



DOI:10.12404/j.issn.1671-1815.2402367

引用格式:万志国,王治国,赵伟,等.基于CBAM-STCN的齿轮箱故障智能诊断方法[J].科学技术与工程,2025,25(9):3760-3768.

Wan Zhiguo, Wang Zhiguo, Zhao Wei, et al. Intelligent fault diagnosis method for gearboxes based on CBAM-STCN[J]. Science Technology and Engineering, 2025, 25(9): 3760-3768.

基于CBAM-STCN的齿轮箱故障智能诊断方法

万志国,王治国,赵伟,窦益华

(西安石油大学机械工程学院,西安710000)

摘要 针对齿轮箱在多种工况下故障特征存在差异,故障诊断易受噪声干扰,导致故障诊断模型泛化性差和识别准确率低的问题,提出一种端到端的具有混合注意力机制和软阈值化特点的时间卷积神经网络(convolutional block attention module-sparse temporal convolutional network with soft thresholding, CBAM-STCN)齿轮箱故障诊断模型识别分类方法。首先,利用希尔伯特变换将齿轮故障振动信号转换为包络谱信号;然后,将其输入CBAM-STCN故障诊断模型中;该模型嵌入的混合注意力机制模块(convolutional block attention module, CBAM),能够自适应学习通道和空间注意力的权重,提取与故障特征相关的敏感信息;嵌入的软阈值函数能够最小化模型输出和原输入之间的差异;最后,利用所提出的方法对两种工况、不同类型的齿轮故障进行识别分类。结果表明:CBAM-STCN故障诊断模型对齿轮故障智能诊断的平均准确率为98.95%。该方法对于齿轮箱故障的智能诊断具有一定的参考价值。

关键词 齿轮箱;故障智能诊断;混合注意力机制;软阈值化;时间卷积神经网络

中图分类号 TP272; **文献标志码** A

Intelligent Fault Diagnosis Method for Gearboxes Based on CBAM-STCN

WAN Zhi-guo, WANG Zhi-guo, ZHAO Wei, DOU Yi-hua

(School of Mechanical Engineering, Xi'an Shiyou University, Xi'an 710000, China)

[Abstract] For gearboxes, variations in fault characteristics under different operating conditions, susceptibility to noise interference in fault diagnosis, lead to poor generalization and low recognition accuracy of fault diagnosis models. An end-to-end convolutional block attention module-sparse temporal convolutional network with soft thresholding (CBAM-STCN) was proposed for gearbox fault diagnosis. Firstly, the Hilbert transform was employed to convert the gear fault vibration signal into an envelope spectrum signal. Then, this signal was input into the CBAM-STCN fault diagnosis model. The model integrates a hybrid attention mechanism module, the convolutional block attention module (CBAM), which adaptively learns the weights of channel and spatial attention to extract information sensitive to fault features. The embedded soft thresholding function minimizes the discrepancy between the model's output and the original input. Finally, the proposed method was utilized to identify and classify various types of gear faults under two different conditions. The results indicate that the CBAM-STCN model achieves an average accuracy of 98.95% in intelligent gear fault diagnosis, demonstrating its potential value for gearbox fault diagnosis.

[Keywords] gearbox; intelligent fault diagnosis; convolutional block attention module (CBAM); soft thresholding; temporal convolutional network (TCN)

齿轮箱是现代机械设备的重要组成部分,已广泛应用于各个工业领域^[1]。齿轮箱以其重量轻、传动范围大、效率高等特殊性能在机械传动系统中发挥着重要作用^[2]。齿轮箱包含齿轮、滚动轴承、传动轴和其他部件^[3],在长时间运转时,尤其是在高速重载的条件下,十分容易产生故障^[4]。对于规模较大、自动化程度较高的设备,当齿轮箱发生故障,会造成严重的停工停产。因此,为了有效减少设备

维修成本并防止突发性事故发生,对齿轮箱故障诊断的研究是至关重要的^[5]。

目前,具有自适应学习特性的深度学习神经网络模型已经被广泛应用于齿轮箱故障诊断^[6],而且基于深度学习的网络模型具有较强的直接特征提取能力^[7]。其中基于深度学习的网络模型,如时间卷积网络(temporal convolutional network, TCN)、深度自编码器(denoising auto encoder, DAE)、循环神

收稿日期:2024-04-02 修订日期:2024-12-06

基金项目:陕西省自然科学基金研究计划(2022JQ-412)

第一作者:万志国(1988—),男,汉族,山东泰安人,博士,副教授。研究方向:机械设备状态检测与故障诊断、油气井管柱力学及完整性评价。E-mail:wanzhiguo168@163.com。

神经网络(recurrent neural network, RNN)、深度信念网络(deep belief network, DBN)、卷积神经网络(convolutional neural network, CNN)等已经成功应用到旋转机械的故障诊断^[8]。

Li 等^[9]提出了一种将时间卷积神经网络与软阈值算法(spatial attention module-temporal convolutional network with soft thresholding, SAM-TCNST)相结合的齿轮箱故障智能识别方法,识别准确率较高,但是需要对原始振动信号进行处理,不能实现端到端的齿轮箱故障诊断。李莎等^[10]提出了一种软阈值化的时间卷积神经网络(soft thresholding-temporal convolutional network, ST-TCN),网络结构简单,收敛速度快,但是忽略了注意力机制模块,不能处理不同尺度的信息。吕卫民等^[11]提出了一种基于时间卷积神经网络(TCN)和轻量级梯度提升机(light gradient boosting machine, LGBM)的故障诊断网络模型,在故障数据训练时序逻辑的基础上结合 LGBM 模型对特征进行快速分类,但是在注意力机制中只考虑了通道注意力机制,而没有考虑空间注意力机制。张璐莹等^[12]提出了利用基于注意力时间卷积网络和双向门控循环单元对轴承进行故障诊断,能够有效识别轴承故障类型,但是没有利用软阈值化去除所提取特征中的噪声。Zhang 等^[13]提出了一种注意机制增强时间卷积网络(attention mechanism enhanced-temporal convolutional network, AME-TCN)的故障诊断模型,利用注意力机制区分不同监测变量的重要性,提高了 TCN 对故障诊断的性能,但是同样没有利用软阈值化消除噪声的影响。Zhu 等^[14]提出了一种基于加权对抗网络改进的卷积块注意模块(CBAM),用于部分域自适应齿轮箱故障诊断,诊断准确率较高,但是没有考虑软阈值化。Wang 等^[15]提出了一种集成卷积块注意模块(CBAM)的齿轮箱故障诊断方法。将 CBAM 嵌入到深度残差网络中,增强了图像的特征提取,具有较高的准确率,但是同样忽略了软阈值化。Zhan 等^[16]提出了一种基于注意机制和大规模卷积的神经网络模型。将预处理后的图像输入网络中进行特征提取,分类准确率较高,但是也没有考虑软阈值化。

