

DOI: 10.19650/j.cnki.cjsi.J1905178

基于动态特征矩阵的 k 近邻风电机组故障检测方法^{*}

钱小毅, 张宇献

(沈阳工业大学电气工程学院 沈阳 110870)

摘要:受风的间歇性和随机性影响风电机组运行状态频繁切换,导致设备状态异常检测误报和漏报情况严重,风电企业运维成本居高不下。为此,提出了基于动态特征矩阵的 k 近邻故障检测方法,该方法采用基于互信息的动态特征矩阵描述风电机组的动态特性,通过加权 k 近邻同时考虑动态特征矩阵中的特征贡献率与累计互信息的影响,利用动态阈值计算降低运行状态突变造成的误报。分别以美国可再生能源实验室 5 MW 海上风机基准模型的常见传感器和执行器故障以及 SCADA 数据中记录的变桨系统故障为例,将所提方法的故障检测结果分别与 PCA、KPCA、FD-kNN 以及 PC-kNN 故障检测方法进行对比,结果表明所提方法能够准确进行故障信息的检测,所提方法优于其他对比故障检测方法。

关键词: 风力发电机;故障检测;动态特征矩阵;加权 k 近邻;动态阈值

中图分类号: TP277 TM614 **文献标识码:** A **国家标准学科分类代码:** 470.40

Fault detection of wind turbines using k -nearest neighbor based on dynamic feature matrix

Qian Xiaoyi, Zhang Yuxian

(School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, China)

Abstract: The intermittency and randomness of wind make the operation state of wind turbine change frequently. As a result, the false positive ration and false negative rate in anomaly detection of equipment are serious. The costs of operation and maintenance in the wind power industry are high. To solve this problem, one kind of K-nearest neighbor fault detection method based on dynamic feature matrix is proposed in this work. It constructs a dynamic feature matrix based on mutual information to describe the dynamic characteristics of wind turbine. The weighted k-nearest neighbor fault detection method is introduced to address the influence of the characteristic contribution and cumulative mutual information in dynamic feature matrix. The dynamic threshold can help reduce false alarm caused by the sudden change of operation state. This paper takes examples of the common sensor faults and actuator faults in the 5MW offshore benchmark of National Renewable Energy Laboratory and the pitch system faults in SCADA system. The fault detection results of the proposed method are compared with PCA, KPCA, FD-kNN and PC-kNN, respectively. Experimental results demonstrate that the proposed method can accurately detect the fault information. Compared with other methods, it can achieve better fault detection results.

Keywords: wind turbine; fault detection; dynamic feature matrix; weighted-k-nearest neighbor; dynamic threshold

0 引言

能源短缺与日益严重的环境问题,使可再生能源成为各界研究的热点。而风力发电作为绿色的可再生能源,正在全球范围内飞速发展^[1-2]。其中,海上风场所占比例逐年提高,以受益于海面上更强、更均匀的风况。然

而,由于海上风机工作环境恶劣,工作状态的频繁转换,使各部件故障频发,运行维护的费用达到总费用的 15~35%^[3-4]。因此,为使检修人员能够对风力发电机进行及时的检修维护,进而降低停机时间与维修费用,对异常状态及时、准确的检测成为当前亟待解决的问题。

风电机组运行环境恶劣,且受机械、电气等因素影响,导致其具有复杂、多变的运行模式^[5]。因此,基于单

收稿日期:2019-05-21 Received Date:2019-05-21

^{*} 基金项目:国家自然科学基金(61102124)、辽宁省自然科学基金资助项目(20180551032)资助

一模型的故障检测方法,难以全面描述风机运行状态,进而影响对异常状态的识别效果^[6]。为引入“分而治之”的思想,出现了多种基于工况划分,建立分散式模型的故障检测方法。在基于数据驱动的故障检测方法中,相比于根据单特征或多特征,通过网格划分工况的方法^[7-9],对多特征进行聚类分析能够从数据角度实现更合理的工况划分。例如,在文献[10]通过对风力发电机运行状态中的风速、风向、扭矩等特征变量进行 k 均值(k -means)聚类,将运行数据划分为4个工况,随后,在各工况的子空间中分别建立高斯模型,进而提高了检测模型精度。在文献[11]提出粒子群寻优的核主元分析(kernel principal component analysis, KPCA),用于特征提取,之后对特征空间利用模糊C-均值(fuzzy C-means, FCM)聚类算法划分工况。然而,在基于聚类算法的工况划分中,工况个数往往依赖于专家经验,评价过程中各子模型贡献的取舍会对检测结果造成较大影响。此外,随着数据复杂程度的增加,工况划分的个数也随之增多,计算量呈指数增长,影响检测效率。

基于 k 近邻的故障检测方法(fault detection based on k -nearest neighbor, FD- k NN)最早被应用于处理非线性、多模态过程的故障检测问题^[12]。一方面 k NN算法是非线性分类器,FD- k NN能够更好地处理非线性问题。另一方面,FD- k NN对在线数据的异常识别仅依赖在训练数据中与其相似的 k 个运行状态,在避免分散式建模的同时,降低了多工作模式对检测过程的影响。为提高FD- k NN的效率与精度,文献[13]提出了一种基于主元的 k NN规则(principal component-based k nearest neighbor rule, PC- k NN),其将FD- k NN执行于主元空间,提高了故障检测的效率。文献[14]提出了一种基于 k NN主元得分差分的故障检测策略(fault detection strategy based on principal component score difference of k nearest neighbors, k Diff-PCA),该方法通过主元分析(principal component analysis, PCA)方法计算真实得分,应用 k NN均值计算估计得分,通过上述两种得分计算样本的得分差分矩阵和残差矩阵,进而消除数据结构对过程故障检测的影响。然而,在现存的基于 k 近邻故障检测方法中,忽略了不同特征对评价指标的贡献差异,此外,目前仍缺乏 k NN故障检测方法处理数据动态特性以及动态阈值方面的研究。

本文考虑风电机组运行过程的动态特性以及多变的工作状态对故障检测的影响,提出一种基于动态特征矩阵的加权 k NN故障检测(weighted k -nearest neighbor fault detection based on dynamic feature matrix, DFM- Wk NN)方法。在DFM- Wk NN中,首先,通过引入延时特征得到增广特征矩阵,并通过特征间的互信息度量排序获取动态特征矩阵。之后,综合考虑特征贡献率与互信息累计

