

多运营工况下高速列车轴箱轴承的 阈值自适应智能健康监测方法

高扬^{1,2}, 阮鑫懿¹, 杨彬¹, 雷亚国¹, 李响¹, 李乃鹏¹

(1. 西安交通大学现代设计及转子轴承系统教育部重点实验室, 710049, 西安;

2. 中车戚墅堰机车车辆工艺研究所股份有限公司, 213011, 江苏常州)

摘要: 为了克服高速列车轴箱轴承的健康监测指标随运营线路、速度变化而波动, 解决传统阈值法难以基于统一标准进行轴承健康监测的问题, 提出了多运营工况下高速列车轴箱轴承的阈值自适应智能健康监测方法。首先, 通过分析特征与转速的 Spearman 相关系数, 对多运营速度下的轴承健康指标进行归一化处理, 消除车速变化所产生的健康指标波动, 从而实现监测阈值与列车运营工况的自动适配。然后, 建立基于随机特征子集和主成分方向分割的改进孤立森林 (SPCD-iForest) 算法, 利用多维特征提供的协同信息在数据的主成分方向对轴承的正常与故障状态进行划分, 并在保持监测准确性的同时提高异常检测的计算效率。最后, 通过列车轴箱轴承线路试验数据对提出的智能健康监测方法进行验证。结果表明: 所提方法消除了列车运营工况变化对轴箱轴承健康指标的影响, 其输出的异常因子为 0~1 之间的无量纲指标, 可有效反映轴箱轴承健康状态的变化; 相较于轨旁声学检测方法, 能够提早 10 d 以上发现轴承故障。所提方法可适应列车不同运营速度、不同线路下的监测诊断需求, 对保障高速列车安全运行、实现轴箱轴承的预测性延寿维护具有重要意义。

关键词: 高速列车; 轴箱轴承; 健康监测; 孤立森林; 特征归一化

中图分类号: TH17 **文献标志码:** A

DOI: 10.7652/xjtuxb202507002 **文章编号:** 0253-987X(2025)07-0013-11

Threshold Adaptive Intelligent Health Monitoring Method for Axle Box Bearings of High-Speed Trains Under Multiple Operating Conditions

GAO Yang^{1,2}, RUAN Xinyi¹, YANG Bin¹, LEI Yaguo¹, LI Xiang¹, LI Naipeng¹

(1. Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University,

Xi'an 710049, China; 2. CRRC Qishuyan Institute Co., Ltd., Changzhou, Jiangsu 213011, China)

Abstract: To overcome the fluctuations in health monitoring indicators of axle box bearings in high-speed trains due to variations in operating lines and speeds, and to address the difficulties of traditional threshold methods in monitoring bearing health based on a unified standard, a threshold adaptive intelligent health monitoring method for axle box bearings under multiple operating conditions is proposed. First, by analyzing the Spearman correlation coefficients between features and rotational speed, the health indicators of the bearings at multiple operating

收稿日期: 2024-12-30。 作者简介: 高扬(1988—), 男, 高级工程师; 杨彬(通信作者), 男, 助理教授, 硕士生导师。
基金项目: 国家自然科学基金资助项目(52305129); 中国博士后科学基金资助项目(2023M732789); 中央高校基本科研业务费专项资金资助项目。

网络出版时间: 2025-04-01

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20250331.0956.002>

speeds are normalized to eliminate fluctuations caused by changes in train speed. This enables the automatic adaptation of monitoring thresholds to the train's operating conditions. Next, an improved Isolation Forest algorithm based on random feature subsets and principal component direction segmentation (SPCD-iForest) is established. This algorithm uses the collaborative information provided by multidimensional features to classify the normal and faulty states of the bearings along the principal component direction of the data, enhancing the computational efficiency of anomaly detection while maintaining monitoring accuracy. Finally, the proposed intelligent health monitoring method is validated using data from line tests of train axle box bearings. Results show that the proposed method eliminates the impact of changes in train operating conditions on the health indicators of axle box bearings. The output anomaly factor is a dimensionless indicator ranging from 0 to 1, effectively reflecting changes in the health status of the axle box bearings. Compared to the TADS trackside acoustic detection method, it can detect bearing faults more than 10 days in advance and is adaptable to monitoring and diagnostic needs across different train speeds and lines. This method is significant for ensuring the safe operation of high-speed trains and enabling predictive maintenance for axle box bearings.

Keywords: high-speed trains; axle box bearings; health monitoring; isolated forest; feature normalization

高速铁路作为国家重要基础设施与国民经济大动脉,是推动实现中国式现代化的重要引擎。截至2023年底,全国铁路运营里程达15.9万km,高铁里程达4.5万km,稳居世界第一^[1]。高速度、高密度的铁路运营对列车的安全可靠运行提出了更高要求。2021年,国家铁路局印发《“十四五”铁路科技创新规划》,明确指出要以智能铁路技术突破、安全保障技术明显提升为发展目标,构建形成基于全方位检测监测数据的风险预测预警与防控的一体化主动安全防控技术体系^[2]。

轴箱轴承是高速列车走行部的关键部件,在列车运营过程中,承受了来自车体和转向架传递的静态压力,同时,在牵引、制动工况下还会承受沿轨道方向的动态载荷^[3]。此外,在车轮多边形、踏面擦伤、钢轨波磨等轮轨关系恶化的条件下,轴箱轴承退化过程加剧^[4]。为了及时发现潜在隐患,避免重大事故的发生,轴箱轴承的实时健康监测至关重要^[5]。

当前,列车轴箱轴承的健康监测主要采用计划检修与轨旁声学检测。其中,计划检修难以及时发现轴承的突发性故障^[6-7],轨旁声学检测仅在列车通过检测点时才进行数据获取与分析,难以判断列车轴承的实时健康状况^[8-9]。为提升列车轴承健康监测的实时性,实时分析列车轴箱轴承的振动、温度等数据,提取能够反映轴承健康状态变化的特征指标,并结合阈值法判断轴承正常与否,是当前高速列车车载端轴箱轴承健康监测的迫切需求^[10-11]。

