

面向无标签数据的旋转机械故障诊断方法

陈 菲¹, 杨 峥², 张志成², 罗 巍²

(1. 深圳技术大学 中德智能制造学院, 广东 深圳 518118; 2. 吉林大学 机械与航空航天工程学院, 长春 130022)

摘 要: 针对旋转机械故障诊断算法大多面向有标签的数据, 且参数多需要人为设置的问题, 提出了一种面向无标签数据、参数自适应化的二次异常识别加聚类的无监督诊断算法。该方法通过改进经验小波变换和拉普拉斯分值算法对信号进行特征提取和选择, 并采用二次异常识别结合改进模糊 C 均值聚类的无监督方法进行故障识别。通过电主轴转子系统故障数据验证, 所提方法诊断精度可达 93%, 与传统无监督诊断方法对比, 具有良好的准确性和鲁棒性。

关键词: 故障诊断; 旋转机械; 无标签数据; 参数自适应; 无监督学习

中图分类号: TH133.33; TP183 **文献标志码:** A **文章编号:** 1671-5497(2022)11-2514-09

DOI: 10.13229/j.cnki.jdxbgxb20210355

Fault diagnosis method of rotating machinery for unlabeled data

CHEN Fei¹, YANG Zheng², ZHANG Zhi-cheng², LUO Wei²

(1. Sino-German College of Intelligent Manufacturing, Shenzhen Technology University, Shenzhen 518118, China;
2. School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China)

Abstract: Most fault diagnosis algorithms for rotating machinery are labeled and need to be set manually, an unsupervised fault diagnosis algorithm with adaptive parameters for unlabeled fault data by introducing twice anomaly recognitions and clustering algorithms was proposed. The method extracts and selects signal features by improving empirical wavelet transform and Laplace score algorithm, and adopts the unsupervised method of quadratic anomaly identification combined with improved fuzzy C-means clustering for fault identification. Through the verification of the fault data of the rotor system of the electric spindle, the diagnostic accuracy of the proposed method can reach 93%. Compared with the traditional unsupervised diagnostic method, it has good accuracy and robustness.

Key words: fault diagnosis; rotating machinery; unlabeled data; adaptive parameters; unsupervised learning

收稿日期: 2021-04-22.

基金项目: 广东省普通高校特色创新类项目(2020KTSCX127); 深圳技术大学研究生校企合作研究基金项目(XQHZ202003).

作者简介: 陈菲(1970-), 女, 教授, 博士. 研究方向: 智能制造装备故障预测与健康管理的 E-mail: chenfei@sztu.edu.cn

通信作者: 罗巍(1986-), 女, 副研究员, 博士. 研究方向: 数控装备可靠性的 E-mail: luoweicn@jlu.edu.cn

0 引 言

旋转机械作为绝大多数机械设备的核心部件,在工业生产中有广泛的应用,其一旦发生故障,会造成重大经济损失甚至人员伤亡等严重后果。因此,需要进行适应实际工况的智能故障诊断研究,提高旋转机械故障诊断的准确性、减少人为干预和人工成本,以保障其乃至整个设备、生产线的安全、稳定运行^[1]。

目前,机器学习、传感器、大数据等技术的发展越来越成熟,旋转机械故障诊断方法的发展也朝向自动化、智能化。旋转机械的智能故障诊断方法主要包括 3 个步骤:数据预处理、特征提取和故障识别。其中,特征提取工作可以得到原始信号的特征量并表征信号潜在的故障信息,再将提取的特征作为支持向量机^[2]、BP 神经网络^[3]等有监督智能算法的输入进行故障识别。

由 Gilles^[4] 根据经验模态分解 (Empirical mode decomposition, EMD) 和小波变换 (Wavelet transform, WT) 有关思想提出的经验小波变换 (Empirical wavelet transform, EWT) 作为一种较为成熟的时频域特征提取方法,可以很好地解 EMD 的模式混叠和端点效应问题,并用于旋转机械振动信号的特征提取。Yu 等^[5]、何洋洋等^[6] 利用 EWT 筛选特征信息丰富的振动成分,实现旋转机械振动信号的特征提取;叶益丰^[7] 通过 EWT 和核主成分分析法 (Kernel principal component analysis, KPCA) 对电主轴的振动数据进行特征提取和降维,以实现机械与电气耦合故障的识别。针对旋转机械振动信号在不同工况下的非平稳特征,叶柯华等^[8]、赵若好等^[9] 分别将关联维数和多尺度数学形态谱与 EWT 进行结合。为了有效诊断出滚动轴承早期微弱故障,乔志城等^[10] 针对传统 EWT 频谱划分过多的问题,基于互信息值对频谱进行重新划分与合并以有效减少频带的数量。

学者们已经基于 EWT 对旋转机械信号的特征提取方法进行了大量研究,并主要将所提取特征输入到有监督学习模型中进行故障识别,用训练好的模型进行故障诊断工作。但有监督学习具有局限性:①模型复杂、效率低,需要借助大量数据进行训练,迭代次数多;②实际工况所获得的故障样本数目往往较少。无监督学习提供了解决该局限的途径,即通过分析样本内在结构对样本进

