

# 基于大数据深度迁移模型的机械故障诊断\*

曾德贵<sup>1</sup> 赵建明<sup>2</sup>

(1. 广安职业技术学院 电子与信息工程学院, 四川 广安 638000; 2. 电子科技大学 电子科学与工程学院, 成都 610054)

**摘要:** 针对现有机械故障诊断方法在小样本条件下检测率低的不足, 提出一种基于深度迁移学习模型的机械大数据故障诊断方法研究。构建深度学习模型, 计算模型的稀疏特性和分类错误率指标, 并基于此提取机械大数据的故障特征类型; 针对实际检测中有效样本较少的不足, 利用迁移学习方法将实验数据用于辅助机械故障特征大数据的训练与测试, 不断地调整输出结果并提高对故障点的定位与诊断精度。实验结果表明, 提出诊断方法的 G-Mean 指标优于现有方法, 在故障比为 1:1000 的条件下, 故障查准率仍可达到 96.34%。

**关键词:** 机械大数据; 故障诊断; 深度学习; 迁移学习; G-Mean 指标

**中图分类号:** TH162; TG506

**文献标识码:** A

## Mechanical Fault Diagnosis Based on Big Data Deep Migration Model

ZENG De-gui<sup>1</sup> ZHAO Jian-ming<sup>2</sup>

(1. Department of Electronics and Information Engineering, Guang'an Vocational & Technical College, Guang'an Sichuan 638000, China; 2. School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)

**Abstract:** In view of the low detection rate of existing mechanical fault diagnosis methods under the condition of small samples, this paper proposes a research on mechanical big data fault diagnosis method based on deep transfer learning model. Build deep learning model, calculate the sparse characteristics and classification error rate index of the model, and extract the fault feature types of mechanical big data based on this; in view of the shortage of less effective samples in the actual detection, use the migration learning method to use the experimental data to assist the training and testing of Mechanical fault feature big data, constantly adjust the output results and improve the location of fault points Diagnostic accuracy. The experimental results show that the g-mean index of the proposed diagnosis method is better than that of the existing method. Under the condition that the fault ratio is 1:1000, the accuracy of the fault can still reach 96.34%.

**Key words:** mechanical big data; fault diagnosis; deep learning; transfer learning; G-Mean index

## 0 引言

在工业互联网、机械自动化及无线传感技术的推动下, 工业机械设备的功能性和单机作业效率得到持续改善, 设备复杂程度也不断提高<sup>[1]</sup>。大型机械设备往往包括上千个的机械零部件, 设备的关键部件大都安装多源传感器, 用于监控设备的运行状态, 但机械故障监控大数据的海量性特征, 给设备的在线监控带来极大的难度<sup>[2-3]</sup>, 传统以人工设备点检为主的方式, 无法实现设备运行状态的在线监测, 一旦设备发生故障会给企业带来难以估量的损失<sup>[4]</sup>。工业机械设备的故障监测数据价值密度低、重复性高<sup>[5-6]</sup>, 设备在长期运转过程中形成的故障监测数据库中仅有极少量有用信

息, 长时间频繁地停机重复检测代价过高, 而且直接影响到工厂的生产效率。近年来随着计算机人工智能网络技术的发展<sup>[7]</sup>, 大量机器学习和人工智能方法被广泛地应用到机械故障诊断领域当中, 并取得了阶段性的研究成果。

如集成经验模态方法<sup>[8]</sup>、包络数据拟合方法<sup>[9]</sup>、局部均值信号分解方法等<sup>[10]</sup>, 其中集成经验模态分析方法的主要优势在与能够自适应分析故障类别, 不需要再另外选择与之匹配的基函数<sup>[11]</sup>; 包络数据拟合方法的适用性要更强, 针对信号特征的提取效果要优于集成经验模态方法; 局部均值分解方法在处理非线性故障信号时具有优势, 能够有效减少模态混叠带来的信

收稿日期: 2020-02-12

\* 基金项目: 教育部科技中心基金项目“天诚汇智”创新促教基金(2018B01005)