度量对矩阵中的特征加权,形成加权 k NN故障检测方法。最后,以固定阈值与近邻样本的距离的均值获取在线动态阈值。分别利用5 MW benchmark模型中的10种传感器故障与执行器故障以及风电机组SCADA系统采集的变桨系统故障,验证所提算法的有效性。

1 基本方法

1.1 互信息

互信息(mutual information, MI)是变量间相互依赖性的度量,在多变量统计过程它常被用于测量2个特征间的相关性^[15]。2个离散变量的互信息可表示为:

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \lg \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

其中, X 和 Y 为2个离散型随机变量,其边缘概率分别为 $p(x)$ 和 $p(y)$,联合分布为 $p(x, y)$, $I(X, Y)$ 为 X 和 Y 的MI度量。

MI度量值随 X 和 Y 之间的相关性变化,相关性越大,则MI值越大,若 X 和 Y 之间不存在重叠的信息,MI值为0。对于求解过程中需要得各离散变量的密度分布概率,通常使用核密度估计方法来确定。

1.2 k 近邻故障检测策略

FD- k NN的对故障数据的主要辨识依据为:对于训练样本集,故障样本的近邻距离大于正常样本的近邻距离^[12]。近邻距离通过欧氏距离计算如下:

$$D^2 = \sum_{i=1}^k d_i^2 \quad (2)$$

其中, d_i 为样本 x 到第 i 个近邻的欧氏距离。根据训练样本集中的 D^2 覆盖率,以给定置信度 α 确定阈值 D_α^2 ^[12],例如,95%的置信度的表示阈值可覆盖训练集中95%的样本。

对于在线样本,若其到训练集中 k NN欧氏距离之和低于阈值则判定为正常状态,否则为故障状态。

2 DFM- Wk NN故障检测方法

本文提出的DFM- Wk NN方法对FD- k NN做出3方面改进,分别为基于互信息的动态特征矩阵转化,加权FD- k NN的实现,以及针对FD- k NN的动态阈值设定。

2.1 基于互信息的动态特征矩阵

采用基于MI排序策略,将每个时刻的样本拓展为动态特征矩阵,矩阵中的即时特征能够反映当前时刻的相关特征,延时特征反映系统运行的动态特性。同时,可降低不相关延时特征对检测的干扰影响。

首先,为降低 k NN距离的计算复杂度,进而提高故障检测效率,可根据实际应用需求与数据特点,选择合适

的无监督特征选择或特征提取方法,对原数据进行降维。

在本文的实验中,数据集通过 SPSS 软件的非参数校验,最小置信水平均大于 0.05,符合高斯分布。因此,本文采用 PCA 特征提取算法获取主元空间 $\mathbf{X}$,对 $\mathbf{X}$ 引入前 l 个时刻的主元特征,构成如下增广矩阵^[16]:

$$\mathbf{X}_a = \begin{bmatrix} \mathbf{x}_1(t-l) & \mathbf{x}_1(t-l+1) & \cdots & \mathbf{x}_1(t) \\ \mathbf{x}_2(t-l) & \mathbf{x}_2(t-l+1) & \cdots & \mathbf{x}_2(t) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n(t-l) & \mathbf{x}_n(t-l+1) & \cdots & \mathbf{x}_n(t) \end{bmatrix} \quad (3)$$

式中: $\mathbf{X}_a$ 为对 $\mathbf{X}$ 引入延时特征的增广矩阵; $\mathbf{x}_i(t)$ 表示第 i 个特征在 t 时刻之前的所有数值。

计算每个即时特征与所有延时特征之间的 MI,以衡量特征间的序列相关性,即:

$$C(i, j) = I(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

以 $n=3$ 为例,以特征 $\mathbf{x}_1$ 对所有备选特征之间的 MI 计算为例,解释对延时特征的互信息度量过程,如图 1 所示。

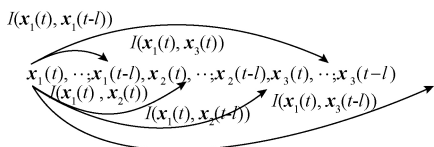


图 1 延时特征的互信息度量

Fig.1 Mutual information measurement of delay features

其中: $\mathbf{x}_i(t-j)$ 为训练集中第 i 个特征在 $t-j$ 时刻所有数值组成的延时特征向量; $I(\mathbf{x}_1(t), \mathbf{x}_i(t-j))$ 为 t 时刻特征 $\mathbf{x}_1$ 对 $t-j$ 时刻特征 $\mathbf{x}_i$ 的 MI 度量。

根据 $\mathbf{x}_1(t)$ 与其它特征的互信息排序,获取 s 个相关即时特征与延时特征,形成特征 $\mathbf{x}_1$ 的特征矩阵。不重复的保留各特征矩阵中的延时特征与即时特征,作为动态特征矩阵。为便于 k NN 故障检测中的欧氏距离运算,进一步将动态特征矩阵重构为向量形式。图 2 所示为以延时变量个数 $l=2$,相关特征个数 $s=3$ 为例,说明动态特征矩阵的形成过程。

2.2 基于特征贡献率与累计互信息的加权 k 近邻故障检测

在 FD- k NN 中, D^2 为在线样本到离线 k NN 的欧氏距离之和,所有特征被“公平的对待”。然而,在实际应用中,无论是即时特征还是延时特征,对模型的重要性皆存在差异性。本文同时考虑即时特征与延时特征的重要程度,提出一种对 FD- k NN 的加权策略。

首先,在采用 PCA 对离线样本进行特征提取时,记录各主元特征的贡献率 $W_p(i, j)$ $i=1, 2, \dots, n, j=1, 2, \dots, w$ 。其中 n 为主元特征维数, w 为动态特征向量维数。对各时刻的同一主元特征,其 $W_p(i, j)$ 值均为 PCA 获得

