

本文引用格式：林朝晖,廖奕校,招智铭,等.基于特征分布对齐与多传感器融合的水泵异常检测方法[J].自动化与信息工程,2025,46(1):36-40;46.

LIN Zhaohui, LIAO Yixiao, ZHAO Zhiming, et al. Water pump anomaly detection method based on feature distribution alignment and multi-sensor fusion[J]. Automation & Information Engineering, 2025,46(1):36-40;46.

基于特征分布对齐与多传感器融合的水泵异常检测方法

林朝晖¹ 廖奕校² 招智铭¹ 万智勇² 周松斌²

(1.广州市自来水有限公司石门水厂, 广东 广州 510000

2.广东省科学院智能制造研究所, 广东 广州 510070)

摘要: 水泵是供水系统的重要加压设备, 对其进行异常检测并及时发现运行异常, 对保障供水安全具有重要意义。针对现有人工智能方法在水泵异常检测中存在的异常样本获取困难、检测精度低等问题, 提出一种基于特征分布对齐与多传感器融合的水泵异常检测方法。该方法以多通道传感信号的对数梅尔谱为输入; 利用卷积自编码网络来融合多传感器信息; 以最小化信号重构损失和特征分布损失为目标, 训练卷积自编码网络; 基于重构损失计算样本的异常分数, 实现水泵的异常检测。实验结果表明, 该方法有效提高了水泵异常检测的性能。

关键词: 水泵异常检测; 数据重构; 特征分布对齐; 多传感器融合

中图分类号: TP311.5

文献标志码: A

文章编号: 1674-2605(2025)01-0005-06

DOI: 10.3969/j.issn.1674-2605.2025.01.005

开放获取

Water Pump Anomaly Detection Method Based on Feature Distribution Alignment and Multi-sensor Fusion

LIN Zhaohui¹ LIAO Yixiao² ZHAO Zhiming¹ WAN Zhiyong² ZHOU Songbin²

(1.Shimen Waterworks, Guangzhou Water Supply Co., Ltd., Guangzhou 510000, China

2.Institute of Intelligent Manufacturing, Guangdong Academy of Sciences, Guangzhou 510070, China)

Abstract: The water pump is an important pressurization device in the water supply system. Conducting anomaly detection and timely detection of operational abnormalities is of great significance for ensuring water supply safety. A water pump anomaly detection method based on feature distribution alignment and multi-sensor fusion is proposed to address the problems of difficulty in obtaining anomaly samples and low detection accuracy in existing artificial intelligence methods for water pump anomaly detection. This method takes the logarithmic Mel spectrum of multi-channel sensing signals as input; Using convolutional autoencoder networks to fuse multi-sensor information; Train a convolutional autoencoder network with the goal of minimizing signal reconstruction loss and feature distribution loss; Calculate the anomaly score of samples based on reconstruction loss to achieve anomaly detection of water pumps. The experimental results show that this method effectively improves the performance of water pump anomaly detection.

Keywords: anomaly detection of water pump; data reconstruction; feature distribution alignment; multi-sensor fusion

0 引言

水泵是供水系统的核心设备, 其非计划停机不仅会给水厂造成经济损失, 还会影响人们的日常生活和工业生产秩序^[1-2]。通过对水泵进行异常检测和故障诊断, 可降低因设备故障而引发的非计划停机频率, 提高水厂的运维效率, 保障供水安全^[3]。

近年来, 随着人工智能技术的发展, 学者们针对水泵的异常检测和故障诊断开展了大量的研究工作^[4]。文献^[5]通过稀疏自编码网络重构输入信号, 利用重构信号与原始信号的残差对水泵进行异常检测, 可有效识别水泵的早期异常。文献^[6]利用振动信号频谱和振动时域特征来检测水泵轴承的异常, 实现水泵的实时

