

文章编号:0254-0096(2020)05-0085-06

基于机器学习的风电机组变桨系统故障研究

熊中杰, 邱颖宁, 冯延晖, 程 强

(南京理工大学能源与动力工程学院, 南京 210094)

摘 要: 针对变桨系统的齿形带断裂故障问题, 首先通过分析变桨系统的工作原理, 基于多维 SCADA 信号进行特征数据挖掘, 再利用主成分分析法对数据进行预处理, 并保留时序信息重构数组, 最后利用高斯核支持向量机进行机器学习, 实现对齿形带断裂故障的智能检测。结果表明该方法可准确诊断齿形带断裂故障, 并已通过多台风电机组监测数据进行验证, 准确性可达到 98.8%, 证明该文所用方法和模型的广泛适用性。研究结果可对未来智慧风电场的开发管理提供有利用价值的工具。

关键词: 风电机组; 故障诊断; 主成分分析; 核支持向量机; SCADA 数据; 变桨系统

中图分类号: TM315

文献标志码: A

0 引 言

风电机组在长期运作中会伴随着各部件故障频繁发生, 采用合理高效的监测手段和准确的故障诊断方法对提高风电机组可靠性具有重要意义。

数据采集与监视控制(SCADA)系统是目前应用于风电行业的主要监控系统, 通过采集各项传感器数据反映风力机组的运行状态。相比于状态监测系统(CMS), SCADA 系统采样频率较低, 但信息记录的周期更长, 能够为风电机组故障诊断提供丰富的数据来源^[1-2]。目前常见的故障诊断方法有基于模型分析、专家经验知识定性分析、信号处理分析、数字驱动统计分析等^[3]。郭鹏等^[4]利用 SCADA 数据, 采用非线性估计方法搭建了塔架振动模型, 并通过后续分析反映叶轮的运行状态。姚万业等^[5]利用相似性原理提取 SCADA 数据的特征信息, 并通过搭建数字驱动的非线性估计模型来评估风电机组运行的健康状况。Zimroz 等^[6]基于合理算法模型的观点, 通过监测不稳定空气负载对变桨距系统稳定性的影响来决策相关停机的情况。文献[7]通过对不同故障的预测和隔离, 总结风电机组主轴、齿轮箱、变桨系统及电动机等关键部件的常用故障预测方法。文献[8]通过跟踪检测风电齿轮箱油温温升趋势实现了早期故障诊断, 并结合 SCADA 数据验证诊

断的有效性。文献[9]利用机器学习中的神经网络(ANN)方法, 针对 SCADA 记录中的警报信号识别变桨系统的异常状态。文献[10]通过对 SCADA 记录中警报信号, 提出 2 种利用警报信号评估风电机组可靠性的数学分析方法。

目前变桨系统故障诊断的研究主要针对零部件状态监测数据的直接测量分析, 以及故障模型的搭建。本文以变桨系统中的 SCADA 数据作为研究对象, 基于主成分分析和整机运行分析提出数据特征挖掘方法, 结合支持向量机进行机器学习, 并对同类型多台风电机组进行验证, 实现变桨系统中齿形带断裂故障的有效诊断。

1 齿形带故障模式

风力发电机组变桨系统是由变桨驱动系统(变桨减速器、变桨电机)、桨叶(叶片、变桨轴承)、齿形带、齿形带支架、变桨控制系统组成。当风速超过额定值时, 通过调整叶片桨距角控制风轮转速, 使输出功率保持在额定状态; 而当风速低于额定风速时桨距角保持设定值。变桨系统的常见故障包括变桨安全链故障、变桨速度超限故障、变桨最小角超限故障、变桨电容电压不平衡故障、变桨电机温度过高故障、变桨充电器反馈丢失故障等^[11]。根据变桨系统的工作原理,

