

基于 VMD 和 CBAM-DRSN 的扬声器异常声分类方法研究

李振业¹ 周静雷¹ 路 昌¹
LI Zhenye ZHOU Jinglei LU Chang

摘 要

为解决扬声器异常声非平稳信号特征提取困难及分类识别问题,提出一种基于变分模态分解(variational mode decomposition, VMD)和空间-通道融合注意力机制深度残差收缩网络(convolutional block attention module - deep residual shrinkage networks, CBAM-DRSN)扬声器异常声分类方法。首先,采集扬声器异常声响应信号进行 VMD 分解获取其频域特征;其次,利用深度残差收缩网络提取深层次特征;然后采用通道-空间融合注意力机制对多尺度特征进行自适应加权并进一步提取深层抽象特征;最后,使用 softmax 分类器完成扬声器异常声分类任务。试验结果表明,所提方法具有更优的分类准确性和更稳定的抗噪性能,其平均分类准确率为 99.23%,在不同噪声等级环境下的平均准确率为 94.65%。

关键词

扬声器异常声;故障诊断;深度残差收缩网络;注意力机制

doi: 10.3969/j.issn.1672-9528.2025.01.027

0 引言

随着对更高音频质量日益增长的认知和需求,消费者在智能设备上获得最佳音频体验的期待也在提升。其中扬声器在智能设备中的重要性不言而喻,在扬声器的异常声检测方面,其故障诊断及分类的准确性、时效性,对于扬声器生产过程而言,具有极大影响。工业上传统异常声检测由听音员进行主观听音测试,难以保证扬声器异常声分类的准确性及精度问题,因此对于提升扬声器异常声检测质量、效率及加快工业自动化进程,具有重要的现实意义。

扬声器异常声检测时得到的声音响应信号是非线性多分量信号,其信号组成表现出频段相互重叠和密集紧凑的复杂特征。Huang 等人^[1]提出经验模式分解(empirical mode decomposition, EMD),基于含噪声信号的时频域局部特征,通过迭代筛分的方式分解模态分量。但 EMD 将信号分量重构过程中分解的本征模态函数失真且出现信号模态混叠现象。Wu 等人^[2]为解决信号模态混叠提出一种集合经验模式分解(ensemble empirical mode decomposition, EEMD),通过在原始信号中加入一系列的高斯白噪声改善模式混叠现象,但无法从低信噪比的多分量信号中准确提取分量。Dragomiretskiy 等人^[3]创新型的提出多分量信号处理方法——变分模态分解(variational mode decomposition, VMD),VMD 通过求解变分约束优化问题,实现既定数目分量的提取,在抗模态混叠方面具有一定优势。

针对故障分类辨识问题,Wang 等人^[4]提出一种基于 SVM 的风电机组齿轮箱多故障诊断方法,通过映射方法对特征重构后作为 SVM 输入来提升不同类型故障特征的可辨识度。Pang 等人^[5]基于决策树算法,通过对故障类型与特征变量关联关系的量化来提高故障诊断的准确率。深度学习算法相比传统机器学习有更好的特征提取及非线性表达能力,有效避免对专业知识和人工特征提取的依赖。如卷积神经网络(convolutional neural network, CNN)^[6]、门控循环单元网络(gated recurrent uni, GRU)^[7]等。Jia 等人^[8]提出一种基于 CNN 的齿轮箱故障诊断方法,在故障诊断中表现出较好的特征提取能力。Xu 等人^[9]使用内在特征提取和卷积块注意力模块(convolutional block attention module, CBAM)增强的 InceptionNet 进行故障诊断,获得较高的准确率。

针对模型抗噪性方面,在故障诊断领域,深度学习在实际工业环境中的自适应能力值得关注。Zhang 等人^[10]提出一种具有抗噪能力的卷积神经网络模型进行故障诊断,证明其具有较强的域适应能力,但模型性能仍需进一步挖掘。