行识别。常用的方法有:聚类分析、模糊数学理论、自组织特征映射 (Self-organizing feature mapping, SOM) 网络等。常勇等^[11]、林越等^[12] 分别引入核函数和有效性函数对模糊 C 均值聚类算法进行改进,以实现轴承、电力变压器的非线性故障分类能力。针对采集数据缺失、信息不完整的问题,贺湘宇等^[13]、王庆锋等^[14] 利用 C-均值聚类算法分别实现了液压系统和离心压缩机故障的识别。针对非线性、非平稳振动信号的特点,院老虎等^[15] 提出一种基于 EWT 和密集连接卷积网络的飞行器机械部件故障诊断识别方法。考虑多类型故障在不同因素下呈现一定的规律性及聚集性,徐忠兰^[16] 基于 SOM 神经网络根据不同权重参数的聚集形态,形成对应特定规律的故障分类。

目前针对无监督故障诊断的研究相对较少,且大多数无监督故障诊断算法依然需要借助已知类别的样本或者人工确定参数,不能进行没有任何先验信息样本的故障诊断工作。本文提出一种不需要借助样本先验信息、面向无标签数据、参数自适应化的二次异常识别加聚类的无监督故障诊断算法。该方法通过改进经验小波变换算法 (Empirical wavelet transform based on scale-space, SEWT) 和改进拉普拉斯分值算法 (Laplace score algorithm based on max-relevance and min-redundancy, mRMR-LS) 对振动信号进行特征提取和特征选择,自适应地设定特征数目;并采用了二次异常识别方法结合改进模糊 C 均值聚类算法的策略,提高旋转机械故障在无监督数据情况下的准确性。

1 特征提取和选择算法

1.1 改进经验小波变换(SEWT)算法

EWT 是一种较为成熟的时频域分析方法,其本质是根据原信号的傅里叶频谱进行分割,能避免产生模式混叠及虚假模式的现象,从而提高了模式分解的准确性。EWT 分解信号的具体步骤见文献[4],为寻找有意义的模式,Gilles 提出了将傅里叶频谱转化为尺度空间^[17]表示的自适应频谱分割方法^[18]。本文基于尺度空间理论对 EWT 算法进行了改进:①对频谱进行正则化处理可去除噪声信息;②采用尺度空间理论自适应划分频谱,并通过相关系数来选择最后确定的模式,达到不需要人为干预、参数自适应化的效果,

并避免分割点过多所造成的频带破裂现象。SEWT算法的具体实施流程如图1所示。

(1)根据式(1)获取原始信号的频谱,同时采用移动平均的方法对傅里叶频谱做正则化处理,去除噪声等干扰信息。

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} s(t) e^{-i\omega t} dt \quad (1)$$

式中: $s(t)$ 为原始信号; t 为时间; ω 为频率; $e^{-i\omega t}$ 为复变函数。

(2)将正则化后的频谱函数映射到尺度空间,获取初始极小值点对应的尺度空间曲线,然后用 k 均值法确定阈值,并选择长度大于阈值的尺度空间曲线对应的初始极小值点作为频谱分割的边界点。

(3)确定好频谱分割点后,构建滤波器组将原始信号分解成模态分量。

(4)借助文献[19]中选择模态的方法,将自适应分解后的各模态与原始信号的相关系数作为保留模态的评价标准。

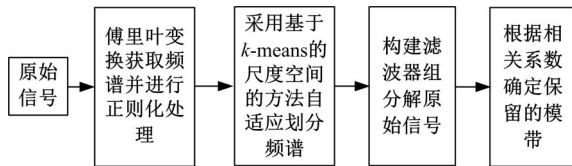


图1 基于SEWT的特征提取流程图

Fig. 1 Feature extraction flow chart based on SEWT

1.2 改进拉普拉斯分值(mRMR-LS)算法

拉普拉斯分值(Laplace score, LS)是一种无监督的特征选择算法^[20],借助图谱理论将高维度的样本集中每一个样本当作一个节点来构造结构图,并输出每个特征的拉普拉斯得分情况,最后根据每个特征的分值来进行故障特征选择,具体步骤见文献[21]。然而,LS算法只考虑了特征的相关性,未考虑特征的冗余性。本文基于Peng等^[22]提出的最大相关最小冗余算法(Max-relevance and min-redundancy, mRMR)进行改进,其具体实现方法如下:

(1)输入高维样本集(特征集) $F\{f_1, f_2, \dots, f_n\}$,设置LS算法的近邻参数 k ,建立选择特征集合 M 。

(2)计算每个特征的拉普拉斯评分值 L_r ,组成评分值向量 $L\{L_1, L_2, \dots, L_n\}$,以所有特征评分值的箱型图中下四分位点的值为阈值 λ ,选择评分值小于阈值 λ 的特征加入到选择特征集

合 M 中。

$$L_r = \frac{\hat{f}_r^T L \hat{f}_r}{\hat{f}_r^T D \hat{f}_r} = \frac{\sum_{ij} (f_{ri} - f_{rj})^2 S_{ij}}{\text{Var}(f_r)} \quad (2)$$

式中: f_r 为第 r 个特征在所有样本中的值所组成的向量; \hat{f}_r^T 为 f_r 的中间变量; D 为对角阵; L 为 D 与样本间权重矩阵的差异; f_{rj} 为第 i 个样本的第 r 个特征; $\text{Var}(f_r)$ 为第 r 个特征的方差; S_{ij} 为样本间有边连接时的权重。