健康监测。文献[7]提出基于自校准注意力机制和分布边缘预测策略的故障诊断方法，在小样本场景下实现了水泵的故障诊断。文献[8]针对小样本下的水泵故障诊断，提出基于深度迁移卷积神经网络和支持向量机的故障诊断方法，可有效识别水泵的异常状态。文献[9]提出基于小波相关分析和卷积神经网络的离心泵故障诊断方法，提高了离心泵故障的识别精度。文献[10]提出改进的辛几何重构数据增强方法，用于水泵的故障诊断，提高了故障诊断模型的精度和收敛速度。上述方法大多需要异常样本参与模型训练，但在实际工业场景中，难以获得大量的异常样本，无法满足模型高精度训练的需求。另外，基于信号重构的异常检测方法虽然不需要异常样本参与模型训练，但未充分考虑隐藏层特征分布与多传感器融合，限制了异常检测的性能。

针对以上问题，本文提出一种基于特征分布对齐与多传感器融合的水泵异常检测方法（以下简称水泵异常检测方法）。该方法以多通道传感信号的对数梅尔谱为输入；利用卷积自编码网络来融合多传感器信息；以最小化信号重构损失和特征分布损失为目标，训练卷积自编码网络；基于重构损失计算样本的异常分数，实现水泵的异常检测。

1 水泵异常检测方法

水泵异常检测方法以多通道传感信号的对数梅尔谱为异常检测模型的输入，并针对模型的输入特征，构建卷积自编码网络来融合多传感器信息，原理如图1所示。

异常检测模型的输入为多通道的二维数据，编码器网络由多个二维卷积层和二维池化层堆叠而成，解码器网络由多个二维反卷积层和二维反池化层堆叠而成。利用最小化输入多通道传感信号的对数梅尔谱 x 和卷积自编码网络输出的重构对数梅尔谱 x' 之间的重构损失 L_{RC} ，以及编码器网络倒数第2层输出特征 z_{n-1} 及其重构特征 z'_{n-1} 之间的特征分布损失 L_{MMD} ，训练异常检测模型。

由于异常样本不参与模型训练，训练后的异常检

测模型仅对正常样本具有良好的重构能力，无法有效重构异常样本与正常样本之间的差异特征。因此，在异常检测模型训练完成后，通过输入 x 和重构 x' 之间的均方误差来实现水泵的异常检测。

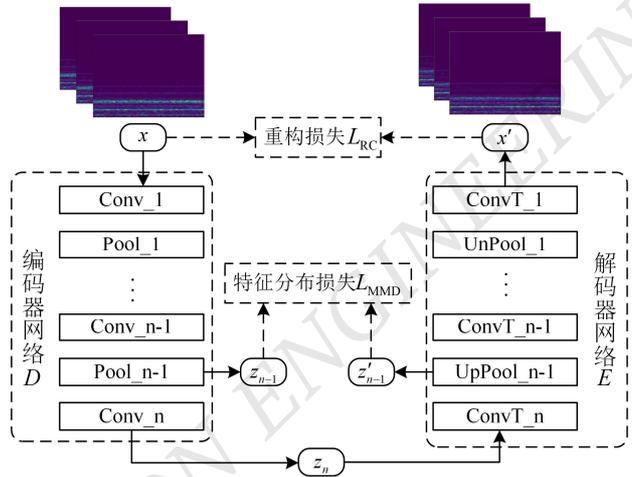


图1 水泵异常检测方法原理图

1.1 对数梅尔谱

对数梅尔谱是基于短时傅里叶变换的一种时频谱图^[11]，在保留信号时域信息的同时，可获取信号的频率分布。在水泵的异常检测中，其常被用作模型的输入特征。对数梅尔谱的计算过程如图2所示。

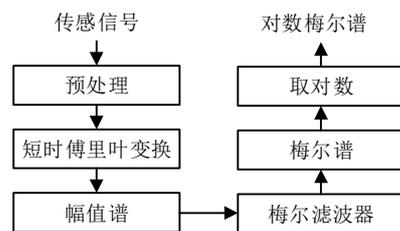


图2 对数梅尔谱的计算过程

首先，根据实际需求，对传感信号进行去直流、预加重、加窗等预处理，并对预处理后的传感信号进行短时傅里叶变换，以获取幅值谱；

然后，利用梅尔滤波器对幅值谱进行梅尔滤波，以获取梅尔谱；

最后，对梅尔谱的幅值取对数，得到对数梅尔谱。

1.2 卷积自编码网络

卷积自编码网络^[12]由编码器网络 D 和解码器网络 E 组成（见图1）。其中，编码器网络 D 以多通道

传感信号的对数梅尔谱 x 为输入，对输入进行特征提取和降维；解码器网络 E 以编码器网络 D 提取的特征 $z_n = D(x)$ 为输入，对输入特征进行恢复和升维，最终输出多通道传感信号的重构对数梅尔谱 $x' = E(z_n) = E(D(x))$ 。