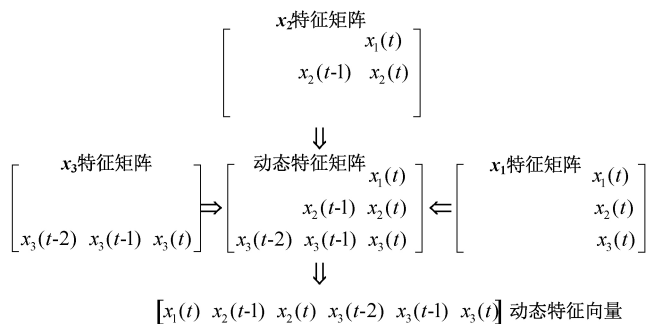


图 2 动态特征矩阵形成过程

Fig.2 The formation of dynamic feature matrix

的贡献率。可见, W_p 代表特征之间贡献率差异,不考虑动态相关性。之后,计算动态特征矩阵中各特征与其它特征之间互信息度量累计值,记为 $W_l(j)$ $j=1, 2, \dots, w$ 。

$$W_l(j) = \sum_{j' \neq j} C(j, j') \quad (5)$$

W_l 同时考虑了延时特征与即时特征之间的重要程度。对 W_p 与 W_l 均进行标准化处理,每个动态特征的实际权值为其对应的 W_p 与 W_l 的乘积,即

$$W = W_p \times W_l \quad (6)$$

在对动态特征矩阵向量化之后,DFM- k NN 方法与传统的 k NN 故障检测方法的实现过程仅存在特征权值上的差异。因此,将权值 W 以与特征 d 乘积的形式加入到 D^2 的计算中,如式(7)所示。

$$D^2 = \sum_{i=1}^k \sum_{j=1}^w (W(i, j) \times d(i, j))^2 \quad (7)$$

式中: $W(i, j)$ 与 $d(i, j)$ 分别为在线样本的第 i 个近邻中,第 j 个特征的权值与特征距离。

2.3 基于近邻距离的动态阈值设定

对于基于 FD- k NN,现有的阈值设定方法为在给置信度下对训练样本的覆盖比率。通常采用定值的阈值设定形式,无法体现 FD- k NN 检测过程“分而治之”的特点。

从 FD- k NN 的检测原理可知,每个在线样本是否被判定异常取决于其与离线样本中 k 个近邻的距离。由于在线样本在离线样本中的近邻需要独立确定,因此,可将每个在线样本的检测看作是一个独立过程,这也给动态阈值的设定提供了可能性与合理性。

本文提出对 FD- k NN 的动态阈值设定方法,以使阈值能够更好的适合具体的运行状态,进而降低误报率与漏报率,具体的阈值设定方法如下。

1) 对在线样本 $\mathbf{x}_i(i=1, 2, \dots, m)$,根据欧氏距离确定其在训练样本中的 k 个近邻样本 $\mathbf{kn}_j(j=1, 2, \dots, k)$;

2) 对每个近邻样本 $\mathbf{kn}_j$,查找其在训练过程中对应的 k NN 距离 $D_i^2(j)$,并计算所有 $D_i^2(j)$ 的平均值 $\bar{D}_i^2$;

3) 以给定置信度下的定值阈值与 $\bar{D}_i^2$ 的均值作为 $\mathbf{x}_i$

的阈值,即:

$$T_D^i = [T_\alpha + (1/k) \sum_{j=1}^k D_i^2(j)]/2 \tag{8}$$

式中: T_D^i 为当前时刻样本 x_i 的阈值; T_α 为置信度 α 下的 FD- k NN 的定值阈值。由于定值阈值仅需计算 1 次,且离线样本的 $D_i^2(j)$ 在训练过程已经获取,因此,在计算量上

并无明显增大。
为便于描述,该节所述方案均对应 FD- k NN 的在线检测过程,在 DFM-W k NN 中仅需将样本序列替换为动态特征序列,将近邻距离替换为加权距离,即可实现动态阈值的设定。
综上所述,本文所提出的 DFM-W k NN 流程如图 3 所示。

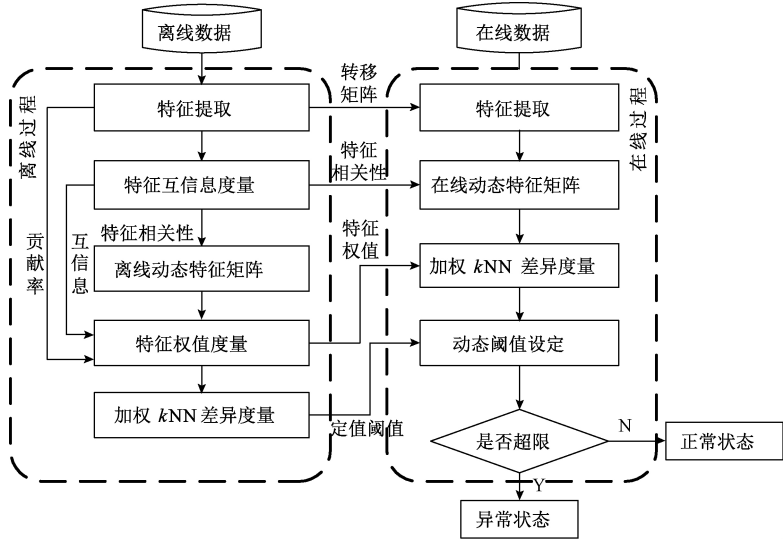


图 3 DFM-W k NN 故障检测方法流程
Fig.3 The flowchart of DFM-W k NN fault detection scheme

3 基于 DFM-W k NN 的风电机组故障检测

3.1 NREL-5 MW 海上风电机组故障检测

1) 故障描述
FAST (fatigue, aerodynamics, structures, and turbulence) 是美国可再生能源实验室 (national renewable energy laboratory, NREL) 重要的工程计算工具^[17],被用于模拟风力发电机的耦合动态响应。FAST 中包括流体动力学模型、空气动力学模型、结构动力学模型、控制和电力系统动力学模型。本文采用三叶片水平变速的 5 MW 海上风机基准模型,获取其正常运行数据以及常见的 10 种传感器与执行器故障数据。风机参数如表 1 所示^[18]。
为使模拟出的湍流风更加真实,采用由 NREL 开发的湍流-风模拟器 TurbSim^[19]。风况的特征参数如下: Kaimal 湍流模型强度:10%;风类型:对数剖面型;轮毂高度平均风速:18.2 m/s;粗糙系数:0.01 m。
在本文的 FAST-Simulink 模型中,布置 15 个传感器用于监测风速 v_w ,发电机转速 w_g ,发电机转矩 τ_g ,转子转速 w_r ,发电功率 p_e ,第 i 个叶片变桨角度 β_i ,叶片 i 根部力矩 M_i ,低速侧方位角 ϕ ,塔顶水平加速度 X_{acc} ,竖直