(3)计算选择特征集合 M 中所有特征的冗余度 $R(f_i)$,并以特征冗余度的箱型图中的上四分位点作为阈值 ϵ ,将冗余度大于阈值 ϵ 的特征从 M 中删除。

$$R(f_i) = \frac{1}{n^2} \sum_{f_j \in F} I(f_i; f_j) \quad (3)$$

式中: n 为特征集中特征的个数; $I(f_i; f_j)$ 为 f_i 和 f_j 的互信息。

(4)输出选择的特征集合 M 。

2 无监督模式识别算法

2.1 改进局部离群因子(LLOF)算法

局部离群因子(Local outlier factor, LOF)算法是一种基于密度的无监督式的异常检测算法^[23],其具体思想为通过计算某个样本的LOF值来判断该样本是否为异常样本。为了自适应确定邻域参数 k ,本文采用了改进LOF(LOF based on the natural nearest neighbor search method, LLOF),其主要计算方法如下:

(1)借助文献[24]中自然最近邻居搜索的思想自适应确定数据集中样本的邻域参数 k 。

(2)计算 X 点的第 k 距离。即距离 X 第 k 远的点的距离,不包括 X ,记为 $d_k(X)$,并且满足:在集合中至少有不包括 X 在内的 k 个点 $o' \in C\{o' \neq X\}$,满足 $d(X, o') \leq d(X, o)$,在集合中最多有不包括 X 在内的 $k-1$ 个点 $o' \in C\{o' \neq X\}$,满足 $d(X, o') < d(X, o)$ 。

(3)计算 X 点的第 k 距离邻域。记为 $N_k(X)$,满足式(4),表示给定 X 点的第 k 距离后, X 点的第 k 距离邻域为 X 的第 k 距离所有点的集合。

$$N_k(p) = \{o' \in C\{o' \neq p\} \mid d(p, o') \leq d_k(p)\} \quad (4)$$

(4)计算点 o 到点 X 的第 k 可达距离。记为 $\text{reach-dist}_k(X, o)$ 。即点 o 到点 X 的第 k 可达距离

为点 o 的第 k 距离与点 o 到点 X 的真实距离中值较大的一方:

$$reach-dist_k(X, o) = \max\{d_k(o), d(X, o)\} \quad (5)$$

(5) 计算点 X 的局部可达密度。表示点 X 的第 k 邻域内所有点到 X 点的平均可达距离的倒数:

$$lrd_k(X) = 1 / \left(\frac{\sum_{o \in N_k(X)} reach-dist_k(X, o)}{|N_k(X)|} \right) \quad (6)$$

(6) 计算点 X 的局部离群因子。表示点 X 的第 k 邻域内所有点的局部可达密度与点 X 的局部可达密度之比的平均数。若点 X 的 LOF 值越大, 则表明点 X 的密度小于其周围点的密度, 点 X 越可能为异常点。

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|} = \frac{\sum_{o \in N_k(p)} lrd_k(o)}{|N_k(p)| * lrd_k(p)} \quad (7)$$

式中: $lrd_k(p)$ 为点 p 的局部可达密度; $lrd_k(o)$ 为点 o 的邻域点的局部可达密度; $N_k(p)$ 为点 p 的第 k 距离邻域点的个数。

2.2 基于核密度估计的改进 DBSCAN 算法

具有噪声的基于密度的聚类方法 (Density-based spatial clustering of applications with noise, DBSCAN) 是由 Ester 等^[25]提出的一种不需要人为指定簇数、基于密度的聚类方法, 可将样本集中具有离群属性的异常点标记为噪声并应用到异常识别中。领域 (Eps)、邻域中最小的样本数量 ($MinPts$) 是 DBSCAN 算法中需要经验确定的两个重要参数, 本文采用基于核密度估计方法的改进 DBSCAN 算法 (DBSCAN based on kernel density estimation method, KDBSCAN), 可以自适应地确定 Eps 、 $MinPts$, 具体流程^[26]如图 2 所示。

(1) 根据数据集的特征矩阵, 依次计算样本集中每个样本与其他样本之间的欧式距离, 生成数据集的距离矩阵。

(2) 计算距离矩阵中每个距离样本的核密度估计值, 并将曲线中簇内密度较高所对应的距离



图 2 KDBSCAN 算法流程图

Fig. 2 Flow chart of KDBSCAN algorithm

值 (即第一个峰值) 设定为 Eps 。

(3) 确定参数 $MinPts$:

$$MinPts = \frac{1}{n} \sum_{i=1}^n P_i \quad (8)$$

式中: P_i 为对象 i 的 Eps 领域内包含的样本数量; n 为数据集样本的总数。

2.3 改进模糊 C 均值聚类 (AFCM) 算法

模糊 C 均值聚类 (Fuzzy C-means, FCM) 算法是模糊聚类中应用最广的无监督学习方法之一^[27], 根据隶属度来衡量各样本点属于某一聚类的程度, 可实现对数据集进行软划分且无需用有标签的数据训练。聚类类别数 c 的设定会影响聚类效果。本文采用了一种自适应模糊 C 均值聚类算法 (Adaptive fuzzy C-means, AFCM), 其具体步骤如下:

(1) 根据样本的异常得分值分布的断档, 初步确定类别数 c_0 。

(2) 引入了划分系数 (Partition coefficient, PC)^[28]作为模糊聚类的评价指标, 根据式 (9) 计算 c_0 左右簇数的 PC 值, 并选取 PC 值最大的簇数为最终簇数 c 。

$$V_{PC} = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \quad (9)$$

(3) 初始化参数, 根据步骤 (2) 设定类别数 c 、停止阈值 ϵ 和最大迭代次数 N , 随机选取 c 个样本作为聚类中心。

(4) 计算数据集中样本到已选聚类中心的隶属度, 并以此为根据划分到隶属度最大的类别。

(5) 计算更新聚类中心。

(6) 计算目标函数值, 与上次迭代目标函数的数值作比较, 若两者误差的绝对值小于停止阈值 ϵ 或达到了最大迭代次数 N , 输出聚类结果, 否则返回步骤 (4) 继续迭代。