作者简介: 曾德贵(1972—), 男, 四川泸县人, 广安职业技术学院教授, 研究方向为电子信息( E-mail: 641803107@qq.com)。通讯作者: 赵建明(1964—), 男, 四川洪雅县人, 电子科技大学教授, 博士, 研究方向为功率半导体、集成电路的研发, 功率电子学控制与应用。

号处理不稳定性<sup>[12]</sup>。但现有的故障诊断方法在数据处理时,要求采集到系统被测试数据与原始训练数据具有相同的分布特征,在实际诊断中训练数据的提取总数量难以满足要求,导致无法针对目标问题建模,影响模型的诊断精度。

为此本文提出一种基于机械大数据深度迁移模型的故障诊断方法研究,深度学习的优势在于能够处理海量机械大数据的非线性问题,准确提取机械设备的故障特征;而迁移学习模型具有高效的知识传递能力,能够达到降低不同领域之间的差异性给数据针对带来的不利影响,提高模型的鲁棒性和泛化能力,以达到提高机械故障在线诊断精度的目的。

## 1 深度学习机械故障建模与特征提取

设备运转中多源传感器采集的设备有效故障样本较少,且故障特征提取难度较大,浅层机器学习方法的故障特征提取效果往往难以满足实际工作中的故障诊断要求。与传统神经网络、支持向量机等方法相比,深度学习的智能化水平、非线性表征数据处理的能力<sup>[13]</sup>、和网络收敛性能更强<sup>[14]</sup>,能够解决传统方法盲目增加网络训练层数的不足,降低诊断网络整体成本开销。建立深度学习模型能够自适应地从振动信号频谱中提取出故障点的特征信息,但基于深度学习方法建模既要保证机械故障特征的稀疏性最低,还要降低机械故障分类的错误率。首先故障的稀疏特征有助于识别出故障信号频谱内部的结构特征<sup>[15-16]</sup>,改善机械故障特征的传达效果。如果用 $\xi$ 表示一种深度学习模型,那么模型的稀疏特征和分类错误率分别表示为 $\eta(\xi)$ 和 $\chi(\xi)$ :

$$\eta(\xi) = \sum_{j=1}^n \xi \cdot (\tau \parallel \delta) \quad (1)$$

其中, $n$ 为隐含层的节点数量, $\xi$ 为深度学习模型的相对熵, $\tau$ 表示为模型的稀疏参数, $\delta$ 为给定的故障样本集 $(z_m)_{m=1}^M$ 编码器第 $j$ 个节点的激活度参数, $M$ 为故障样本的总数量。从理论上来说稀疏参数值 $\tau$ 趋近于零,且相对熵值会随着激活度参数偏离程度的增加而变大。深度学习模型 $\eta(\xi)$ 的稀疏性特征,深度学习模型中所提取的机械故障稀疏性越低,证明模型对故障特征的表达越强,越有利于故障诊断结果准确率的提高,稀疏表达式中的 $\xi \cdot (\tau \parallel \delta)$ 部分和第 $j$ 个节点的激活度参数可以分别表示为:

$$\begin{cases} \xi \cdot (\tau \parallel \delta_j) = \tau \log \frac{\tau}{\delta_j} + (1 - \tau) \log \frac{(1 - \tau)}{(1 - \delta_j)} \\ \delta_j = \frac{1}{M} \sum_{m=1}^M \xi(x_m) \end{cases} \quad (2)$$

为了便于后续的故障特征比较,在 $[0, 1]$ 的区间范围内对稀疏特征函数 $\eta(\xi)$ 做归一化处理。基于深度学习模型训练故障数据时,如果故障数据被错分类的概率过高证明故障诊断模型存在问题,故障样本集 $M$ 中被错分的样本数量为 $M_e$ ,则深度学习模型的错误分类率 $\chi(\xi)$ 表示为:

$$\chi(\xi) = \frac{M_e}{M} \quad (3)$$

错误分类率的取值区间也为 $[0, 1]$ ,取值越小证明模型的更好。令 $G(\xi)$ 为机械大数据故障诊断模型的目标函数,按照稀疏特征 $\eta(\xi)$ 最低化和分类错误率 $\chi(\xi)$ 的值最小化的原则,建立机械大数据故障特征深度提取模型:

$$\begin{cases} \min [G(\xi)] \\ G(\xi) = [\eta(\xi) \chi(\xi)] \end{cases} \quad (4)$$