收稿日期: 2017-10-24

基金项目: 国家自然科学基金(51505225); 江苏省六大人才高峰(ZBZZ-045); 江苏省自然科学基金(BK20131350); 中央高校基本科研业务费专项(30915011324)

通信作者: 邱颖宁(1977—) 女, 博士、副教授, 主要从事风电机组故障诊断方面的研究。yingning.qiu@njust.edu.cn

叶片变桨次数与风速变化频率有关。风电机组的频繁变桨使连接桨叶和变桨电机中的传动部件长期受到交变载荷作用,从而材料疲劳的累计速度很快,最终导致齿形带过早断裂。

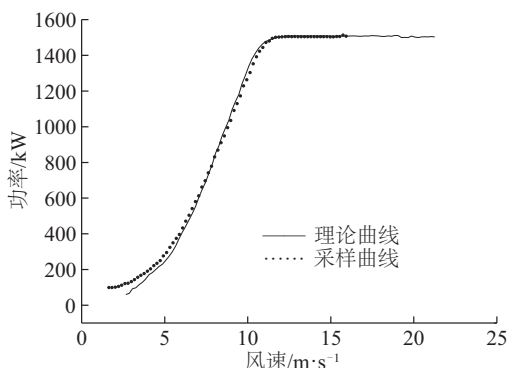
2 SCADA 数据分析

2.1 运行状态分析

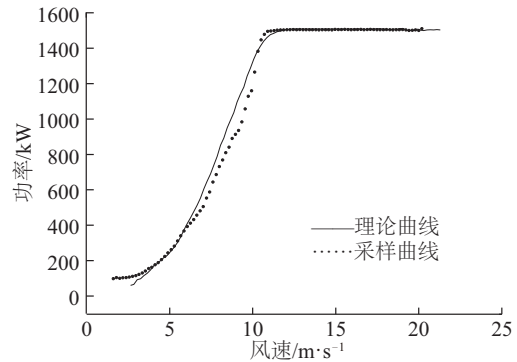
本文选取某风场多台 1.5 MW 风电机组海量 SCADA 数据,主要针对变桨电机拖动桨叶旋转的中间部件齿形带断裂进行研究。由于此类故障无法由 SCADA 系统原始数据直接表征,本文提取故障风电机组将近 1 个月(9 月 9~30 日)的数据记录,每条记录均包括风速、发电机转速、输出有功功率、桨叶角度、桨叶速度、变桨电机温度等 26 个监测项。通常越是接近故障发生前,信号特征表现就越明显。考虑到 SCADA 数据之间的相关性,分别对风电机组正常运行状态和故障运行状态进行风速-功率曲线绘制,如图 1 所示。由于风速波动的随机性以及采样数据的离散性,本文将所有数据按照天数划分为不同区间。首先,提取各区间最大风速 V_{\max} 和最小风速 V_{\min} ,依照 IEC 标准^[12],将 $[V_{\max}, V_{\min}]$ 划分成 N 等份(N^* 为正整数),则:

$$N \approx \frac{V_{\max} - V_{\min}}{0.5}, N \in N^* \quad (1)$$

再将每等份内的数据点求平均值^[13]。图 1 为 9 月 14~17 日和 9 月 18~21 日采样数据经上述处理之后的风速-功率曲线,并与理论曲线作比较。其中,图 1a 所示风电机组在实际风速区间下的输出功率与理论曲线较吻合,而图 1b 中显示实际输出功率曲线低于理论曲线,即风电机组运行异常。考虑到风电机组各故障之间的相互影响,针对齿形带断裂故障形式,需对 SCADA 数据中的隐含特征做进一步挖掘,实现时域上的及时准确诊断。



a. 9 月 14~17 日处理结果



b. 9 月 18~21 日处理结果

图 1 风速-功率曲线

Fig. 1 Wind speed and power curve

2.2 样本提取及预处理

结合变桨系统的工作原理,以及 2.1 节划分的区间风电机组的运行状态,选取 SCADA 数据中的风速、发电机转速、叶片的速度与角度、变桨电机温度、机舱温度、变桨直流电源温度与电流这几个特征量进行分析。考虑到原始样本较大(21000 多条记录),以及各参数之间的相关性,采用主成分分析方法进行特征选择。主成分分析法对数据降维表现良好,主要根据方差最大化原理,通过求解一组线性无关且相互正交的向量表征原数据矩阵来寻找累积贡献率较大的特征值,以此利用几个综合指标代表原来众多特征。