3 无监督、参数自适应故障诊断算法

针对现有故障诊断算法需要借助样本先验信息, 并且大部分参数需要人为指定, 实际可操作性不强的缺点, 本文对上述所涉及到的算法进行了改进, 提出了一种面向无标签数据的无监督、参数自适应化的故障诊断算法, 可对没有任何先验信息的样本集自适应的完成故障诊断工作, 尽可能地减少了人为的主观干扰。

所采用的方案大致分为 4 个步骤: 原始振动信号的采集和预处理、特征提取与特征选择、异常识别、故障识别。其详细诊断流程如图 3 所示, 具

体步骤如下:

(1)对原始振动信号组成的样本集 M 中的样本进行小波降噪预处理。

(2)利用SEWT自适应提取出降噪处理后 M 中的样本的最优IMF分量,得到 M 的特征矩阵。

(3)利用所采用mRMR-LS算法对步骤(2)得到的特征矩阵进行特征选择。

(4)采用两次异常识别的手段,初次识别采用KDBSCAN算法,将得到的噪声点和少类簇作为初次异常样本,然后用LLOF进行二次识别,只计算初次异常样本的LOF值,将LOF值大于1的样本确定为最后的异常样本。

(5)为了考虑异常样本的全局离群属性,定义GLOF值作为异常样本最后的异常得分值, $GLOF=LOF+GB$,其中GB为某一异常样本到所有正常样本欧式距离的平均值。由此利用无监

督的方式得到了异常样本及其异常得分值,便于后续的针对性分析。

(6)针对得到的异常样本和其异常得分值(GLOF),利用AFCM算法进行聚类。

4 故障诊断算法验证

4.1 电主轴状态监测试验系统

转子系统是旋转机械最重要的部件之一,本文利用实验室电主轴状态监测试验系统对所提算法进行验证,其主要由加载单元、控制单元、采集单元、抱夹装置和冷却装置组成,如图4所示。

(1)加载单元

为了模拟主轴真实工况,设计了可以进行轴向加载、径向加载和扭矩加载的加载单元。轴向加载是借助导向轴承等连接机构将伺服电机的驱动力传递到前段的加载叉上。径向力加载部分加载分为两部分:通过伺服电机提供驱动力以钢丝牵引的方式提供静态的均值加载;通过电磁激振器及其控制器进行动态的幅值加载。扭矩加载借助电力测控机和联轴器进行加载。

(2)控制单元

控制单元通过工控机实现,由硬件电路和软件平台两部分构成。硬件电路控制主轴的启停、转速、加载信号的模数转换和冷却系统等,软件平台由VB语言开发设计并控制加载情况。

(3)采集单元

采集单元由传感器、采集卡和上位机的软件平台构成。传感器分为振动、电流、电压、温度和激光位移传感器。实际采集振动信号时,将两个IEPE类型的振动传感器吸附在电主轴的前、后两

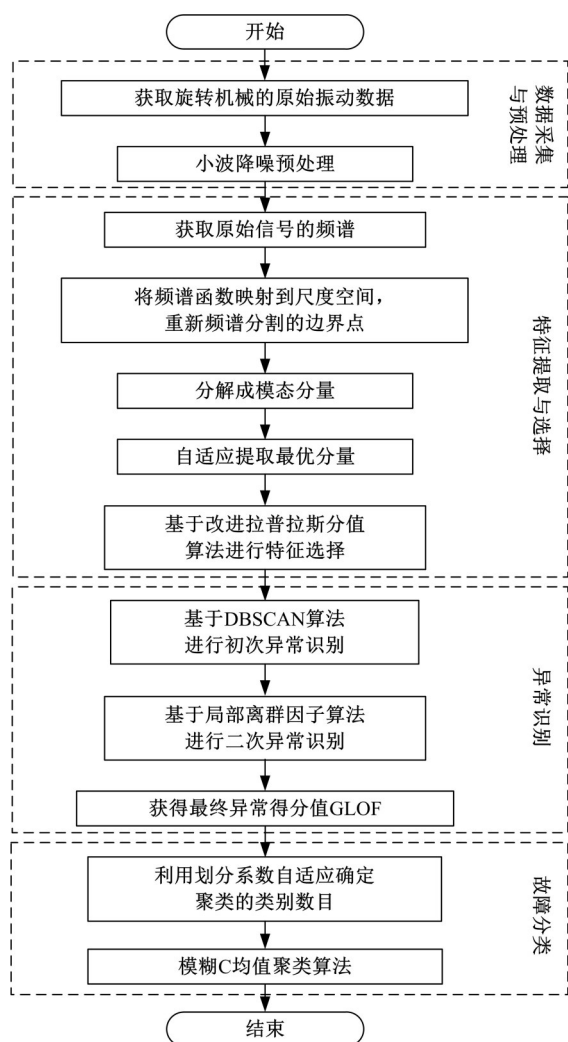


图3 无监督、参数自适应故障诊断算法流程图

Fig. 3 Flow chart of unsupervised fault diagnosis algorithm with adaptive parameters