在模型求解的过程中利用深度学习模型和分解优化的方案,同时对多个单目标故障诊断问题求解,求解的过程可以描述为:

$$\begin{cases} \min G(J | \kappa, h^*) = \max [\kappa_i | \eta(J) - h^*] \\ \text{s. t. } s \in T \end{cases} \quad (5)$$

其中, $J$ 为深度学习模型的决策向量, $T$ 决策向量的空间,其中 $n, \varphi$ 为影响到决策向量取值的模型参数,与决策向量的关系为 $J = (n, \varphi)$ , $h^*$ 是深度学习模型的参考向量,满足如下条件:

$$h^* = \min [\eta(J) | J \in T] \quad (6)$$

机械大数据模型的故障特征提取算法流程如下所示:

- (1) 将机械故障特征的样本集 $[z_m]_{m=1}^M$ 输入深度学习模型;
- (2) 优化深度学习模型 $\xi$ 的全局各项参数,根据多个单目标子集的数量确定出原始种群的规模,并计算出激活度参数、稀疏参数、及异步方差值;
- (3) 初始化种群得出权重矢量,并得到新的决策向量。通过反复迭代得到更新后的决策向量 $J'$ 的适应度函数,当满足 $\eta(J) < h^*$ 时更新目标参考点,当 $\eta(J) = h^*$ 时参考点的位置不变;
- (4) 重复步骤(3)逐步增加迭代循环的次数,直到得到最优的决策向量,得到全局最优解,并从海量故障特征数据中定位故障点,并准确识别出机械故障特征类型。

利用深度学习模型提取机械大数据的故障特征并确定故障类型和故障点的位置,还需要解决实际故障诊断中有效样本量过小的问题。在实验室环境下可以模拟故障信号的特征,得到更多的有效样本数据,用于实际故障检测的故障特征匹配。本文在构建机械大数据深度学习模型的基础上,利用迁移学习的方法解决深度特征提取中对有效样本数量过于依赖的问题,并改善诊断网络的性能。深度迁移学习模型可以将实验室中的大量有效辅助数据,应用到机械大数据的智能诊断领域,并基于深度神经网络模型反复训练数据,即可以显著改善对于复杂目标故障问题的诊断精度。

## 2 基于大数据迁移学习的机械故障诊断

在机械大数据深度迁移学习过程中,深度学习模型数据库中已知的故障数据集被称为源域,从实验室测试系统中获取的待迁移故障数据集被称为目标域。深度迁移学习的目的,就是找寻到实验室故障数据集

与实际的故障数据集之间的共性特征,利用两者之间的共同点弥补提取到的有效故障样本特征过少的不足,为机械大数据故障数据集的训练提供比对数据上的支持。在源域与目标域的数据转换中,迁移学习方法的目的是要找到两个域中的共同特征来辅助大数据故障特征的训练与筛选,两个域中所包含的共同特征越多,就越适合深度迁移学习;如果源域和目标域中的数据共同性特征过少,证明实验数据集的筛选存在问题或参数设置存在错误,没有复制和利用的价值。此外,在数据深度迁移过程中深度神经网络算法发挥着核心的作用,只有利用神经网络训练特性,才能实现目标故障数据集迁移精度与适用性的准确判断。定义源域  $Z_s$  的样本空间为  $\{z_i | i = 1, 2, \dots, m_s\}$ ,其中  $z_i = \{z_{i1}, z_{i2}, \dots, z_{im}\}$  服从  $P$  分布;集合  $\{y_{si} | i = 1, 2, \dots, m_s\}$  为源域样本空间的标记空间,其中  $m_s$  和  $m$  为源域样本空间的特征维度和样本数量总数。定义目标域  $B_s$  中的模拟数据样本空间为  $\{b_i | i = 1, 2, \dots, n_t\}$ ,其中  $b_i = \{b_{i1}, b_{i2}, \dots, b_{in}\}$  服从  $Q$  分布,  $P \neq Q$ ,集合  $\{y_{ti} | i = 1, 2, \dots, n_t\}$  为目标域样本空间的标记空间。对于源域空间上的故障数据任务,通过任一个标记空间  $y_{si}$ ,都可以在源域样本空间上获得一个用于故障样本分类的函数满足条件  $y_{si} = g_s(z_i)$ ,同理在任一个目标域空间内,也可以得到一个实验故障样本分类标记空间  $y_{ti}$  满足条件  $y_{ti} = g_t(b_i)$ 。迁移学习方法的思想是利用从实验室目标域空间  $B_s$  中训练学习到故障识别与训练的相关知识,并将其应用到源域中的故障诊断当中,以神经网络和提高故障样本分类器的性能,增加有效样本的数量,深度迁移学习的模拟过程如图 1 所示。