基于此,由于 TCN 在不同数据集上的表现比传统神经网络更优秀,而且内存利用率高,可以同时执行卷积,拥有可调整的感受野以及梯度平稳性。因此,现基于 TCN 模型,在网络结构中引入 Dropout 层,减少卷积神经网络神经元数量,提升模型的鲁棒性。对 TCN 的残差结构进行优化,选用 Activation 和 LeakyReLU 作为激活函数,引入正则化损失函

数,防止过拟合,增强模型的泛化能力。其次,在 TCN 中嵌入混合注意力模块(CBAM),包括通道注意力模块和空间注意力模块。混合注意力模块不仅结构简单,而且能通过动态调整每个通道内和空间位置上的加权系数,突出关注目标的细节信息,减少特征提取中的冗余信息,提高模型的拟合精度和自适应性。同时,利用软阈值化消除所提取特征相关的噪声信息,提高模型故障诊断的准确率。

1 基本原理

1.1 时间卷积神经网络(TCN)

时间卷积神经网络(TCN)是一种可以对时间序列数据进行处理神经网络架构,与传统的卷积神经网络相比,它可以更有效地提取时序数据的特征^[17]。在 TCN 的网络结构中,采用因果卷积提取时间序列数据中的特征,能够使网络层间具有因果关系,实现时序建模。通过扩张因果卷积扩大感受野,一个卷积能够学习到更多的特征,可以适应不同尺度的时间依赖关系。引入残差连接能够增加 TCN 层数,有效解决梯度消失和爆炸问题,保持稳定的网络性能^[18]。

1.1.1 因果卷积

与传统卷积神经网络相比,因果卷积是单向结构,必须利用下一层 t 时刻之前的数据得到上一层 t 时刻的数据,这样不仅能保证 TCN 因果性,而且也能保证 TCN 具有严格的时间约束^[19]。在输入网络中的序列 $\mathbf{X} = \{x_0, x_1, \dots, x_{t-1}, x_t, \dots\}$ 中, t 时刻的输出 $y_t (i=0, 1, \dots, t)$ 只能利用当前时刻的 x_t 和之前时刻输入 x_0, x_1, \dots, x_{t-1} 计算得到。此外,在输入序列左侧进行零填充可以保证网络结构中的输入张量与输出张量具有相同的长度^[20]。因果卷积原理如图 1 所示。

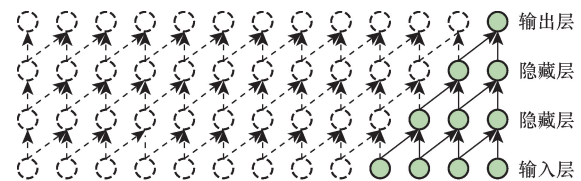


图 1 因果卷积原理

Fig. 1 Causal convolution principle

1.1.2 扩张因果卷积

传统卷积神经网络对于长时序数据的特征信息提取能力不强,必须要增加卷积层数。为了有效地解决依赖关系问题,使得 TCN 网络具有更大的感受野,扩张因果卷积可以在不增加参数和模型复杂度的前提下,利用指数增长的间隔采样来增大感受野^[21]。扩张因果卷积原理如图 2 所示。

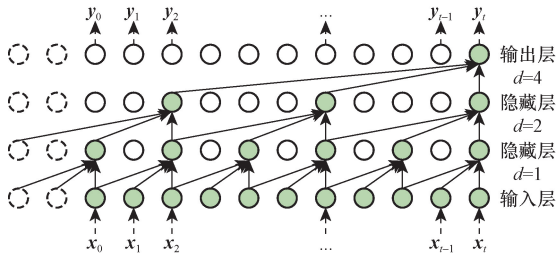


图2 扩张因果卷积原理

Fig. 2 Extended causal convolution principle

其中,扩张因子 d 是采样率的影响因素。最底层 $d=1$ 表示每个时间点进行采样作为输入,隐藏层 $d=2$ 表示每隔一个时间点进行采样作为输入,最顶层 $d=4$ 表示每隔 3 个时间点进行采样作为输入。对于输入的齿轮箱数据集序列 $\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t, \dots\}$, 其中 \mathbf{x}_i 为列向量, $i \in [1, n]$; 过滤器 $F = (f_1, f_2, \dots, f_k)$, 在 t 时刻的卷积运算 F 为

$$F(\mathbf{x}_t) = (\mathbf{X}dF)(\mathbf{x}_t) = \sum_{k=1}^K f_k \mathbf{x}_{t-dk} - d(K-k) \quad (1)$$

式(1)中: d 为扩张因子; k 为卷积核大小, $k = 1, 2, \dots, K$; 上一层第 $t - dk$ 个元素可用 $t - dk$ 表示,感受野 N 为

$$N = 1 + (k-1) \frac{b^n - 1}{b - 1} \quad (2)$$

式(2)中: b 为扩张基数;扩张因子 $d = b^i$, $i = 1, 2, \dots, n$, n 为层数。

根据图 2,当卷积核大小为 3 时,扩张因子为 $[1, 2, 4]$ 时 t 时刻的输出 y_t 依赖输入 $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t\}$, 此时感受野能够完全涵盖输入序列中的所有值。

1.1.3 TCN 残差结构优化

TCN 接受域受扩张因子、滤波器大小和网络深度的影响,而且在网络层数增多的过程中,会出现梯度消失或梯度爆炸的情况。所以必须要加入残差模块,简化深层网络的训练,使深层网络运行稳定,保持较好的性能^[22]。残差模块如图 3 所示。

通过残差连接,将输入数据 x 与模型输出 $F(x)$ 进行加权融合,生成 TCN 输出 o 。其计算公式如式(3)所示。

$$o = \text{Activation}[\mathbf{x} + F(\mathbf{x})] \quad (3)$$

式(3)中:Activation 为激活函数,并且选用 LeakyReLU 激活函数对 TCN 残差结构进行优化,将所有的负值施加一个非零斜率。其表示方式如式(4)所示。

$$y_i = \begin{cases} x_i, & x_i > 0 \\ \frac{x_i}{a_i}, & x_i \leq 0 \end{cases} \quad (4)$$