针对以上问题,本文提出一种基于 VMD 和 CBAM-DRSN 的扬声器异常声分类方法。首先,采集多种状态的扬声器声响应信号并进行 VMD 分解获取特征信息;其次,通过深度残差收缩网络提取深层次特征;然后,构造空间-通道融合注意力机制对提取到的特征信息进行自适应加权,减小信息冗余;最后,对特征信息进行整合与分类,完成扬声器异常声分类任务。试验结果表明,本研究方法与其他研究方法通过试验数据对比分析,有效提升扬声器异常声分类准确性,证实该方法具有更优的抗噪性与泛化能力。

1. 西安工程大学电子信息学院 陕西西安 710600

[基金项目] 陕西省科技厅重点研发计划项目(2024GX-YBXM-231)

1 基本原理简介

1.1 VMD 算法原理

VMD 是一种完全非递归多分量自适应信号分解方法。该方法以 Wiener 滤波为基础，在变分模型框架内对输入信号搜索最优解，能够自适应地更新中心频率、频带宽度与相应的子信号，从而实现固定个数模态分量的分解，能够有效解决 EMD 中存在的模态混叠现象。VMD 首先将异常声信号的频谱分割为多个区间，然后根据边界频段特征构造 Wiener 滤波器组，最后利用 Wiener 滤波器实现声响应信号的分解。基于 VMD 的异常声信号分解流程如图 1 所示。

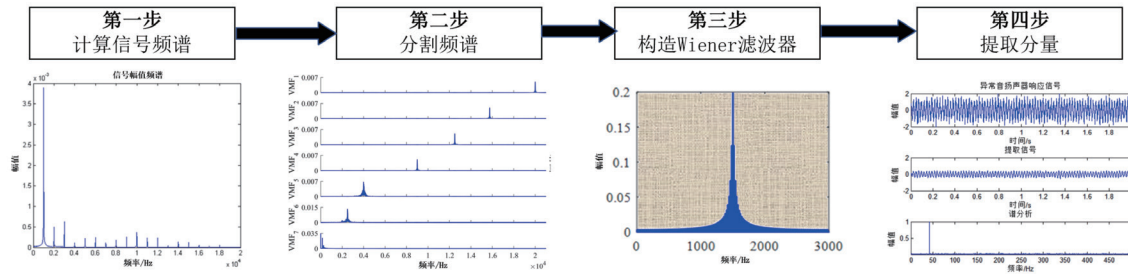


图 1 扬声器异常声信号分解流程

该算法具体构造步骤如下：

(1) 初始化参数 $\{\hat{u}_k\}$ 、 $\{\hat{\omega}_k\}$ 、 $\hat{\lambda}^1$ 与 $n=0$ ；其中 $\{\hat{u}_k\}$ 为模态分解后的分量， $\{\hat{\omega}_k\}$ 是模态分解后的中心频率， $\hat{\lambda}^1$ 为拉格朗日乘子， n 表示循环次数；

(2) 令 $n=n+1$ ；

(3) 遍历 $k=1 \rightarrow K$ ，分别以如下公式更新 $\hat{u}_k^{n+1}$ 和 ω_k^{n+1} ：

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (1)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (2)$$

式中： α 表示二次惩罚因子。

(4) 通过式 (3) 更新 $\hat{\lambda}^{n+1}(\omega)$ ：

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega) \right] \quad (3)$$

(5) 循环第 2~4 步，直至满足公式 (4) 中所示的收敛条件。

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{(n+1)} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon \quad (4)$$

1.2 DRSN 分类网络

深度残差收缩网络 (deep residual shrinkage networks, DRSN)，主要用来在特征学习的过程中，剔除冗余信息。典型的 DRSN 整体结构由输入层、卷积层、一定数量的残差收缩模块 (residual shrinkage building unit, RSBU)、批标准化 (batch normalization, BN)、ReLU 函数、全局均值池化层 (GAP) 和全连接输出层 (FC) 组成的深度神经网络结构。