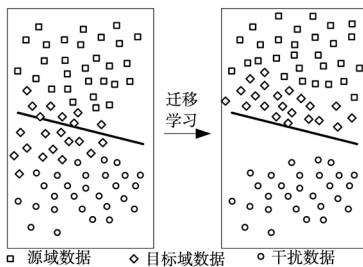


图 1 机械故障大数据迁移学习过程描述

在目标与源域的大数据故障特征样本中寻找相同的典型特征,并保证空间内两者的故障特征差异最小。深度迁移学习方法基于实验室大量的有效仿真数据来训练神经网络模型,并将有效数据迁移到机械大数据的诊断中,丰富实际的故障样本,可以提升机械大数据故障诊断模型的泛化能力。故障诊断中以迁移后的期望数据和原始数据的混合样本,作为深度迁移诊断模型的输入项,并得到基于神经网络模型训练后信号归一化处理后  $z'$  的故障状态特征,对混合故障样本信号做预处理:

$$z' = \frac{z_i - \bar{z}}{\sigma} \quad (7)$$

其中,  $z_i$  为原始样本、 $\bar{z}$  为混合后样本的均值、 $\sigma$  为方

差。对归一化处理后的混合样本进行分类,具体包括训练用样本和测试用样本,将处理好的样本数据输入深度迁移学习模型。然后调整迁移模型的各种参数,包括神经网络的层次数量和权重比例。保持训练数据集和测试数据集数量的稳定和平衡,如果有效训练样本的数量过少,再通过增加源域的大数据比例的方式,平衡两种数据集的数量结构比例。先将少量的大数据故障训练样本输入深度学习网络模型,通过不断地反复训练微调网络模型参数,具体的诊断流程如图 2 所示。

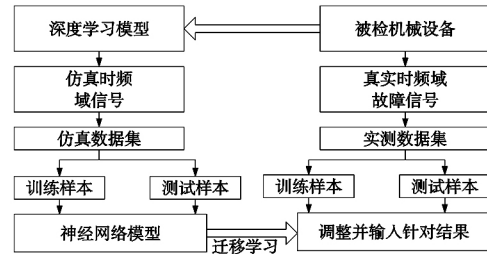


图 2 深度迁移网络模型大数据故障诊断基本流程

在深度学习模型的神经网络中,利用数据的迁移学习优化了权重比例和参数偏置调整效率,解决实际机械故障诊断中有效样本过少的问题,并提高模型在线故障诊断的能力,在小样本或样本缺失的条件下也能够实现对海量大数据的故障点的识别和诊断。

### 3 实验部分

#### 3.1 实验准备

##### (1) 实验平台

本文实验用平台选择美国 SQ 公司的综合机械故障诊断 DDS 实验平台,主体结构由变速驱动电机、直齿轮箱、粉末冶金制动器、FPGA 电路模块等基础部分构成。在 DDS 实验平台的基础上以电机轴故障检测为例,验证深度迁移学习诊断方法的性能,实验平台的主要构成如图 3 所示。