围绕轴承智能健康监测,国内外学者开展了大量工作。例如,Randall等系统总结了轴承故障诊断中的信号处理方法,为后续研究奠定了重要基础^[12]。Xu等揭示了不同径向间隙的轴承特性,提出了一种基于调制信号双谱边带估计器的轴承间隙监测指标^[13]。Abhilash等提出了一种改进的小波互谱方案用于滚动轴承的健康监测,借助异常值技术从统计指标中选择贡献带宽,提高其他互相关尺度的指标^[14]。Wang等对谱峰度、谱负熵、谱基尼指数和谱平滑度指数进行了研究,揭示了它们之间的密切关系,分别给出了基于领域知识的权重设计和数据驱动方法^[15]。雷亚国等开发了基于深度学习的轴承智能诊断方法,实现了故障特征的自适应提取^[16]。Wang等将卷积神经网络与双向门控循环单元模型相结合,利用自适应噪声和主成分分析的完全集合经验模态分解自动构建健康指标^[17]。

不同线路的运营条件存在差异,尤其在在不同运营速度下,轴箱轴承的健康指标随着外部工况的变化而显著波动。高速行驶时,列车的动态载荷远高于低速行驶时,这直接影响了轴箱轴承的受力情况,使得固定阈值法难以准确判断轴承健康状态。针对这些问题,传统的轴承健康监测方法往往依赖于经验丰富的专家手动设置多个固定阈值,这不仅耗时耗力,需要大量的数据积累,而且灵活性差,难以适应复杂多变的实际工况^[18-19]。

为了消除工况变化对轴承健康监测的影响,An

等提出了基于循环神经网络的轴承监测框架,通过长短期记忆网络中门单元的作用,忽略工况状态^[20]。Wang等提出了一种新的多源域特征自适应网络,使用特征提取器从不同的源域和目标域对中学习可迁移特征,然后构建特定于域的分布对齐模块,减少所有域对之间的偏移^[21]。Zhou等建立混合系统响应模型来描述设备的运行过程,以神经网络参数作为健康状态的表示,并通过对神经网络参数进行降维,构建健康指标进行实时监测^[22]。Moshrefzadeh等提出频谱幅度调制方法,计算信号包络谱的脉冲性以量化其循环平稳性水平,用于变载荷及变转速的健康监测^[23]。这些方法虽然可以在变工况下进行,但通常需要复杂的模型设计和参数调整,这种复杂性限制了它们的实际应用,尤其是在高速列车的实时监控中,难以实现快速、精准的响应。

针对上述问题,本文通过寻找最优阶次,实现多运营速度下的特征归一化,并通过结合随机维度特征子集和主成分方向划分数据的改进孤立森林算法(SPCD-iForest),使用异常分数来进行轴承的健康监测,适应不同运营速度、不同线路下的监测诊断需求。本文所提出的方法模型简单、计算效率高,可为高速列车轴箱轴承的实时健康监测提供一种新的解决方案。

1 多运营速度下的特征归一化

在高速列车的行驶过程中,不同的运营速度会导致轴箱轴承的振动信号在频域和时域上的特征值产生差异,从而影响振动监测阈值的设置。为了克服这一挑战,本文提出了多运营速度下的特征归一化处理,方法流程如图1所示,具体步骤如下。

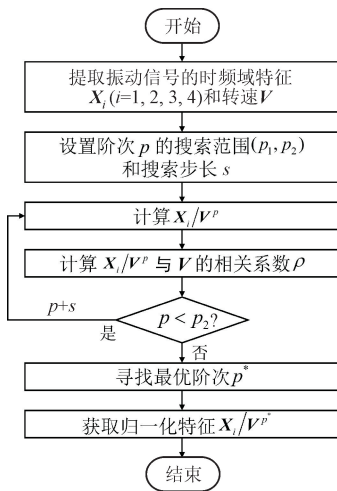


图1 多运营速度下的特征归一化方法流程图

Fig.1 Flowchart of feature normalization method under multiple operating speeds

步骤1 获取特征和转速信号。从列车振动信号中提取出若干时域、频域特征 X_i , 同时获取转速信号 V 。本文所选用的特征包括时域均方根、峰峰值、频域均方根和谱能量, 即 $i = 1, 2, 3, 4$ 。这些特征能够有效反映振动信号的强度, 对于后续分析具有重要意义。

步骤2 设置阶次搜索参数。在确定相关特征和转速信号后, 接下来需要设置阶次搜索参数。具体而言, 需要确定阶次 p 的搜索范围 (p_1, p_2) 以及搜索步长 s 。本文设置搜索范围为 $(0.01, 10)$, 搜索步长设置为 0.01 。设置初始阶次 $p = p_1$ 作为起始值进行后续的阶次优化过程。

步骤3 计算归一化特征。基于当前的阶次 p , 计算每个特征与转速之间的归一化特征形式, 即 X_i / V^p 。此归一化处理旨在消除转速变化对振动特征的影响, 以便更准确地提取反映轴箱轴承健康状态的特征信息。

步骤4 计算相关性。在获得归一化特征后, 计算每个归一化特征 X_i / V^p 与转速信号 V 之间的相关系数 ρ 。由于轴箱轴承的健康指标与转速之间的关系通常呈现非线性且不完全符合正态分布, 并且健康指标的异常波动较大, 因此本研究中采用 Spearman 相关系数来捕捉这种单调的变化趋势^[24]。

步骤5 判断是否完成搜索。计算完相关系数后, 判断当前阶次 p 是否已经达到了预定的搜索上限 p_2 。如果当前阶次小于搜索上限, 则调整阶次为 $p + s$, 并返回步骤3与步骤4继续进行归一化处理和相关性计算。如果当前阶次已经达到或超过搜索上限, 则结束搜索。

步骤6 寻找最优阶次。在完成所有阶次的搜索后, 基于计算得到的相关性结果, 选择相关系数最小的阶次 p^* 作为最优阶次。此时, 归一化后的特征 X_i / V^{p^*} 与转速 V 的相关系数达到最低值。换言之, 此阶次下的归一化特征可以最大限度地消除转速变化对监测阈值的影响。

步骤7 获取归一化特征。利用最优阶次 p^* 对所有特征进行归一化处理, 得到特征 X_i / V^{p^*} 。这些特征可以作为输入数据, 用于进一步的健康监测任务。归一化后的特征不再受转速波动的干扰, 能够更加准确地反映轴箱轴承的状态。

图2展示了4种时域、频域特征(时域均方根、峰峰值、频域均方根、谱能量)在阶次搜索过程中的相关性变化。通过观察这些图形可以明显看到, 每个特征在某一特定阶次下, 归一化后的特征与转速