其中 RSBU 作为 DRSN 的核心部分，其结构主要包含深度残差网络、软阈值函数以及注意力机制。深度残差收缩网络结构如图 2 所示。

1.2.1 深度残差网络

残差单元 (residual building unit, RBU) 通过恒等映射的方式将深度神经网络的难度极大程度地降低，残差收缩模块 (residual shrinkage block unit, RSBU) 是 DRSN 网络的核心，RSBU 模块保留了 RBU 模块中恒等映射的方法。通过引入跨层恒等连接，避免模型出现梯度消失和网络退化现象。由多个残差单元构成了深度残差网络的主体部分，在进行反

向传播的模型训练时，其损失既能通过卷积层等进行逐层反向传播又能够通过恒等映射的方式进行更为便捷的反向传播，这种方式允许神经网络

的某些层跳过与下一层神经元的直接连接，从而降低深层神经网络的训练难度。并且在两个卷积隐层后添加了软阈值和注意力模块，结构如图 2 (a) 所示。

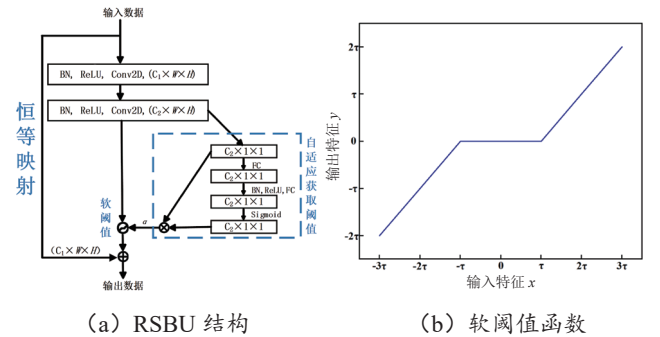


图 2 深度残差收缩网络

1.2.2 软阈值函数

软阈值化是降噪的核心步骤。软阈值函数图像如图 2 (b) 所示。具体方式为：将绝对值大于某阈值的特征趋于零进行收缩，并将绝对值小于某个阈值的特征进行去除。软阈值函数中阈值的选取需要两个基本条件：一是阈值的取值为整数；二是阈值不大于输入信号的最大值，否则输出为零。软阈值化的每个样本应根据自身的噪声含量，具有独立阈值。软阈值函数表达式为：

$$y = \begin{cases} x - \tau & x > \tau \\ 0 & -\tau \leq x \leq \tau \\ x + \tau & x < -\tau \end{cases} \quad (5)$$

式中： x 、 y 分别为输入、输出特征； τ 式阈值。

1.2.3 注意力机制

注意力机制有助于模型聚焦局部的关键信息。在深度

残差收缩网络的子网络中,首先对输入样本的全部特征求取绝对值。通过全局均值池化和平均得到全局平均特征,在另一通道将全局均值池化的特征图输入到全连接网络作为 Sigmoid 函数,实现输出的归一化。后将归一化后的输出值与之前所得特征相乘,得到自适应的阈值。自适应阈值能够根据输入样本的特征进行动态调整。深度残差收缩网络在通过子网络学习一组阈值后,对各个特征通道进行软阈值化,注意并保留与当前分类任务有关的特征,抑制冗余特征。

1.3 CBAM 空间 - 通道融合注意力机制

注意力机制本质上可以看作输入信号动态调整权重的过程,通过计算每个特征信息的重要程度,再重新分配权重让模型更关注与目标重要特征相关的信息,并忽视其他无关信息,以提高模型效率。CBAM 空间 - 通道融合注意力机制结构如图 3 所示。