表 1 NREL-5 MW 海上风力发电机 Table 1 NREL-5 MW offshore WT	
额定功率 (P_n)	5 MW
叶片数	3
塔架高度/m	87.6
风轮直径/m	126
切入、额定、切出风速/($\text{m}\cdot\text{s}^{-1}$)	3, 11.4, 25
齿轮箱减速比	98
发电机转速($w_{g,n}$)/($\text{r}\cdot\text{min}^{-1}$)	1 173.7

加速度 Y_{acc} 与偏航误差 Ξ 。对所有传感器的模拟均加入白噪声处理。图 4 为传感器在风力发电机中的位置。
除正常运行状态外,在预先设定的 630 s 时间间隔内,分段模拟 10 种兆瓦级风力发电机常见的传感器故障与执行器故障,其中故障 1~6 为传感器故障,包括测量值与真实值之间存在微小偏差,比例误差以及数值偏移问题,各故障描述如表 2 所示。故障 7~10 为执行器故障,包括故障 7 与 8 的变桨系统故障,故障 9 的发电机转矩故障和故障 10 的偏航系统故障。其中故障 7 与 8 通过改变传递函数参数模拟,即分别修改固有频率与阻尼系数。故障 9 的模拟通过向发电机添加偏移量,故障 10 的模拟通过偏航角速度的设置。各故障描述如表 3 所示。

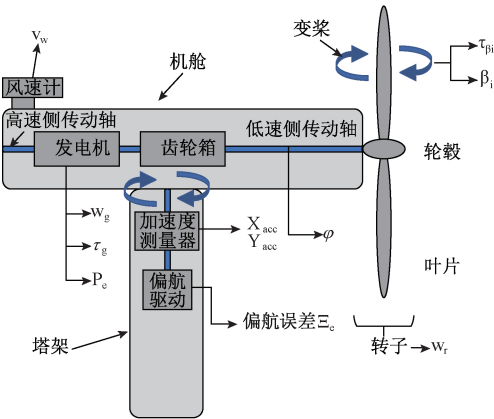


图4 传感器位置

Fig.4 The location of sensors

表2 传感器故障生成说明

Table 2 The instruction of sensor fault generation			
No.	故障位置	故障表现	参数设置
Fault 1	叶片根部弯矩传感器	比例偏差	$M_2 \times 0.95$
Fault 2	加速度传感器	数据偏移	$\ddot{x} - 0.5 \text{ m/s}^2, \ddot{y} - 0.5 \text{ m/s}^2$
Fault 3	电机转速传感器	比例偏差	$\omega_g \times 0.95$
Fault 4	变桨角度传感器	数据偏移	保持 β_i 为 1 deg
Fault 5	发电机功率传感器	比例偏差	$P_e \times 1.1$
Fault 6	低速侧位置编码器	微小误差	ϕ +随机补偿

表3 执行器故障生成说明

Table 3 The instruction of actuator fault generation			
No.	故障位置	故障表现	参数设置
Fault 7	变桨系统执行器	空气含量增加	$\omega_{n,2} = 5.73, \zeta_{n,2} = 0.45$
Fault 8	变桨系统执行器	液压下降	$\omega_{n,2} = 3.42, \zeta_{n,2} = 0.9$
Fault 9	发电机转矩	数据偏移	$\tau_g + 1\,000 \text{ NM}$
Fault 10	偏航驱动	驱动卡死	设置偏航角速度 0 rad/s

$$\mathbf{X}_d = \begin{bmatrix} \mathbf{x}_1(t) & \mathbf{x}_2(t) & \mathbf{x}_3(t) & \mathbf{x}_4(t) & \mathbf{x}_5(t) & \mathbf{x}_6(t) & \mathbf{x}_7(t) \\ \mathbf{x}_1(t-1) & & \mathbf{x}_3(t-1) & \mathbf{x}_4(t-1) & \mathbf{x}_5(t-1) & \mathbf{x}_6(t-1) & \mathbf{x}_7(t-1) \\ & \mathbf{x}_2(t-2) & & & & \mathbf{x}_7(t-2) & \end{bmatrix} \quad (9)$$

在权值的计算中,根据式(6),首先计算 $\mathbf{X}_d$ 中各特征与其他特征的MI度量的累计值,作为 $\mathbf{W}_l$,对由PCA获取的特征贡献值按照 $\mathbf{X}_d$ 中元素进行拓展,即对不同时刻的同一特征按照PCA获取的特征贡献值赋予相同权值,作为 $\mathbf{W}_p$,最后将 $\mathbf{W}_l$ 与 $\mathbf{W}_p$ 对应项相乘,获得DFM-WkNN中的特征权值 $\mathbf{W}$,结果如表4所示。

将DFM-WkNN对各故障的检测结果与PCA、KPCA、FD-kNN以及PCA-kNN进行对比,其中各算法的阈值设定均为95%的置信度,对PCA与KPCA均同时采用Q统

2) 结果与分析

在实验过程中,对每个故障截取故障发生前的500个正常样本与故障发生后的700个样本作为测试数据,在剩余所有的正常运行数据中随机选择10 000组样本作为训练数据。

首先,对训练数据采用PCA进行主元特征提取,选取85%的累计贡献,获得主元空间 $\mathbf{X}$ 与7个主元特征,如图5所示,特征 $x_1 \sim x_7$ 的贡献值分别为0.28, 0.18, 0.12, 0.11, 0.11, 0.10, 0.09。

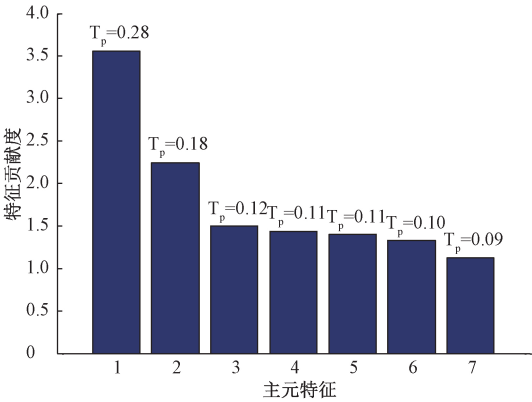


图5 主元特征贡献度

Fig.5 Contribution rate of principal component features

考虑延时特征的引入会使增广空间成比例增加,进而对故障检测的效率造成一定的影响,本文参考文献[15]向主元空间引入 $l=2$ 的延时特征,构造增广矩阵。计算各即时特征到所有延时特征的互信息度量,保留每个即时特征前 $w=3$ 个相关特征,即按照2.1节所提方法形成动态特征矩阵 $\mathbf{X}_d$,如式(9)所示。由式(9)可见, $t-1$ 时刻的状态与当前时刻具有较高动态相关性,而 $t-2$ 时刻与当前时刻相关性明显降低,也验证了选取 $l=2$ 的合理性。