式(4)中: a_i 为 $(1, +\infty)$ 区间内的固定参数。在负

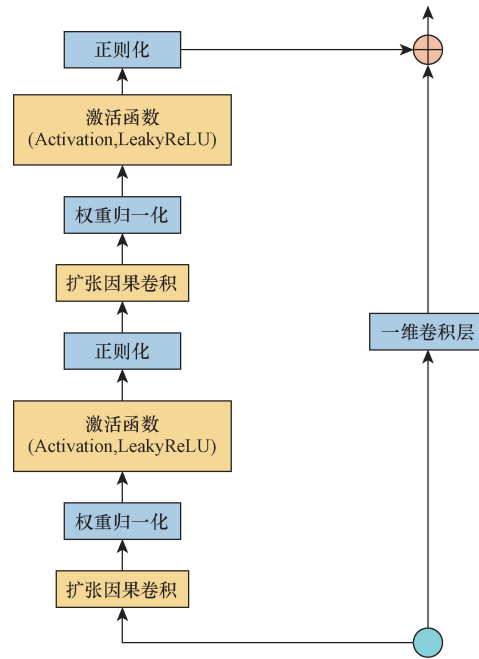


图3 残差模块

Fig. 3 Residual module

值区域 LeakyReLU 拥有比较小的正斜率,这样就会使得尽管输入负值,同样可以进行反向传播,并且同时保留了 ReLU 激活函数的优点。

1.2 CBAM 注意力机制

混合注意力机制(CBAM)包括通道注意力机制和空间注意力机制,CAM(channel attention module)模块执行通道注意力,SAM(spatial attention module)模块执行空间注意力^[23]。CBAM 通过串联方式将通道和空间注意力机制结合在一起,全方位关注输入特征的通道和空间两个方面。它使用空间注意力机制来定位目标区域,获取权重进行调整,通过通道注意力机制优化卷积通道之间的资源分配,提升目标区域的特征表现能力,提升 TCN 对输入数据的关注程度,从而提高模型性能^[24]。CBAM 网络框架如图 4 所示。

由图 4 可以看出,通道注意力机制就是将向量化卷积层输出的特征 F ,经过全局最大池化和全局平均池化后分别得到 F_{\max} 和 F_{avg} ,之后将 F_{\max} 和 F_{avg} 经过多层感知器(multilayer perceptron,MLP)处理得到两个特征向量,将两个特征向量相加并经过激活函数 Sigmoid 重新分配一个新的权重,获得输入特征层中每一个通道的权值,其表达方式如式(5)所示。

$$M_c(F) = \beta \{ \text{MLP}[\text{AvgPool}(F)] + \text{MLP}[\text{MaxPool}(F)] \} = \beta \{ W_1[W_0(F_{\max})] + W_1[W_0(F_{\text{avg}})] \} \quad (5)$$

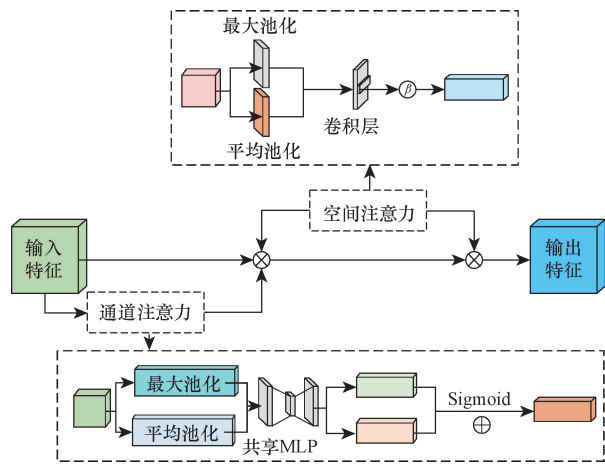


图4 CBAM网络框架图

Fig. 4 CBAM network frame diagram

式(5)中: W_0 和 W_1 为 MLP 的权重, 并且共享输入; MLP 为多层感知机; β 为 Sigmoid 函数, 表达方式如式(6)所示。

$$\beta(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

对于空间注意力机制的输入是通道注意力模块特征 F' , 通过全局最大池化和全局平均池化后分别得到 F'_{\max} 和 F'_{avg} 两个特征向量, 然后对 F'_{\max} 和 F'_{avg} 两个特征向量进行卷积操作, 通过激活函数 Sigmoid 将每个特征重新分配权重。其表达方式如式(7)所示。

$$\begin{aligned} M_s(F) &= \beta(f^{6 \times 6} \{ [\text{AvgPool}(F); \\ &\quad \text{MaxPool}(F)] \}) \\ &= \beta(f^{6 \times 6} \{ [F'_{\text{avg}}; F'_{\max}] \}) \end{aligned} \quad (7)$$

式(7)中: $f^{6 \times 6}$ 表示滤波器大小为 6×6 的卷积操作。混合注意力机制就是将通道注意力模块中的输入 F 与经过通道注意力机制得出的特征图 M_c 相乘, 并串联空间注意力模块中的输入 F' 。最终, 将其与经过空间注意力机制得到的特征图 M_s 相乘, 得到最终的 F'' 。其表达方式如式(8)所示。

$$F'' = M_s(F M_c + F') \quad (8)$$

1.3 软阈值化

软阈值化是一种非线性变化, 它能够保留正面和负面的特征, 并将近似于零的特征值设为零。所以关键信息都得以保留, 并可去除与噪声相关的特征。软阈值函数是将输入 x 绝对值小于阈值 τ 的特征删除, 将绝对值大于 τ 的特征朝着零的方向进行收缩^[25]。其表达方式如式(9)所示。

$$y = \begin{cases} x - \tau, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ x + \tau, & x < -\tau \end{cases} \quad (9)$$

对软阈值化函数求导得

$$\frac{\partial y}{\partial x} = \begin{cases} 1, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ 1, & x < -\tau \end{cases} \quad (10)$$

式中: x 为卷积层在网络中提取的特征; y 为通过后端输出的特征; τ 为阈值。由式(10)可以看出, 软阈值函数的导数只有 0 和 1 两种取值。软阈值函数如图 5 所示。