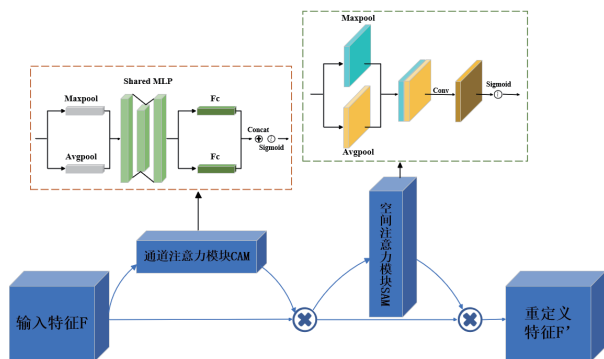


图 3 空间 - 通道融合注意力机制

CBAM 通过对卷积块的通道和空间信息进行注意力加权来提高网络的表征能力及泛化能力,通道注意力用于捕捉不同通道之间的互相依赖关系,它通过计算输入特征获得每个通道的权重,然后对每个通道进行加权,以减少通道之间的不相关信息,从而提高特征提取的质量。空间注意力用于捕捉特征图中不同位置之间的依存关系。在通道注意力的基础上,空间注意力计算每个位置的重要性权重,然后对特征图进行加权,以减少不相关位置的信息。CBAM 通过与 DRSN 网络结构结合,能够更高效的实现特征表达能力,以提高模型的性能。

2 基于 VMD 和 CBAM-DRSN 的扬声器异常声分类

为了提取扬声器异常声信号的深层次特征以提高异常声分类准确性及效率,本文提出了一种基于 CBAM-DRSN 的扬

声器异常声分类方法。模型架构如图 4 所示。该模型的输入数据为经 VMD 分解所得多分量信号样本。除去每个 RSBU 模块中的两个卷积层,在模型中交替设置 7×7 、 5×5 、 3×3 三个不同大小的卷积核来提取多维特征,并分别在 5×5 和 3×3 卷积层后进行最大池化。在第三个 RSBU 模块后,通过 CBAM 注意力模块提取重要特征,降低冗余特征重要性,然后添加卷积层和池化层,最后通过 Flatten 层、Dropout 层,全连接层使用 Softmax 激活函数进行故障分类,得到最后的预测结果。