计量与 T^2 统计量。在对各故障的检测中,计算测试数据前500样本的误报率以及后700个样本出现的漏报率。表5所示为对比算法对各故障的漏报率对比结果,表6所示为对比算法对各故障的误报率对比结果。

由表5可见,当故障发生时,由于故障4的数据偏移比较明显,5种故障检测方法均能得到较好的检测结果。由于PCA故障检测方法对非线性处理的能力较差,因此,对其余9种故障的检测结果均不理想,漏报率较高。从故障2、5、7、8、10以及漏报率的平均值的对比结果

表 4 DFM-WkNN 算法权值
Table 4 The Weights of DFM-WkNN

	$x_1(t)$	$x_1(t-1)$	$x_2(t)$	$x_2(t-2)$	$x_3(t)$	$x_3(t-1)$	$x_4(t)$	$x_4(t-1)$	$x_5(t)$	$x_5(t-1)$	$x_6(t)$	$x_6(t-1)$	$x_7(t)$	$x_7(t-1)$	$x_7(t-2)$
W_p	0.28	0.28	0.18	0.18	0.12	0.12	0.11	0.11	0.11	0.11	0.10	0.10	0.09	0.09	0.09
W_I	0.35	0.08	0.11	0.5	1	0.25	1.18	0.45	0.45	1.18	0.4	0.8	0.22	1.22	0.22
W	0.09	0.02	0.02	0.09	0.12	0.03	0.13	0.05	0.05	0.13	0.04	0.08	0.02	0.11	0.02

表 5 故障漏报率对比
Table 5 The comparison of alarm failure rate

No.	PCA		KPCA		FD- k NN	PC- k NN	DFM-WkNN
	Q	T ²	Q	T ²			
故障 1	72.29	84.86	44.31	4.29	3.37	1.52	0
故障 2	38.82	96.29	39.23	10.74	8.62	8.85	3.72
故障 3	91.86	100	29.00	5.25	3.83	1.52	1.07
故障 4	64.71	0	19.79	0	0	1.1	0
故障 5	89.29	95.00	45.36	15.25	14.02	8.87	3.22
故障 6	51.00	99.57	67.64	16.05	0	0	0
故障 7	35.43	99.71	61.03	25.93	11.62	9.27	2.65
故障 8	53.14	98.20	47.49	9.93	5.60	6.17	1.07
故障 9	82.43	92.29	61.17	16.97	3.62	1.77	0.53
故障 10	95.86	55.86	44.84	4.25	5.86	3.76	0
平均值	67.48	82.17	45.98	10.87	5.65	4.28	1.23

表 6 故障误报率对比
Table 6 The comparison of false alarm rate

No.	PCA		KPCA		FD- k NN	PC- k NN	DFM-WkNN
	Q	T ²	Q	T ²			
故障 1	15.00	3.00	3.39	3.51	3.73	1.58	0.65
故障 2	13.45	2.52	7.28	3.09	5.65	3.27	0
故障 3	0.62	15.82	2.56	9.92	8.87	5.51	0.07
故障 4	0	12.46	16.65	12.84	0	3.09	0.64
故障 5	25.80	19.90	15.92	25.39	2.56	3.99	1.52
故障 6	12.00	0	15.35	0	7.07	2.76	0
故障 7	5.20	28.40	41.80	32.60	6.52	4.32	0.87
故障 8	24.00	2.59	20.69	6.38	6.26	7.81	0
故障 9	30.80	27.40	13.38	46.24	4.03	3.51	1.25
故障 10	20.40	0	15.49	3.84	3.25	4.78	0
平均值	14.72	18.29	15.25	14.38	4.79	4.06	0.50

可见,本文所提出的 DFM-WkNN 故障检测方法与 KPCA 故障检测方法以及其他 2 种基于 k NN 的故障检测方法相比,可以显著降低漏报率。说明本文提出的动态特征矩阵转化与加权 k NN 方法能够更有效地实现故障数据与正常数据的分离。

从表 6 中对正常样本的误报率统计可见,本文所提出的 DFM-WkNN 对正常数据的识别具有较高的精度,对 10 种故障均得到了较低的误报率。这是由于传统定值形式的阈值设定方法只能根据训练数据的总体特征决定阈值的大小,依赖于数据的完备性,无

法处理工况突变给评价指标带来的波动,进而造成误报。本文依据 k 近邻故障检测方法的特点,提出动态阈值设定方法,其中阈值随训练数据中的相似状态变化,即相似运行状态的近邻距离较大,则在线状态的阈值也相应提高,进而降低由于运行状态突变引起的误报现象。

为更明显的对比 5 种故障检测方法对实验中故障数

据的检测效果,图 6 与图 7 分别给出各故障检测方法对故障 5 与故障 8 的检测过程。从对两个故障的检测过程可见,DFM-W k NN 方法能够更有效的将故障数据与正常运行数据分离。从放大的部分正常数据部分可见,DFM-W k NN 通过阈值的动态变化将阈值保持在较低水平,有效降低由于运行状态突变引起的误报。从而验证了本文所提出改进方法的有效性。

