

计算机集成制造系统  
*Computer Integrated Manufacturing Systems*  
ISSN 1006-5911, CN 11-5946/TP

## 《计算机集成制造系统》网络首发论文

题目: 融合分形特征的风机运行状态辨识方法  
作者: 郭怡, 王荣喜, 高建民  
收稿日期: 2021-09-06  
网络首发日期: 2021-12-28  
引用格式: 郭怡, 王荣喜, 高建民. 融合分形特征的风机运行状态辨识方法[J/OL]. 计算机集成制造系统.  
<https://kns.cnki.net/kcms/detail/11.5946.TP.20211228.1356.014.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

## 融合分形特征的风机运行状态辨识方法

郭 怡, 王荣喜<sup>+</sup>, 高建民

(西安交通大学 机械制造系统工程国家重点实验室, 陕西 西安 710049)

**摘 要:** 为保障风机的安全可靠运行, 提出一种融合分形特征的 CNN 卷积神经网络风机运行状态辨识模型。首先将分形等统计特征分析结果建立为初始样本特征数据集; 用 DBSCAN 无监督聚类方法对样本状态进行标记, 用正则化特征选择方法确定最终的样本特征数据集; 最后建立 CNN 卷积神经网络状态辨识模型。结果表明, 模型准确率达到 98.925%, 为实现风机“事前维修”模式提供科学参考和理论指导, 可以有效地应用于复杂机电系统状态辨识领域, 为风力发电机组以及其他复杂机电系统的数据挖掘、模式识别方面提供了基础。

**关键词:** 风电机组; 分形特征; 状态辨识; 特征提取

中图分类号: TP391.4; TM623.7

文献标识码: A

## Operation state recognition method based on fractal features of wind turbines

GUO Yi, WANG Rongxi<sup>+</sup>, GAO Jianmin

(State Key Laboratory for Manufacturing Systems Engineering,  
Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** To ensure the safe and reliable operation of wind turbines, a model based on CNN convolution neural network with fractal features for wind turbine operation state recognition is proposed. Firstly, initial sample feature data set is established based on fractal and other statistical features. Then, DBSCAN unsupervised clustering method is used to mark the sample states, and regularization feature selection method is used to determine the final sample feature data set. Finally, state recognition model is established based on convolutional neural network (CNN). The experiment results show that the accuracy of the model is 98.925%. The model provides scientific reference and theoretical guidance for “prior maintenance” mode of wind turbines, and can be effectively applied for state recognition of complex electromechanical systems, and provides a basis for data mining and pattern recognition of wind turbines and other complex electromechanical systems.

**Key words:** Wind turbine; Fractal feature; State recognition; Feature extraction

## 0 引言

为应对全球气候变化, 我国提出做好“碳达峰、碳中和”工作、推动绿色低碳发展<sup>[1]</sup>。以风力发电

收稿日期: 2021-09-06; 修订日期: 2021-11-25。Received 06 Sep. 2021; accepted 25 Nov. 2021.

基金项目: 国家自然科学基金资助项目(51905409)。**Foundation items:** Project supported by the National Natural Science Foundation, China (No. 51905409)。

---

为代表的清洁能源发电方式，是除水力发电之外技术最成熟的一种可再生能源发电方式，随着清洁能源的需求不断增长<sup>[2]</sup>，非水力可再生能源发电比例的扩大，建立和完善智慧运维系统是保证风电机组工作效率、降低系统维护成本的基础，也是保障以风机为代表的新一代电力系统健康发展的必然要求。然而，为满足风力需求，风电机组常被安装在山区、海上<sup>[3]</sup>和沙漠，工作环境较为恶劣，在追求更大功率、更高效率的同时，风电机组系统的复杂程度也随之增大。长时间在变风载、大温差等极端工况下工作，任何微小异常的积累或部件受到异常扰动都有可能導致系统故障的发生，增加系统维修维护成本。风力发电机组本质上是一种由多种结构机构、电气设备，通过能量、控制信号等多介质网络耦合而成的复杂机电系统<sup>[4]</sup>，其海量、高维的监测数据<sup>[5]</sup>蕴含着丰富的状态信息<sup>[6]</sup>，运行状态具有复杂多变的特点。然而，风机智慧运维在处理海量、高维的风机监测数据时仍然存在数据挖掘不足的问题<sup>[7]</sup>，且由于风电机组运行状态具有时变性，如何充分利用监测数据、挖掘有效信息，实时监测风电机组的运行状态，是风机智慧运维中亟待解决的问题。

为实时监测风机运行状态的异常<sup>[8-9]</sup>，从而有针对性的指导风机维修，降低风机检修难度，合理的风机运行状态辨识方法<sup>[10]</sup>至关重要。风机运行状态辨识是实现风机运行状态实时监测的核心步骤，而采用有效的特征提取方法<sup>[11]</sup>挖掘风机海量、高维监测数据中的信息，是风机运行状态辨识的关键环节<sup>[12]</sup>。现有的时间序列特征提取方法主要分为以下两类：基于统计的特征提取方法和于人工智能的特征提取方法<sup>[13]</sup>。基于人工智能的特征提取方法能够直接处理时间序列，通过神经网络的多隐层结构<sup>[14]</sup>提取和抽象时间序列的分布式特征。曹大理等人<sup>[15]</sup>提出一种基于卷积神经网络的刀具磨损量在线监测模型，采用深度网络自适应地提取特征，并通过加深网络进一步挖掘信号中隐藏的微小特征；熊红林等人<sup>[16]</sup>提出一种基于多尺度卷积神经网络（MCNN）图像识别模型，在划痕缺陷和杂质缺陷识别方面的准确率较高；戴稳等人<sup>[17]</sup>建立了一种基于深度学习特征降维及特征后处理的布谷鸟优化参数的最小二乘支持向量机预测模型，相比于传统的特征选择，摆脱了对先验知识和经验需求的依赖，大幅提升了效率。因此，基于人工智能的特征提取方法泛化能力较强，应用较广泛，可以有效避免人为特征提取的局限，但是容易忽略时间序列不同维度之间的相关性，可解释性较差<sup>[18]</sup>；而基于统计分析的特征提取方法能够兼顾时间序列的形状特征和时间依赖特征<sup>[19]</sup>，因此，该类方法可以有效的弥补人工智能方法在特征提取方面的不足。吴江波等人<sup>[20]</sup>提出了一种基于显著度和统计特征的光谱信号检测与提取算法，降低信号提取过程中背景噪声、基线畸变等不利因素的影响；Wang R 等人<sup>[21]</sup>应用耦合去趋势波动分析（CDFA）研究复杂机电系统的多变量耦合关系，分析并提取分形特征；何涛等人<sup>[22]</sup>针对个性化机械零件的不规则和自相似性，将分形方法与去趋势波动分析相结合，