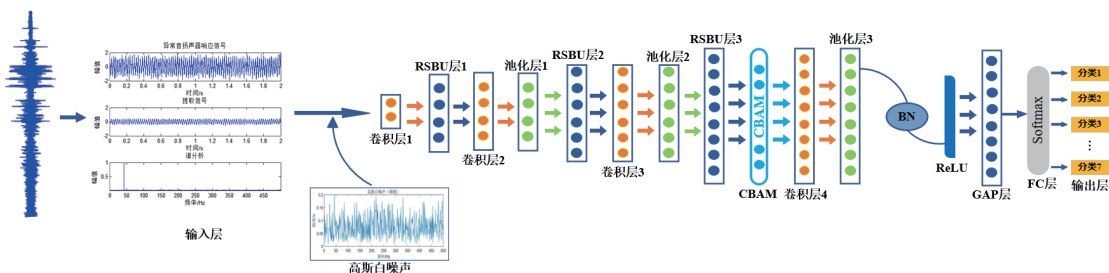


图 4 CBAM-DRSN 模型架构图

基于 VMD 和 CBAM-DRSN 的扬声器异常声分类具体流程如图 5 所示。

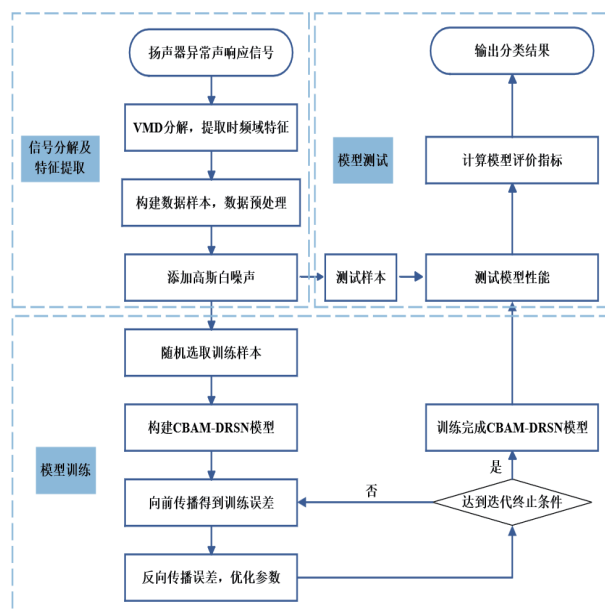


图 5 扬声器异常声分类流程

依据扬声器异常声信号在全频段内难以准确提取到对健康状态或故障敏感的特征信息,首先采用 VMD 方法筛选已采集音频数据,对其频谱进行分割,其次依据频带特征重构分量,完成信号特征提取。然后对数据样本添加不同分量的高斯白噪声,通过将输入样本按照一定比例划分训练集和测试集,训练集输入 CBAM-DRSN 网络进行训练得到分类诊断模型。最终将测试集输入模型得到健康扬声器与不同故障类型扬声器的分类结果,根据分类准确性、迭代收敛速度及抗噪声干扰能力来衡量模型的有效性。

3 试验分析

3.1 数据采集平台

本次试验采用动圈式扬声器，其型号为 3580-3 型，谐振频率为 180 Hz。扬声器激励信号为 1800-20Hz 的连续对数扫频信号，时长 1 s。本次试验共测用健康、定脱、异物、帽子失风、碰圈、脱盆架、纸盆声 7 类，每种类型的扬声器单元各 24 个。扬声器异常声采集平台如图 6 所示。

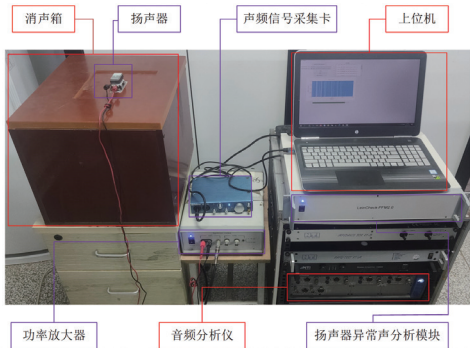


图 6 扬声器异常声采集平台

3.2 分类网络模型参数

本文采用 Google 公司研发的 tensorflow2.0 框架。试验所用硬件配置为 AMD Ryzen7 3700X 8-Core Processor CPU@3.59 GHz，GPU 为 NVIDIA GeForce GTX 1660 SUPER，16 GB 运行内存。

3.3 试验结果及分析

为展示基于 CBAM-DRSN 模型的异常声分类诊断的优越性，各类异常声数据量均衡采集 95 条，共采集 665 组数据，将异常声响应信号进行 VMD 分解后，设置多组试验以验证。选取 70% 为训练样本，为避免试验产生随机性及试验结果误差较大影响，每组试验均进行 10 次，取样本平均测试结果为最终结果。

3.3.1 分类试验

为了验证该分类方法的有效性，进行对比试验。使用支持向量机(SVM)、随机森林(RF)、门控循环神经网络(GRU)、残差网络(ResNet)网络与本文模型进行异常声分类试验。

表 1 展示了选用传统机器学习模型 SVM 和 RF，及深度学习模型 GRU、ResNet、DRSN 和 CBAM-DRSN 得到的不同模型评价结果。传统机器学习诊断准确性依赖于特征提取与选择且提取特征的有效性很大程度上取决于专业知识，从所示结果可知，传统机器学习模型具有较高的准确性，但其损失函数值相对于深度学习模型较高，分别为 0.693 5 和 0.538 9。深度学习模型具有更好的特征提取和非线性表达能力，可以更好的挖掘隐藏在非线性、非稳定数据深处的条件相依关系，从所示结果可知，深度学习模型依然有较高的准确性，DRSN 的故障分类准确率为 98.12%，比 GRU、ResNet 分别提高了 5.53% 和 0.05%，模型误差为 1.47%，比 GRU、

ResNet 分别减少了 3.47% 和 4.22%。且本文提出的 CBAM-DRSN 准确率为 99.03%，模型误差为 0.65%，相比 DRSN 的准确性提升 0.91%，模型误差降低 0.82%。本文模型损失函数值为 0.036 0。