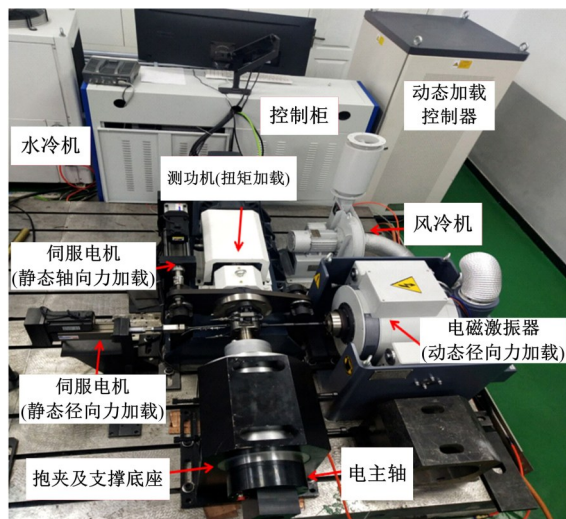


图4 电主轴状态监测试验系统

Fig. 4 Condition monitoring system of motor spindle

端轴上,其参数如表 1 所示。

数据采集部分选用 CompactDAQ 机箱(型号 NI cDAQ-9174)和配套的采集卡(型号 NI 9234),并通过电主轴状态监测软件平台完成数据的采集、存储和导出。

(4)抱夹、冷却装置

抱夹装置由支撑底座和轴套构成,用来固定和支撑电主轴。冷却装置由水冷机和风冷机构成,在电主轴运行期间可以起到冷却保护作用。

表 1 振动传感器参数

Table 1 Vibration sensor parameters	
参数	数值
量程/g	±50
灵敏度/(mV·g ⁻¹)	100
采样频率/Hz	0.5~9000
分辨率/g	0.0002

4.2 转子数据集介绍

转子系统故障诊断基于电主轴状态监测试验系统,通过人工模拟了 3 种转子系统的故障:①电主轴与测功机之间不对中;②不同程度冲击的磨损故障;③转速突变(500~1000 r/min)。试验设置的采样频率为 5000 Hz,取 5000 个点即 1 s 的信号作为一个样本,样本配置如表 2 所示(50 个正常样本,43 个异常样本,样本的编号从 0 开始依次递增)。

表 2 样本配置

Table 2 Sample configurations	
样本类型	样本数量
正常样本,编号:0~49	50
X 方向连续冲击程度 1,编号:56~62	7
X 方向连续冲击程度 2,编号:63~66	4
Y 方向连续冲击程度 1,编号:67~72	6
Y 方向连续冲击程度 2,编号:73~77	5
Z 方向连续冲击程度 1,编号:78~84	7
Z 方向连续冲击程度 2,编号:85~89	5
不对中,编号:50~55	6
转速突变,编号:90~92	3

4.3 转子系统故障诊断结果

4.3.1 异常识别算法结果对比

首先,对经过小波降噪预处理的原始振动信号用 SEWT 算法进行特征提取,用 mRMR-LS 算法进行特征选择,得到样本的特征矩阵。应用 KDBSCAN 算法加 LLOF 算法的两次异常识别算法(DLLOF)进行异常识别工作,对样本的特征矩阵进行标准化处理后,分别采用 LLOF 算法和

DLLOF 算法进行异常识别和对比验证。其识别效果对比如图 5(用 PCA 算法把特征矩阵降到 2 维)和表 3 所示。

由图 5 和表 3、表 4 所示,采用 LOF 算法对转子数据进行异常识别后,输出了 48 个异常样本。其中正确识别出 41 个异常样本,有两个样本点未能识别出(编号 78、80),准确率为 0.95;额外输出了 7 个异常样本(编号 0、5、7、14、32、35、41),误识别率为 0.16;算法的运行时间为 6.6 s。采用 DLLOF 算法对转子数据进行异常识别后,输出了 41 个异常样本。其中正确识别出 40 个异常样本,未能识别出 3 个异常样本,其中一个异常样本(编号 92)是因初次识别未识别出,两个异常样本(编号 78、80)是因二次识别未识别出,准确率为 0.93;额外输出了一个正常样本点(编号 5),误识别率为 0.02;算法的运行时间为 2.1 s。由对比可看出,DLLOF 算法在保证不同工况情况下准确

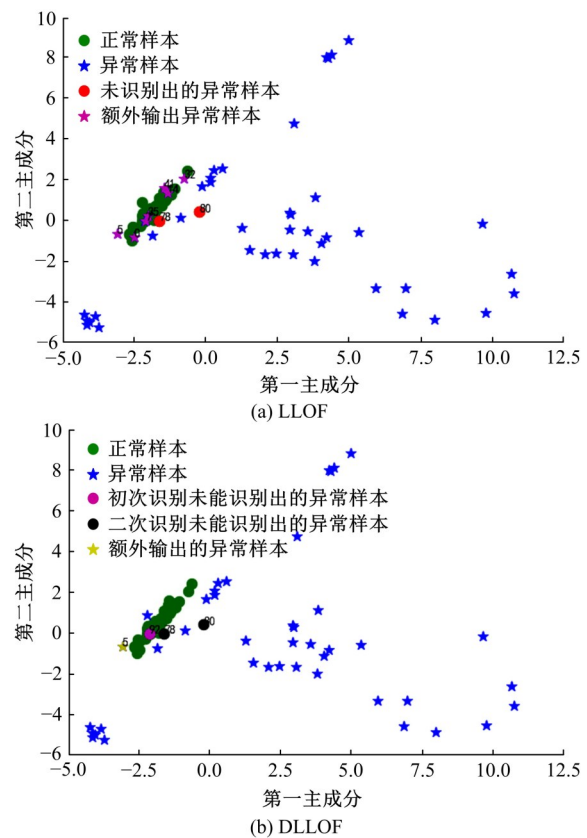


图 5 特征矩阵经算法识别的可视图

Fig. 5 Visualization of feature matrix identified

表 3 两种异常识别算法对比

Table 3 Comparison of LOF and DLLOF			
算法	准确率 α	误识别率 δ	运行时间/s
LOF	0.95	0.16	6.6
DLLOF	0.94	0.02	2.1

