根据表 6 中所示预测模型对测试样本的 MAE,由式(8)计算各预测模型的权重,计算结果为[0.28, 0.27, 0.23, 0.22]。根据式(9)可得异常辨识结果为[0, 0.24, 0.76]。根据最大隶属度原则可知该状态参数处于异常状态。而若仅采用以本机组近期数据为训练样本模型的预测残差计算异常程度指标,发电机轴承 a(前)温度的异常辨识结果为“注意”。实例分析表明:基于多预测模型的的风电机组状态参数异常辨识方法具有更高的准确性。

8 结论

a. 本文提出了基于预测模型与模糊综合理论的风电机组状态参数异常辨识方法。基于 3 类样本数据建立了风电机组状态参数的预测模型;采用预测模型对测试样本的预测精度量化其异常辨识的有效性,提出了预测模型的选择方法;基于预测残差统计分布模型,提出利用预测残差异常程度指标描述残差序列的异常程度。采用模糊综合评判方法融合基于不同样本数据建立的预测模型的异常辨识结果。

b. 通过 2 个实例对辨识方法进行了验证,分析结果表明:短时风速波动对预测模型预测残差具有较大影响,本文通过建立状态参数的异常程度指标,所提出的异常辨识方法比传统的阈值方法更能准确有效地反映参数的异常;当机组处于异常状态时,基于本机组近期数据建立的预测模型的异常辨识灵敏度明显降低,综合考虑多个预测模型异常辨识结果

可提高状态参数异常辨识的准确性。

参考文献:

- [1] 李辉,胡姚刚,李洋,等. 大功率并网风电机组状态监测与故障诊断研究综述[J]. 电力自动化设备,2016,36(1):6-16.
LI Hui, HU Yaogang, LI Yang, et al. Overview of condition monitoring and fault diagnosis for grid connected high power wind turbine unit[J]. Electric Power Automation Equipment, 2016, 36(1):6-16.
- [2] KUSIAK A, ZHANG Z, VERMA A. Prediction, operations, and condition monitoring in wind energy[J]. Energy, 2013, 60(7):1-12.
- [3] YANG W, COURT R, JIANG J. Wind turbine condition monitoring by the approach of SCADA data analysis[J]. Renewable Energy, 2013, 53(9):365-376.
- [4] LAPIRA E, BRISSET D, ARDAKANI H D, et al. Wind turbine performance assessment using multi-regime modeling approach[J]. Renewable Energy, 2012, 45(3):86-95.
- [5] KUSIAK A, VERMA A. Monitoring wind farms with performance curves[J]. IEEE Trans on Sustainable Energy, 2013, 4(1):192-199.
- [6] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey[J]. ACM Computing Surveys, 2009, 41(3):1-58.
- [7] SCHLECHTINGEN M, SANTOS I F, ACHICHE S. Using data-mining approaches for wind turbine power curve monitoring: a comparative study[J]. IEEE Trans on Sustainable Energy, 2013, 4(99):671-679.
- [8] SCHLECHTINGEN M, SANTOS I F, ACHICHE S. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description[J]. Applied Soft Computing, 2013, 13(1):259-270.
- [9] ATA R. Artificial neural networks applications in wind energy systems: a review[J]. Renewable & Sustainable Energy Reviews, 2015, 49:534-562.
- [10] THIAE L, FALL S S, KASSE M, et al. A neural network based approach for wind resource and wind generators production

assessment[J]. Applied Energy, 2010, 87(5):1744-1748.

- [11] LIU Y, SHI J, YANG Y, et al. Short-term wind-power prediction based on wavelet transform-support vector machine and statistic-characteristics analysis[J]. IEEE Trans on Industry Applications, 2012, 48(48):1136-1141.
- [12] CHEN B, PETER C M, PETER J T. Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition[J]. IET Renewable Power Generation, 2015, 9(5):503-513.
- [13] GUO P, INFELD D, YANG X. Wind turbine generator condition-monitoring using temperature trend analysis[J]. IEEE Trans on Sustainable Energy, 2012, 3(1):124-133.
- [14] SCHLECHTINGEN M, SANTOS I F. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection[J]. Mechanical Systems & Signal Processing, 2011, 25(5):1849-1875.
- [15] KUSIAK A, LI W. Virtual models for prediction of wind turbine parameters[J]. IEEE Trans on Energy Conversion, 2010, 25(1):245-252.
- [16] KUSIAK A, VERMA A. Analyzing bearing faults in wind turbines: a data-mining approach[J]. Renewable Energy, 2012, 48(6):110-116.
- [17] SPERA D A. Wind turbine technology: fundamental concepts of wind turbine engineering[M]. New York, USA: ASME, 1994:67-69.
- [18] SILVERMAN B W. Density estimation for statistics and data analysis[M]. New York, USA: Chapman and Hall, 1986:152-153.

作者简介:



孙 鹏

孙 鹏(1989—),男,河南新乡人,博士,主要研究方向为风电机组状态监测与故障诊断(E-mail:sunpeng@cqu.edu.cn);

李 剑(1971—),男,陕西宝鸡人,教授,博士,主要研究方向为电力设备状态监测与新型绝缘材料(E-mail:lijian@cqu.edu.cn)。

Wind turbine status parameter anomaly detection based on prediction models and fuzzy theory

SUN Peng¹, LI Jian², KOU Xiaokuo¹, LÜ Zhongbin¹, YAO Degui¹, WANG Ji¹, WANG Leilei¹, TENG Weijun¹

(1. State Grid Henan Electrical Power Research Institute, Zhengzhou 450000, China; 2. State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing 400030, China)

Abstract: In order to improve the ability of early WT(Wind Turbine) outage warning, a method for the anomaly detection of WT status parameters is proposed based on the SCADA(Supervisory Control And Data Acquisition) system data of wind farm. The parameters are grouped and the neural networks are applied to develop the prediction models for the environmentally-sensitive status parameters. The resent and historical SCADA data samples of this WT and the resent SCADA data samples of other WTs are used as the training data and the prediction accuracies of the prediction models based on the three sample types are compared. The MAE(Mean Absolute Error) is used to select the prediction models trained by the historical SCADA data sample of this WT and the resent SCADA data samples of other WTs. An anomaly index is proposed to quantify the anomaly level of the residual error of status parameter prediction. In order to improve the accuracy of anomaly detection, the fuzzy synthetic evaluation is applied to integrate the results of anomaly detection by the selected prediction models. The proposed anomaly detection method is applied to a domestic 1.5 MW WT and the results show that it has higher accuracy than the traditional methods.

Key words: wind turbine; wind farm SCADA system; prediction models; fuzzy synthetic evaluation; anomaly detection