表 1 不同模型评价结果

模型	准确率 (%)	损失函数值	模型误差 (%)
SVM	93.14	0.693 5	6.39
RF	97.42	0.538 9	3.65
GRU	92.59	0.357 7	4.94
ResNet	98.05	0.068 2	5.69
DRSN	98.12	0.074 6	1.47
CBAM-DRSN	99.23	0.036 0	0.65

图 7 表明，通过对比不同深度学习模型的准确率及损失函数曲线，CBAM-DRSN 收敛速度最快且准确性最高，在进行 20 次训练之后该模型的波动最为平稳，相比其他模型表现出更好的预测性能。

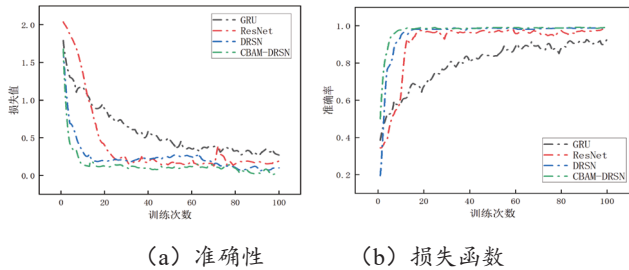


图 7 不同深度学习模型的准确率和损失函数值

3.3.2 抗噪声试验

异常扬声器发出的异常声是指非正常状态下的声音，这种异常声与扬声器的物理状态和运行状态有关。为了验证本文所提方法在不同噪声水平下进行扬声器异常声分类的准确性，通过在输入样本中加入不同功率的高斯噪声以验证该模型在不同信噪比情况下的故障诊断准确性。

图 8 分别给出了信噪比为 -3 dB、-2 dB、-1 dB、0 dB、2 dB、4 dB、6 dB 和 10 dB 八种情况下，与卷积神经网络(CNN)和门控循环神经网络(GRU)及残差网络(ResNet)模型分类的平均准确性对比。根据图 8 所示结果，添加的噪声对各模型的诊断准确性有着明显的影响。当异常声信号的信噪比较大时，除 CNN 之外模型都有很好的分类准确性。对于加入的噪声功率大于原始信号功率时(信噪比低于 0 dB)所有模型的性能都明显下降。ResNet 对比于 CNN、GRU，在噪声功率较小时，模型有高于 91% 的准确率，其主要原因是模型采用跨层恒等映射的方式，在加入高斯噪声的情况下依然有效提取深层网络的抽象特征信息。DRSN 对比于 ResNet，虽然模型随着信噪比的下降，准确性在不断降低，但在信噪比达到 -3 dB 时，DRSN 仍然有着高于 84% 的准确率，其主要原因是软阈值化机制可以有效过滤部分噪声。

对比 DRSN 和 CBAM-DRSN 随着噪声功率增大的准确性变化可以发现, CBAM-DRSN 的平均分类准确率要比 DRSN 高 0.13% ~ 2.18%, 且在 -3 dB 高噪声环境下 CBAM-DRSN 仍有 86.79% 的准确率, 其主要原因是注意力机制具有增强有用信息弱化无用信息的作用, 噪声作为一种特殊的无用信息, 也能够起到一定的过滤作用。

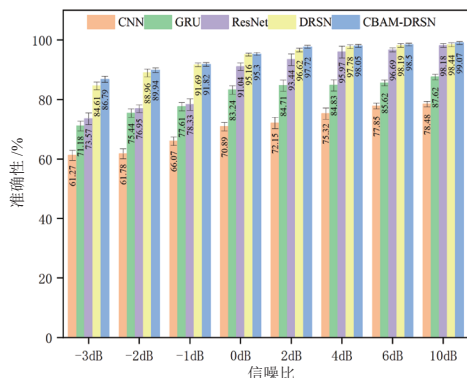


图 8 不同信噪比情况下的诊断准确性对比

表 2 给出了不同模型抗噪试验评价比较, 在 8 种不同噪声等级下, 对比平均准确率 (AVG) 可以看出, DRSN 相比于其他模型具有更稳定的抗噪性能, 其 AVG 达 93.93%, 而改进的 CBAM-DRSN 性能更优, AVG 达 94.65%。对比平均误差率 (MAPE) 可以看出, 改进的 CBAM-DRSN 达 0.6536%, 相比于其他模型具有更准确的分类预测结果。