异常检测模型通过最小化重构损失 L_{RC} 来训练卷积自编码网络，使 x' 能够最大程度地还原 x 。本文采用均方误差作为重构损失 L_{RC} ，其计算公式为

$$L_{RC} = \frac{1}{N} \sum_{i=1}^N \|x_i - x'_i\|_2^2 \quad (1)$$

式中： N 为训练样本数量， x_i 和 x'_i 分别为第 i 个样本及其重构。

1.3 特征分布对齐

为了使异常检测模型更好地重构正常样本，本文除重构损失 L_{RC} 之外，还增加了对编码器网络倒数第 2 层输出特征 z_{n-1} 的约束。通过最小化特征分布损失 L_{MMD} ，对齐 z_{n-1} 与其重构特征 z'_{n-1} 之间的特征分布。采用最大均值差异（maximum mean discrepancy, MMD）^[13-14] 作为特征分布损失 L_{MMD} ，其计算公式为

$$L_{MMD} = \sup_{f \in \mathcal{F}} \left(\mathbb{E}_{z_{n-1} \sim P} [f(z_{n-1})] - \mathbb{E}_{z'_{n-1} \sim Q} [f(z'_{n-1})] \right) \quad (2)$$

式中： P 和 Q 分别为 z_{n-1} 和 z'_{n-1} 的概率分布； f 为函数空间 \mathcal{F} 中的一个函数， f 将概率分布 P 和 Q 映射到两者均值差异最大的特征空间中。

从公式(2)可以看出， L_{MMD} 计算的是 2 个概率分布 P 和 Q 在两者均值差异最大的特征空间中的均值差异，即 MMD。

1.4 水泵异常检测方法流程

水泵异常检测方法通过同时最小化重构损失 L_{RC} 和特征分布损失 L_{MMD} 来训练异常检测模型，总体损失函数为

$$L = L_{RC} + \lambda L_{MMD} \quad (3)$$

式中： λ 为权衡系数。

水泵异常检测方法主要包括模型训练和模型测试，流程如图 3 所示。

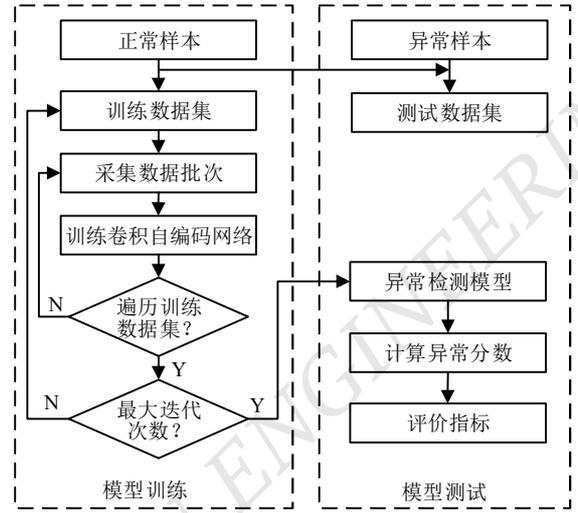


图 3 水泵异常检测方法流程图

1.4.1 模型训练流程

- 1) 对正常和异常样本进行分割和去直流预处理，获取对数梅尔谱样本，并对该样本进行标注；
- 2) 将部分正常样本划分为训练数据集，用于异常检测模型的训练，剩余的正常样本与异常样本组成测试数据集，用于异常检测模型的测试；
- 3) 从训练数据集中选取一批次的训练数据，通过前向传播计算该批次数据的损失值，并利用梯度下降法反向传播损失值训练卷积自编码网络；
- 4) 重复步骤 2)，以遍历整个训练数据集；
- 5) 重复步骤 2)、3)，直至迭代次数达到设定值，获得最终的异常检测模型。