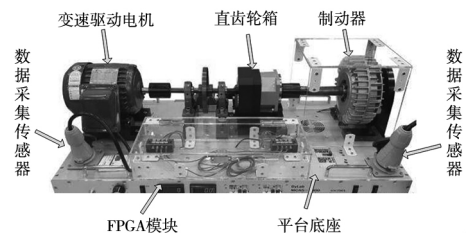


图 3 机械大数据故障诊断实验平台结构

##### (2) 实验数据集

在电机驱动轴为 0Hp 时开始采集原始的机械故障数据,并构建不平衡故障数据集,故障数据与非故障数据的比值设定标准和训练信息如表 1 所示。

表 1 非均衡的机械大数据故障信息集合

故障数据所占比例	总体样本数量	少数样本数量	多数样本数量
1:3	9980	520	1560
1:20	11340	540	10800
1:50	9945	195	9750
1:100	9595	95	9500
1:300	12642	42	12600
1:1000	12012	12	12000

机械设备电机轴的常见 4 种故障包括不同轴、偏磨、擦伤和弯曲,在 1:50 和 1:300 的故障数据比例下检测样本的分布情况,如图 4 所示。

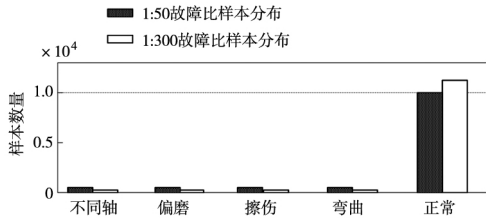


图 4 混合故障比率下的样本分布情况

3.2 实验结果与分析

(1) G-Mean 值对比分析

分别在 6 种故障数据比例关系下,验证文中提出深度迁移模型的故障诊断查准率  $PR$ 、查全率  $RR$  和  $G$ -Mean 值情况,为了更公平、更准确地衡量出不同诊断方法的在性能上差异,引入了传统的集成经验模态方法、包络数据拟合方法和局部均值信号分解方法参与对比。其中查准率  $PR$ 、查全率  $RR$  和  $G$ -Mean 值计算方法如下所示:

$$\begin{cases} PR = \frac{TP}{TP + FN} \times 100\% \\ PP = \frac{TN}{TN + FP} \times 100\% \\ G - Mean = \sqrt{PR \times PP} \end{cases} \quad (8)$$

式中,各公式变量含义解释,如表 2 所示。

表 2 G-Mean 值计算中指标变量的解释说明

$TP$	$FN$	$TN$	$FP$
正确样本数量	错误样本数量	多数类样本数量	少数类样本数量

首先在不同的样本故障比例关系下分析 4 种故障检测方法  $PR$  指标和  $RR$  指标的变化情况,如表 3 和表 4 所示。

表 3 不同样本故障比例关系下  $PR$  指标值分布

故障数据比例	$PR$ 指标值/%			
	集成经验模态	包络数据拟合	局部均值分解	深度迁移学习
1:3	98.12	97.98	98.14	98.65
1:20	97.74	95.58	96.45	98.14
1:50	96.25	93.34	94.14	98.11
1:100	92.41	93.10	92.51	97.14
1:300	90.14	89.41	88.47	97.01
1:1000	86.54	84.14	79.14	96.34

表 4 不同样本故障比例关系下  $PP$  指标值分布

故障数据比例	$PP$ 指标值/%			
	集成经验模态	包络数据拟合	局部均值分解	深度迁移学习
1:3	97.21	98.11	96.77	98.82
1:20	96.11	95.54	93.41	98.10
1:50	93.25	94.03	92.36	97.13
1:100	90.12	91.12	87.52	97.01
1:300	84.15	86.14	85.41	96.67
1:1000	73.52	82.24	76.15	95.52

在故障样本比例较高的条件下各种故障诊断算法的查准率  $PR$  指标和查全率指标  $RR$  的差距较小,而随着故障样本比例值的不断下降,只有基于深度迁移学习故障诊断方法的查准率和查全率仍旧保持在较高的水平,在机械故障数据比例为 1:1000 时,指标值还能够保持在 95% 以上。其次,以查准率指标  $PR$  和查全率指标  $RR$  为基础计算的  $G$ -Mean 指标值分布情况,如表 5 所示。