主成分分析可能舍弃贡献率小的成分造成样本信息丢失,同时对数据的含义解释模糊,缺乏表达特征完整性的能力。因此本文主要针对 3 个桨叶的角度、速度、变桨电机温度、充电器温度和电流的不对称数据进行主成分分析,并二次组合以上指标的均值和标准差进行特征提取。为消除不同变量量纲和变化幅度不同带来的影响,原始数据都应该做标准化处理,具体步骤:

步骤 1:采集 SCADA 在线数据,剔除采样值缺省的记录。首先提取除时间之外的 18 个特征参数,即风速、发电机转速、机舱温度、3 个桨叶的速度和角度、3 个变桨电机的温度、3 个变桨系统直流充电器的温度和直流电流,作 $[-1, 1]$ 归一化处理;

步骤 2:设定风速阈值 $V_0 = 10$ m/s,划分变桨角度的变化区间 $[\beta_{\min}, \beta_{\max}]$,其中 β_{\min} 为风电机组变桨角度初设值 0, β_{\max} 为正常工作能达到的最大变桨角度;

步骤 3:利用主成分方法对不对称数据分析,综合标准化处理的风速、发电机转速、机舱温度数据,最终提取 8 个特征参量,并按照时序排列;

步骤4:以每分钟内连续10个采集点划分区间,计算各区间的平均值 $x(i)$ 和标准差 $\sigma(i)$, i 代表不同特征参量序号;

步骤5:重复步骤1~步骤4整理风电场多台风电机组数据,按照有无故障发生分类处理。

通过上述方法可有效剔除原始数据中的冗余部分,并在保留数据时序上信息的同时提取表征故障特性的参数。

3 故障识别方法

支持向量机是在传统机器学习理论上发展起来的一种机器学习算法,以统计学理论的 VC(Vapnik-Chervonenkis dimension)维和结构风险最小化原则为理论依据。其重点在于从训练样本数据中构造最优分类超平面的支持向量,以及样本分类问题的线性或非线性处理,在数学上可归结为求解一个凸二次优化问题。其基本思想可以由图2说明^[14]。黑点和白点分别表示2种不同类别的样本点, H 为空间上区分这2类的边界分类线,虚线 H_1 和 H_2 分别是通过2类样本中离 H 最近的样本点且平行于 H 的直线。支持向量机的目的是寻找合适的边界线 H ,并确保 H 到 H_1 、 H_2 之间的分类距离(Margin)最大,由此实现最优分类。其中定义边距 H_1 、 H_2 的最近样本数据点称为支持向量。图中的 b 代表参数向量,即超平面的法线, w 代表分类的阈值。

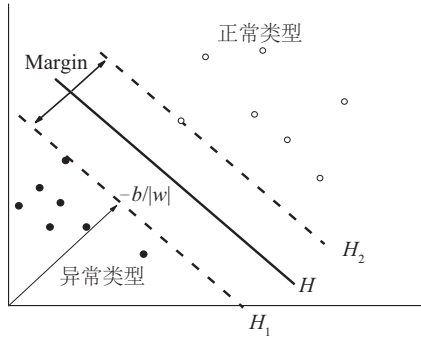


图2 最优分类超平面

Fig. 2 Optimal classification hyperplane

由于本文提取的训练样本数据在初始空间中是非线性不可分的,因此需要利用非线性变换 φ 将输入空间数据映射到更高维的特征向量空间中再构造出最优分类超平面进行分类,最后映射回原始空间中。为解决维数增加导致向量内积运算的困难性,支持向量机引入核函数来代替高维特征空间的内积运算,其结构形式见图3^[15]。图中 x_1, x_2, \dots, x_n 代表支持向量机的输入特征; $K(x_1, x), K(x_2, x), \dots, K(x_n, x)$ 代表高斯核函数;