1.4.2 模型测试流程

- 1) 将测试数据集输入卷积自编码网络，计算各样本的异常分数 S_a ：

$$S_a(x) = \|x - x'\|_2^2 \quad (4)$$

- 2) 根据样本标签计算异常检测模型的评价指标，以评估模型性能。

2 实验设置

本实验采用公开数据集 MIMI^[15] 中的水泵数据

集对本文方法进行验证。该数据集包含 4 台水泵（编号分别为 00、02、04、06）运行状态的声音数据。每个声音数据样本包含 8 通道声音信号，采样频率为 16 kHz，样本长度为 10 s。水泵数据集的具体描述如表 1 所示。

表 1 水泵数据集的具体描述 单位：个

水泵编号	正常样本	异常样本
00	1 006	143
02	1 005	111
04	702	100
06	1 036	102

2.1 数据集划分与预处理

将水泵数据集划分为训练集和测试集，划分详情如表 2 所示。

表 2 数据集划分详情 单位：个

水泵编号	训练集		测试集
	正常样本	正常样本	异常样本
00	906	100	143
02	905	100	111
04	602	100	100
06	936	100	102

实验中，将水泵的原始声音信号转换为对数梅尔谱，设置傅里叶变换的窗口长度为 2 048 个样本点，滑动窗口的步长为 1 580 个样本点，梅尔特征维度为 128。对对数梅尔谱进行逐通道的最大-最小归一化，得到异常检测模型的输入维度为 $8 \times 128 \times 100$ 。

2.2 网络结构参数

本文采用的卷积自编码网络结构参数如表 3 所示。

表 3 卷积自编码网络结构参数

编码器网络	解码器网络
Conv2D(8, 64, (9, 9))	ConvT2D(32, 64, (3, 3))
MaxPool(2, 2)	MaxUnPool(2, 2)
Conv2D(64, 64, (5, 5))	ConvT2D(64, 64, (5, 4))
MaxPool(2, 2)	MaxUnPool(2, 2)
Conv2D(64, 64, (5, 4))	ConvT2D(64, 64, (5, 5))
MaxPool(2, 2)	MaxUnPool(2, 2)
Conv2D(64, 32, (3, 3))	ConvT2D(64, 8, (9, 9))

除最后一个二维反卷积层 ConvT2D(64, 8, (9, 9)) 采用线性激活函数外，其他卷积层均采用 Leaky_ReLU 激活函数。

2.3 评价指标

本实验采用 ROC 曲线下与坐标轴围成的面积（area under curve, AUC）、精确度（precision）、召回率（recall）、F1 分数（F1-score）作为异常检测模型的性能评价指标。

AUC 的计算公式为

$$S_{AUC} = \frac{1}{N_+ N_-} \sum_{i=1}^{N_+} \sum_{j=1}^{N_-} H(S_a(x_i^+) - S_a(x_j^-)) \quad (5)$$

式中： $H(\cdot)$ 为阈值函数，输入大于 0 时函数值为 1，否则为 0； $S_a(x_j^-)$ 和 $S_a(x_i^+)$ 分别为正常样本 x_j^- 和异常样本 x_i^+ 的异常分数， N_- 和 N_+ 分别为正常、异常样本的数量。

本实验采用 DCASE2020 比赛提出的阈值确定方法对训练样本的异常分数进行分布统计，取最佳分布的 90% 分位点对应的数值作为异常阈值。基于此阈值判定样本为正常或异常，并结合样本实际标签计算精确度、召回率、F1 分数等评价指标。

$$P_{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$R_{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{P_{Precision} \cdot R_{Recall}}{P_{Precision} + R_{Recall}} \quad (8)$$

式中： TP 为异常样本被正确判定为异常的数量， FP 为正常样本被误判为异常样本的数量， FN 为异常样本被误判为正常样本的数量。

3 实验结果分析

为验证本文方法的有效性，与文献[16]方法(DAE)进行对比实验，结果如表 4~7 所示。

表 4 对比实验结果（水泵编号 00）

水泵编号	评价指标	DAE	本文方法
00	AUC	0.810 7±0.066 4	0.857 8±0.030 4

续表

水泵编号	评价指标	DAE	本文方法
00	precision	0.845 8±0.060 6	0.857 6±0.034 9
	recall	0.695 8±0.076 8	0.767 8±0.044 2
	F1-score	0.761 3±0.058 4	0.809 4±0.030 5

表 5 对比实验结果 (水泵编号 02)