之间的相关系数达到最低值,而这个最低值非常接近于0。这表明,采用最优阶次后,归一化特征与转速几乎无关,从而实现了排除转速干扰的目的。

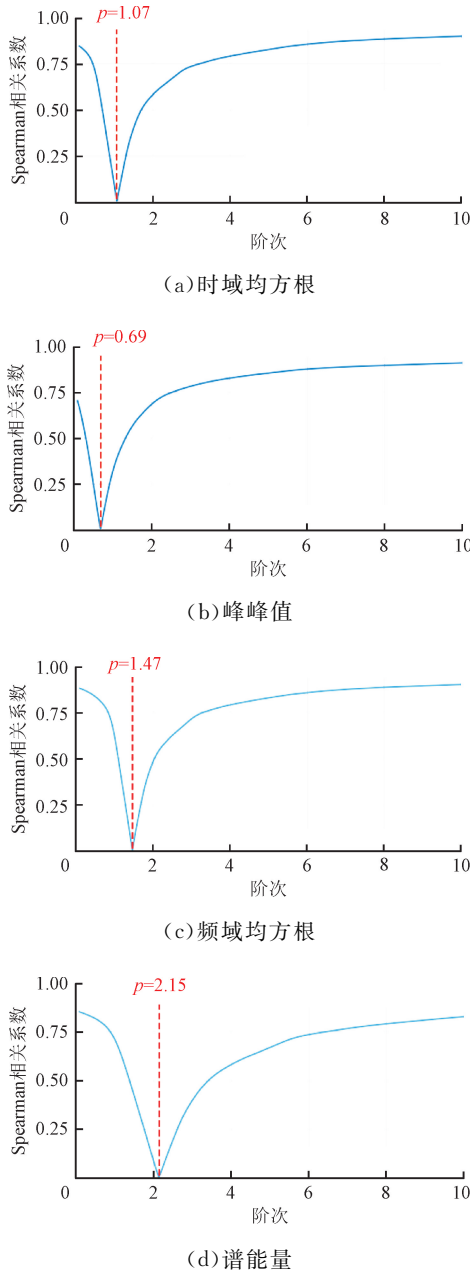


图2 各特征阶次搜索过程中的相关性

Fig. 2 Correlation during the order search process of each feature

2 SPCD-iForest 智能健康监测模型

针对不同高速列车的轴箱轴承故障阈值难以设定的问题,本文在孤立森林(iForest)算法的基础上,设计了改进算法——SPCD-iForest,利用滑动窗口获取轴箱轴承实时振动数据的异常得分,实现轴

箱轴承的健康监测。

2.1 孤立森林算法

孤立森林是一种无监督异常检测算法,不需要借助类似距离、密度等指标去描述样本与其他样本的差异,而是直接通过分割数据的难易去刻画疏离程度^[25],算法简单、高效,在工业界应用较多。算法包括训练和评分两个阶段。

在训练阶段进行之前,需要明确构建的孤立树数量 k ,以及每棵孤立树的最大深度 l 。训练阶段的步骤如下。

步骤1 随机采样。假设数据集 $D = \{x_1, x_2, \dots, x_n\}$,包含 n 个样本,每个样本 $x_i \in \mathbf{R}^d$,即每个样本包含 d 个特征值。从数据集中随机抽取一个子集 $D' \in D$,子集大小为 φ (通常 $\varphi \ll n$)。

步骤2 划分数据。对每个子集 D' ,递归构造孤立树;随机选择一个特征 $f = \{1, 2, \dots, d\}$ 。在该特征范围内随机选择一个分割点 c ,满足 $c \in [\min D'_f, \max D'_f]$,将数据集 D' 分成两部分: $D'_L = \{x \in D' : x < c\}$ 和 $D'_R = \{x \in D' : x \geq c\}$ 。

步骤3 多棵树的构造。对 D'_L 和 D'_R 递归执行上述步骤,直到以下停止条件之一成立:①数据已经无法再分割,即只包含一条数据;②达到预设的树高度 l 。此时得到一棵孤立树 T_1 。重复上述步骤,构建 k 棵孤立树 $T_1, T_2, \dots, T_k$ 。

孤立森林通过计算样本在所有孤立树中的平均路径长度,并利用其与预期路径长度的对比来评估异常分数。异常分数计算公式为

$$s(x, \varphi) = 2^{-E(h(x))/c(\varphi)} \quad (1)$$

式中: x 为要隔离的样本; $h(x)$ 为指样本通过该孤立树,从树的根节点到达该样本被孤立的节点所需要的边数; $E(h(x))$ 为该样本在森林中所有树的路径长度的平均值; $c(\varphi)$ 为树的平均路径长度,作为标准化的基准,用于将路径长度标准化,具体如下

$$c(\varphi) = \begin{cases} 0, & \varphi < 2 \\ 1, & \varphi = 2 \\ 2H(\varphi - 1) - \frac{2(\varphi - 1)}{\varphi}, & \varphi > 2 \end{cases} \quad (2)$$

式中: $H(\varphi - 1)$ 为调和数, $H(\varphi - 1) = \ln(\varphi - 1) + 0.58$ 。

由式(1)和式(2)可知,当样本的平均路径长度与树的平均路径长度相等,即 $E(h(x)) = c(\varphi)$,异

常分数 $s(x, \varphi) = 0.5; E(h(x))$ 越小, $s(x, \varphi)$ 越趋近于 1; $E(h(x))$ 越大, $s(x, \varphi)$ 趋近于 0。也就是说,样本的平均路径长度越大,异常分数越低,且通常可以认为异常分数不超过 0.5 时,对应样本为轴承正常状态下采集。判定为异常的阈值,则需要根据实际需求进行设置。

2.2 SPCD-iForest 算法

在实际应用中,振动数据通常具有高维、分布复杂的特点,而轴承故障信号往往埋藏在正常信号的噪声中。传统的孤立森林算法在处理高维数据时,以单维度特征的随机划分为基础,这种方法难以捕捉特征之间的相互关系,导致对异常点的检测能力有限。尤其是在故障信号表现为多特征联合变化时,单维划分的方式很容易忽略关键的异常模式,存在明显不足。