---

提取零件特征进行缺陷识别。相较于人工智能方法, 统计分析、信号分析为代表的特征提取方法能够提高特征的可解释性, 但精度和泛化能力较差<sup>[23]</sup>。因此, 为充分风机利用监测数据, 以较高的精度提取数据与状态之间的复杂特征关系<sup>[24]</sup>, 同时提高模型的泛化能力, 本文提出一种融合分形特征<sup>[21]</sup>的 CNN 卷积神经网络风机运行状态特征分析方法, 即在人工智能特征提取之前, 先以分形等统计方法建立样本特征数据集, 再以 DBSCAN 无监督聚类<sup>[25]</sup>结果为状态标签, 选择重要程度较高的特征训练 CNN 卷积神经网络模型, 建立一种风机运行状态辨识模型。该模型能够实现对风机阶段运行状态的实时辨识, 同时也为风机状态维修和“事前维修”模式<sup>[26]</sup>提供科学参考和理论指导, 提高风电机组安全性和经济性, 保障风电机组的安全可靠运行, 为风机运行阶段性状态监测<sup>[27]</sup>与故障诊断<sup>[28]</sup>提供科学依据。

本文其余部分结构如下: 第 1 节主要介绍分形特征分析和 DBSCAN 无监督聚类方法原理; 第 2 节提出一种统计与分形特征融合的风机运行状态特征分析方法, 并介绍风机运行状态辨识模型的方法框架; 第 3 节通过实例验证得出本文方法应用于风机运行状态辨识和状态监测的有效性; 第 4 节得出结论。

## 1 分形特征分析和 DBSCAN 无监督聚类原理

为了对风机监测数据进行全面、准确的特征分析, 首先对风机各测点监测数进行整理。其中, 齿轮箱是风电机组传动系统中的重要部件, 其故障发生率较高且故障诊断难度较大<sup>[13]</sup>, 监测风机齿轮箱的运行状态对风机服役质量状态辨识具有重要意义。因此, 本文以风机齿轮箱中间轴<sup>[29]</sup>测点处监测数据作为研究对象, 进行数据预处理, 建立监测数据时间序列样本数据集。由于风机监测数据具有自相似性, 故对数据集中的各时间序列样本进行分形等统计特征分析, 建立样本特征数据集, 以 DBSCAN 无监督聚类方法对样本状态进行标记。

### 1.1 分形特征分析

分形特征分析包括分形维数特征分析和多重分形特征分析<sup>[21]</sup>, 表示组成系统的部分与整体的相似性。以风机为代表的复杂机电系统中虽然存在复杂的耦合关系, 但其确定性特征和自组织结构可以通过自我分形挖掘出来, 而多重分形通过对概率分布函数及其各阶矩的计算, 能够对分形的复杂性和不均匀性进行更细致的刻画。因此, 本节分别对风机监测数据集计算分形维数和 Hurst 指数, 用来描述该时间序列样本的分形特征, 同时采用多重分形谱函数从多个尺度全面反映系统的行为特征差异。

对监测数据的时间序列样本分别计算分形维数和 Hurst 指数, 使用 A.Chhabra 和 R.V.Jensen 提出

的标准盒计数法计算分形维数，其计算原理如式 1 所示。

$$D = Dim(F) = \liminf_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \quad (1)$$

式中， $Dim(F)$  为盒维数； $N_\delta(F)$  为集合  $F$  和坐标网格立方体的相交个数； $D$  为无标度区间内  $(-\log \delta, \log N_\delta(F))$  的双对数图的斜率；在无标度区间内， $-\log \delta$  与  $\log N_\delta(F)$  满足线性回归方程，如式 2 所示。

$$\log N_\delta(F) = -D \log \delta + b \quad (2)$$

因此，盒维数  $D$  可以由  $-\log \delta$  与  $\log N_\delta(F)$  的关系唯一确定。

对监测数据的时间序列样本分别进行多重分形分析，根据如下公式计算得到多重分形特征谱：

$$f(q) = -\lim_{N \rightarrow \infty} \frac{1}{\ln N} \sum_{i=1}^N \mu_i(q, L) \ln[\mu_i(q, L)] \quad (3)$$

$$\alpha(q) = -\lim_{N \rightarrow \infty} \frac{1}{\ln N} \sum_{i=1}^N \mu_i(q, L) \ln[P_i(L)] \quad (4)$$

多重分形特征谱如图 1 所示。

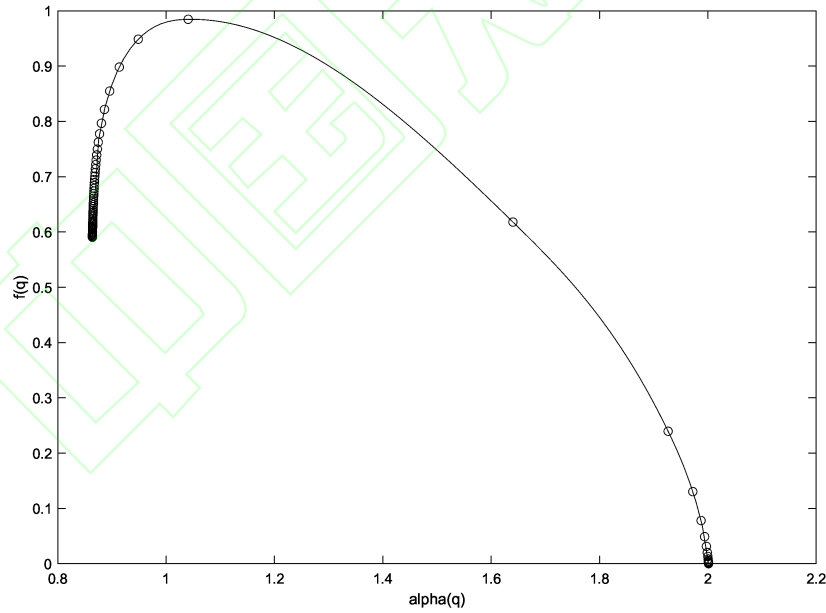


图 1 多重分形特征谱

在图 1 中，横坐标为分形指数  $\alpha$ ，纵坐标为多重分形谱  $f(\alpha)$ 。根据分析结果，记录  $\Delta f$ 、 $\alpha_{\max}$ 、 $\alpha_{\min}$  和  $\Delta\alpha$  等多重分形特征的变化情况，其中， $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$  描述多重分形谱覆盖的局部 holder 指数的范围，反映多重分形程度； $\Delta\alpha$  小则信号趋于单一分形，反之则趋向于多重分形， $\alpha = 1$  表示均匀的点分布， $\alpha < 1$  表示“内密外稀”类型的点分布， $\alpha > 1$  表示“内稀外密”类型的点分布。

1.2 基于密度的无参时间序列聚类

通过运行状态特征分析得到风机监测数据的内在特征，相较于原始监测数据，分析后的结果更容易区分出正常和异常的特征样本，从而通过异常特征样本定位到具有异常的监测时间序列段。