$a_1 y_1, a_2 y_2, \dots, a_n y_n$ 代表优化约束条件。

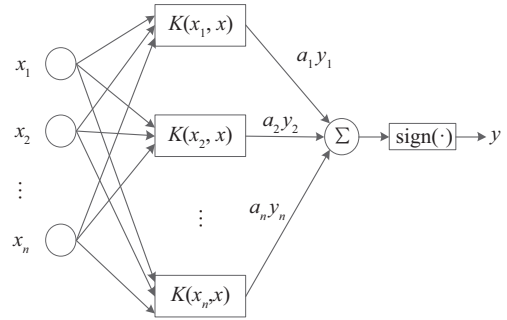


图3 支持向量机的结构形式

Fig. 3 Structural form of SVM

支持向量机的最大特征是利用核函数计算非线性变换的内积,规避了目标函数直接非线性变化的复杂性,从而精简了计算资源。高斯核函数在数据存在噪声时有抗干扰能力,并且可通过调节内部函数参数表现数据的局部特征。在模型训练过程中表现为学习速度快,泛化能力强,不易陷入局部极小值,已成功应用于小样本数据分类、模式判别和图像处理等方面^[16]。

高斯核函数的表达式为:

$$K(x_i, x_j) = \exp\left(-\|x_i, x_j\|^2 / 2\delta^2\right) \quad (2)$$

式中, (x_i, x_j) ——特征变量,下标 i, j 代表特征变量的序号,其中 $i \neq j$ 。

引入 Lagrange 因子 $\alpha_i \geq 0$ 和惩罚项 $C \sum \xi_i$ (ξ_i 为松弛变量,起到容错作用)将文本中的齿形带故障识别转化为带约束条件的优化问题:

$$\begin{cases} \max L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \exp\left(-\|x_i, x_j\|^2 / 2\delta^2\right) \\ \text{s.t. } 0 \leq \alpha_i \leq C, i = 1, 2, 3, \dots, n \\ \sum_{i=1}^n y_i \alpha_i = 0 \end{cases} \quad (3)$$

由此解出最优分类函数为:

$$f(x) = \text{sign}\left(\sum_{i,j=1}^n \alpha_i y_i \exp\left(-\|x_i, x_j\|^2 / 2\delta^2\right) + b\right) \quad (4)$$

由式(4)表明只要找到合适的惩罚系数 C 及核函数半径 δ 即可实现分类识别。

4 模型验证

4.1 同一风电机组故障识别

本文选取同风场的4台同类型风电机组 A1~A4 的 SCADA 数据,其中 A1、A2 为已发生齿形带断裂故障机组, A3、A4 为未发生故障机组。首先提取 A1 风电机组

9 月份的 21000 条记录,利用上述预处理步骤提取故障特征,最终得到 4300 组数据作为模型输入。其中预处理后的齿形带工作异常区间为 $[2729, 3299]$,其他区间变桨系统均运行正常。将变桨正常状态设定标签为 0,故障异常状态设为标签 1。取惩罚系数 C 的变化范围为 $[-0.5, 0.5]$,高斯核函数半径 δ 的变化范围为 $[-2, 2]$ 。