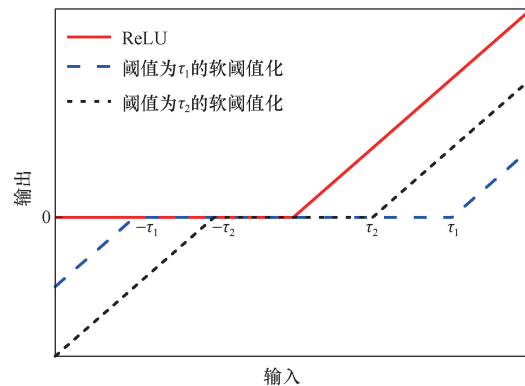


图5 软阈值函数

Fig. 5 Soft threshold function

由图 5 可以看出, 在区间 $[-\tau, \tau]$ 设置的值为零, 通过调节 τ , 软阈值可以将任意区间的特征值设置为零。因此, 绝对值小于阈值 τ 的特征和噪声等相关特征能够被软阈值化有效去除。

2 齿轮故障诊断模型

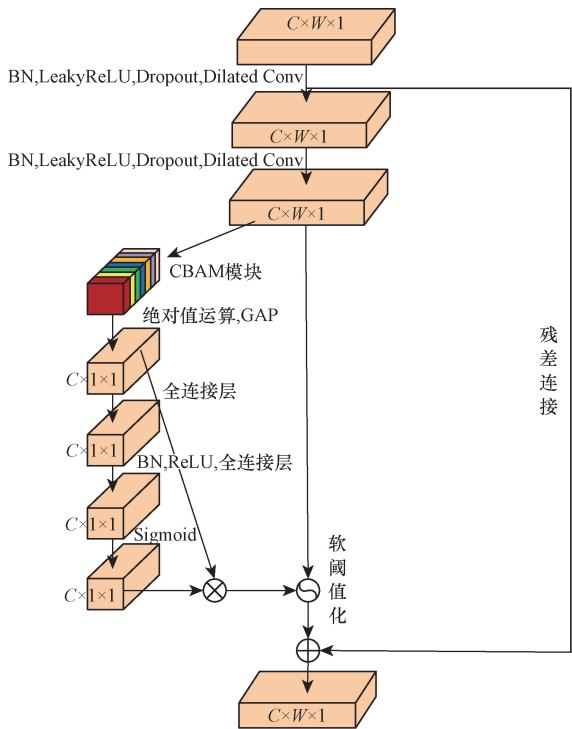
不同工况下齿轮箱的故障特征不同, 并且在采集齿轮振动信号时会受到噪声干扰。这些噪声会对齿轮箱故障诊断产生影响, 导致故障特征难以准确提取, 降低识别效果和准确率。因此, 过滤无效信息显得尤为重要, 对于特征提取的注意力程度不同, 需要设置不同的阈值来满足提取特征的不同通道。因此本文提出了一种端到端具有混合注意力机制和软阈值化特点的时间卷积神经网络 (CBAM-STCN) 齿轮箱故障诊断模型。

2.1 TCN 模块

在基本 TCN 模块中, 因为有效的故障特征会被全局平均池化层忽略, 造成局部故障特征信息丢失。当全连接层接收到所获得信息时, 不能将与故障有关的信息进行有效的合理组合。因此, 为了提高故障诊断准确率, TCN 需要增强对局部特征信息的提取能力, 通过充分表示特征, 更准确地捕获数据中的信息, 提高模型的性能和泛化能力。

在基本 TCN 模块中第二个扩张因果卷积层后嵌入 CBAM 注意力机制模块, 通过 CBAM 模块自适应地学习通道注意力和空间注意力的权重, 提高诊断模型的特征表达能力, 进一步获取局部有效信

息,抓取与故障特征相关的信息。其可以看成是在不同维度上捕获特征之间的相关性,从而提高模型性能。在基本 TCN 模块中引入软阈值函数,保留有效的特征。阈值函数利用注意力机制训练的自网络进行自适应学习,阈值不是确定值。在 TCN 的结构中创建一个子网络,该子网络经过注意力机制进行训练。在训练过程中,通过优化来动态调整阈值函数的值,以减少模型输出与原始输入之间的差异。基本 TCN 模块如图 6 所示。



C 为通道数;W 为特征图的宽度

图 6 基本 TCN 模块

Fig. 6 Basic TCN module

TCN 模块主要包含两个扩张因果卷积层,一个 CBAM 模块,一个 GAP(global average pooling)层,两个全连接层和软阈值化。该模块将 BN 和 LeakyReLU 函数置于扩张因果卷积层之前,采用了预激活残差方法,LeakyReLU 激活函数,优化了零梯度问题,使得优化过程更加稳定。

2.2 网络整体结构

提出的 CBAM-STCN 网络整体结构由输入层、卷积层、最大池化层、3 个基本 TCN 模块、全局平均池化层和全连接层组成,CBAM-STCN 网络整体结构如图 7 所示。

提出的 CBAM-STCN 网络各个结构层参数如表 1 所示。

2.3 模型故障诊断流程

基于 CBAM-STCN 的端到端智能诊断方法,模型的诊断流程如图 8 所示,具体步骤如下。

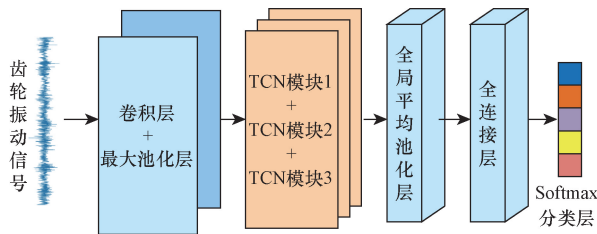


图 7 CBAM-STCN 网络整体结构

Fig. 7 CBAM-STCN network overall structure

表 1 各结构层参数

Table 1 Parameters of each structural layer

网络结构	参数
卷积层	卷积核数量 16,大小 64,步长 8
最大池化层	1 × 4 最大池化
TCN 模块 1	32 × 13,泄露率 0.2,膨胀率 1
TCN 模块 2	64 × 13,泄露率 0.2,膨胀率 2
TCN 模块 3	128 × 13,泄露率 0.2,膨胀率 4
GAP	全局平均池化
全连接层	—
Softmax	—