水泵编号	评价指标	DAE	本文方法
02	AUC	0.771 6±0.043 6	0.800 0±0.013 9
	precision	0.807 8±0.040 7	0.829 6±0.018 9
	recall	0.588 3±0.049 9	0.641 4±0.023 2
	F1-score	0.678 6±0.027 9	0.723 0±0.011 9

表 6 对比实验结果 (水泵编号 04)

水泵编号	评价指标	DAE	本文方法
04	AUC	0.932 7±0.014 3	0.981 2±0.010 4
	precision	0.922 7±0.042 3	0.930 2±0.033 4
	recall	0.790 0±0.053 3	0.916 0±0.063 3
	F1-score	0.849 3±0.027 4	0.922 5±0.031 9

表 7 对比实验结果 (水泵编号 06)

水泵编号	评价指标	DAE	本文方法
06	AUC	0.966 5±0.009 1	0.974 6±0.008 2
	precision	0.963 2±0.011 4	0.960 2±0.014 2
	recall	0.815 7±0.079 1	0.890 2±0.051 6
	F1-score	0.881 3±0.048 0	0.923 0±0.026 7

由表 4~7 可以看出,本文方法的 AUC、precision、recall 和 F1-score 比 DAE 均有明显提升,说明基于特征分布对齐与多传感器融合,能有效提高水泵异常检测的性能。

4 结论

本文针对水泵异常检测存在的异常样本获取困难、检测精度低等问题,提出了一种基于特征分布对齐与多传感器融合的水泵异常检测方法。经实验验证,本文方法通过特征分布对齐与多传感器融合,有效提高了水泵异常检测的性能。该方法的主要贡献如下:

1) 异常检测模型在训练过程中,仅需要正常样本参与训练,无需异常样本参与训练,降低了训练数

据获取难度,有利于模型在实际工业现场的推广应用;

2) 融合多传感器信息,全面获取水泵的健康状态信号,提高了异常检测模型的检测能力;

3) 结合隐藏层特征分布对齐策略,使卷积自编码网络能更好地重构正常样本,增大正常、异常样本的判定边界。未来可以结合迁移学习开展水泵异常检测研究,利用现有的水泵数据和异常检测模型,在仅采集少量或不采集水泵正常样本的情况下,构建目标水泵的异常检测模型,以提高异常检测精度。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

参考文献

- [1] 顾祖坤.基于无线传感网络的水泵振动状态监测系统设计与[D].江苏:江苏大学,2018.
- [2] 刘煜辉,高艳强.多传感器融合的工业生产线故障预警技术[J].自动化与信息工程,2024,45(2):41-44.
- [3] 郑日晖,岑健,陈志豪,等.基于 EMD 样本熵与改进 DS 证据理论的故障诊断方法[J].自动化与信息工程,2020,41(2):19-26.
- [4] 范国栋,李博涵.基于机器学习的工业机械设备故障预测方法[J].自动化与信息工程,2023,44(4):13-18;50.
- [5] LIANG X, DUAN F, BENNETT I, et al. A sparse autoencoder-based unsupervised scheme for pump fault detection and isolation[J]. Applied Sciences, 2020,10(19):6789.
- [6] 舒相挺,杨璋,徐逸哲,等.某循环海水泵振动故障诊断与趋势分析[J].振动.测试与诊断,2022,42(4):791-796;831.
- [7] ZOU F, SANG S, JIANG M, et al. A few-shot sample augmentation algorithm based on SCAM and DEPS for pump fault diagnosis[J]. ISA Transactions, 2023,142:445-453.
- [8] 吴佳,李明宸,唐文妍.基于 DTCNN-SVM 的工业循环水系统供水泵故障诊断[J].振动与冲击,2023,42(13):226-234.
- [9] ULLAH N, AHMAD Z, SIDDIQUE M F, et al. An intelligent framework for fault diagnosis of centrifugal pump leveraging wavelet coherence analysis and deep learning[J]. Sensors, 2023,23(21):8850.
- [10] LIU S, YIN J, HAO M, et al. Fault diagnosis study of hydraulic pump based on improved symplectic geometry reconstruction data enhancement method[J]. Advanced Engineering Informatics, 2024,61:102459.

(下转第 46 页)