本节选用 DBSCAN 无监督聚类方法对时间序列样本进行状态标记，此方法为基于密度的聚类算法，即某点邻域内的样本点数不小于最小样本数  $MinPts$  时，则以该点为聚类中心进行聚类。对于特征分析得到的样本特征数据集，规定在距离阈值  $\varepsilon$  内的样本点为正常，超出距离阈值  $\varepsilon$  的样本点异常。DBSCAN 聚类方法相较于传统的 K-means 聚类方法，不需要预先设定需要形成的簇类数量，只需给出合适的最小样本数  $MinPts$  和距离阈值  $\varepsilon$ ，就能够找到密度相连对象的最大集合，有效处理噪声点并发现任意形状的空间簇类。选择合适的参数  $MinPts$  和  $\varepsilon$  后，其具体过程及算法描述见表 1，DBSCAN 聚类结果如图 2 所示。

表 1 DBSCAN 算法过程描述

算法名称: DBSCAN	
初始化参数: Minpts=6; $\varepsilon$ =0.2	
输入: 样本特征数据集 D, 最小样本数 MinPts, 距离阈值 $\varepsilon$	
输出: 簇集合 (样本状态标签)	
1	首先将样本特征数据集 D 中的所有对象标记为未处理状态 5%-Co-Zn-B
2	for (数据集 D 中每个对象 p) do
3	if (p 已经归入某个簇或标记为噪声) then
4	Continue;
5	else
6	检查对象 p 的 Eps 邻域 $\varepsilon(p)$ ;
7	if ( $\varepsilon(p)$ 包含的对象数小于 MinPts) then
8	标记对象 p 为边界点或噪声点;
9	else
10	标记对象 p 为核心点, 将 p 邻域内所有点加入新簇 C
11	for ( $\varepsilon(p)$ 中所有尚未被处理的对象 q) do
12	若其 Eps 邻域 $\varepsilon(q)$ 包含至少 MinPts 个对象, 则将 $\varepsilon(q)$ 中未归入任何一个簇的对象加入 C;
13	end for
14	end if
15	end if
16	end for



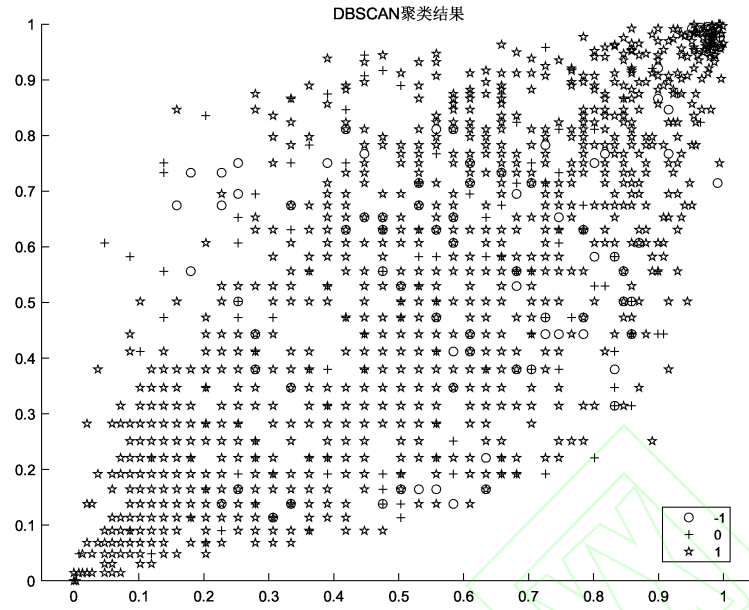


图 2 DBSCAN 无监督聚类结果

图 2 中，横纵坐标分别代表数据归一化之后的测量值，因此没有单位和具体的意义，通过 DBSCAN 方法聚类之后，没有明显的聚类中心。图中不同的点形代表了不同的状态，其中，星形样本点 1 为正常类样本，圆形-1 和十字形样本点 0 分别代表聚类结果中的噪声点和边界点，均归为异常类样本点。分别以 0 和 1 表示正常和异常两类样本，完成对样本的状态标记。

## 2 统计和分形特征融合的运行状态辨识框架

为充分风机利用监测数据，从海量高维的监测数据中准确的学习到内在特征，本节建立风机服役质量状态辨识模型。首先对风机原始监测数据进行预处理，建立时间序列样本的监测数据集；然后提出一种新的风机运行状态特征分析方法，建立初始样本特征数据集；以初始样本特征数据集为依据，采用 DBSCAN 无监督聚类方法标记状态标签，同时，通过正则化特征选择<sup>[30]</sup>，从初始样本特征数据集中，选择重要程度最高的特征建立最终的样本特征数据集；最后，以最终特征选择以及状态标记结果对 CNN 状态辨识模型进行训练和测试，实现风机运行状态辨识。研究路线如图 3 所示。

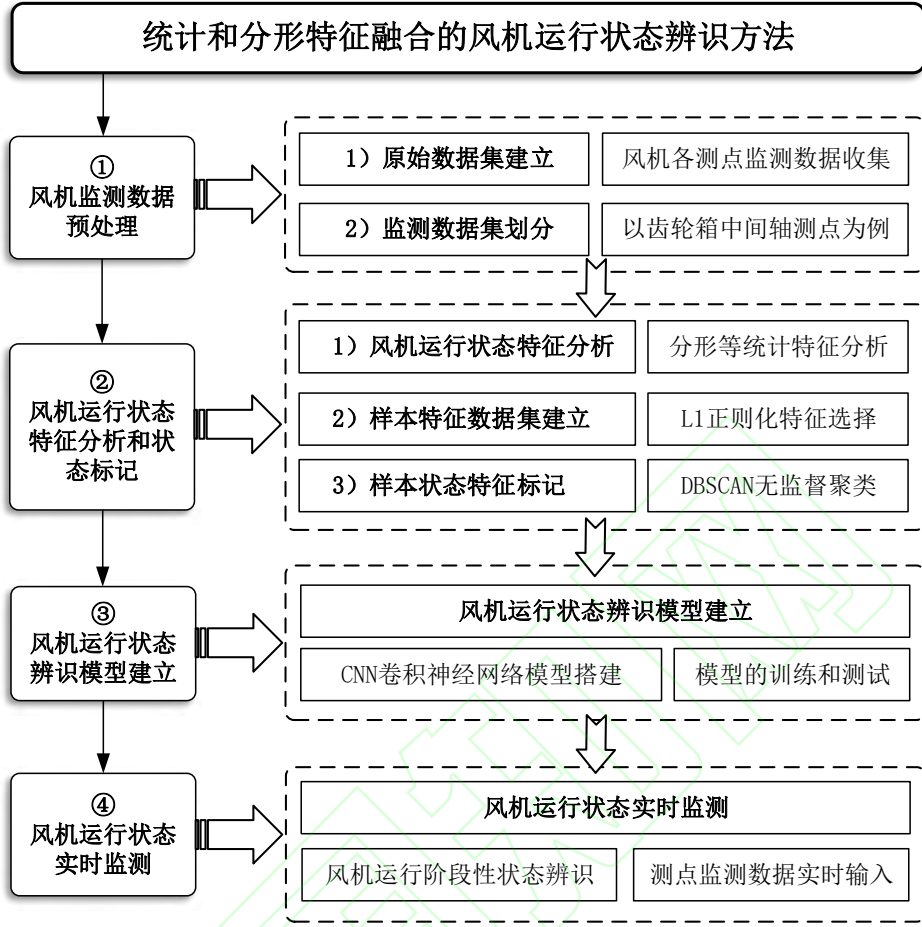


图3 研究路线图

## 2.1 风机监测数据预处理

按照采样频率和采样时间采集并收集风力发电机组齿轮箱中间轴测点对应的监测数据，将该测点的监测数据建立原始数据集，其中，每一列为一个时间序列样本，最终建立共包含  $n$  个时间序列样本的监测数据集  $D$ ：

$$D = [X_1, X_2, \dots, X_L]^T = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L,1} & x_{L,2} & \cdots & x_{L,n} \end{bmatrix}$$