传统的孤立森林算法基于单特征划分的方式存在“伪影”问题,导致部分数据的异常检测出现错误,而扩展孤立森林算法采用二维特征的随机斜率划分,缓解了单维特征划分的局限性^[26]。然而,这种算法依然存在不足。一方面,仅考虑二维特征仍然难以全面刻画高维数据中复杂的特征关联性;另一方面,使用随机斜率作为划分依据缺乏对数据分布的有效利用,划分方向往往与数据的主要变化方向不一致。这种随机性在高维场景下更加明显,难以准确描述多特征间的联合变化,限制了对复杂异常模式的检测能力。

针对上述问题,本文提出了一种基于随机特征子集和主成分方向分割的孤立森林算法,能够更高效地利用特征间的协同信息,提升高维复杂数据的异常检测能力。

改进后的孤立森林算法在训练阶段采用了新的分割策略,主要步骤如下。

步骤 1 随机维度子集采样。假设数据集 $D = \{x_1, x_2, \dots, x_n\}$, 包含 n 个样本,每个样本 $x_i \in \mathbf{R}^d$ 。从数据集中随机抽取一个子集 $D' \in D$, 子集大小为 φ (通常 $\varphi \ll n$)。从 d 个特征中随机选择 t 个特征,构成随机特征子集 $F = \{x_1, x_2, \dots, x_\varphi\} \in \mathbf{R}^t$, 其中 $1 \leq t \leq d$ 。从所有可用的特征中随机挑选出一小部分特征,用来帮助决策树进行数据划分。这种方式不仅可以避免模型过度依赖于某些特定特征,提高算法的泛化能力,而且能更高效地利用高维数据中不同特征的组合所包含有用的信息,从而更好地识别异常点。

步骤 2 计算主成分方向。对随机特征子集 F 进行主成分分析(若后续树的构造选取到同样的特征子集,则无需重复进行),获得 t 个主成分方向向量 $\{v_1, v_2, \dots, v_t\}$ 及其对应的重要性比例(解释的方差) $\{r_1, r_2, \dots, r_t\}$ 。主成分的排序按照其对应的解释的方差由大到小排列,其中 v_i 代表第 i 个主成分方向, r_i 代表第 i 个主成分方向解释的方差的比,表示该主成分对数据方差的贡献。例如 v_1 表示第一主成分方向(解释方差最大)。

步骤 3 基于主成分方向划分数据。每次划分基于对应主成分构造超平面:超平面方程为 $v_i^T x = b$, $i = 1, 2, \dots, t$, 其中 b 为随机截距,取值范围为该特征子集在 v_i 方向上投影的最小值与最大值之间。根据主成分方向的重要性比例,动态调整划分方向的频率。例如,对于二维特征子集,若 v_1 的重要性是 v_2 的 2 倍,则划分过程中超平面 $v_1^T x = b$ 的出现频率是超平面 $v_2^T x = b$ 的 2 倍。

步骤 4 多棵树的构造。按照主成分方向的加权比例递归执行划分,直到满足停止条件,得到孤立树 T_1 。重复上述步骤,构建 k 棵孤立树 $T_1, T_2, \dots, T_k$ 。评分阶段与传统孤立森林算法相同。

图 3 展示了本文 SPCD-iForest 算法在不同维度特征下的划分策略。图 3(a)中,以时域均方根单一特征为例,数据具有唯一的主成分方向,此时 SPCD-iForest 与传统孤立森林算法完全一致,因而传统孤立森林算法可以视为 SPCD-iForest 的一种特例。图 3(b)中,以二维特征(时域均方根、峰峰值)为例,红色划分线由第一主成分方向确定,蓝色划分线由第二主成分方向确定。由于使用了主成分重要性加权划分策略,红色划分线的数量多于蓝色划分线。图 3(c)进一步将分割情况扩展至三维特征(时域均方根、峰峰值及频域均方根),其中红色划分面由第一主成分方向确定(与第一主成分垂直)。该平面优先捕捉数据的主要变化趋势,随着划分的递归进行,其他主成分方向生成的划分面进一步细化数据分布区域,从而更精准地实现异常点的隔离检测。

相比传统孤立森林算法和现有的扩展孤立森林算法,本文 SPCD-iForest 算法通过选择随机特征子集,增强了对复杂异常模式的检测能力,同时利用主成分分析,快速聚焦于数据的主要变化方向,降低了划分随机性,使划分方向更加贴合数据的实际分布。

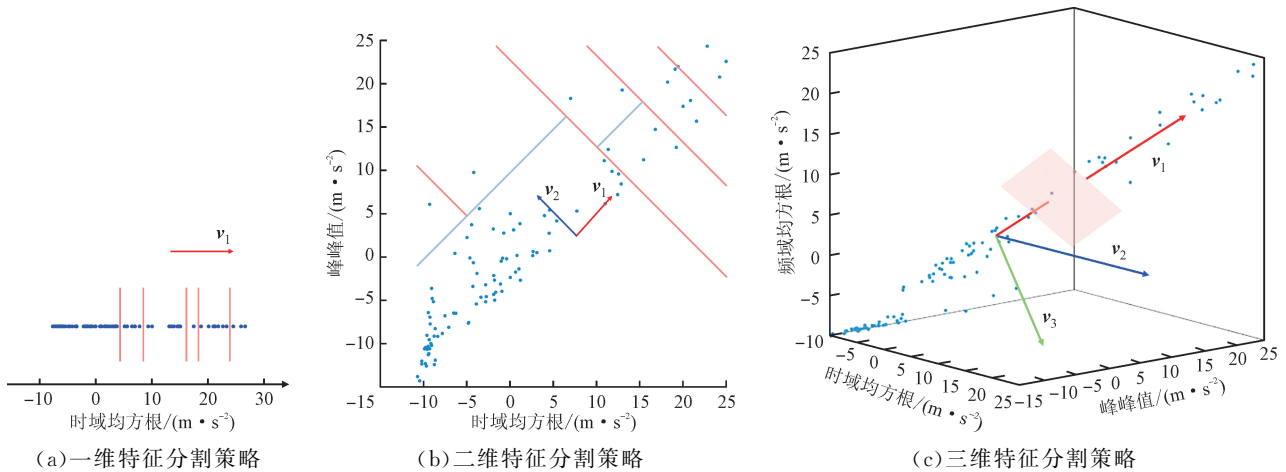


图 3 不同维度特征时 SPCD-iForest 健康监测的分割策略示意图

Fig. 3 Schematic diagram of segmentation strategy for SPCD-iForest with different dimensional features

