

DOI:10.3969/j.issn.1001-4551.2021.12.013

基于随机森林的滚动轴承故障辨识方法研究^{*}

王 兰 兰¹, 朱 捷², 周正平³, 常兆庆⁴

(1. 郑州铁路职业技术学院, 河南 郑州 451460; 2. 郑州航空工业管理学院 管理工程学院, 河南 郑州 450046;
3. 南京航空航天大学 自动化学院, 江苏 南京 211106; 4. 江苏曙光光电有限公司, 江苏 扬州 225000)

摘要:滚动轴承工作时的故障信息难以获取,且其故障信息的辨识也存在困难,针对这一问题,提出了一种基于随机森林(RF)的滚动轴承故障辨识方法。首先,采集了滚动轴承的原始振动信号,并基于时域统计指标提取出了其原始振动数据的特征向量;然后,建立了基于随机森林的轴承故障辨识模型,同时利用测试集验证了故障分类结果,给在测试集分类过程中识别率较高的决策树赋予了较大的权重,使得对应的决策树在未来的分类过程中可以发挥更大的作用;最后,用验证集验证了最终的分类结果,通过多域多通道的滚动轴承故障特征数据集验证了所提方法的有效性。研究结果表明:在不同转速和变工况条件下,基于随机森林的滚动轴承故障诊断方法都能取得良好的轴承故障辨识效果,其分类准确率达到 96%;与采用 BP 神经网络(BPNN)、K-近邻算法(KNN)、支持向量机(SVM)传统分类器的辨识结果相比,采用随机森林辨识方法的分类准确率明显更高。

关键词:滚动轴承;故障辨识;特征提取;随机森林

中图分类号:TH133.33;TH165.3

文献标识码:A

文章编号:1001-4551(2021)12-1599-06

Fault identification method of rolling bearing based on random forest

WANG Lan-lan¹, ZHU Jie², ZHOU Zheng-ping³, CHANG Zhao-qing⁴

(1. Zhengzhou Railway Vocational & Technical College, Zhengzhou 451460, China; 2. School of Management Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450046, China; 3. School of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; 4. Jiangsu Shuguang Optoelectronics Co., Ltd., Yangzhou 225000, China)

Abstract: Aiming at the problem of difficulty in obtaining and identifying the fault information when the rolling bearing was working, a random forest (RF) based fault identification method for rolling bearings was proposed. First, the rolling bearing vibration signal was collected and the feature vector of the original vibration data based on the time-domain statistical indicators was extracted. Then, the fault identification model based on random forest is established, and the test set was used to verify the classification results. The decision tree with a higher recognition rate was given a greater weight, so that the corresponding decision tree could play a greater role in the future classification process. Finally, the verification set was used to verify the final classification result. The effectiveness of the proposed method was verified through a multi-domain and multi-channel rolling bearing fault feature data set. The research results show that the fault diagnosis of rolling bearings based on random forest can achieve good identification results under different speeds and variable operating conditions, and the classification accuracy is as high as 96%; comparing with the traditional classification of back propagation (BP), k-nearest neighbor classification (KNN), and support vector machines (SVM), the comparison results of the detectors show that the accuracy of random forest classification is significantly higher than that of traditional classifiers at different speeds.

Key words: rolling bearing; fault identification; feature extraction; random forest (RF)

0 引 言

滚动轴承是旋转机械的关键部件,一旦其发生故

障,会严重影响设备的安全稳定运行,并可能会造成安全事故,以及难以估计的经济损失。因此,准确识别滚动轴承的故障状态,并对轴承进行故障诊断具有重要

收稿日期:2021-06-15

基金项目:国家自然科学基金资助项目(51775272)

作者简介:王兰兰(1981-),女,江苏徐州人,硕士,讲师,主要从事机械设计、制造及 CAD/CAM 方面的研究。E-mail:23045468@qq.com

的意义^[1,2]。

滚动轴承故障辨识属于典型的多分类问题。随着当前机械设备(旋转机械)逐步向大型化和系统化方向发展,轴承的故障特征也往往具有多样性;并且在轴承的故障进行信息处理过程中,还往往存在进化的知识,因此,在处理这类复杂数据时,采用传统的单一分类器往往会显得力不从心。