表 5 不同样本故障比例关系下  $G$ -Mean 指标值分布

故障数据比例	$G$ -Mean 指标值/%			
	集成经验模态	包络数据拟合	局部均值分解	深度迁移学习
1:3	93.41	92.41	93.14	97.71
1:20	88.14	90.13	89.97	97.12
1:50	86.17	88.47	82.41	96.81
1:100	77.91	80.15	79.64	95.62
1:300	75.41	74.67	73.44	94.71
1:1000	69.15	72.34	68.89	94.50

经过计算后得出  $G$ -Mean 指标值结果显示,深度迁移学习模型在多种不同的故障比例场景下,都具有良好的综合性能,机械大数据故障诊断效果良好,从统计结果上看,明显优于三种传统故障诊断方案。

(2) 故障特征分类能力分析

电机轴具体的故障类型具体包括同轴度低、轴身弯曲、轴体偏磨及擦伤等,每一种故障特征的信号表现不一致,加上有效的故障样本数据较少,给故障点的定位及故障程度的判断带来极大的难度。深度迁移学习模型的优势在于将实验室中的故障类别分析与诊断的有效样本,迁移到实际的故障诊断当中,在提高故障诊断精度的同时也改善了不同类别故障分类的能力。利用 MATLAB 仿真分析软件,将多维故障大数据的故障特征点反馈在一个  $10 \times 10$  的二维平面坐标系当中,原始未经过分类处理的不同故障信号,按照频谱特征的分布情况,如图 5 所示。

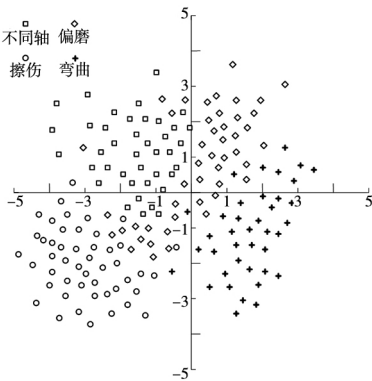


图 5 原始故障不同故障信号的特征分布图

由于不同类别的机械电机轴故障特征在信号频谱的差异较小,导致了 4 种不同的故障特征点出现了混叠的现象。而经过深度迁移学习模型的处理后,利用

从实验室数据中得到的样本资源,丰富了实际检测中有效故障样本的数量特征分类效果,如图6所示。

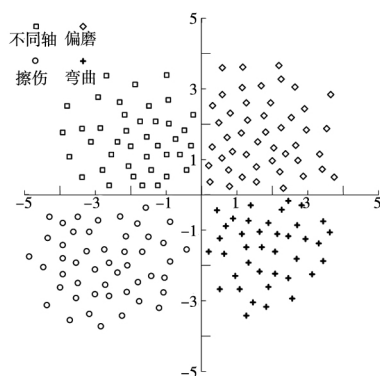


图6 深度迁移学习故障诊断方法的分类性能

经过文中提出方法分类处理后,4种不同的故障特征点分属于4个不同的象限,且故障点之间的界限较为清晰,这是提高机械大数据故障诊断的查准率的前提条件。选取机械故障数据比例为1:1000时G-Mean指标值相对较高的包络数据信号分解方法参与对比,故障特征分类性能结果如图7所示。

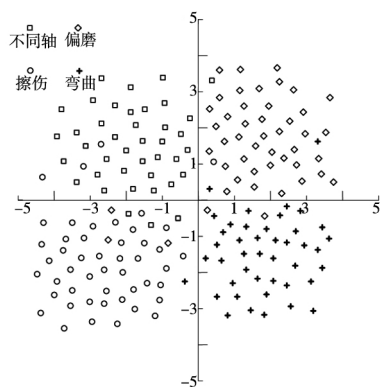


图7 包络数据故障诊断方法的分类性能

相对于原始故障特征的分布状态,包络数据信号分解诊断方法在一定程度上区分了不同类别的故障特征类型,但仍旧存在个别故障点分类失效和分类界限模糊的问题,给电机轴故障诊断带来一定难度。