2.3 整体模型框架

为了实现高速列车轴箱轴承实时健康监测,本文

设计了基于 SPCD-iForest 的健康监测模型,模型框架如图 4 所示,分为离线建模和在线测试两个阶段。

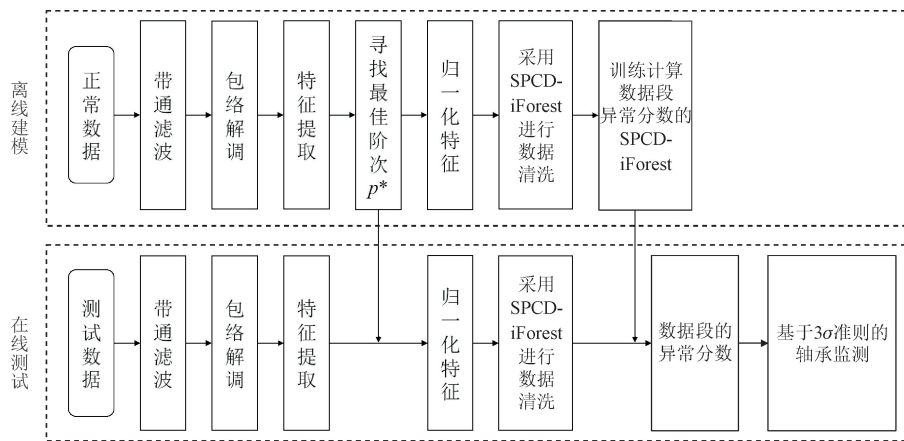


图 4 基于 SPCD-iForest 健康监测模型整体框架

Fig. 4 Overall framework of the health monitoring model based on SPCD-iForest

2.3.1 离线建模

在离线建模阶段,模型使用历史监测数据进行训练,流程包括以下步骤。

数据预处理与特征提取:首先使用带通滤波器去除无关频段的噪声干扰,保留轴承谐振频率范围内的有效信号。对滤波后的信号进行包络解调分析,提取包络信号中的特征频率成分。从包络解调后的信号中提取多种统计特征,这里包括 4 种时域和频域特征:时域均方根(反映时域信号的整体能量水平)、峰峰值(表示信号的振动幅度大小)、频域均方根(表示频域中信号的能量强度)、谱能量(描述信号的整体能量分布)。

特征归一化:结合轴承的运行速度,利用图 1 所

示的多运营速度下的特征归一化方法,确定最佳特征阶次 p^* ,从而实现速度归一化,尽可能降低速度对特征的影响。

数据清洗及模型训练:由于转向架和轮轨直接接触,一旦轨道不平顺就会产生冲击,这些冲击可能会导致误报,因此采用箱线图剔除其中存在的异常点^[27],通过计算每个特征的四分位数,剔除大于上限的异常值,以减少瞬时冲击带来的干扰。为了进一步增强模型的鲁棒性,将滑窗内数据点的异常分数取平均值,作为该数据段的综合异常分数。随后,利用这些数据段的异常分数训练 SPCD-iForest 模型,用于后续的健康监测。

2.3.2 在线测试

在测试阶段,模型用于实时监测高速列车轴承的健康状态,流程包括以下步骤。

数据预处理与特征提取:与离线建模阶段相同,需要对轴承实时振动数据进行带通滤波和包络解调,然后再提取时域和频域特征。

特征归一化:利用离线阶段寻找到的最优阶次 p^* 将提取的特征进行归一化处理,减小转速变化对特征产生的影响。

数据清洗及模型评价:首先利用采用箱线图对实时特征进行清洗,剔除瞬时冲击带来的干扰,增强模型的鲁棒性。清洗后的实时数据被存储于缓冲区,缓冲区存储一定数量(数量由滑窗决定)的数据点后,输入离线阶段训练好的 SPCD-iForest 模型,计算数据段的异常分数,最后结合模型输出的异常分数,基于 3σ 准则确定阈值,对列车轴承的运行状态进行实时监测。

3 数据集及验证结果

3.1 数据集介绍

采用两组高速列车轴箱轴承的线路试验数据对本文所提方法进行验证。试验中设置采样频率为 40 kHz,每次采集轴承转动 20 圈的振动信号,并使用轨旁声学系统(TADS)进行检测和报警。

图 5 为某天内线路试验列车的轮轴转速变化。对于采集到的振动数据,首先,提取 4 种时域和频域特征,包括时域均方根、峰峰值、频域均方根和谱

能量。以第一组试验轴承为例,绘制了时域均方根和谱能量的特征变化图,如图 6(a)、(d)所示。由于不同车速下的特征数据混杂在一起,同时受到瞬时冲击干扰,因此特征变化趋势不明显。若直接依据这些特征划定阈值,将增加漏报和误报的风险。

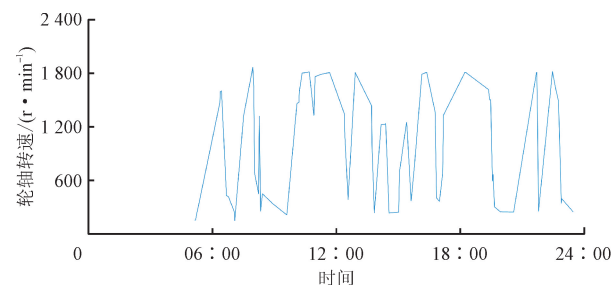


图 5 某天内列车轮轴转速变化图

Fig. 5 Wheel axle speed change chart within one day

3.2 特征归一化分析

针对运营速度多变的问题,对特征数据进行了归一化处理,消除不同运营条件下的量纲差异。根据图 2 所示的分析结果,确定了时域均方根和谱能量的最优阶次分别为 1.07 和 2.15。如图 6(b)、(e)所示,归一化后的特征变化能够有效减少速度变化的影响,从而更清晰地揭示特征变化趋势。为了剔除由于瞬时冲击干扰带来的噪声,采用了箱线图方法进行数据清洗。如图 6(c)、(f)所示,清洗后的特征表现出更加稳定的变化趋势,较好地反映了轴承的实际退化过程。