为了提高最终的故障辨识效率,利用融合算法整合多个分类器,将是今后轴承故障辨识领域研究的关键之一^[3,4]。

随机森林(random forests, RF)是集成学习的代表算法之一。该算法通过将随机性引入到决策树中,改善了决策树易过拟合的现象,解决了人工神经网络(artificial neural networks, ANN)收敛速度过慢,且易陷于过拟合的问题;同时,也克服了支持向量机(support vector machines, SVM)在处理大样本数据时能力不足的缺点。随机森林算法具有较强的抵抗噪声干扰能力,通过集成多种特征向量,可以有效地提高故障诊断的准确率。

吴海滨等人^[5]通过改进多尺度幅值感知排列熵与随机森林,解决了滚动轴承故障类别辨识问题,以及对故障严重程度进行分类的问题。张西宁等人^[6]提出了一种基于多维缩放和随机森林的轴承故障诊断方法;该方法先利用多维缩放对故障特征集进行了降维处理,再利用随机森林算法对轴承故障进行了辨识。

为了解决传统的随机森林算法存在的故障特征数目过大、复杂度过高及易产生过拟合等问题,李兵等人^[7]提出了一种基于改进随机森林的故障诊断方法;该方法分别改进了决策树 bagging 方式和综合投票方法,可以很好地对电机轴承故障状态进行识别。者娜等人^[8]先利用变分模态分解(VMD)方法,对轴承故障原始信号进行了预处理,并提取出了其特征向量,然后利用随机森林模型对反应堆金属撞击信号进行了识别。陈石等人^[9]提出了一种基于小波包能量熵和随机森林的级联 H 桥多电平逆变器故障诊断方法;该方法利用 PCA 对故障特征进行了降维,然后再利用随机森林算法对逆变器故障进行了分类,提高了逆变器故障识别的准确率。

由此可见,随机森林算法目前已广泛应用于故障诊断、数据挖掘和图像处理等领域^[10-12]。

基于上述分析,笔者提出一种基于随机森林的滚动轴承故障模式辨识方法。首先,提取出滚动轴承振动信号的时域统计指标,作为特征向量;然后,利用随

机森林算法对滚动轴承故障进行诊断;最后,利用综合故障模拟实验台的轴承数据,将其与传统分类器的诊断结果进行对比,以验证该模式辨识方法的有效性。

1 原理介绍

随机森林(RF)^[13,14]是以 CART 决策树作为弱分类器的一种集成学习算法。该算法在决策树的基础上引入随机属性选择,其本质是一个包含多个决策树的组合分类器,由 Leo Breiman 和 Adele Cutler 于 2001 年提出。

随机森林由多个决策树 $\{h(x, \theta_m), m = 1, 2, \dots, M\}$ 组成,其中, θ_m 是相互独立的随机向量,其最终分类结果由多个决策树的综合投票来决定。相比于单个分类器,随机森林算法具有更好的分类结果,可以有效地提升学习系统的泛化能力。

假定一数据集为 $D = \{X_i, Y_i\}$, $X_i \in R_k$, $Y_i \in \{1, 2, \dots, c\}$, 随机地建立一个由多个决策树 $\{h(x, \theta_m), m = 1, 2, \dots, M\}$ 组成的森林,各决策树之间没有关联,当输入样本 x_i 时,随机森林的基决策树都会给出各自的识别结果,最后将投票数多的类别作为该输入的最终分类。

随机森林算法流程如下:

(1) 利用 bootstrap 重采样方法^[15]从原始数据集中随机抽取 n 个样本,组成单决策树的训练集(依次为每一个 bootstrap 训练集训练组建分类数,共产生 n 棵决策树构成一片“森林”,这些决策树均不进行剪枝),每个训练集大小约为原始数据集的 $2/3$ 。虽然随机的有放回采样存在一定的重复率,但可以避免决策树陷入局部最优的情况;

(2) 定义训练样本的输入特征个数为 M 。随机森林中的单棵决策树从根节点开始自上而下递归分裂,每棵决策树在每个节点上分裂时,从 M 个特征向量里随机(无放回)选择 m 个特征向量(m 取 M 的向下整数),然后按照分裂节点不纯度最小的原则从上述特征中挑选出一个最好的特征进行分裂生长。 m 在整个森林生长的过程中保持不变,重复上述过程依次分裂,直至该决策树遍历所有的特征属性;