其中， $X_i$  为矩阵中每个样本的第  $i$  个数据点； $x_{i,j}$  为矩阵中第  $j$  个时间序列样本的第  $i$  个数据点；

$L$  为每个时间序列样本的长度； $n$  为时间序列样本个数。

## 2.2 风机运行状态特征分析和状态标记

风机运行状态特征分析包括分形等统计特征分析，将特征分析结果整理成初始样本特征数据集，并以此为对象进行 DBSCAN 无监督聚类，完成状态标记。同时，初始样本特征数据集中可能包含冗



余特征，为降低后续训练模型的复杂程度，故按照特征对模型的重要程度进行特征选择，建立最终的样本特征数据集。

### 2.2.1 风机运行状态特征分析

通过分别分析风机齿轮箱中间轴测点处监测数据集  $D$  中多个时间序列样本的分形等统计特征。

在统计特征分析中，时域特征值是衡量监测数据特征的重要指标。对风机齿轮箱中间轴测点处数据集中的各时间序列样本进行时域统计分析，分别计算其均值、方差、标准差、均方值、均方误差、峰峰值、方根幅值、平均幅值等统计指标和有量纲参数指标。其中，均方值为信号的二阶矩，反映信号的能量；方差为二阶中心矩，反映信号能量的动态分量以及数据的分散程度。

经计算，共得到  $c_1$  个除分形特征外的统计特征。

分形特征分析包括分形维数特征和多重分形特征。不同时间序列样本的分形维数之间有的存在差异，有的较为接近。其中，存在较大差异的样本，其分形维数和 Hurst 指数能够作为特征反映风机服役质量状态的变化情况，而分形维数较为接近的样本，则需要通过进一步提取多重分形特征进行分析。因此，计算监测数据集  $D$  中的  $n$  个时间序列样本分形维数和 Hurst 指数，经计算，得到上述两个分形特征，即  $c_2=2$ ；对时间序列样本进行多重分形分析，得到包括  $\Delta f$ 、 $\alpha_{\max}$ 、 $\alpha_{\min}$  和  $\Delta\alpha$  在内的  $c_3$  个多重分形特征。

### 2.2.2 初始样本特征数据集建立

综合考虑分形等统计特征，将上述特征建立为初始的样本特征数据集  $f$ :

$$f = [f_1, f_2, \dots, f_n]^T = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,c} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,c} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n,1} & f_{n,2} & \cdots & f_{n,c} \end{bmatrix}$$

其中， $f_i$  为矩阵中的第  $i$  个时间序列样本的全部特征； $f_{i,j}$  为矩阵中第  $i$  个时间序列样本的第  $j$  个特征； $c$  为分析得到的全部特征的个数， $c = c_1 + c_2 + c_3$ ； $n$  为时间序列样本个数。

### 2.2.3 样本状态特征标记

选择 DBSCAN 无监督聚类方法对时间序列样本进行状态特征标记，选取适当的最小样本数  $MinPts$  和邻域的距离阈值  $\varepsilon$  参数，按照邻域内点的个数不少于  $MinPts$  的原则进行聚类，得到正常和异常两类样本，分别以 0 和 1 表示，样本状态特征标签为  $y$ 。

#### 2.2.4 风机运行状态特征选择

在以神经网络为代表的人工智能方法中，如果参数太多，则容易造成过拟合，使得模型在训练样本中表现较好，而在实际测试样本中表现较差。为解决上述问题，需要通过特征选择<sup>[28]</sup>降低模型的复杂程度。本节应用正则化特征选择方法，该方法作为结构风险最小化的策略实现，能够有效抑制过拟合。可以从初始特征集中找到包含信息最多的特征，从而以更紧凑的形式表示原始数据，提高模型的泛化能力。

通过  $L_1$  正则化对初始化样本特征数据集中的全部特征进行筛选，提取重要程度较高的特征，得到最终的样本特征数据集，用于训练状态辨识模型。 $L_1$  正则化方法计算 LASSO（最小绝对收缩选择算子），其优化目标函数如式 3 所示。

$$L = \min_{\omega} \sum_{i=1}^m (y_i - \omega^T x_i)^2 + \lambda \|\omega\|_1 \quad (5)$$

式中， $x_i$  和  $y_i$  为训练样本及其相应标签； $\omega$  为权重系数向量，通过最小化目标函数实现权重系数向量  $\omega$  的求解； $\lambda$  为正则化参数，且  $\lambda > 0$ 。

由式(3)可知，优化目标函数中包含训练样本误差项和正则化项两部分，其中，正则化参数  $\lambda$  起权衡作用。

通过  $L_1$  正则化特征选择对当前所分析的全部  $c$  个特征进行筛选，提取重要程度较高的特征，用于训练状态辨识模型，提取出对模型贡献度最高的 8 个特征，分别为：分形维数  $FD$ 、Hurst 指数  $h$ 、均方幅值  $X_{rms}$ 、平均值  $X_{mean}$ 、峰峰值  $X_p$ 、平均幅值  $X_r$ 、分形谱  $\Delta f$  和分形谱  $\Delta \alpha$ 。上述特征构成  $C$  维样本特征数据集  $F$ ：

$$F = [F_1, F_2, \dots, F_8]^T = \begin{bmatrix} F_{1,1}, & F_{1,2}, & \dots & F_{1,C} \\ F_{2,1}, & F_{2,2}, & \dots & F_{2,C} \\ \vdots & \vdots & \vdots & \vdots \\ F_{8,1}, & F_{8,2}, & \dots & F_{8,C} \end{bmatrix}$$

其中  $C=8$ ，对应的状态标签为  $Y$ 。至此，最终的样本特征数据集建立完成。

### 2.3 建立风机运行状态辨识模型

为充分挖掘和提取风机监测数据特征，表征特征与状态之间的对应关系，本节以样本特征数据集训练 CNN 卷积神经网络模型，通过 filter 不断提取局部和总体特征，进而建立样本特征与状态标签的对应关系，实现风机运行状态辨识。

CNN 卷积神经网络具有局部权值共享结构，在处理图像和高维数据过程中可以并行学习，从而降低网络和特征提取过程中数据重建的复杂度。本节将  $C$  维样本特征数据与其对应状态标签  $Y$  划分成训练集和测试集，将训练集样本特征数据集  $F$  作为状态辨识模型的输入，对应的状态标签为  $Y$  作为状态辨识模型的输出，训练基于 CNN 卷积神经网络的状态辨识模型，如图 4 所示。