表 4 经 DLLOF 算法识别后异常样本的异常得分值
Table 4 GLOF value of outlier samples identified by DLLOF algorithm

编号	LOF	GLOF	编号	LOF	GLOF
5	1.21	7.46	70	2.11	6.79
42	0.97	无	71	2.54	7.94
50	2.78	10.92	72	1.97	8.38
51	3.17	12.9	73	2.92	5.58
52	2.91	11.58	74	2.3	11.41
53	2.74	10.73	75	2.63	14.5
54	4.11	14.64	76	3.25	15.02
55	2.72	10.61	77	2.6	13.39
56	3.05	7.12	78	0.98	无
57	2.04	8.12	79	1.9	5.14
58	1.65	11.05	80	0.97	无
59	1.63	7.75	81	1.92	6.02
60	2.77	13.96	82	1.86	5.46
61	1.82	8.07	83	2.18	6.32
62	3.49	14.29	84	4.08	11.53
63	4.46	15.1	85	2.82	10.63
64	4.31	18.05	86	4.05	15.2
65	4.45	17.89	87	3.68	16.24
66	4.6	18.04	88	3.68	15.03
67	2.21	12.23	89	2.41	13.62
68	1.74	8.66	90	1.48	3.98
69	2.02	7.64	91	1.78	5.87

率和鲁棒性的前提下,显著降低了误识别率和运行时间。

4.3.2 聚类算法实验结果对比

针对 DLLOF 算法识别出的异常样本采用 AFCM 算法进行聚类。异常样本的 GLOF 值分布如图 6 所示,其中有 6 个明显断档,因此初步设

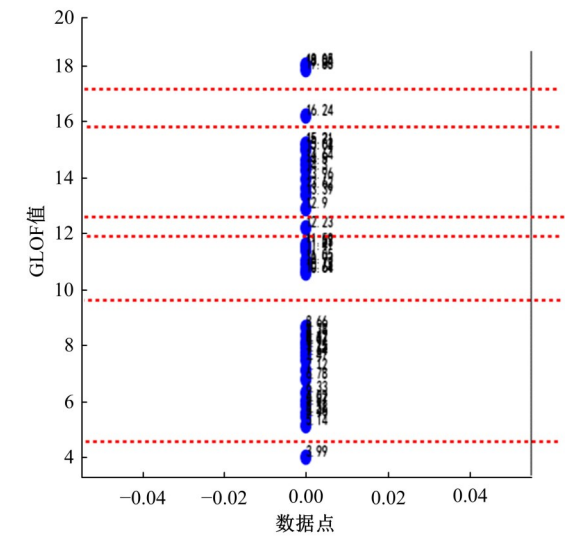


图 6 异常样本 GLOF 值分布

Fig. 6 GLOF value distribution of abnormal sample

置类别数 $c_0=7$,并在其左右取 $c=5,6,8,9$ 作为候选簇数。针对不同的簇数 c 计算划分系数 (PC),如表 5 所示,最终选取最大的 PC 值所对应簇数 $c=8$ 为类别数。

为了进行对比,设置一组对照:直接用 FCM 算法进行聚类(类别数设置为 9)。最终异常样本的聚类结果如图 7 所示(用 PCA 算法将特征矩阵降到 2 维)。AFCM 聚类算法能正确地将大部分 (93%) 转子异常样本聚类,仅聚类错误 3 个样本 (编号 68、72、85)。其准确率和运行时长与直接设置 $c=9$ 的传统 FCM 算法的对比如表 6 所示,可看出在转子数据上,AFCM 算法聚类的准确率和运行时长都优于直接采用 FCM 算法。

表 5 不同簇数下异常样本 AFCM 聚类的 PC 值
Table 5 PC values of AFCM clustering for abnormal samples under different cluster numbers

簇数	PC 值
5	0.08
6	0.12
7	0.09
8	0.18
9	0.13

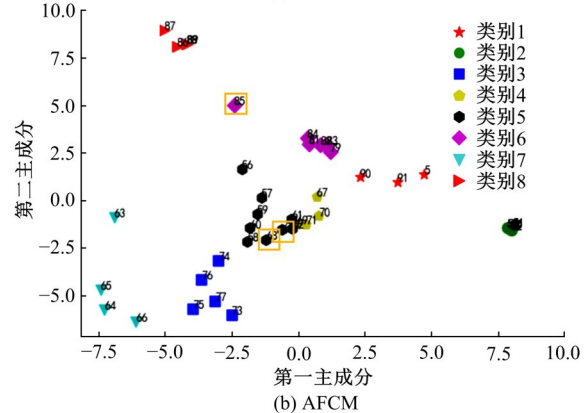
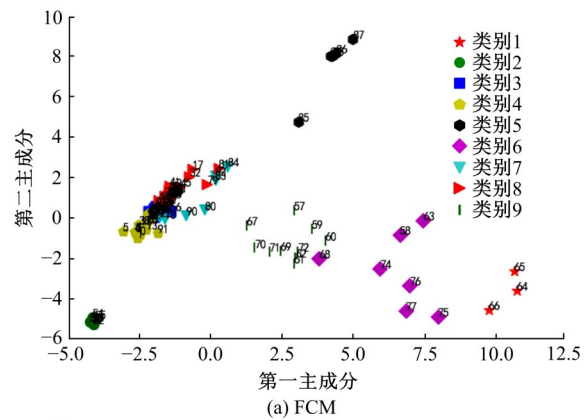