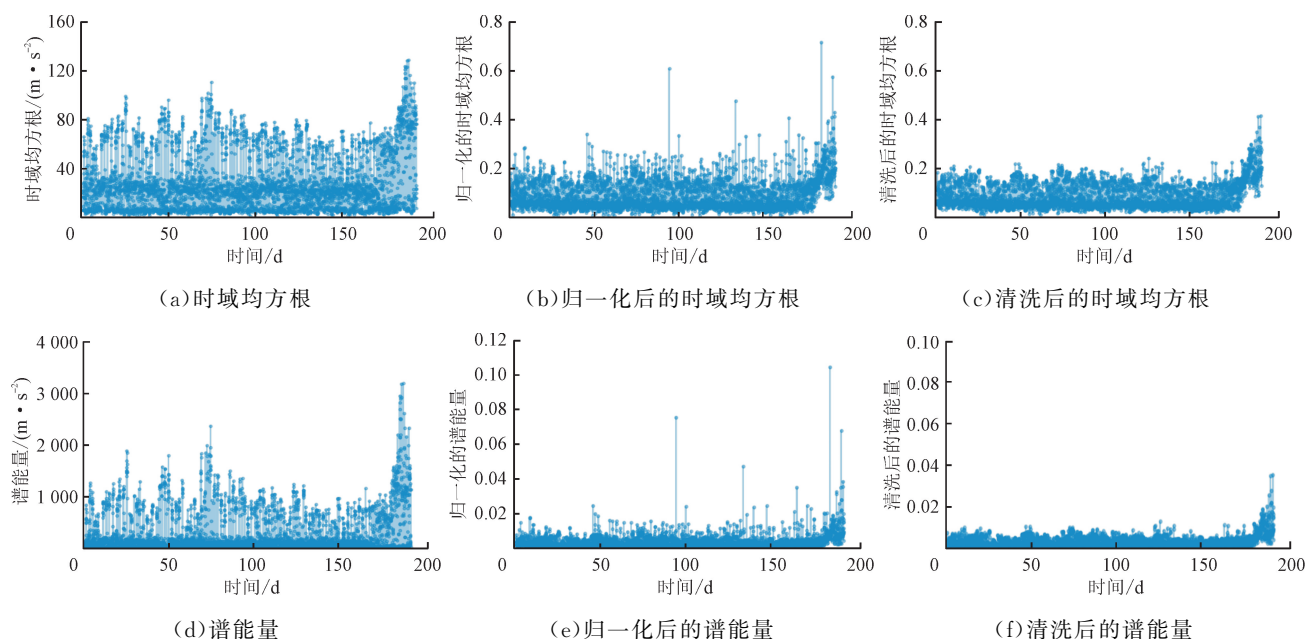


图 6 第一组轴承试验数据均方根、谱能量的各阶段特征变化图

Fig. 6 Characteristic changes of RMS of test data and spectral energy in each stage of the first bearing

3.3 健康监测结果分析

使用离线正常数据训练 SPCD-iForest 模型,并对清洗后的特征计算数据段的异常分数。在模型训练和测试过程中,关键参数的设置如表 1 示,其中,模型中森林树的数量设为 100,即使用 100 棵树来构建模型,以保证存在足够的决策树来提高检测精度。训练阶段假设数据集无异常值,因此异常点比例参数接近于 0;在测试阶段,为识别出异常状态,设置异常点比例参数为 0.1,即数据集中约有 10% 的异常点。每棵树使用的特征数量设置为 1~4 之间的随机值,即通过选择随机特征子集,增强模型对复杂异常模式的检测能力。此外,为增强模型鲁棒性,计算一定滑窗大小范围内数据异常分数的平均值,以此作为该数据段的综合异常分数,考虑到滑窗大小取值过大会导致报警日期的准确度降低,因此本文中选取滑窗大小为 20。

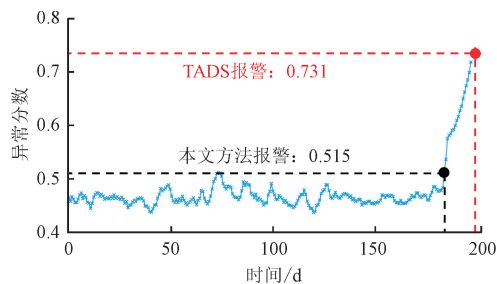
表 1 模型关键参数

Table 1 Model key parameters

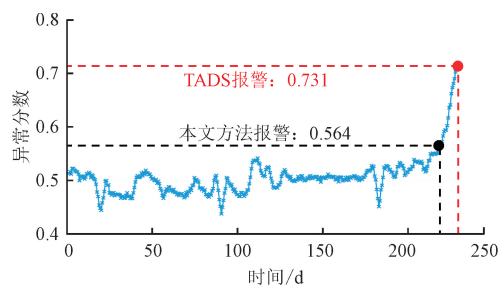
参数	训练阶段取值	测试阶段取值
孤立森林树的个数	100	100
异常点比例参数	0.001	0.100
特征数	1~4 的随机值	1~4 的随机值
滑窗大小	20	20

图 7 展示了两组轴承的异常分数变化情况。从图中可以观察到:轴承运行的初期阶段,异常分数维持在 0.5 上下浮动,与理论分析结果一致,表明轴承处于健康状态;在报警前的一段时间内,异常分数逐渐上升;两组轴承分别经过 196 d 10 h 和 234 d 16 h, TADS 系统发生报警,报警时两组轴承的异常分数分别为 0.731 和 0.713。

基于异常分数的前 75% 数据,并应用 3σ 准则计算得出的阈值分别为 0.515 和 0.564。对应的报警时间分别为 181 d 4 h 和 223 d 16 h,相较于 TADS 系统的报警时间,分别提前了 15 d 6 h 和 11 d,表明通过结合异常分数和 3σ 准则进行健康监测,可以早于 TADS 系统感知轴承的潜在故障,从而降低运行风险并减少维护成本。



(a) 第一组轴承的异常分数变化

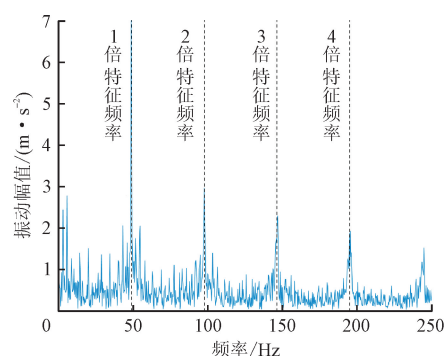


(b) 第二组轴承的异常分数变化

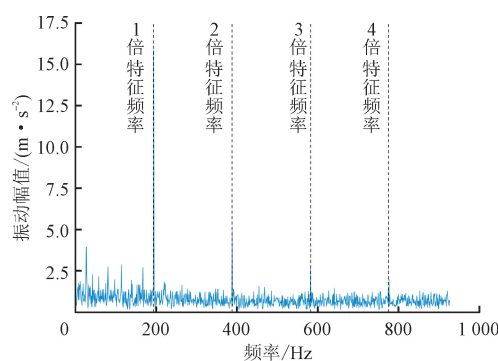
图 7 两组列车的异常分数变化

Fig. 7 Abnormal score changes of two sets of trains

两组数据在本方法报警时的包络谱如图 8 所示。由图可以明显看到,在报警时,轴承的外圈故障特征频率及其倍频突出,说明此时轴承发生了外圈故障。因此,基于本文算法的诊断结果能够准确反映出轴承的异常状态,并为故障诊断提供有效依据。