(3) 在分类阶段,通过集合 n 棵决策树的分类结果,采用相对多数投票原则来决定样本的类别。

随机森林的分类原理决定了它在处理数据时,能够避免由于模型输入数据维数上升而使得分类精度下降的情况,保证了在原始信号特征维数较大的情况下,仍能获得良好的分类效果。

随机森林分类过程如图 1 所示。

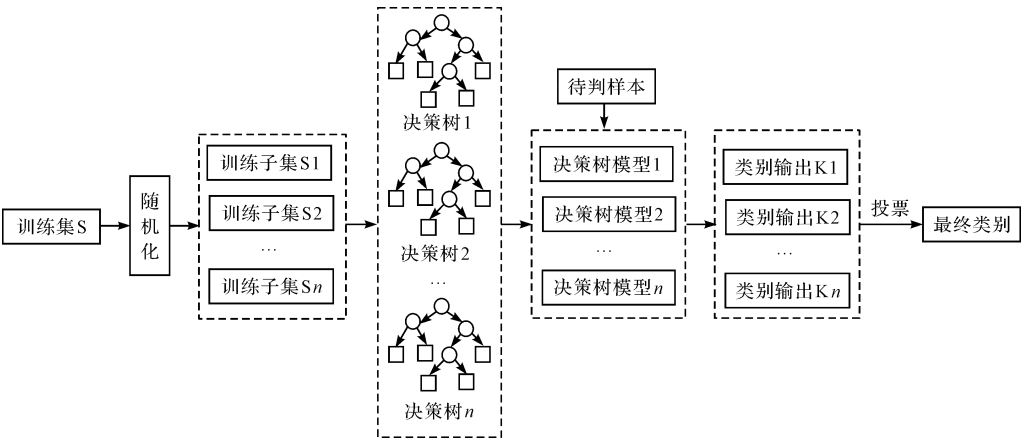


图 1 随机森林分类过程

2 随机森林分类模型

2.1 振动信号采集

此处的原始数据来源于一双跨双转子综合故障模拟平台—HZXT-DS-001 型综合故障模拟实验台。该综合故障模拟平台实物图如图 2 所示。

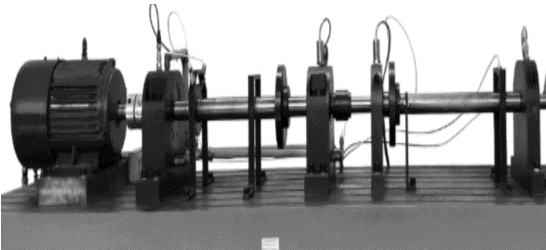


图 2 双跨双转子综合故障模拟平台

该综合故障模拟平台的轴承安装端实物图,如图 3 所示。

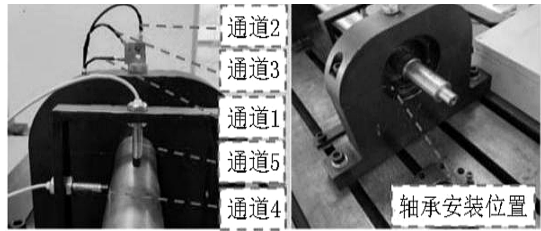


图 3 故障轴承安装端

由图 3 左半部分可见,采样通道共 5 个。实验中,笔者测得滚动轴承 NSK6308 在滚动体故障、保持架故障、内圈故障、外圈故障、正常情况共 5 种状态下的振动信号各 80 组。其中,训练样本 30 组,测试样本 20 组,剩下 30 组作为验证样本。