本文利用以上数据建立基于高斯核函数的支持向量机模型。机器学习中往往会出现训练模型准确率很高,但测试集验证准确性不高的现象。其主要原因主要有训练数据中 2 类情况的特征不明显,预处理数据结果空间分布不平衡及测试集中 2 类分布比例相较于训练集不均,由此导致模型训练过拟合或欠拟合。为优化数据构建的训练模型,本文针对预处理得到的训练数据采用 K 折交叉验证处理。其主要思想为:对数据样本 X 分割成 K 个子样本,每个单独的子样本 K 被保留作为验证模型的数据,其余 $K-1$ 个样本重构训练组来产生模型;重复 K 次交叉验证,通过判断各子样本验证准确率来衡量模型的稳定性。对训练数据组分层抽样,随机分组引入 10 折交叉验证,并同时保证同组正常标签与故障标签的比例相同。经多次训练计算,选取惩罚系数 $C=1.4$,核函数半径 $\delta=4$ 。最后测试集引入模型进行检验,故障预测标签如图 4 所示。

图 4 表明利用高斯核函数支持向量机模型能有效识别 A1 风电机组变桨系统齿形带故障区间。预测输出

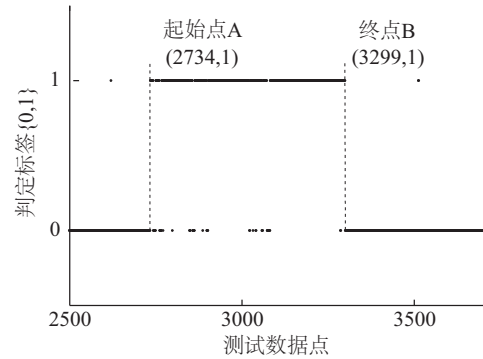
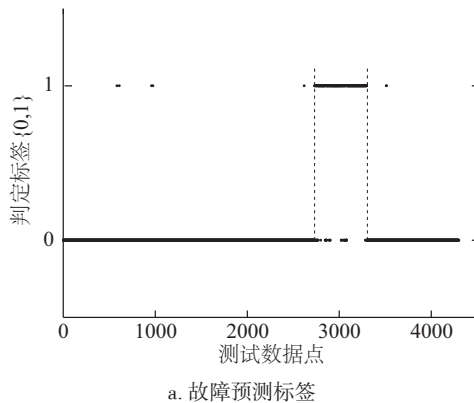


图 4 A1 风电机组变桨故障识别

Fig. 4 Fault identification of variable pitch of A1 WT

区间为 $[2734, 3299]$,这与上述实际转化区间 $[2729, 3299]$ 齿形带断裂。从图 4b 图局部放大中可看出部分实际情况为异常的数据点被误判断为正常,而在变桨正常工作区间内也有少数测试点诊断错误。这是由于数据本身在故障状态下与正常状态特征不明显所致,同时在预处理阶段采用求取平均值的方法,一定程度上弱化了 2 种情况的特征差异。

4.2 不同风电机组故障识别

为考察模型的泛化能力以及稳定性,对同风场其他 3 台风电机组 A2、A3、A4 的监测数据进行整理验证。将预处理步骤 5 中得到的数据分层抽样并按照时序排列。采用相同的惩罚系数和核函数半径,同时用 A1 风电机组机器学习模型进行交叉验证。模型训练各测试数组的结果如表 1 和图 5 所示。图 5 中的结果均为识别标签,为判断故障诊断模型的预测准确性,将标签区间回归到各机组原始 SCADA 数据中,并与实际断裂故障发生点作比较。预测准确率及误判样本点总数见表 1。结合图 5 和表 1 中的结果表明本文对 SCADA 数据的处理方法能较好地预测出不同风电机组的变桨系统故障发生区间,并且通过不同机组之间的交叉验证证明模型具有一定的泛化能力。在计算速度方面,该方法构建训练模型迅速,占用计算机资源较少。

表 1 支持向量机预测结果

Table 1 SVM predict results

机组	实际故障区间	预测故障区间	误判数量	准确率/%
A1	$[126854, 133722]$	$[127033, 133722]$	179	97.4
A2	$[54138, 55656]$	$[54157, 55656]$	18	98.8
A3	无故障	无故障	0	100
A4	无故障	无故障	0	100