(a) 第一组轴承



(b) 第二组轴承

图 8 两组外圈故障轴承报警时包络谱

Fig. 8 Envelope spectrum when two sets of outer ring faulty bearings alarm

3.4 算法对比

为验证 SPCD-iForest 算法的性能优越性,将其与现有的扩展孤立森林算法进行了对比。针对同一段数据集进行异常值检测,扩展孤立森林从 4 个特

征(时域均方根、峰峰值、频域均方根和谱能量)中选择两个特征,并基于这两个特征进行随机斜率划分;SPCD-iForest 算法随机选择特征,并利用主成分分析确定数据的主要变化方向,基于这些主成分方向进行加权划分。图 9 直观地展示了上述 2 种算法基于时域均方根和峰峰值时的划分策略。

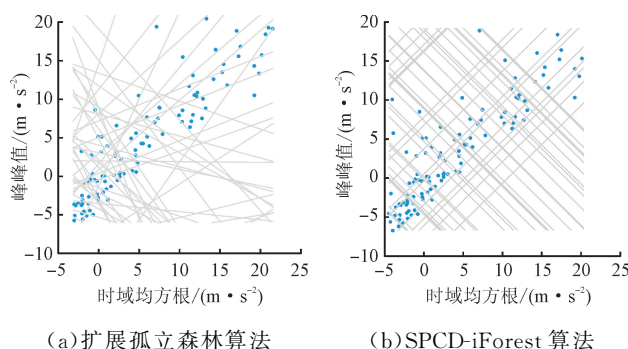


图 9 扩展孤立森林与 SPCD-iForest 的特征划分策略对比
Fig. 9 Comparison of feature partitioning strategies between extended iForest and SPCD-iForest

从第一组线路试验数据中选取包含正常点和异常点(比例约 10%)的数据集,并利用 3σ 准则计算出的阈值(即 0.515)进行异常检测,分别使用扩展孤立森林和 SPCD-iForest 算法进行对比。结果表明,2 种算法在异常点识别方面都具有较好效果,检出率均达到 95% 以上。考虑到高速列车轴箱轴承健康监测要求实时性强,进一步比较了 2 种算法的计算效率。当森林中树的数量均设为 100、每棵树的样本数量设为 200 时,预设的树高度为远大于样本总数,直到数据无法分割才会停止划分。

图 10 绘制了不同特征组合下扩展孤立森林和 SPCD-iForest 算法的划分次数,前 6 种算法为特征两两组合时的扩展孤立森林。结果表明,扩展孤立森林算法在遍历所有特征组合下的划分次数均多于 SPCD-iForest,表明 SPCD-iForest 在划分数据时能够更好地利用数据的内在结构,特别是考虑到特征间的联合变化和主成分方向,能够减少划分次数并

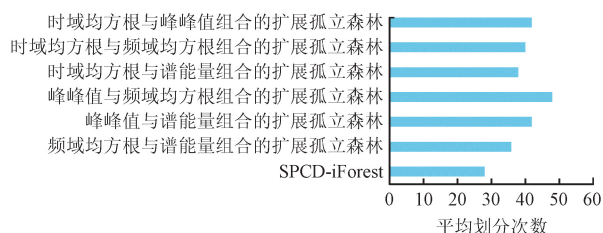


图 10 扩展孤立森林和 SPCD-iForest 算法的划分次数统计
Fig. 10 Statistics on the number of partitions for extended isolation forest and SPCD-iForests

提高计算效率。这说明 SPCD-iForest 算法在保证轴承异常检测准确性的同时,能够更快地完成检测任务,满足适合实际工程中实时监测的需求。

4 结 论

本文分析了多运营工况下高速列车轴箱轴承的振动数据特点,并针对轴承健康监测阈值难以统一的问题,提出了基于特征归一化和 SPCD-iForest 的阈值自适应智能健康监测方法,在线路试验数据上进行了测试,得到以下结论。

(1)提出了多运营速度下的轴箱轴承特征归一化方法,通过 Spearman 相关系数评估不同阶次下归一化特征与转速的相关性,选取相关性最小的阶次作为最优阶次,从而消除了转速变化对监测特征的影响,使不同运营速度下的特征具有相同尺度,为健康监测提供了统一的特征基础。

(2)设计了基于随机特征子集和主成分方向分割的改进孤立森林(SPCD-iForest)算法,结合多维特征的协同信息及其主成分方向进行划分。相比传统的单维度特征划分和扩展孤立森林的随机斜率划分,SPCD-iForest 算法能更准确地捕提高维数据中的异常模式。对比实验表明,SPCD-iForest 算法在保持检测准确性的同时减少了划分次数,具有更高的计算效率。

(3)通过线路试验数据验证,所提方法输出的异常分数为 0~1 之间变化的无量纲指标,能够有效反映轴箱轴承的健康状态变化。通过 3σ 准则计算出的阈值,可自动适配不同列车、不同运营速度下的轴箱轴承健康监测任务。

尽管所提方法在线路试验数据中具有良好的效果,但仍然存在待优化的地方。例如,如何结合更多的实车数据和复杂工况,以提高模型的泛化能力。此外,所提方法能够较早诊断出轴承故障,有助于提前制定维护计划,但在工程实际中,过早的报警或导致不必要的停机和维护。对此,下一步可通过以下方式进行优化:一是动态调整报警阈值,结合实时监测数据,避免在故障初期就发出报警;二是结合故障演化模型,通过分析故障发展趋势来推迟报警时机;三是采用多层次报警系统,在不同阶段触发不同级别的报警,以便在轴承故障对系统产生实际影响之前发出预警。未来研究可以探讨如何合理地平衡早期诊断和实际应用中的需求,从而提升整体的监测效果和工程应用中的实用性。