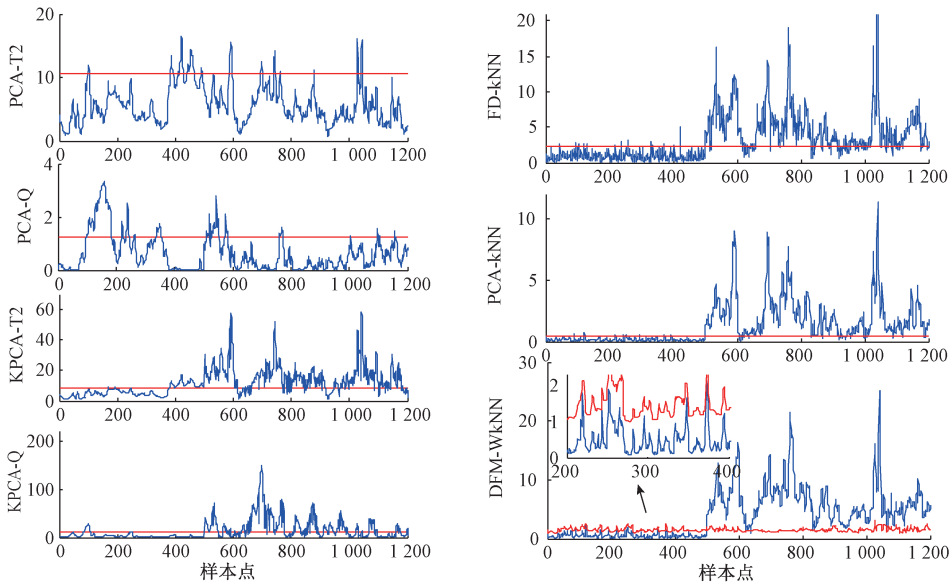


图 6 故障 5 检测结果对比
Fig.6 Results comparison of fault 5

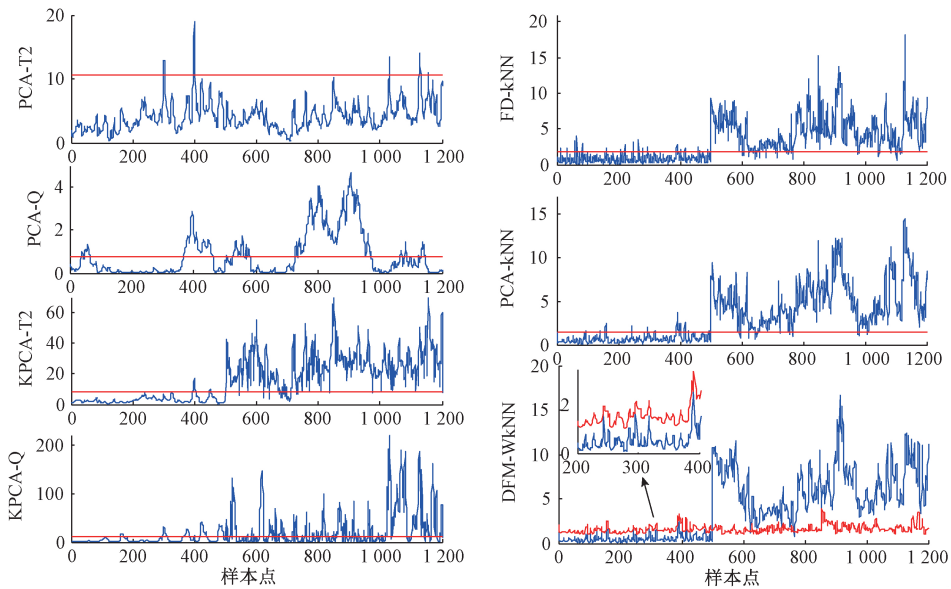


图 7 故障 8 检测结果对比
Fig.7 Results comparison of fault 8

3.2 基于 SCADA 数据的变桨系统故障检测

大型风电机组均配备有数据采集与监视控制 (supervisory control and data acquisition, SCADA) 系统^[20],以获取风电机组实时运行状态参数,包括风速、转速、电压、电流、功率、振动、温度等。通过对实时状态的监测能够实现对风电机组实时预警与故障诊断^[21-22],进而保证风电机组的安全运行。

为进一步验证所提 DFM-WkNN 方法对风电机组实际运行故障检测的有效性,将所提算法应用于 SCADA 系统中记录的变桨系统故障检测过程。实验数据采集自大连驼山风电场,型号为金风 GW77/1 500 kW 的直驱型风电机组,其风轮直径 76.84 m,切入风速 3 m/s,额定风速 11.5 m/s,额定转速 17.3 r/min。根据部分 SCADA 数据绘制其风速-功率散点关系,如图 8 所示。

在 SCADA 系统中选择与变桨系统相关的 15 个监测变量用于故障检测,如表 7 所示。首先以采样频率 1 次/min 采集系统记录的 7 号机组从 2 月 1 日~15 日中的 22 300×15 组

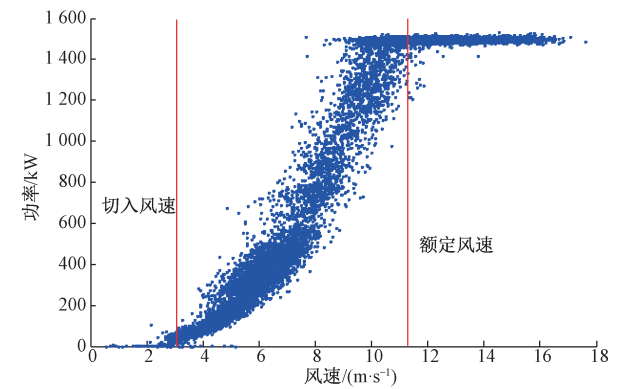


图 8 SCADA 数据风速-功率关系
Fig.8 Relationship between wind speed and power in SCADA data

数据。为保证训练数据均处于风机正常运行状态,剔除期间故障性维护停机时间的数据和风速小于切入风速的数据,剩余 15 360×15 组数据,用作实验训练数据。

表 7 变桨系统相关 SCADA 监测变量

Table 7 The monitoring variables related to pitch system in SCADA

No.	变量	No.	变量	No.	变量
1	风速	6	3#变桨电机温度	11	2#编码器位置
2	功率	7	1#变桨电机电流	12	3#编码器位置
3	发电机转速	8	2#变桨电机电流	13	1#冗余编码器位置
4	1#变桨电机温度	9	3#变桨电机电流	14	2#冗余编码器位置
5	2#变桨电机温度	10	1#编码器位置	15	3#冗余编码器位置

测试数据选择风场中 7 号机组于 2014 年发生的 2 次变桨系统故障,故障描述如表 8 所示。同样以 1 次/min 的采样频率采集系统报警前的 400 组样本与报警后的 200 组样本。应用 DFM-WkNN 方法对 2 次故障进行检测,将结果分别与 FD-kNN 以及 PCA-kNN 故障检测方法进行对比,其中由于 SCADA 系统发出报警前状态参数有可能已经发生异常,因此,仅统计前 300 组样本的误报率以及后 200 组样本的漏报率。表 9 所示为 3 种方法对