步骤 1 利用希尔伯特变换对采集的齿轮箱振动信号进行转换,由时域数据转换为包络谱数据。

步骤 2 将包络谱数据划分为训练集和测试集,并且将它们输入到本文模型中进行训练。

步骤 3 利用本文模型提取的故障特征验证测试数据集,评估故障识别分类的性能。

3 实验与分析

3.1 数据集介绍

实验采用的是东南大学齿轮箱数据集。该数据集是基于动力传动故障诊断试验台(drivetrain dynamic simulator, DDS)收集的齿轮箱故障数据集^[26]。

该数据集包含 5 种类型的齿轮数据,包括 1 种健康类型和 4 种故障类型。故障种类有缺陷(Chipped, 齿轮上有裂纹)、断齿(Miss, 齿轮上有断齿)、齿根磨损(Root, 齿根上有裂纹)和齿面磨损(Surface, 齿轮表面有磨损),故障类型如表 2 所示。

在该数据集中,包括两种工况,分别为速度 20 Hz(1 200 r/min)-空载 0 V(0 N·m)和速度 30 Hz(1 800 r/min)-负载 2 V(7.32 N·m)^[27],用于验证故障诊断模型在不同工况下的泛化能力。

对于每一个工况取 2 048 个采样点作为一个采样样本,每个工况采集 5 000 个独立样本,其中每个工况下每一种类型的齿轮采集 1 000 个独立样本。此数据集的齿轮故障诊断是一个五分类任务,对于采集的每个工况下 5 000 个独立样本,将时域数据利用希尔伯特变换^[28]转换为包络谱数据,突出故障频率、减少噪声干扰、增强故障特征以及简化故障诊

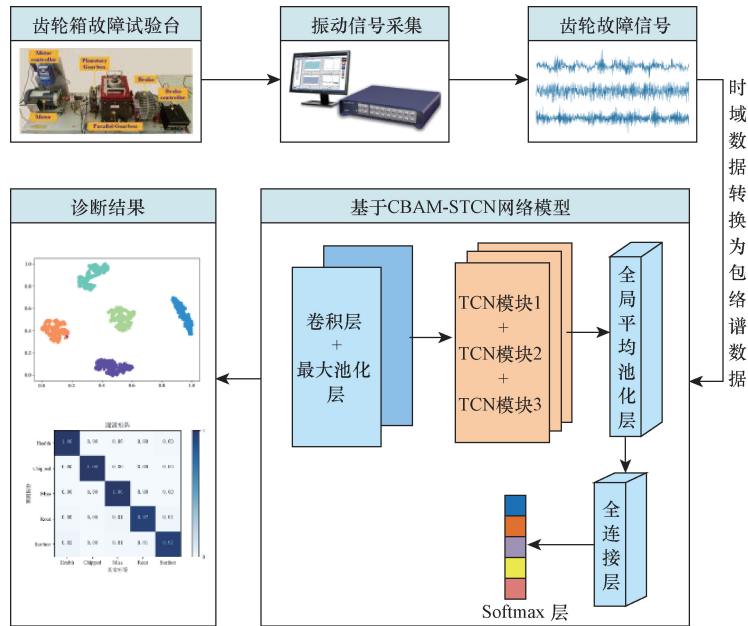


图8 本文模型诊断流程

Fig. 8 This paper model diagnostic flow

表2 齿轮故障类型

Table 2 Gear fault type

种类	详细情况
Chipped	齿轮上有裂纹
Miss	齿轮上有断齿
Root	齿根上有裂纹
Surface	齿轮表面有磨损

断分析过程。将4 000个样本用作训练集,1 000个样本用作测试集,进行100轮迭代训练。

3.2 实验结果分析

将本文模型(CBAM-STCN)与卷积神经网络(CNN)、深度残差网络(residual network, ResNet)、深度学习模型(LeNet)以及改进卷积神经网络(deep convolutional neural network, DCNN)进行对比。由实验结果可以得到,本文模型在识别分类准确率方面高于其他4种方法。在速度20 Hz(1 200 r/min)-空载0 V(0 N·m)的工况下,本文模型比CNN和DCNN分别提高了5.1%和8%,比ResNet和LeNet分别提高了2.6%和8.7%。在速度30 Hz(1 800 r/min)-负载2 V(7.32 N·m)的工况下,本文模型比CNN和DCNN提高了分别2.9%和2.6%,比ResNet和LeNet分别提高了0.3%和2.9%。两种工况下本文模型比CNN和DCNN平均提高了4%和5.3%,比ResNet和LeNet平均提高了1.45%和5.8%。实验结果如表3所示。

通过本文模型与CNN、ResNet、LeNet和DCNN在两种工况下齿轮故障分类准确率的对比,可以得出本文模型与上述4种诊断模型相比具有更好的自适应性和泛化能力。

表3 齿轮故障分类准确率

Table 3 Gear fault classification accuracy

方法	分类准确率/%	
	20 Hz-0 V	30 Hz-2 V
CNN	93.6	96.3
ResNet	96.1	98.9
LeNet	90.0	96.3
DCNN	90.7	96.6
本文模型	98.7	99.2

3.2.1 齿轮故障诊断模型性能对比

本文模型与CNN、ResNet、LeNet和DCNN在两种工况下测试集的分类准确率如图9和图10所示。

由图9、图10可以看出,在20 Hz-0 V工况下,相比于其他4种诊断模型,本文模型在第16次迭代,测试集的准确率就达到了91.4%,并且在第50次迭代,测试集的准确率达到了98.7%并趋于平稳。