图 7 最终异常样本的聚类结果

Fig. 7 FCM clustering results of final abnormal samples

表 6 AFCM 与传统 FCM 准确率和运行时长对比
Table 6 Comparison of accuracy and running time of
AFCM and conventional FCM

算法	准确率	运行时长/s
传统FCM聚类	0.78	6.7
AFCM聚类	0.93	2.4

5 结束语

针对现有旋转机械故障诊断算法大多数基于有监督学习方法并需要借助样本先验信息,本文提出了一种面向无标签数据的无监督、参数自适应化的故障诊断算法。本文算法改进了经验小波变换算法(SEWT)和改进拉普拉斯分值算法(mRMR-LS),以实现振动信号故障特征的提取和选择,克服了需要人为设置参数的弊端。在故障识别阶段,本文采用二次异常识别结合自适应模糊C均值聚类算法的策略,提高了无标签旋转机械数据的故障诊断准确性。利用电主轴状态监测系统所获取的转子系统数据分别对“二次识别”和“故障聚类”算法在无监督情况下的准确率和鲁棒性进行验证,结果表明本文算法可精准识别无标签数据中绝大多数异常样本以及故障样本,并相比于传统无监督诊断方法具有良好的准确性。

参考文献:

[1] Liu R, Yang B, Zio E, et al. Artificial intelligence for fault diagnosis of rotating machinery: a review[J]. Mechanical Systems & Signal Processing, 2018, 108: 33-47.

[2] 黄海松,魏建安,任竹鹏,等. 基于失衡样本特性过采样算法与SVM的滚动轴承故障诊断[J]. 振动与冲击, 2020, 39(10): 65-74,132.

Huang Hai-song, Wei Jian-an, Ren Zhu-peng, et al. Rolling bearing fault diagnosis based on imbalanced sample characteristics oversampling algorithm and SVM[J]. Journal of Vibration and Shock, 2020, 39(10): 65-74,132.

[3] 何雷,刘溯奇,蒋婷,等. 基于改进LMD与BP神经网络的变速箱故障诊断[J]. 机械传动, 2020, 44(1): 171-176.

He Lei, Liu Su-qi, Jiang Ting, et al. Gearbox fault diagnosis based on improved LMD and BP neural network[J]. Journal of Mechanical Transmission, 2020, 44(1): 171-176.

[4] Gilles J. Empirical wavelet transform[J]. IEEE Transactions on Signal Processing, 2013, 61(16):

3999-4010.

[5] Yu J, Hua Z, Li Z. A new compound faults detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator[J]. Chaos, Solitons & Fractals, 2016, 89: 8-19.

[6] 何洋洋,王馨怡,董晶. 基于经验小波变换与谱峭度的船舶轴系故障特征提取方法[J]. 中国舰船研究, 2020, 15(增刊1): 98-106.

He Yang-yang, Wang Xin-yi, Dong Jing. Fault feature extraction method for marine shafting based on empirical wavelet transform-spectral kurtosis[J]. Chinese Journal of Ship Research, 2020, 15(Sup. 1): 98-106.

[7] 叶益丰. 基于MEWT-KPCA的电主轴故障诊断技术研究[D]. 长春: 吉林大学机械与航空航天工程学院, 2018.

Ye Yi-feng. Fault diagnosis technology of motorized spindle based on MEWT-KPCA[D]. Changchun: School of Mechanical and Aerospace Engineering, Jilin University, 2018.

[8] 叶柯华,李春,胡璇. 基于经验小波变换和关联维数的风力机齿轮箱故障诊断[J]. 动力工程学报, 2021, 41(2): 113-120.

Ye Ke-hua, Li Chun, Hu Xuan. Fault diagnosis of a wind turbine gearbox based on empirical wavelet transform and correlation dimension[J]. Journal of Chinese Society of Power Engineering, 2021, 41(2): 113-120.

[9] 赵若好,马宏忠,魏旭,等. 基于EWT及多尺度形态谱的高压并联电抗器故障诊断研究[J]. 电力系统保护与控制, 2020, 48(17): 68-75.

Zhao Ruo-yu, Ma Hong-zhong, Wei Xu, et al. Research on fault diagnosis of a high voltage shunt reactor based on EWT and multiscale spectral spectrum [J]. Power System Protection and Control, 2020, 48 (17): 68-75.

[10] 乔志城,刘永强,廖英英. 改进经验小波变换与最小熵解卷积在铁路轴承故障诊断中的应用[J]. 振动与冲击, 2021, 40(2): 81-90,118.

Qiao Zhi-cheng, Liu Yong-qiang, Liao Ying-ying. Application of improved wavelet transform and minimum entropy deconvolution in railway bearing fault diagnosis[J]. Journal of Vibration and Shock, 2021, 40 (2): 81-90,118.