2 次故障的漏报率与误报率的统计结果。图 9 与 10 所示分别为 3 种故障检测方法对 2 种故障的检测过程对比。

表 8 变桨系统故障描述

Table 8 Fault description of pitch system

No.	故障类型	发生时间
故障 1	变桨位置比较故障	2014 3/10 23:29:00
故障 2	变桨电机温度过高故障	2014 5/18 16:11:00

表 9 故障漏报率与误报率对比

Table 9 The comparison of alarm failure rate and false alarm rate

漏报率	FD-kNN	PC-kNN	DFM-WkNN	误报率	FD-kNN	PC-kNN	DFM-WkNN
故障 1	19.26	4.00	0	故障 1	7.89	6.60	2.11
故障 2	25.36	17.27	0.80	故障 2	4.85	6.20	2.42

由表 9 可见,所提算法能够得到较低误报率与漏报率。一方面说明了所提出的动态特征选择方案与特征加权策略能够将正常样本与异常样本更好地分离,进而使异常状态能够更好的被识别。另一方面,由于实际数据

中存在着更多的“坏数据”,并且在数据筛选中很难被剔除,导致动态阈值无法完全跟随近邻距离的变化,但是并不影响在大趋势上对近邻距离变化的趋势跟随,因此所提算法能够有效避免故障检测中由于状态突变造成的误

报现象。此外,由图 9 与 10 的故障检测过程可见,相比于其他 2 种对比 k NN 故障检测方法,本文方法能够更早地发现异常状态,也进一步验证了本文方法对异常状态具有更好的敏感性。

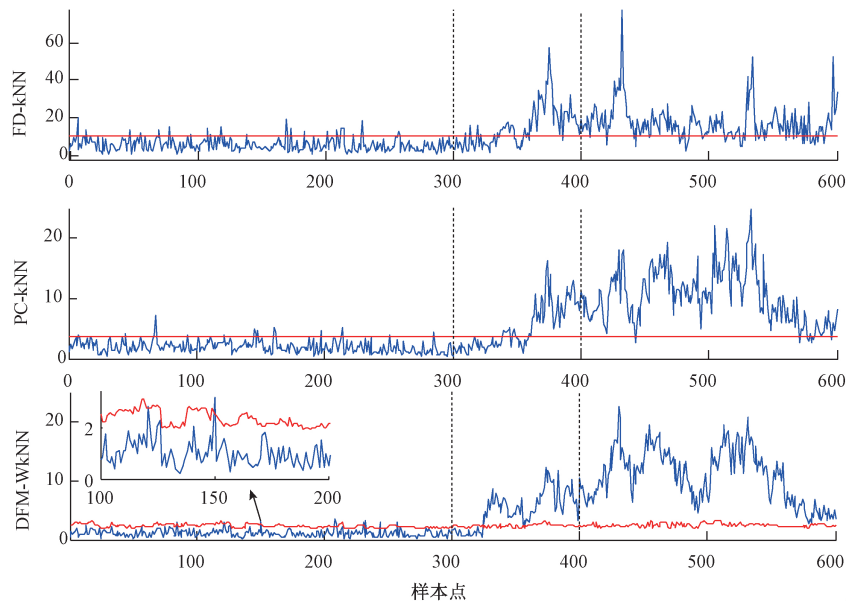


图 9 变桨系统故障 1 检测结果对比
Fig.9 Results comparison of fault 1 in pitch system

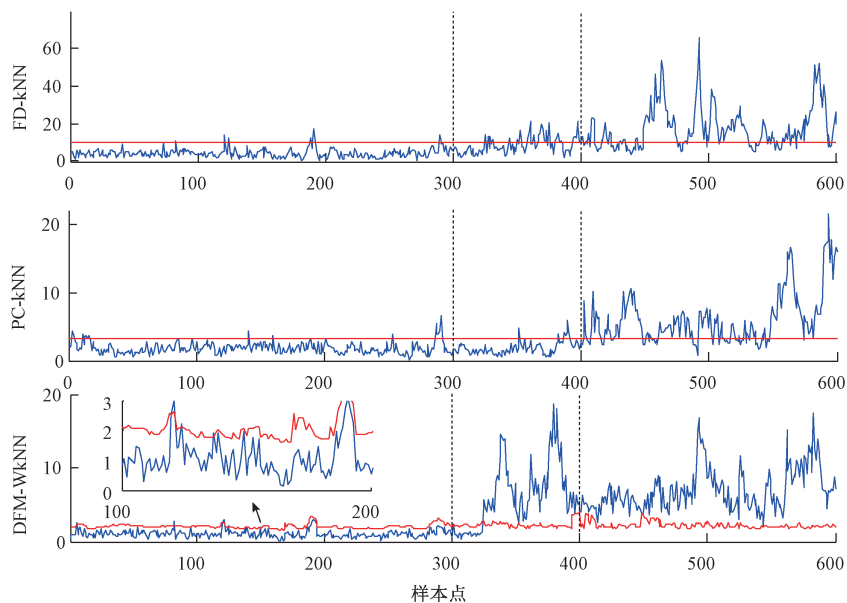


图 10 变桨系统故障 2 检测结果对比
Fig.10 Results comparison of fault 2 in pitch system

4 结 论

本文提出 DFM-W k NN 方法,分别以美国可再生能源实验室 5 MW 海上风机基准模型的常见传感器和执行器故障,以及 SCADA 系统采集到的变桨系统故障为例进行

故障检测,实验结果分析表明:1)采用所提动态特征提取方法生成风电机组动态特征矩阵,能够反映风机运行中状态特征间的动态相关性,同时降低了无关特征对检测过程的干扰影响;2)所提出的对 k NN 故障检测的加权方案更合理的处理了特征间的差异性,能够更好的将正常状态与异常状态分离;3)所提出的动态阈值设定能够适