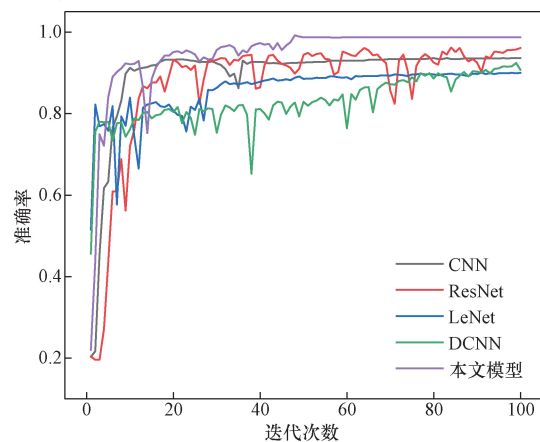


图9 20 Hz-0 V 工况下测试集准确率

Fig. 9 Test set accuracy under 20 Hz-0 V condition

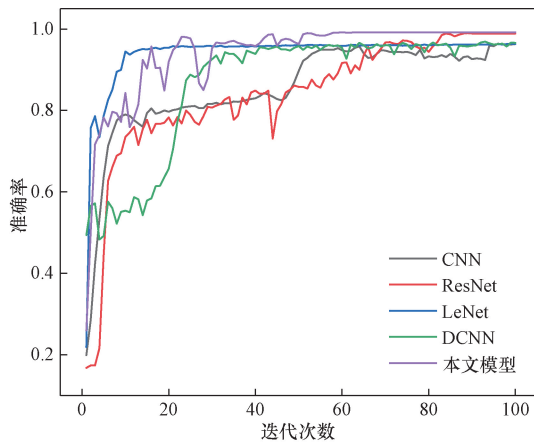


图 10 30 Hz-2 V 工况下测试集准确率
Fig. 10 Test set accuracy under 30 Hz-2 V condition

在 30 Hz-2 V 工况下,本文模型在第 14 次迭代,测试集的准确率就达到了 92.4%,并且在第 60 次迭代,测试集的准确率达到 99.2%并趋于平稳。在两种工况下,本文模型与 CNN、ResNet、LeNet 和 DCNN 相比收敛速度更快,并且拥有更好的泛化能力以及更高的故障识别分类准确率。

本文模型与 CNN、ResNet、LeNet 和 DCNN 在两种工况下测试集的平均损失如图 11、图 12 所示。

由图 11、图 12 可以看出,在 20 Hz-0 V 工况下,相比于其他四种诊断模型,本文模型在第 22 次迭代,测试集的平均损失达到了 12.1%,并且在第 48 次迭代,测试集的平均损失降低至 2.3%并趋于平稳。在 30 Hz-2 V 工况下,本文模型在第 24 次迭代,测试集的平均损失达到了 5.5%,并且在第 58 次迭代,测试集的平均损失降低至 2.8%并趋于平稳。本文模型在两种工况下相比于 CNN、ResNet、LeNet 和 DCNN 测试集的平均损失下降速率明显更快,而且最终测试集的平均损失明显低于上述 4 种故障诊断模型。

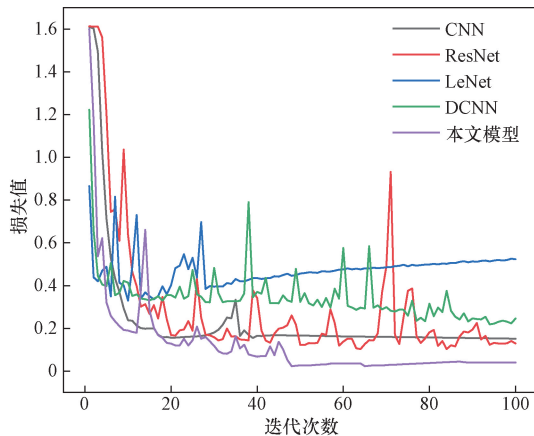


图 11 20 Hz-0 V 工况下测试集平均损失
Fig. 11 The average loss of the test set under 20 Hz-0 V condition

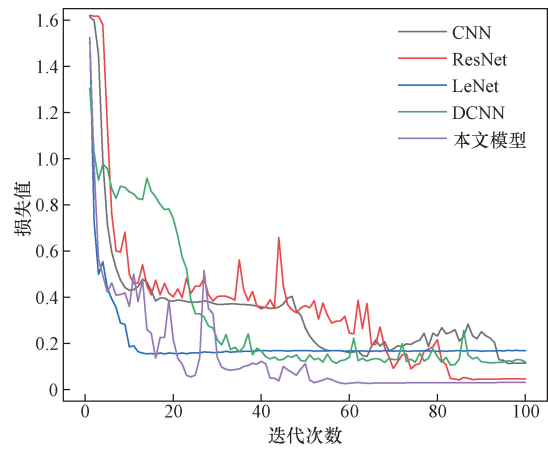


图 12 30 Hz-2 V 工况下测试集平均损失
Fig. 12 The average loss of the test set under 30 Hz-2 V condition

3.2.2 特征的提取、降维与可视化

通过特征提取、降维与可视化,能够从原始数据中提取出关键特征、降低数据维度,并且实现对数据的有效可视化。利用 *t*-SNE 对原始数据和经过模型训练特征进行可视化处理,其中未训练特征可视化和训练后特征可视化如图 13、图 14 所示。图中蓝色表示 Health(齿轮健康),绿色表示 Chipped(齿轮上有裂纹),青色表示 Miss(齿轮上有断齿),

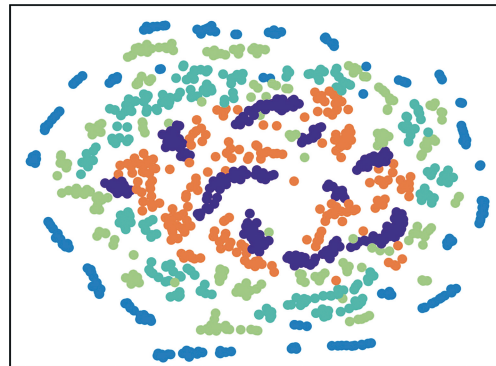


图 13 未训练特征可视化
Fig. 13 Visualization of untrained features

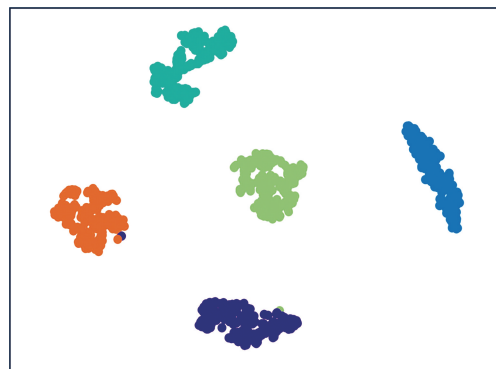


图 14 训练后特征可视化
Fig. 14 Visualization of trained features

紫色表示 Root (齿根上有裂纹), 橙色表示 Surface (齿轮表面有磨损)。

由图 13、图 14 可以看出, 未经训练的 5 种类型齿轮数据分布混乱, 经过本文模型训练之后, 齿轮的五种类型分类整洁, 几乎无分类错误。说明本文模型能够成功捕捉两种工况下齿轮在正常与故障状态下的差异, 能够准确地将相似的样本聚集在一起, 具有较好的泛化能力, 并且拥有较高的齿轮箱故障识别准确率。