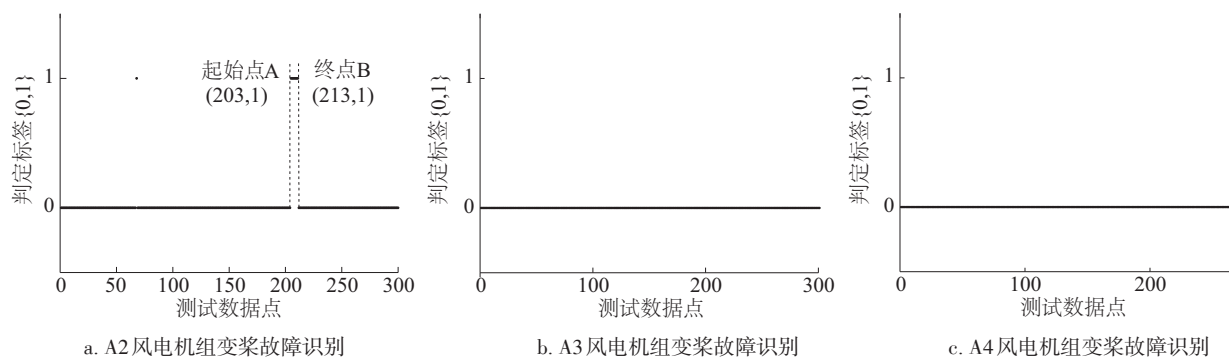


图5 不同风电机组变桨故障识别

Fig. 5 Fault identification of variable pitch of different WTs

5 结 论

本文利用 SCADA 检测数据,利用机器学习方法实现风力机组变桨系统齿形带断裂故障监测。通过主成分分析法提取原始海量数据主要特征,再结合变桨系统工作机理重组训练样本,以高斯核支持向量机建立预测模型,实现由检测数据判别齿形带是否发生断裂故障。本文采用的方法特点为:

1)利用主成分分析对多尺度数据降维,结合风电机组运行状态分析,重组 SCADA 数据,同时保留故障特征信息和信号时序上隐含的信息。相较于仅采用主成分分析的方法,弱化了剔除贡献率小但包含样本差异重要信息数组的不足,为时域连续数据预处理提供有效方法。

2)采用高斯核支持向量机监督学习,验证齿形带断裂故障预测的可行性,并论证该方法具有计算速度快,占用资源少等优点。

3)本文实现的故障诊断方法,首先通过某一故障机型实现数据训练,其后通过多台同类型故障风电机组和健康机组进行横向验证,量化了监测的准确性,监测结果证明模型具有一定的泛化能力。

该研究结果对降低故障诊断成本,实现未来智慧风电场的管理提供了一定的参考方法和工具。

[参考文献]

- [1] YANG W X, COURT R, JIANG J H. Wind turbine condition monitoring by the approach of SCADA data analysis [J]. Renewable energy, 2013, 53 (9): 365-376.
- [2] ZAHER A, MCARTHUR S D J, INFELD D G, et al. Online wind turbine fault detection through automated SCADA data analysis [J]. Wind energy, 2009, 12(6): 574-593.