参考文献:

- [1] 丁叁叁, 陈大伟, 刘加利. 中国高速列车研发与展望 [J]. 力学学报, 2021, 53(1): 35-50.
DING Sansan, CHEN Dawei, LIU Jiali. Research, development and prospect of China high-speed train [J]. Chinese Journal of Theoretical and Applied Mechanics, 2021, 53(1): 35-50.
- [2] “十四五”铁路科技创新规划 [J]. 铁道技术监督, 2022, 50(1): 9-15.
Railway science and technology innovation plan in the 14th Five-year Plan [J]. Railway Quality Control, 2022, 50(1): 9-15.
- [3] WANG Baosen, LIU Yongqiang, ZHANG Bin, et al. Development and stability analysis of a high-speed train bearing system under variable speed conditions [J]. International Journal of Mechanical System Dynamics, 2022, 2(4): 352-362.
- [4] JIN Xuesong. Research progress of high-speed wheel-rail relationship [J]. Lubricants, 2022, 10(10): 248.
- [5] 顾晓辉, 杨绍普, 刘文朋, 等. 高速列车轴箱轴承健康监测与故障诊断研究综述 [J]. 力学学报, 2022, 54(7): 1780-1796.
GU Xiaohui, YANG Shaopu, LIU Wenpeng, et al. Review of health monitoring and fault diagnosis of axle-box bearing of high-speed train [J]. Chinese Journal of Theoretical and Applied Mechanics, 2022, 54(7): 1780-1796.
- [6] SUN Haimeng, HE Deqiang, ZHONG Jiecheng, et al. Preventive maintenance optimization for key components of subway train bogie with consideration of failure risk [J]. Engineering Failure Analysis, 2023, 154: 107634.
- [7] ZHANG Huixian, WEI Xiukun, GUAN Qingluan, et al. Joint maintenance strategy optimization for railway bogie wheelset [J]. Applied Sciences, 2022, 12(14): 6934.
- [8] DING Xiaoxi, LI Yulan, XIAO Jiawei, et al. Parametric Doppler correction analysis for wayside acoustic bearing fault diagnosis [J]. Mechanical Systems and Signal Processing, 2022, 166: 108375.
- [9] GONZALO A P, ENTEZAMI M, WESTON P, et al. Railway track and vehicle onboard monitoring: a review [C]//International Conference on Management Science and Engineering Management (ICMSEM 2023). Les Ulis, France: EDP Sciences, 2023: 02014.
- [10] CHEN Hongtian, JIANG Bin, DING S X, et al. Data-driven fault diagnosis for traction systems in high-speed trains: a survey, challenges, and perspectives [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(3): 1700-1716.
- [11] CHENG Chao, WANG Jiuhe, CHEN Hongtian, et al. A review of intelligent fault diagnosis for high-speed trains: qualitative approaches [J]. Entropy, 2020, 23(1): 1-33.
- [12] RANDALL R B, ANTONI J. Rolling element bearing diagnostics: a tutorial [J]. Mechanical Systems and Signal Processing, 2011, 25(2): 485-520.
- [13] XU Minmin, HAN Yaoyao, SUN Xiuquan, et al. Vibration characteristics and condition monitoring of internal radial clearance within a ball bearing in a gear-shaft-bearing system [J]. Mechanical Systems and Signal Processing, 2022, 165: 108280.
- [14] ABHILASH S, PRADEEP R, REJITH R, et al. Health monitoring of rolling element bearings using improved wavelet cross spectrum technique and support vector machines [J]. Tribology International, 2021, 154: 106650.
- [15] WANG Dong, PENG Zhike, XI Lifeng. The sum of weighted normalized square envelope: a unified framework for kurtosis, negative entropy, Gini index and smoothness index for machine health monitoring [J]. Mechanical Systems and Signal Processing, 2020, 140: 106725.
- [16] 雷亚国, 贾峰, 周昕, 等. 基于深度学习理论的机械装备大数据健康监测方法 [J]. 机械工程学报, 2015, 51(21): 49-56.
LEI Yaguo, JIA Feng, ZHOU Xin, et al. A deep learning-based method for machinery health monitoring with big data [J]. Journal of Mechanical Engineering, 2015, 51(21): 49-56.
- [17] WANG Zhiyuan, GUO Junyu, WANG Jiang, et al. A deep learning based health indicator construction and fault prognosis with uncertainty quantification for rolling bearings [J]. Measurement Science and Technology, 2023, 34(10): 105105.
- [18] ENTEZAMI M, ROBERTS C, WESTON P, et al. Perspectives on railway axle bearing condition monitoring [J]. Proceedings of the Institution of Mechanical Engineers: Part F Journal of Rail and Rapid Transit, 2020, 234(1): 17-31.
- [19] LIU Lei, SONG Dongli, GENG Zilin, et al. A real-time fault early warning method for a high-speed EMU axle box bearing [J]. Sensors, 2020, 20(3): 823.
- [20] AN Zenghui, LI Shunming, WANG Jinrui, et al. A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network [J]. ISA Transactions, 2020, 100: 155-170.

- [21] WANG Rui, HUANG Weiguo, WANG Jun, et al. Multisource domain feature adaptation network for bearing fault diagnosis under time-varying working conditions [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-10.
- [22] ZHOU Haoxuan, WANG Bingsen, ZIO E, et al. Hybrid system response model for condition monitoring of bearings under time-varying operating conditions [J]. Reliability Engineering & System Safety, 2023, 239: 109528.
- [23] MOSHREFZADEH A. Condition monitoring and intelligent diagnosis of rolling element bearings under constant/variable load and speed conditions [J]. Mechanical Systems and Signal Processing, 2021, 149: 107153.
- [24] DE WINTER J C F, GOSLING S D, POTTER J. Comparing the Pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data [J]. Psychological Methods, 2016, 21(3): 273-290.
- [25] LIU Fei, TING Kaiming, ZHOU Zhihua. Isolation forest [C]//Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. Piscataway, NJ, USA: IEEE, 2008: 413-422.
- [26] HARIRI S, KIND M C, BRUNNER R J. Extended isolation forest [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(4): 1479-1489.
- [27] MAZAREI A, SOUSA R, MENDES-MOREIRA J, et al. Online boxplot derived outlier detection [J]. International Journal of Data Science and Analytics, 2025, 19(1): 83-97.
- (编辑 亢列梅)