3.3 量化故障分类

为了更准确地评估齿轮在两种不同工况下故障的分类准确率, 引入混淆矩阵进一步量化和可视化故障诊断模型的性能。两种工况下故障诊断模型五分类输出的混淆矩阵如图 15、图 16 所示。

其中, 横坐标与纵坐标的 Health (齿轮健康)、Chipped (齿轮上有裂纹)、Miss (齿轮上有断齿)、Root

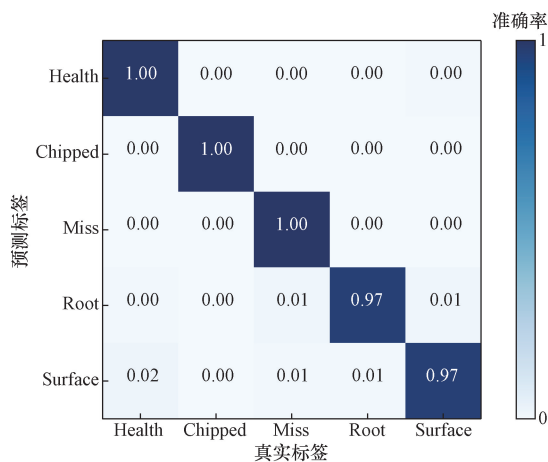


图 15 20 Hz-0 V 工况下故障分类混淆矩阵

Fig. 15 Confusion matrix for fault classification under 20 Hz-0 V condition

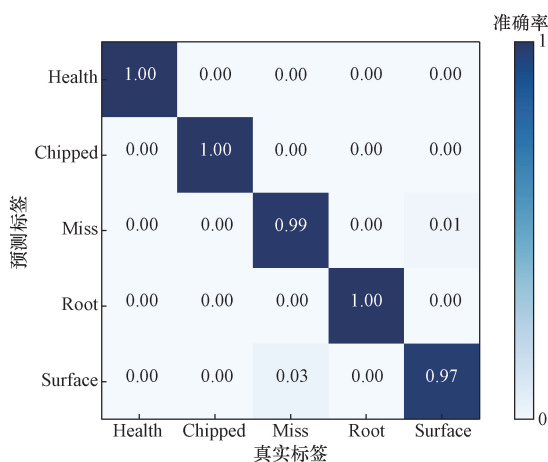


图 16 30 Hz-2 V 工况下故障分类混淆矩阵

Fig. 16 Confusion matrix for fault classification under 30 Hz-2 V condition

(齿根上有裂纹)、Surface (齿轮表面有磨损) 表示齿轮的 5 种类型, 色标刻度轴 0 ~ 1 表示类别识别准确率为 0% ~ 100%。由图 15、图 16 可以看出, 本文模型在两种工况 5 种齿轮类型的分类中, 对于齿轮健康状态与齿轮上有裂纹类型的分类准确率为 100%, 对于其余 3 种齿轮类型分类的准确率几乎为 100%, 因此本文模型具有良好的故障诊断性能。

4 结论

针对齿轮箱在多种工况下故障特征存在差异, 故障诊断受噪声干扰, 导致故障诊断模型泛化性差和识别准确率低的问题。提出了一种具有混合注意力机制和软阈值化特点的时间卷积神经网络 (CBAM-STCN) 齿轮箱故障诊断模型。通过理论与实验研究得出如下结论。

(1) 利用本文模型对齿轮箱进行故障特征提取与智能故障诊断, 实现了齿轮箱端到端的智能诊断。

(2) 利用本文模型进行齿轮箱故障诊断, 故障诊断平均准确率达到 98.95%, 相比于其他诊断模型, 本文模型诊断准确率和效率最高。

(3) 本文模型在齿轮箱不同工况下展现出了良好的适用性和泛化能力, 为齿轮箱故障诊断提供了可靠的解决方案。

参考文献

- [1] Zheng X Y, Ye Z Y, Wu J L. A CNN-ABiGRU method for gearbox fault diagnosis[J]. International Journal of Circuits, Systems and Signal Processing, 2022, 16: 440-446.
- [2] Zhuang Y, Wang S Y, Shang Y, et al. Virtual-real fusion-based transfer learning with limited data for gearbox fault diagnosis[J]. IEEE Sensors Journal, 2024, 24(3): 3420-3430.
- [3] Zhang X F, Xu Q H, Jiang H, et al. Application of deep neural network in gearbox compound fault diagnosis[J]. Energies, 2023, 16(10): 4164.
- [4] 程旺, 郝如江, 段泽森, 等. 基于参数优化变分模态分解与支持向量机的齿轮箱故障诊断[J]. 科学技术与工程, 2022, 22(15): 6099-6105.
Cheng Wang, Hao Rujiang, Duan Zesen, et al. Gearbox fault diagnosis based on parameter optimization variational modal decomposition and support vector machine[J]. Science Technology and Engineering, 2022, 22(15): 6099-6105.
- [5] Lei Y G, Yang B, Jiang X W, et al. Applications of machine learning to machine fault diagnosis: a review and road map[J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.
- [6] Sun G D, Wang Y R, Sun C F, et al. Intelligent detection of a planetary gearbox composite fault based on adaptive separation and deep learning[J]. Sensors, 2019, 19(23): 5222.
- [7] Zhang J Q, Zhang Q, Qin X R, et al. 2D characterization based on MSGMD and its application in gearbox fault diagnosis[C]//2023 IEEE International Conference on Prognostics and Health Management (ICPHM). Montreal: IEEE, 2023: 328-334.