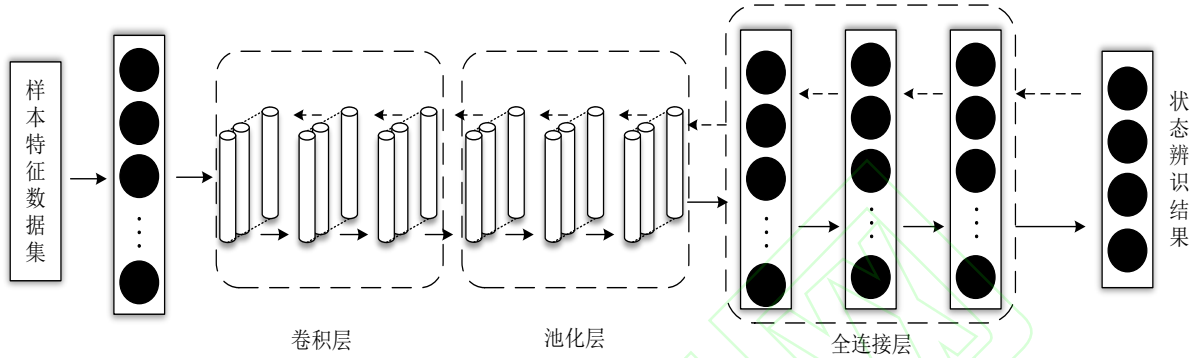


图 4 CNN 卷积神经网络状态辨识模型

如图 4 所示，该模型包含卷积层、池化层、全连接层，最后一层为分类器。以训练集数据作为模型的输入，对应的状态标签作为输出，训练 CNN 模型分类器，输出状态辨识结果；以测试集数据作为 CNN 状态辨识模型的输入，经过分类器输出状态标签，输出为 0 表示模型认为当前输入数据对应的状态为正常状态，输出为 1 表示模型认为当前输入数据对应的状态为异常状态。

### 3 实例分析

收集风机各测点的监测数据，取齿轮箱中间轴测点 2020 年 10 月 17 日的 24 组监测数据，将其划分成包含 1536 个时间序列样本的监测数据集  $D$ ，每个时间序列样本的长度为 128；样本数据均为风机齿轮箱中间轴的振动监测量，其物理含义为振幅，单位为 mm。分析并选择监测数据集  $D$  中各样本的特征值，建立样本特征数据集并标记状态标签，用来训练和测试上述风机服役质量状态辨识模型。

#### 3.1 风机状态辨识模型实例验证

对每个时间序列样本计算样本特征，并采用  $L_1$  正则化特征选择方法选出重要程度最高的 8 个特征，建立最终的样本特征数据集。经计算，时间序列样本的特征值见表 2 至表 4。其中，表 2 为统计特征分析结果，包括均方幅值  $X_{rms}$ 、平均值  $X_{mean}$ 、峰峰值  $X_p$ 、平均幅值  $X_r$ ，列出部分计算结果见表 2。

表 2 时间序列样本的统计特征分析计算结果

样本	均方幅值	平均值	峰峰值	平均幅值	样本	均方幅值	平均值	峰峰值	平均幅值
1	14.634	9.081	64.348	6.868	16	7.574	5.852	24.481	4.904
2	12.229	8.236	43.835	6.365	17	8.392	6.505	27.676	5.397
3	8.443	6.546	34.087	5.623	18	12.997	8.546	44.580	6.578
4	11.838	8.372	49.203	6.936	19	7.200	5.531	24.862	4.524
5	12.172	9.166	49.215	7.691	20	8.944	6.757	35.569	5.599
6	9.393	7.569	29.301	6.485	21	16.517	10.293	73.192	7.816
7	11.240	7.988	53.449	6.594	22	9.744	7.971	33.172	7.016
8	17.581	9.798	107.483	6.811	23	7.464	6.258	17.360	5.487
9	8.389	6.198	28.571	5.056	24	13.060	9.518	49.627	7.801
10	8.783	6.722	37.054	5.618	25	18.273	10.415	98.536	7.826
11	8.921	6.918	28.896	5.593	26	10.622	7.681	35.811	6.066
12	8.342	6.349	26.299	5.249	27	7.648	5.902	28.638	4.986
13	9.328	7.449	24.152	6.262	28	10.658	8.226	38.324	7.072
14	12.748	8.355	55.499	6.558	29	9.509	7.908	22.793	6.915
15	19.569	11.487	122.327	8.579	30	8.212	6.466	22.305	5.392

表 3 和表 4 为分形特征分析结果。采用 Box 维数（盒维数）方法计算分形维数，包括分形维数  $FD$ 、Hurst 指数  $h$ ，列出部分分析结果见表 3。

表 3 时间序列样本的分形维数特征分析结果

样本	分形维数	Hurst 指数	样本	分形维数	Hurst 指数	样本	分形维数	Hurst 指数
1	2.254	-0.254	11	1.966	0.034	21	2.502	-0.502
2	2.224	-0.224	12	2.091	-0.091	22	2.178	-0.178
3	2.117	-0.117	13	2.239	-0.239	23	2.236	-0.236
4	2.315	-0.315	14	2.465	-0.465	24	2.125	-0.125
5	2.286	-0.286	15	2.447	-0.447	25	2.396	-0.396
6	1.977	0.022	16	2.049	-0.049	26	2.279	-0.279
7	2.115	-0.115	17	2.124	-0.124	27	2.162	-0.162
8	2.404	-0.404	18	2.331	-0.331	28	2.387	-0.387
9	2.474	-0.474	19	2.126	-0.126	29	2.236	-0.236
10	2.205	-0.205	20	2.255	-0.255	30	2.242	-0.242

由表 3 可知，不同时间序列样本的分形维数之间有的存在差异，有的较为接近，继续对时间序列样本进行多重分形特征分析，其中部分时间序列样本的特征谱如图 5 所示。

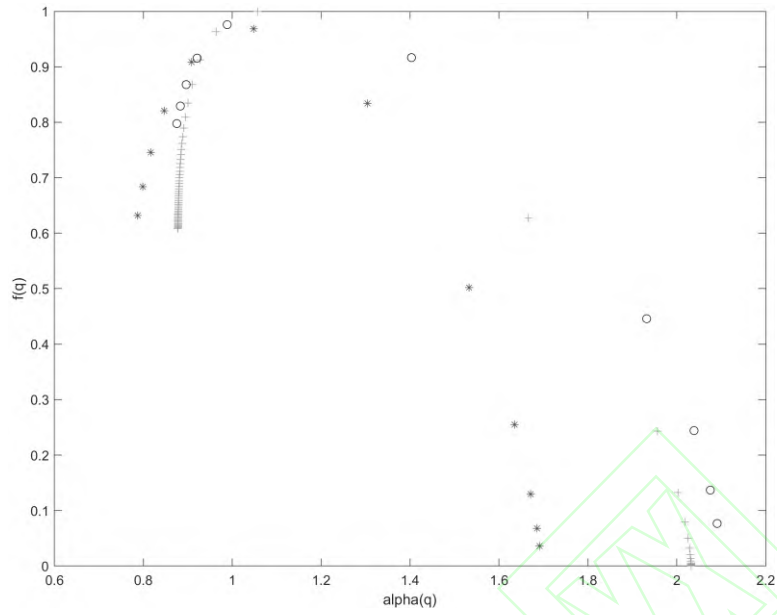


图 5 时间序列样本的多重分形特征谱

图 5 中，不同时间序列样本的多重分形谱的形状和范围存在差异，经过特征选择，提取多重分形特征包括分形谱  $\Delta f$  和分形谱  $\Delta \alpha$ ，部分计算结果见表 4。