应运行状态的变化,进而有效避免由于运行状态突变引起的误报现象。

参考文献

- [1] TAUTZ-WEINERT J, WATSON S J. Using scada data for wind turbine condition monitoring-a review [J]. Iet Renewable Power Generation, 2017, 11(2):382-394.
- [2] 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.
JIN X H, SUN Y, SHAN J H, et al. Fault diagnosis and prognosis for wind turbines: An overview [J]. Chinese Journal of Scientific Instrument, 2017, 38(5):1041-1053.
- [3] 龙霞飞,杨苹,郭红霞,等.大型风力发电机组故障诊断方法综述[J].电网技术,2017,41(11):3480-3491.
LONG X F, YANG P, GUO H X, et al. Review of fault diagnosis methods for large wind turbines [J]. Power System Technology, 2017, 41(11):3480-3491.
- [4] 曾军,陈艳峰,杨苹,等.大型风力发电机组故障诊断综述[J].电网技术,2018,42(3):849-860.
Zeng J, CHEN Y F, YANG P, et al. Review of fault diagnosis methods of large-scale wind turbines [J]. Power System Technology, 2018,42(3):849-860.
- [5] 贾锋,蔡旭,李征,等.风电机组精细化建模及硬件在环实时联合仿真[J].中国电机工程学报,2017,37(4):308-320.
JIA F, CAI X, LI ZH, et al. Refined modeling of wind energy conversion systems and real-time co-simulation with hardware-in-loop [J]. Proceedings of the Chinese Society for Electrical Engineering, 2017, 37(4):308-320.
- [6] 李刚,张建付,文福拴.计及信息不确定性的风电机组健康状态实时评估方法[J].电力系统自动化,2017,41(18):100-107.
LI G, ZHANG J F, WEN F SH. Evaluation of real-time health condition for wind turbines considering information uncertainty [J]. Automation of Electric Power Systems, 2017, 41(18):100-107.
- [7] CHEN Y Q, FANG Y F, GUO SH. Optimization and design for wind turbine airfoil at multiple working conditions based on genetic algorithm [J]. Applied Mechanics and Materials, 2014, 668-669:230-235.
- [8] 胡姚刚.大功率风电机组关键部件健康状态监测与评估方法研究[D].重庆:重庆大学,2017.
HU Y G. Study of health condition monitoring and assessment for critical components of large wind turbine generator systems [D]. Chongqing: Chongqing University, 2017.
- [8] 刘永前,王飞,时文刚,等.基于支持向量机的风电机组运行工况分类方法[J].太阳能学报,2010,31(9):1191-1197.
LIU Y Q, WANG F, SHI W G, et al. Operation condition classification method for wind turbine based on support vector machine [J]. Acta Energiæ Solaris Sinica, 2010, 31(9):1191-1197.
- [10] 董玉亮,李亚琼,曹海斌.基于运行工况辨识的风电机组健康状态实时评价方法[J].中国电机工程学报,2013,33(11):88-95.
DONG Y L, LI Y Q, CAO H B. Real-time health condition evaluation on wind turbines based on operational condition recognition [J]. Proceedings of the Chinese Society for Electrical Engineering, 2013, 33(11):88-95.
- [11] 郑小霞,李美娜,王靖.基于PSO优化核主元分析的海上风电机组运行工况分类[J].电力系统保护与控制,2016,44(16):28-35.
ZHENG X X, LI M N, W J. Operational conditions classification of offshore wind turbines based on kernel principal analysis optimized by PSO [J]. Power System Protection and Control, 2016, 44(16):28-35.
- [12] HE Q P, WANG J. Fault detection using the k -nearest neighbor rule for semiconductor manufacturing processes [J]. IEEE Transactions on Semiconductor Manufacturing, 2007, 20(4):345-354.
- [13] HE Q P, WANG J. Large-scale semiconductor process fault detection using a fast pattern recognition-based method [J]. IEEE Transactions on Semiconductor Manufacturing, 2010, 23(2):194-200.
- [14] 张成,高宪文,李元.基于 k 近邻主元得分差分的故障检测策略[J].自动化学报,2019,1(4):1-10.
ZHANG CH, GAO X W, LI Y. Fault detection strategy based on principal component score difference of k nearest neighbors [J]. Acta Automatica Sinica, 2019, 1(4):1-10.
- [15] 唐勇波,桂卫华,彭涛,等.基于互信息变量选择的变压器油中溶解气体浓度预测[J].仪器仪表学报,2013,34(7):1492-1498.
TANG Y P, GUI W H, PENG T, et al. Prediction method for dissolved gas concentration in transformer oil based on variable selection of mutual information [J]. Chinese Journal of Scientific Instrument, 2013, 34(7):1492-1498.
- [16] 童楚东,史旭华.基于互信息的分散式动态PCA故障检测方法[J].化工学报,2016,10(67):4317-4323.
TONG CH D, SHI X H. Fault detection by decentralized

- dynamic PCA algorithm on mutual information [J]. Journal of Chemical Industry and Engineering, 2016, 10(67): 4317-4323.
- [17] JONKMAN J, BUTTERFIELD S, MUSIAL W, et al. Definition of a 5 MW reference wind turbine for offshore system development [J]. Office of Scientific and Technical Information Technical Reports, 2009, 2(1): 1-75.
- [18] ODGAARD P F, JOHNSON K E. Wind turbine fault detection and fault tolerant control-an enhanced benchmark challenge [C]. Proceeding of the American Control Conference, Washington, USA, 2013: 4447-4452.
- [19] JONKMAN B J. Turbsim user's guide: Version 1.50 [J]. Nrel Report National Renewable Energy Laboratory, 2009, 7(1):1-58.
- [20] TAUTZ-WEINERT J, WATSON S J. Using SCADA data for wind turbine condition monitoring - A review [J]. IET Renewable Power Generation, 2017, 11(4):382-394.
- [21] CHEN N, YU R, CHEN Y, et al. Hierarchical method for wind turbine prognosis using SCADA data [J]. IET Renewable Power Generation, 2017, 11(4):403-410.
- [22] 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(1):550-567.

作者简介



钱小毅, 2012 年于渤海大学取得学士学位, 2016 年于沈阳工业大学取得硕士学位, 现为沈阳工业大学博士研究生, 主要研究方向包括数据挖掘、智能优化、故障检测与故障识别。

Qian Xiaoyi received his B. Sc. degree from Bohai University in 2012 and received his M. Sc. degree from Shenyang University of Technology in 2016. He is currently pursuing his Ph. D. degree at Shenyang University of Technology. His research interests include data mining, intelligent optimization, fault detection and identification.



张宇献 (通信作者), 于东北大学控制理论与控制工程专业取得硕士与博士学位, 清华大学自动化系博士后。目前为沈阳工业大学副教授, 主要研究方向包括数据挖掘、智能控制与智能优化。

Zhang Yuxian (Corresponding author) received his M. Sc. degree and his Ph. D. degree in control theory and control engineering from Northeastern University. He was a postdoctoral with the department of automation at Tsinghua University. He is currently an associate professor at Shenyang University of Technology. His current research interests include data mining, intelligent control and intelligent optimization.