- [3] VENKATASUBRAMANIAN V, RENGASWAMY R, KAVURI S N. A review of process fault detection and diagnosis (Part II): Quantitative models and search strategies [J]. Computers & chemical engineering, 2003, 27 (3): 313-326.
- [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 P, XU M, BAI N, et al. Wind turbine tower vibration modeling and monitoring driven by SCADA data [J]. Proceedings of the CSEE, 2013, 33(5): 128-135.
- [5] 姚万业, 李新丽. 基于状态监测的风电机组变桨系统故障诊断 [J]. 可再生能源, 2016, 34(5): 437-440.
- YAO W Y, LI X L. Pitch system fault diagnosis for wind turbine based on condition monitoring [J]. Renewable energy resources, 2016, 34(5): 437-440.
- [6] ZIMROZ R, BARTELMUS W, BARSZCZ T, et al. Diagnostics of bearings in presence of strong operating conditions non-stationarity—A procedure of load-dependent features processing with application to wind turbine bearings [J]. Mechanical systems & signal processing, 2014, 46(1): 16-27.
- [7] QIAN H, JINAG D X. Application of wavelet neural networks on vibration fault diagnosis for wind turbine gearbox [C]//International Symposium on Neural Networks: Advances in Neural Networks, Wroclaw, Poland, 2008:313-320.
- [8] QIU Y N, FENG Y H, SUN J, et al. Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data [J]. IET renewable power generation, 2016, 10(5): 661-668.
- [9] CHEN B D, QIU Y B, FENG Y H, et al. Wind turbine SCADA alarm pattern recognition [C]//IET Conference

- on Renewable Power Generation (RPG 2011), Edinburgh, UK, 2011: 1-6.
- [10] QIU Y N, FENG Y H, TAVNER P, et al. Wind turbine SCADA alarm analysis for improving reliability [J]. Wind energy, 2012, 15(8): 951-966.
- [11] 张红义, 呼木吉乐图, 王冰, 等. 金风 1.5 MW 风机 Vensys 变桨系统安全链故障处理[J]. 内蒙古石油化工, 2015(9): 71-73.
- ZHANG H Y, HU M J L T, WANG B, et al. Analysis of gold wind 1.5 MW wind turbine Vensys pitch system safety chain fault treatment [J]. Inner Mongolia petrochemical industry, 2015(9): 71-73.
- [12] The International Electrotechnical Commission. IEC61400 Part II : Wind turbine generator systems-acoustic noise measurement techniques [S]. IEC international standard, 2002.
- [13] YANG W X, COURT R, JIANG J H. Wind turbine condition monitoring by the approach of SCADA data analysis[J]. Renewable energy, 2013, 53(9): 365-376.
- [14] WIDODO A, YANG B S. Support vector machine in machine condition monitoring and fault diagnosis [J]. Mechanical system and signal processing, 2007, 21(6): 2560-2574.
- [15] 高文军. 基于人工免疫算法优化支持向量机的电力变压器故障诊断研究[D]. 太原: 太原理工大学, 2012.
- GAO W J. Study on fault diagnosis for power transformer based on support vector machine of artificial immune algorithm [D]. Taiyuan: Taiyuan University of Technology, 2012.
- [16] 刘德鹏, 李刚, 张宏亮, 等. 基于 RBF 核支持向量机的轴箱磨损状态诊断系统[J]. 电子科技, 2017(5): 150-153.
- LIU D P, LI G, ZHANG H L, et al. Fault diagnosis system of axle box wear condition based on support vector machines with RBF kernel [J]. Electronic science and technology, 2017(5): 150-153.

FAULT ANALYSIS OF WIND TURBINE PITCH SYSTEM BASED ON MACHINE LEARNING

Xiong Zhongjie, Qiu Yingning, Feng Yanhui, Cheng Qiang

(School of Energy and Power Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: To solve the problem of timing belt fracture in pitch system, feature signals are mined from large amounts of SCADA data by the analysis on the working principles. The signal data are processed by principal component analysis method next, retaining the timing information. Finally, the Gaussian kernel support vector machine is used for machine learning, which achieves the intelligent detection of the fault. The results show that the method can diagnose the fracture fault of timing belt accurately, and the accuracy of the fault detection algorithm can reach 98.8%, which proves the extensive applicability by checking the data from multi wind turbines. Furthermore, the study provides a useful tool for the development and management of the future wisdom wind field.

Keywords: wind turbine; fault diagnosis; PCA; kernel SVM; SCADA data; pitch system