表 4 时间序列样本的多重分形特征分析结果

样本	分形谱 $\Delta f$	分形谱 $\Delta \alpha$	样本	分形谱 $\Delta f$	分形谱 $\Delta \alpha$	样本	分形谱 $\Delta f$	分形谱 $\Delta \alpha$
1	0.900	0.902	11	0.886	0.904	21	0.894	0.899
2	0.892	0.906	12	0.918	0.953	22	0.914	0.926
3	0.891	0.903	13	0.893	0.897	23	0.936	0.958
4	0.909	0.925	14	0.912	0.926	24	0.893	0.893
5	0.904	0.924	15	0.900	0.909	25	0.890	0.905
6	0.919	0.956	16	0.879	0.858	26	0.893	0.907
7	0.916	0.922	17	0.866	0.880	27	0.933	0.962
8	0.905	0.911	18	0.907	0.922	28	0.901	0.920
9	0.884	0.885	19	0.924	0.971	29	0.926	0.952
10	0.920	0.927	20	0.916	0.926	30	0.917	0.934

将上述样本特征作为状态辨识模型的输入，对应标签作为模型的输出，训练并测试状态辨识模型。CNN 状态辨识模型的测试结果如图 6 所示。

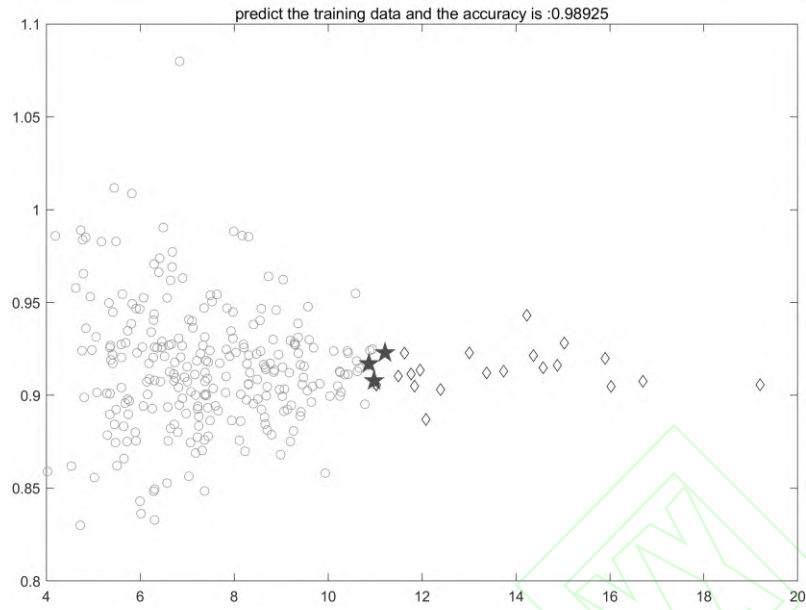


图 6 状态辨识模型的测试结果图

在图 6 中，圆形图案表示输入的特征经过状态辨识模型后输出的状态为正常状态，菱形图案表示输入的特征经过状态辨识模型后输出的状态为异常状态，星形图案表示状态辨识模型输出的结果与实际状态标签不相符。由图 6 可知，模型的错分率较低。

将输出结果与真实状态标签对比，记录模型状态辨识的准确率。经多次实验训练，状态辨识模型准确率记录见表 5。

表 5 本文状态辨识模型准确率

实验次数	1	2	3	4	5	6	7	8	9	10
准确率%	98.505	98.694	98.718	98.925	98.925	99.067	99.283	99.283	98.925	98.925

由表 5 可知，经测试，状态辨识模型的准确率随实验次数上升，最高为 99.283%，最终状态辨识模型的准确率稳定在 98.925%。

综上所述，为验证上述模型的可靠性，从风力发电机组对应测点获取实时测点监测数据，进行预处理和阶段性特征分析，得到样本特征数据集。对重要程度最高的  $C$  个特征作为上述状态辨识模型的输入，并记录模型输出的标签，如果输出的状态标签为 0，表示当前阶段状态为正常状态，则重复上述步骤，持续监测风机运行状态；如果输出的状态标签为 1，表示当前阶段状态出现异常，则应立即采取停机、维修等措施，避免重大事故发生。

### 3.2 状态辨识模型效果比较

本文建立的风机服役质量状态辨识模型不同于直接使用监测数据作为输入的状态辨识模型，该模型通过统计分析和分形分析，计算监测数据中的相应特征，能够从风机海量、高维数据中提取有



效信息，基于时间序列样本特征，训练 CNN 卷积神经网络模型；而以监测数据作为输入的状态辨识模型只能通过 CNN 卷积神经网络提取出有限的分布式特征，且可解释性较差。因此，本文对上述两种状态辨识方法进行对比分析。

将风机齿轮箱中间轴测点的监测数据划分成包含 1536 个时间序列样本的监测数据集  $D$ ，每个时间序列样本的长度为 128；对时间序列样本，采用 DBSCAN 无监督聚类方法，标记状态标签；将时间序列样本数据直接作为 CNN 卷积神经网络模型，样本对应的状态标签作为模型的输出，训练并测试状态辨识模型，测试结果如 7 图所示。

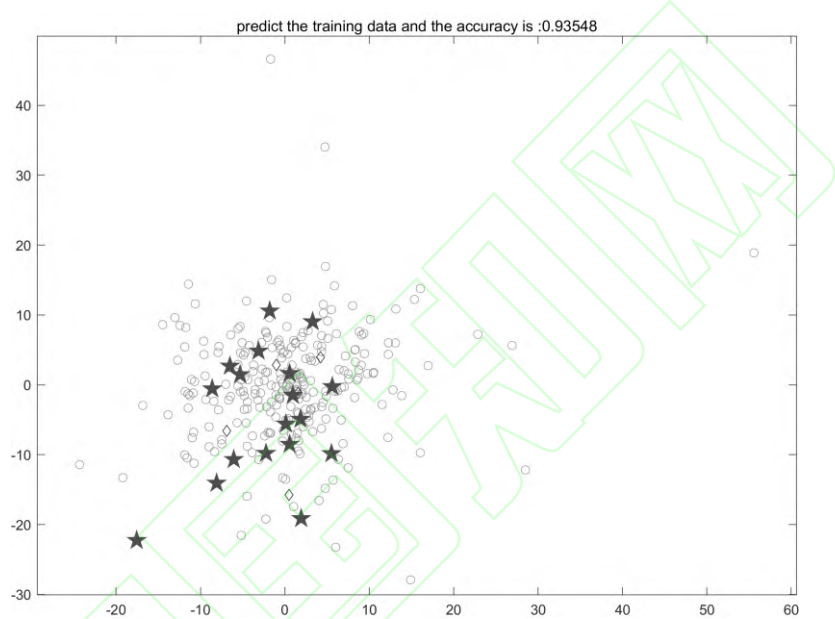


图 7 以监测数据作为输入的状态辨识模型的测试结果图

在图 7 中，圆形和菱形图案分别表示状态辨识模型输出的结果为正常和异常状态，星形图案表示状态辨识模型输出的结果与实际状态标签不相符。由图 7 可知，以测点监测数据直接输入模型进行状态辨识时，相较于本文提出的状态辨识模型，错分率更高。记录该模型的测试准确率见表 6 所示。