采样转速分别为 2 600 r/min、2 800 r/min、3 000 r/min、3 200 r/min,采样频率为 8 kHz。

2.2 振动信号处理及特征提取

采样转速在 3 000 r/min 下,在通道 1 中采集到滚

动轴承的 5 种运行状态结果,如图 4 所示。

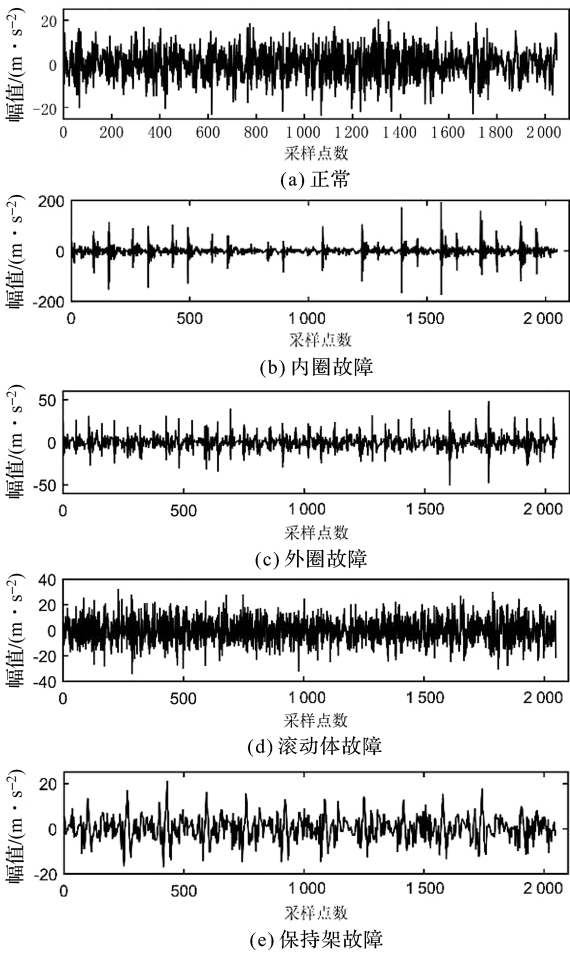


图 4 滚动轴承在部分故障下的振动信号

为了降低采集到的原始信号中环境噪声等干扰信息,笔者先将原始振动信号经一维小波消噪,然后再提取信号中所包含的时域、频域和时频域的多域特征。

最后笔者提取到的 16 个多域特征如表 1 所示。

由于有 5 个通道,此处共可得到 $5 \times 16 = 80$ 维的特征,笔者将原始的特征数据集归一化处理至 $[-1, 1]$ 的区间。

表 1 为 chi 通道构造的原始特征参数 ($i=1,2,3,4,5$)

序号	特征	序号	特征
1	峰峰值	7	频谱二阶矩
2	波形指标	8	标准偏差频率
3	峰值指标	9	频率峭度
4	脉冲指标	10	均方根频率
5	裕度指标	11	中心频率
6	峭度指标	12~16	IMF 分量奇异值

2.3 基于随机森林的故障辨识模型建立

基于随机森林算法的故障辨识具体流程如下:
输入:高维特征训练集 T 和测试集 U ;
输出:故障类别集 S ;
(1)采集振动数据,并进行处理;
(2)从时域、频域和时频域方面提取特征指标,构建原始数据特征集并进行数据归一化处理;
(3)随机森林故障辨识
1) For $i=1:N$ (i 为基学习器数目,此处的基学习器为决策树)
①对训练集 T 采用 Bagging 方法中的重采样算法,得到训练子集 T_i ;
②利用训练子集 T_i 训练出决策树,为 H_i ;
③将测试集 U 输入至 H_i ,得到结果 S_i ;
2) 对 S_i 利用相对多数投票法得到最终分类结果 S ;
End

随机森林中基决策树数目与辨识精度的关系曲线图如图 5 所示。

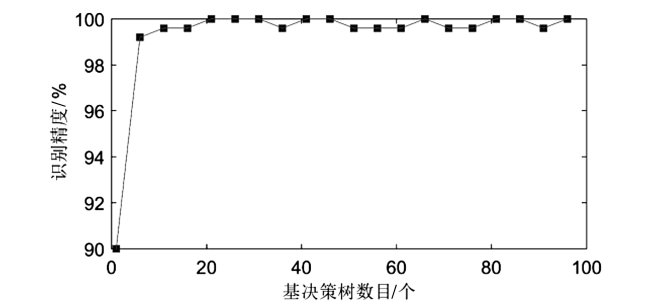


图 5 随机森林中基决策树数目与辨识精度的关系

从图 5 中可以看出:随着基决策树数目增多,随机森林的分类准确率逐渐增加。

图 5 中,由于基决策树数目为 30,笔者在保证单个决策树的识别精度不低于 65 的基础上,将每个输入值各决策树的特征数设为 35。

3 实验与结果分析

3.1 不同转速下随机森林法辨识结果

为了对随机森林故障辨识模型的故障分类能力进

行验证,在 2 600 r/min、2 800 r/min、3 000 r/min、3 200 r/min 几种采样转速下,笔者分别对几种故障状态进行故障辨识。

不同转速下,随机森林的分类识别率如表 2 所示。

表 2 随机森林不同转速下故障辨识结果				
转速/ ($r \cdot \min^{-1}$)	滚动体 故障	保持架 故障	内圈 故障	外圈 故障
分类准确率/(%)				
2 600	100	98	100	100
2 800	98	99	100	97
3 000	100	99	97	100
3 200	96	100	100	99