[11] 常勇,包广清,程思凯,等. 基于VMD和KFCM的轴承故障诊断方法优化与研究[J]. 西南大学学报: 自然科学版, 2020, 42(10): 146-155.

Chang Yong, Bao Guang-qing, Cheng Si-kai, et al.

- Optimization and research of a bearing fault diagnosis method based on VMD and KFCM[J]. Journal of Southwest University(Natural Science Edition), 2020, 42(10): 146-155.
- [12] 林越, 刘廷章, 唐侃. 基于自适应模糊聚类与核主元分析混合模型的变压器异常检测[J]. 科技通报, 2020, 36(9): 56-60.
- Lin Yue, Liu Ting-zhang, Tang Kan. Anomaly detection of power transformer based on KFCM-KPCA hybrid model[J]. Bulletin of Science and Technology, 2020, 36(9): 56-60.
- [13] 贺湘宇, 何清华. 基于有源回归模型与模糊 C-均值聚类的挖掘机液压系统故障诊断[J]. 吉林大学学报: 工学版, 2008, 38(1): 183-187.
- He Xiang-yu, He Qing-hua. Fault diagnosis for excavator hydraulic system based on auto-regressive with extra inputs model and fuzzy C-means clustering[J]. Journal of Jilin University(Engineering and Technology Edition), 2008, 38(1): 183-187.
- [14] 王庆锋, 刘家赫, 卫炳坤, 等. 数据驱动的聚类分析故障识别方法研究[J]. 机械工程学报, 2020, 56(18): 7-14.
- Wang Qing-feng, Liu Jia-he, Wei Bing-kun, et al. Research on data-driven clustering analysis fault identification method[J]. Journal of Mechanical Engineering, 2020, 56(18): 7-14.
- [15] 院老虎, 连冬杉, 张亮, 等. 基于密集连接卷积网络和支持向量机的飞行器机械部件故障诊断[J]. 吉林大学学报: 工学版, 2021, 51(5): 1635-1641.
- Yuan Lao-hu, Lian Dong-shan, Zhang Liang, et al. Fault diagnosis of key mechanical components of aircraft based on densenet and support vector machine[J]. Journal of Jilin University(Engineering and Technology Edition), 2021, 51(5): 1635-1641.
- [16] 徐忠兰. 基于 SOM 神经网络的煤矿用防爆柴油机故障诊断[J]. 煤矿机械, 2021, 42(4): 175-177.
- Xu Zhong-lan. Fault diagnosis of mine explosion-proof diesel engine based on SOM neural network[J]. Coal Mine Machinery, 2021, 42(4): 175-177.
- [17] Lindeberg T. Scale-Space Theory in Computer Vision[M]. Berlin: Springer, 1994.
- [18] Gilles J, Heal K. A parameterless scale-space approach to find meaningful modes in Histograms-application to image and spectrum segmentation[J]. International Journal of Wavelets Multiresolution & Information Processing, 2014, 12(6): 1450044.
- [19] 蔡艳平, 李艾华, 王涛, 等. 基于 EMD-Wigner-Ville 的内燃机振动时频分析[J]. 振动工程学报, 2010, 23(4): 430-437.
- Cai Yan-ping, Li Ai-hua, Wang Tao, et al. I. C. engine vibration time-frequency analysis based on EMD-Wigner-Ville[J]. Journal of Vibration Engineering, 2010, 23(4): 430-437.
- [20] He X, Cai D, Niyogi P. Laplacian score for feature selection[C]//Advances in Neural Information Processing Systems 18, Vancouver, British Columbia, Canada, 2005: 507-514.
- [21] 欧璐, 于德介. 基于拉普拉斯分值和模糊 C 均值聚类的滚动轴承故障诊断[J]. 中国机械工程, 2014, 25(10): 1352-1357.
- Ou Lu, Yu De-jie. Rolling bearing fault diagnosis based on laplacian score and fuzzy C-means clustering[J]. China Mechanical Engineering, 2014, 25(10): 1352-1357.
- [22] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(8): 1226-1238.
- [23] Breunig M M, Kriegel H P, Ng R T, et al. LOF: identifying density-based local outliers[C]//ACM Sigmod International Conference on Management of Data, Dallas, United States, 2000: 93-104.
- [24] 朱庆生, 唐汇, 冯骥. 一种基于自然最近邻的离群检测算法[J]. 计算机科学, 2014, 41(3): 282-284, 311.
- Zhu Qing-sheng, Tang Hui, Feng Ji. Outlier detection algorithm based on natural nearest neighbor[J]. Computer Science, 2014, 41(3): 282-284, 311.
- [25] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//International Conference on Knowledge Discovery and Data Mining, Orlando, USA, 1996: 226-231.
- [26] 王光, 林国宇. 改进的自适应参数 DBSCAN 聚类算法[J]. 计算机工程与应用, 2020, 56(14): 45-51.
- Wang Guang, Lin Guo-yu. Improved adaptive parameter DBSCAN clustering algorithm[J]. Computer Engineering and Applications, 2020, 56(14): 45-51.
- [27] Bezdek J C, Ehrlich R, Full W. FCM: the fuzzy C-means clustering algorithm[J]. Computers & Geosciences, 1984, 10(2): 191-203.
- [28] Wang W, Zhang Y. On fuzzy cluster validity indices[J]. Fuzzy Sets & Systems, 2007, 158(19): 2095-2117.