#### 4 结论

随着工业企业机械化程度的不断提高,利用计算机人工智能和机器学习方法进行机械故障诊断的应用场景不断增多。在小故障样本条件下,有价值的机械设备故障信息较少,而频繁地提高检测次数增加了设备调试和维护的成本,为此本文将深度学习神经网络与迁移学习方法相结合,构建一种经过优化处理的深度学习模型,改善故障样本的稀疏特征。定义源域和目标域的样本空间,找出共同故障源特征,并利用神经网络算法提高深度学习模型的泛化能力,改善了传统机器学习方法在小样本条件下故障诊断率低、分类性能差的不足,提高了对复杂机械设备在线检测的精度。在未来,人工智能和机器学习算法在机械故障检测领

域的应用深度与广度还将持续提高,可以更好地保证机械设备的安全稳定运行。

#### [参考文献]

- [1] 雷亚国,杨彬,杜兆钧. 大数据下机械装备故障的深度迁移诊断方法[J]. 机械工程学报, 2019, 67(7): 1-8.
- [2] Sanaz Salem, Hamed Zandevakili, Ali Mahani, et al. Fault-tolerant delay cell for ring oscillator application in 65 nm CMOS technology [J]. Iet Circuits Devices & Systems, 2018, 12(3): 233-241.
- [3] 郭亦文,耿林霄,胡勇. 大数据智能控制一体化平台及其体系架构[J]. 热力发电, 2019, 48(9): 22-27.
- [4] 王鑫,孙冶,周晓剑. 高风险移动放射源在线监控系统设计[J]. 核电子学与探测技术, 2018, 38(6): 818-822.
- [5] 赵颖,侯俊杰,于成龙. 面向生产管控的工业大数据研究及应用[J]. 计算机科学, 2019, 46(B6): 45-51.
- [6] 许骏,付浩海,张素莉. 基于大数据和AHP-DEMATEL分析法的VR产业知识基础评价研究[J]. 工业技术经济, 2019, 38(7): 101-107.
- [7] 徐启圣,白琨,徐厚昌. 风电齿轮箱状态监测和故障诊断的研究现状及发展趋势[J]. 润滑与密封, 2019, 34(8): 138-147.
- [8] 汤健,柴天佑,丛秋梅. 基于EMD和选择性集成学习算法的磨机负荷参数软测量[J]. 自动化学报, 2014, 40(9): 1853-1866.
- [9] 何坤,李国龙,蒋林. 基于数字法的成形砂轮廓形计算及包络面仿真[J]. 机械工程学报, 2018, 66(1): 205-213.
- [10] 魏勇召,赵明元,杨江天. 信号自适应分解对比研究及其在机车轴承故障诊断中的应用[J]. 机车电传动, 2017, 58(4): 91-94.
- [11] Christopher McMaster, David Liew, Claire Keith, et al. Correction to: A Machine-Learning Algorithm to Optimise Automated Adverse Drug Reaction Detection from Clinical Coding[J]. Drug Safety, 2019, 42(6): 807.
- [12] 杨琳,王衍学,何水龙. 基于变分模态分解和广义Warblet变换的齿轮故障诊断[J]. 机械传动, 2018, 42(7): 157-161.
- [13] 符史仲,杨国来. 一种自紧厚壁圆筒非线性混合硬化模型及残余应力分析[J]. 兵工学报, 2018, 39(7): 1277-1283.
- [14] 李元香,项正龙,夏界宁. 模拟退火算法的动力系统模型及收敛性分析[J]. 计算机学报, 2019, 42(6): 1161-1173.
- [15] 李红,孙冬梅,沈玉成. EEMD降噪与倒频谱分析在风电轴承故障诊断中的应用[J]. 机床与液压, 2018, 46(13): 156-159.
- [16] 田立勇,吴境林. 基于Zoom-FFT的排污泵故障诊断与分析[J]. 机械设计与研究, 2018, 35(3): 171-174.

(编辑 李秀敏)