从表 2 中看出:在不同转速下,随机森林对几种故障状态的分类准确率都比较高。

3.2 不同分类器的辨识精度对比

为了验证随机森林法在故障辨识精度上的优越性,笔者将该算法获得的分类准确率与分类器,即 BP 神经网络 (back propagation neural network, BPNN)、k 近邻分类 (k-nearest neighbor classification, KNN)、支持向量机 (support vector machines, SVM) 获得的分类准确率进行了对比。

不同分类器的分类结果如表 3 所示。

表 3 不同分类器的分类准确率				
转速 ($r \cdot \min^{-1}$)	RF	BP	KNN	SVM
分类准确率/(%)				
2 600	99	85	95	90
2 800	100	92.5	95	90
3 200	100	90	100	95

从表 3 中可以看出:
在不同转速下,随机森林的分类准确率高于其他 3 种分类方式。这是因为相比传统的分类器 BP、KNN、SVM,随机森林不需要繁琐的参数寻优过程和最优的特征向量选择;

同时,随机森林算法可以利用简单的弱分类器 (决策树),以投票的形式产生最终的分类结果,从而可以有效地提高诊断的精度。

3.3 变工况下随机森林法的辨识精度

在变工况和转速波动条件下,为了验证随机森林方法的辨识精度,笔者分别选择 2 600 r/min ~ 2 800 r/min,2 800 r/min ~ 3 000 r/min,3 000 r/min ~ 3 200 r/min,2 600 r/min ~ 2 800 r/min ~ 3 000 r/min 的 4 个转速范围对随机森林方法的辨识性能进行验证。

在变工况下,随机森林分类识别率如表 4 所示。

表 4 表明,在转速波动时,随机森林方法仍然具有较高的辨识精度。

表 4 随机森林在变工况下分类准确率

转速/ ($r \cdot \min^{-1}$)	滚动体 故障	保持架 故障	内圈 故障	外圈 故障
	分类准确率/(%)			
2 600 ~ 2 800	100	100	100	100
2 800 ~ 3 000	100	100	100	100
3 000 ~ 3 200	100	100	100	100
2 600 ~ 2 800 ~ 3 000	100	100	100	100

3.4 不同训练样本数下随机森林法的辨识精度

为了进一步测试不同训练样本数目对随机森林法故障辨识精度的影响,笔者随机设置训练样本数目分别为 10、15、20、25、30,经测试,得到了不同训练样本数目与故障识别率的关系,如图 6 所示。

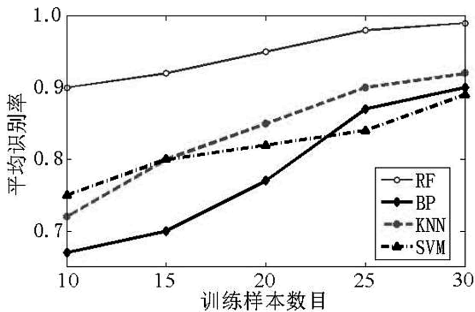


图 6 不同训练样本比例对应的平均识别正确率

从图 6 可以看出:

- (1)随着训练样本数目的增加,几种分类器故障识别准确率不断增大;
- (2)在训练样本数较少的情况下,随机森林分类识别准确率一直保持稳定,而其他传统分类器的故障分类准确率则受训练样本数目的影响较大。

3.5 不同分类器的抗噪性对比

由于在实际情况下采集到的滚动轴承振动信号往往伴随噪声等干扰信息,为了验证随机森林法的抗干扰能力,笔者在所建立的测试集中加入随机扰动。

扰动矩阵定义为 $M=0.2 * F(120,16)$, F 为随机函数,旨在产生区间为 $[0,1]$ 的随机数;120 表示测试样本的数量为 120 个,16 表示测试集特征参数的个数为 16 个。