- [8] Pei X L, Zheng X Y, Wu J L. Rotating machinery fault diagnosis through a transformer convolution network subjected to transfer learning[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-11.
- [9] Li D, Qing L. Fault diagnosis of rotating machinery using novel self-attention mechanism TCN with soft thresholding method[J]. Measurement Science and Technology, 2024, 35(4): 047001.
- [10] 李莎, 陈泽华, 刘海军. 基于 ST-TCN 的太阳能光伏组件故障诊断方法[J]. 电子技术应用, 2022, 48(12): 79-88.
Li Sha, Chen Zehua, Liu Haijun. Fault diagnosis method of solar panel module based on ST-TCN[J]. Application of Electronic Technique, 2022, 48(12): 79-88.
- [11] 吕卫民, 孙晨峰, 任立坤, 等. 一种基于 TCN-LGBM 的航空发动机气路故障诊断方法[J]. 兵工学报, 2024, 45(1): 253-263.
Lü Weimin, Sun Chenfeng, Ren Likun, et al. A gas path fault diagnosis method for aero-engine based on TCN-LGBM model[J]. Acta Armamentarii, 2024, 45(1): 253-263.
- [12] 张璐莹, 侯立群. 基于注意力时间卷积网络和双向门控循环单元的轴承故障诊断[J]. 电力科学与工程, 2023, 39(6): 62-70.
Zhang Luying, Hou Liqun. Bearing fault diagnosis based on Attention temporal convolutional network and bidirectional gated recurrent unit[J]. Electric Power Science and Engineering, 2023, 39(6): 62-70.
- [13] Zhang J Y, Chang Y, Zou J X, et al. AME-TCN: attention mechanism enhanced temporal convolutional network for fault diagnosis in industrial processes[C]//2021 Global Reliability and Prognostics and Health Management Conference. Nanjing: IEEE, 2021: 1-6.
- [14] Zhu Y Y, Pei Y, Wang A Q, et al. A partial domain adaptation scheme based on weighted adversarial nets with improved CBAM for fault diagnosis of wind turbine gearbox[J]. Engineering Applications of Artificial Intelligence, 2023, 125: 106674.
- [15] Wang Z H, Tao Y X, Du Y P, et al. Optimization of gearbox fault detection method based on deep residual neural network algorithm[J]. Sensors, 2023, 23(17): 7573.
- [16] Zhan S N, Shao R P, Men C J, et al. Fault diagnosis method for planetary gearbox based on intrinsic feature extraction and attention mechanism[J]. Measurement Science and Technology, 2023, 35(3): 035116.
- [17] 赵星宇, 吴泉军, 朱威. 基于 CEEMDAN 和 TCN-LSTM 模型的短期电力负荷预测[J]. 科学技术与工程, 2023, 23(4): 1557-1564.
Zhao Xingyu, Wu Quanjun, Zhu Wei. Short-term power load forecasting based on CEEMDAN and TCN-LSTM model[J]. Science Technology and Engineering, 2023, 23(4): 1557-1564.
- [18] 项新建, 张颖超, 许宏辉, 等. 基于 CEEMDAN-VMD-TCN-lightGBM 模型的水质预测研究[J]. 中国农村水利水电, 2023(11): 1-22.
Xiang Xinjian, Zhang Yingchao, Xu Honghui, et al. Research on water quality prediction based on CEEMDAN-VMD-TCN-lightGBM model[J]. China Rural Water and Hydropower, 2023(11): 1-22.
- [19] 余琼芳, 王联港, 杨艺. 基于 LSTM-TCN 的综采工作面顶板压力预测[J]. 煤炭技术, 2023, 42(6): 5-9.
Yu Qiongfang, Wang Liangang, Yang Yi. Pressure prediction of top plate of comprehensive mining working face based on LSTM-TCN[J]. Coal Technology, 2023, 42(6): 5-9.
- [20] 马佳成, 王晓霞, 杨迪. 基于 Attention 机制的 TCN-LSTM 非侵入式负荷分解[J]. 电力信息与通信技术, 2023, 21(8): 43-51.
Ma Jiacheng, Wang Xiaoxia, Yang Di. Non-intrusive load decomposition based on TCN-LSTM model with Attention mechanism[J]. Electric Power Information and Communication Technology, 2023, 21(8): 43-51.
- [21] 王世杰, 王兴芬, 岳婷. 基于 XGBoost 和 TCN-Attention 的棉花价格多影响因素选择及预测[J]. 计算机系统应用, 2023, 32(10): 10-21.
Wang Shijie, Wang Xingfen, Yue Ting. Selection and prediction of multiple influencing factors of cotton price based on XGBoost and TCN-Attention[J]. Computer Systems & Applications, 2023, 32(10): 10-21.
- [22] 王焱, 丁华, 孙晓春, 等. 基于改进 ECANet-TCN 和迁移学习的轴承剩余寿命预测[J]. 振动与冲击, 2023, 42(21): 149-159.
Wang Yan, Ding Hua, Sun Xiaochun, et al. Bearing residual life prediction based on improved ECANet-TCN and transfer learning[J]. Journal of Vibration and Shock, 2023, 42(21): 149-159.
- [23] 孙敏, 成倩, 丁希宁. 基于 CBAM-CGRU-SVM 的 Android 恶意软件检测方法[J]. 计算机应用, 2024, 44(5): 1539-1545.
Sun Min, Cheng Qian, Ding Xining. CBAM-CGRU-SVM based malware detection method for Android[J]. Journal of Computer Applications, 2024, 44(5): 1539-1545.
- [24] 李筱玉, 张乾, 周遵富, 等. 融合 CBAM 注意力机制的区域归一化图像修复[J]. 信息技术与信息化, 2023(10): 136-143.
Li Xiaoyu, Zhang Qian, Zhou Zunfu, et al. Region normalization image inpainting with CBAM attention module[J]. Information Technology and Informatization, 2023(10): 136-143.
- [25] 刘高辉, 宋博武. DRSN 与集成融合的 OFDM 辐射源个体识别方法[J]. 信号处理, 2024, 40(6): 1062-1073.
Liu Gaohui, Song Bowu. DRSN and integrated fusion OFDM radiation source individual identification method[J]. Journal of Signal Processing, 2024, 40(6): 1062-1073.
- [26] Shao S Y, McAleer S, Yan R Q, et al. Highly accurate machine fault diagnosis using deep transfer learning[J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2446-2455.
- [27] Dai Y, Ou Y G, Hu J W, et al. Few-shot gearbox fault diagnosed based on meta-learning and time convolution network[C]//2022 China Automation Congress (CAC). Xiamen: IEEE, 2022: 3341-3346.
- [28] 沙云东, 陈兴武, 栾孝驰, 等. 基于小波包分解-峭度值指标-希尔伯特包络解调融合方法处理声发射信号的滚动轴承故障诊断[J]. 科学技术与工程, 2023, 23(21): 9315-9323.
Sha Yundong, Chen Xingwu, Luan Xiaochi, et al. Fault diagnosis of rolling bearing based on acoustic emission signal analysis by WPD-KIHED combination method[J]. Science Technology and Engineering, 2023, 23(21): 9315-9323.