表 6 以监测数据作为输入的状态辨识模型准确率

实验次数	1	2	3	4	5	6	7	8	9	10
准确率%	92.115	93.548	92.115	91.398	89.616	92.832	93.548	93.190	93.190	89.964

由表 6 可知，以监测数据作为输入的状态辨识模型，其准确率在 89.6%~93.6%之间波动，准确率为最高为 93.548%，后逐渐稳定在 93.190%，但随着实验次数的上升，准确率开始下降。

综上所述，以监测数据作为输入的状态辨识模型，与本文建立的以数据特征作为输入的风机服役质量状态辨识模型不同，由于输入数据较多，模型复杂度提高，造成模型泛化能力较差，此外，

---

参数增加容易造成模型过拟合，使得测试样本表现下降；而本文基于特征分析和特征提取的结果，建立样本特征数据集，用于状态辨识模型的训练，经对比，该方法对海量、高维监测数据处理更有效，对风机服役质量状态辨识的效果更好，模型准确率更高。

## 4 结束语

本文提出一种新的风机服役质量状态辨识方法，首先基于非人工智能方法提取的时间序列样本特征建立数据集，然后在此基础上应用 CNN 卷积神经网络模型进一步提取分布式特征。针对风机齿轮箱中间轴测点处的监测数据，应用本文构造的基于特征分析和特征提取的状态辨识模型，结果表明，将特征提取结果作为模型输入的状态辨识方法准确率明显高于传统的以监测数据作为输入的状态辨识方法准确率，且模型表现更稳定。

（1）在风机服役质量状态辨识过程中，本文提出的基于特征分析和特征提取的状态辨识模型准确率为 98.925%，而以监测数据作为输入的状态辨识模型准确率不高于 93.548%，且准确率波动较大；

（2）本文提出的风机服役质量状态辨识方法更加适应监测数据海量、高维的特点，能够有效提取信息，将统计特征、分形特征与神经网络分布式特征相融合，相较于以监测数据作为输入的状态辨识模型，准确率提升 5.377%；

（3）本文提出的风机状态辨识方法为实现风机“事前维修”模式提供科学参考和理论指导，可以有效地应用于以风机为代表的复杂机电系统服役质量状态辨识领域，为风力发电机组以及其他复杂机电系统的数据挖掘、模式识别方面提供了基础。

为兼顾时间序列样本的时间跨度和样本特征数据集的容量，本文将风机原始监测数据合并后直接分段划分，该划分方法虽然能够通过异常特征定位异常段样本，但忽略了时间序列样本的连续性，可能会造成一定的信息损失。后续会针对该问题改进时间序列样本的划分方法，进一步提高风机服役质量状态辨识的效果。

## 参考文献:

- [1] ZHOU Xiaoxin, CHEN Shuyong, LU Zongxiang, et al. Technology features of the new generation power system in china[J]. Proceedings of the CSEE, 2018,38(07):1893-1904+2205 (in Chinese) . [周孝信,陈树勇,鲁宗相,等. 能源转型中我国新一代电力系统的技术特征[J]. 中国电机工程学报, 2018,38(07):1893-1904+2205.]
- [2] SHANG Nan, ZHANG Xiang, SONG Yihang, et al. Design of capacity market mechanism adapting to clean energy development and spot market operation[J/OL]. Automation of Electric Power Systems: 1-13[2021-08-11] (in Chinese) . [尚楠,张翔,宋艺航,等. 适应清洁能源发展和现货市场运行的容量市场机制设计[J/OL]. 电力系统自动化: 1-13[2021-08-11].]
- [3] FU Yang, XU Weixin, LIU Lujie, et al. Optimization of preventive opportunistic maintenance strategy for offshore wind turbine considering weather conditions[J]. Proceedings of the CSEE, 2018,38(20):5947-5956 (in Chinese) . [符杨,许伟欣,刘璐洁,等. 考虑天气因素的海上风电机组预防性机会维护策略优化方法[J]. 中国电机工程学报, 2018,38(20):5947-5956.]
- [4] LIANG Y, GAO Z, GAO J, et al. A new method for multivariable nonlinear coupling relations analysis in complex electromechanical system[J]. Applied Soft Computing Journal, 2020,94:106457.
- [5] LI Gang, ZHANG Bo, ZHAO Wenqing, et al. Data science issues in state evaluation of power equipment: challenges and prospects[J]. Automation of Electric Power Systems, 2018,42(21):10-20+177 (in Chinese) . [李刚,张博,赵文清,等. 电力设备状态评估中的数据科学问题:挑战与展望[J]. 电力系统自动化, 2018,42(21):10-20+177.]
- [6] LI Gang, QI Ying, LI Yinliang, et al. Research progress on fault diagnosis and state prediction of wind turbine[J]. Automation of Electric Power Systems, 2021,45(04):180-191 (in Chinese) . [李刚,齐莹,李银强,等. 风力发电机组故障诊断与状态预测的研究进展[J]. 电力系统自动化, 2021,45(04):180-191.]
- [7] HAN Bin, WANG Zhongjie, ZHAO Yong, et al. Development status and planning suggestions of smart wind farms[J]. Thermal Power Generation, 2019, 48(9): 34-39 (in Chinese) . [韩斌,王忠杰,赵勇,等. 智慧风电场发展现状及规划建议[J]. 热力发电, 2019,48(9):34-39.]
- [8] LIU Yunpeng, XU Ziqiang, LI Gang, et al. Review on applications of artificial intelligence driven data analysis technology in condition based maintenance of power transformers[J]. High Voltage Engineering, 2019,45(02):337-348 (in Chinese) . [刘云鹏,许自强,李刚,等. 人工智能驱动的数据分析技术在电力变压器状态检修中的应用综述[J]. 高电压技术, 2019,45(02):337-348.]
- [9] PENG J, Kimmig A, NIU Z, et al. Wind turbine failure prediction and health assessment based on adaptive maximum mean discrepancy[J]. International Journal of Electrical Power and Energy Systems, 2022,134:
- [10] JIANG H, WANG R, GAO J, et al. Zhiyong Gao, Xu Gao. Evidence fusion-based framework for condition evaluation of complex electromechanical system in process industry[J]. Knowledge-Based Systems, 2017,124(MAY15):176-187.
- [11] WANG R, GAO X, GAO J, et al. Uncertain texture features fusion based method for performance condition evaluation of complex electromechanical systems[J]. ISA Transactions, 2021(112):108-121.
- [12] TANG Baoping, LUO Lei, DENG Lei, et al. Research progress of vibration monitoring for wind turbine transmission system[J]. Journal of Vibration, Measurement & Diagnosis, 2017,37(03):417-425+622 (in Chinese) . [汤宝平,罗雷,邓蕾,等. 风电机组传动系统振动监测研究进展[J]. 振动. 测试与诊断, 2017,37(03):417-425+622.]
- [13] SHIHAVUDDIN A, CHEN X, FEDOROV V, et al. Wind turbine surface damage detection by deep learning aided drone inspection analysis[J]. Energies, 2019,12(4):12040676.