各分类方法故障识别率如图 7 所示。

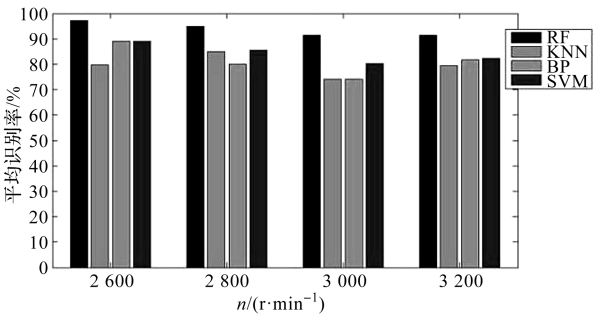


图 7 各分类方法在抗噪性方面的对比

图 7 结果表明:

与传统的分类方法相比,随机森林具有更好的抗噪声等干扰能力。

由此可见,笔者提出的方法在滚动轴承故障辨识方面有良好的工程实用价值。

4 结束语

针对滚动轴承故障诊断中存在的特征选取困难,以及分类精度较低的问题,笔者提出了一种基于随机森林的滚动轴承故障辨识方法。首先,提取出滚动轴承振动信号中的时域特征,将其作为特征向量;然后,以这些特征向量作为随机森林算法的输入,对轴承进行了故障诊断;最后,利用 SQI 实验平台轴承数据,对该方法的可行性与有效性进行了验证。

研究结果表明:

- (1)在不同转速下,随机森林对几种故障状态分类准确率都比较高;并且在变工况和波动转速下,随机森林方法也具有良好的辨识性能;
- (2)在不同转速下,与 BP、KNN、SVM 3 种分类方式相比,随机森林的分类准确率更高;
- (3)在样本数目较少的情况下,随机森林仍然能保持稳定的分类识别率;
- (4)相较于 BP、KNN、SVM 分类器,随机森林算法具有良好的抗噪声干扰能力。

在 VMD 的分解中,分解个数 K 值的选择决定着 VMD 分解的结果,因此,在后续的工作中,笔者将对如何合理有效地选取 K 值做进一步的研究。

参考文献 (References):

[1] 彭 成,王松松,贺 婧,等. 基于离散小波变换和随机森林的轴承故障诊断研究[J]. 计算机应用研究,2021,38(1):101-105.

[2] YANG Y, ZHENG H, LI Y, et al. A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network[J]. ISA Transactions, 2019,91:235-252.

[3] 尹爱军,梁子晓,张 波,等. 基于主曲线相似度的轴承健康状态评估方法[J]. 振动、测试与诊断,2019,39(3):625-630.

[4] 梁治华,曹江涛,姬晓飞. 基于 EEMD 和 CS-SVM 的滚动轴承故障诊断研究[J]. 机电工程,2019,36(6):622-627.

[5] 吴海滨,陈寅生,张庭豪,等. 改进多尺度幅值感知排列熵与随机森林结合的滚动轴承故障诊断[J]. 光学精密工程,2020,28(3):621-631.

[6] 张西宁,张雯雯,周融通,等. 基于多维缩放和随机森林的轴承故障诊断方法[J]. 西安交通大学学报,2019,53

- (8):1-7.
- [7] 李 兵,韩 睿,何怡刚,等.改进随机森林算法在电机轴承故障诊断中的应用[J].中国电机工程学报,2020,40(4):1310-1319.
- [8] 者 娜,刘才学,杨泰波,等.VMD 和随机森林在反应堆金属撞击信号识别中的应用研究[J].振动与冲击,2021,40(5):102-107.
- [9] 陈 石,张兴敢.基于小波包能量熵和随机森林的级联 H 桥多电平逆变器故障诊断[J].南京大学学报:自然科学,2020,56(2):284-289.
- [10] MASETIC Z, SUBASI A. Congestive heart failure detection using random forest classifier[J]. **Computer Methods & Programs in Biomedicine**, 2016, 130:54-64.
- [11] 张 钰,陈 珺,王晓峰,等.随机森林在滚动轴承故障诊断中的应用[J].计算机工程与应用,2018,54(6):100-104.
- [12] ALAM M S, VUONG S T. Random Forest Classification for Detecting Android Malware[C]. Proceedings of 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, Beijing: IEEE, 2013.
- [13] BREIMAN L. Random forests[J]. **Machine Learning**, 2001, 45(1):5-32.
- [14] 周海军,王超伟,周国敬,等.基于随机森林的离心泵滚动轴承故障诊断[J].中国舰船研究,2020,15(3):129-135.
- [15] EFRON B. Bootstrap Methods: Another Look at the Jackknife[M]. Breakthroughs in Statistics. New York: Springer, 1979.