表 2 抗噪试验评价

指标	CNN	GRU	ResNet	DRSN	CBAM-DRSN
AVG/%	70.48	81.28	88.02	93.93	94.65
MAPE/%	2.138 7	1.800 4	1.629 8	0.862 8	0.653 6

4 结语

针对神经网络在处理声响信号时特征学习能力不足, 本文提出了基于 VMD 和 CBAM-DRSN 的扬声器异常声分类方法, 获取经 VMD 分解所得重构信号构成数据样本集, 输入到 CBAM-DRSN 网络中完成分类识别。通过与传统机器学习以及深度学习模型的试验比较, 同一环境下, CBAM-DRSN 相比于其他模型表现出更优的分类诊断准确性, 准确率达 99.03%。不同噪声等级环境下, CBAM-DRSN 的平均分类准确率为 94.65%, 可以实现强噪声环境下扬声器异常声分类更为高效的诊断。近年来, 模态分解方法及深度学习模型在声频信号处理方面展现出极大优势, 但在故障诊断方面仍存在不足, 主要原因是现有采集数据样本较小, 在接下来的工作中, 会在多种不均衡数据样本条件下, 提高提取特征效率并提高扬声器异常声分类的实时性。

参考文献:

[1] HUANG N E, SHEN Z, LONG S R, et al. The empirical

mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[EB/OL].(1998-03-08) [2024-02-22].<https://royalsocietypublishing.org/doi/10.1098/rspa.1998.0193>.

- [2] WU Z H, HUANG N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in adaptive data analysis, 2009, 1(1):1-41.
- [3] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. IEEE transactions on signal processing, 2014, 62(3): 531-544.
- [4] WANG Z Y, LI G S, YAO L G, et al. Data-driven fault diagnosis for wind turbines using modified multiscale fluctuation dispersion entropy and cosine pairwise-constrained supervised manifold mapping[J]. Knowledge-based systems, 2021, 228(9): 107276.
- [5] PANG J S, CHEN Y M, HE S Z, et al. Classification of friction and wear state of wind turbine gearboxes using decision tree and random forest algorithms[J]. Journal of tribology, 2021, 143(9): 091702.
- [6] CHOUDHARY A, MISHRA R K, FATIMA S, et al. Multi-input CNN based vibro-acoustic fusion for accurate fault diagnosis of induction motor[J]. Engineering applications of artificial intelligence, 2023,120(4): 105872.
- [7] LIU Z W .Bearing fault diagnosis of end-to-end model design based on 1DCNN-GRU network[EB/OL].(2022-06-09)[2024-03-05].<https://onlinelibrary.wiley.com/doi/10.1155/2022/7167821>.
- [8] JIA X J, HAN Y, LI Y J, et al. Condition monitoring and performance forecasting of wind turbines based on denoising autoencoder and novel convolutional neural networks[J]. Energy reports, 2021, 7(11): 6354-6365.
- [9] XU S J, YUAN R, LÜ Y, et al. A novel fault diagnosis approach of rolling bearing using intrinsic feature extraction and CBAM-enhanced InceptionNet[J]. Measurement science and technology, 2023,34:105111.
- [10] ZHANG W, LI C H, PENG G L, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load[J]. Mechanical systems and signal processing, 2018, 100(2): 439-453.

【作者简介】

李振业 (1998—), 男, 山东东营人, 硕士研究生, 研究方向: 扬声器的异常声检测、声频信号处理。

(收稿日期: 2024-09-29)