- 
- [14] ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017,40(06):1229-1251 (in Chinese) . [周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报, 2017,40(06):1229-1251.]
- [15] CAO Dali, SUN Huibin, ZHANG Jiduo,et al. In-process tool condition monitoring based on convolution neural network[J]. Computer Integrated Manufacturing Systems,2020,26(01):74-80 (in Chinese) . [曹大理,孙惠斌,张纪铎,等. 基于卷积神经网络的刀具磨损在线监测 [J]. 计算机集成制造系统,2020,26(01):74-80.]
- [16] Xiong Honglin, FAN Chongjun, ZHAO Shan,et al. Glass surface defect detection method based on multiscale convolution neural network[J]. Computer Integrated Manufacturing Systems, 2020,26(04):900-909 (in Chinese) . [熊红林,樊重俊,赵珊,等. 基于多尺度卷积神经网络的玻璃表面缺陷检测方法[J]. 计算机集成制造系统,2020,26(04):900-909.]
- [17] DAI Wen, ZHANG Chaoyong, MENG Leilei,et al. Support vector machine milling wear prediction model based on deep learning and feature re-processing[J]. Computer Integrated Manufacturing Systems, 2020,26(09):2331-2343 (in Chinese) . [戴稳,张超勇,孟磊磊,等. 基于深度学习与特征后处理的支持向量机铣刀磨损预测模型[J]. 计算机集成制造系统,2020,26(09):2331-2343.]
- [18] STETCO A, DINMOHAMMADI F, ZHAO X,et al. Machine learning methods for wind turbine condition monitoring: A review[J]. Renewable Energy, 2019,133:620-635.
- [19] REN Shougang, ZHANG Jingxu, GU Xingjian,et al. Overview of feature extraction algorithms for time series[J]. Journal of Chinese Computer Systems, 2021,42(02):271-278 (in Chinese) . [任守纲,张景旭,顾兴健,等.时间序列特征提取方法研究综述[J]. 小型微型计算机系统, 2021,42(02):271-278.]
- [20] WU Jiangbo, JIA Yunwei, YAO Chengbin,et al. Spectrum signal extraction algorithm and application based on saliency and statistics[J]. Spectroscopy and Spectral Analysis,2021,41(07):2294-2300 (in Chinese) . [吴江波,贾云伟,姚城斌,等. 基于统计特征和显著度的光谱信号提取算法[J]. 光谱学与光谱分析,2021,41(07):2294-2300.]
- [21] WANG R, GAO J, GAO Z,et al. Analysis of multifractality of multivariable coupling relationship of complex electromechanical system in process industry[J]. Proceedings of the Institution of Mechanical Engineers, 2017,231(6):1085-1100.
- [22] HE Tao, WANG Xing, WANG Shaodong,et al. Feature extraction and defect identification of parts based on improved MF-DFA[J]. Modular Machine Tool & Automatic Manufacturing Technique,2021(10):105-110 (in Chinese) . [何涛,王幸,王少东,等. 基于改进 MF-DFA 的零件特征提取与缺陷识别[J]. 组合机床与自动化加工技术,2021(10):105-110.]
- [23] TANG Wenhui, NIU Zhewen, ZHAO Baining,et al. Research and application of data-driven artificial intelligence technology for condition analysis of power equipment[J]. [唐文虎,牛哲文,赵柏宁,等.数据驱动的人工智能技术在电力设备状态分析中的研究与应用 [J]. High Voltage Engineering, 2020,46(09):2985-2999 (in Chinese) . 高电压技术, 2020,46(09):2985-2999.]
- [24] WANG Y, YU Y, CAO S,et al. A review of applications of artificial intelligent algorithms in wind farms[J]. Artificial Intelligence Review, 2019(53):1-54.
- [25] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD-96 Proceedings Second International Conference on Knowledge Discovery and Data Mining. 1996, 226-231.
- [26] CHATTERJEE J, DETHLEFS N. Deep learning with knowledge transfer for explainable anomaly prediction in wind turbines[J]. Wind Energy, 2020,23(8):1693.

- 
- [27] LIANG Tao, QIAN Siqi, JIANG Wen, et al. Wind turbine condition monitoring based on SCADA and support vector regression[J]. Control Engineering of China, 2020,27(8):1317-1323 (in Chinese) . [梁涛, 钱思琦, 姜文, 等. 基于 SCADA 和支持向量回归的风机状态监测[J]. 控制工程, 2020,27(8):1317-1323.]
- [28] WANG R, GAO X, GAO J, et al. An information transfer based novel framework for fault root cause tracing of complex electromechanical systems in the processing industry[J]. Mechanical Systems and Signal Processing, 2018,101:121-139.
- [29] YANG Xiyun, MAIHEMUTI Mierzhati, LIU Siqu, et al. State assessment method of capacity reduction at high temperature for wind turbine based on Vine-Copula Bayesian network model[J]. Proceedings of the CSEE, 2020,40(11):3583-3592 (in Chinese) . [杨锡运, 米尔扎提·买合木提, 刘思渠, 等. 基于 Vine-Copula 贝叶斯网络模型的风机高温降容状态评估方法[J]. 中国电机工程学报, 2020,40(11):3583-3592.]
- [30] MARTI-PUIG P, BLANCO-M A, CÁRDENAS J, et al. Feature selection algorithms for wind turbine failure prediction[J]. Energies, 2019,12(3):453.

### 作者简介:

郭 怡(1998-), 女, 河北沧州人, 硕士研究生, 研究方向: 复杂机电系统服役状态管理, E-mail:yiguo\_gy@163.com;

+王荣喜(1985-), 男, 河南南阳人, 副研究员, 博士, 研究方向: 数据驱动的复杂装备服役质量安全与可靠性控制技术、工业大数据分析技术, 通讯作者, E-mail:rongxiwang@163.com;

高建民(1958-), 男, 陕西西安人, 教授, 博士生导师, 研究方向: 复杂装备制造与服役质量安全、智能制造与智能质量、工业大数据与机器学习, E-mail:gjm@xjtu.edu.cn。