[编辑:雷 敏]

本文引用格式:

王兰兰,朱 捷,周正平,等.基于随机森林的滚动轴承故障辨识方法研究[J].机电工程,2021,38(12):1599-1604.

WANG Lan-lan, ZHU Jie, ZHOU Zheng-ping, et al. Fault identification method of rolling bearing based on random forest[J]. Journal of Mechanical & Electrical Engineering, 2021, 38(12):1599-1604.

《机电工程》杂志;http://www.meem.com.cn

(上接第 1591 页)

参考文献 (References):

- [1] 林思苗,张艳荣,郭丽萍.改进 WMRA 在滚动轴承故障诊断研究中的应用[J].机电工程,2017,34(11):33-36,81.
- [2] 杨文志,张茹军,安文斌.基于 LCD 与峭度-能量比准则的滚动轴承故障诊断研究[J].机电工程,2020,37(5):53-57.
- [3] 王立臣,梁 浩.滚动轴承故障诊断技术现状及发展趋势[J].电子测试,2013,(22):141-143.
- [4] 李道军,李廷锋,刘德平.基于 LMD 与改进 SVM 的轴承故障诊断方法[J].机械制造,2021,59(6):5.
- [5] 李 恒,张 氢,秦仙蓉,等.基于短时傅里叶变换和卷积神经网络的轴承故障诊断方法[J].振动与冲击,2018,37(19):124-131.
- [6] KHAN N A, SANDSTEN M. Time-frequency image enhancement based on interference suppression in Wigner-Ville distribution[J]. **Signal Processing**, 2016, 127:80-85.
- [7] 孙 成.基于 FFT、小波变换和神经网络的电力系统谐波检测方法研究[D].山东理工大学电气与电子工程学院,2018.
- [8] YU Gang, YU Ming-jin, XU Chuan-yan. Synchroextracting transform[J]. **IEEE Transactions on Industrial Electronics**, 2017, 64(10):8042-8054.
- [9] 王 琛.基于 SVD-PE 与神经网络的滚动轴承故障诊断模型[J].自动化与仪表,2021,36(6):85-89.
- [10] 朱 丹,苏燕辰,燕春光.基于 SVD-MOMEDA 的高速列车齿轮箱轴承故障诊断[J].机车电传动,2020,2:149-153,157.
- [11] LU Ling, XU Wei, QIAO San-zheng. A fast SVD for multilevel block Hankel matrices with minimal memory storage[J]. **Numerical Algorithms**, 2015, 69(4):875-891.
- [12] 沈科宇,严 华,申 雨.基于 S 变换与奇异值分解的大型起重机减速箱故障信号降噪方法[J].中国设备工程,2020,6:57-59.
- [13] 张 龙,张 磊,熊国良,等.基于 S 变换和改进 SVD 的滚动轴承智能诊断方法[J].仪表技术与传感器,2016,1:63-66.
- [14] 于 刚.挖掘机振声信号时频分析研究与应用[D].山东大学,2016.
- [15] 张 安.基于 SVD 的滚动轴承故障特征提取[D].石家庄铁道大学,2019.
- [16] HOU Fa-tao, CHEN Jin, DONG Guang-ming. Weak fault feature extraction of rolling bearings based on globally optimized sparse coding and approximate SVD[J]. **Mechanical Systems & Signal Processing**, 2018, 111:234-250.
- [17] GOLAFSHAN R, SANLITURK K Y. SVD and Hankel matrix based de-noising approach for ball bearing fault detection and its assessment using artificial faults[J]. **Mechanical Systems and Signal Processing**, 2016, 70-71:36-50.
- [18] 张文凤,周 俊.基于 Dropout-CNN 的滚动轴承故障诊断研究[J].轻工机械,2019,37(2):62-67.
- [19] 郑 煜,王 凯,杨利红.滚动轴承早期故障优化自适应随机共振诊断法[J].轻工机械,2020,38(2):74-76,83.
- [20] 孟志鹏.基于双树复小波的机电设备故障诊断方法及应用[D].北京工业大学,2014.

[编辑:雷 敏]