

基于 MFMD 和 Transformer-CNN 的滚动轴承故障诊断方法

刘俊锋¹, 俞翔², 万海波², 刘潇¹

(1. 海军工程大学 动力工程学院, 武汉 430033;

2. 海军工程大学 舰船与海洋学院, 武汉 430033)

摘 要: 针对滚动轴承在变工况和跨型号下故障诊断效果不佳、泛化能力较差, 同时在实际训练中样本数量严重不足的问题, 从振动信号序列特性出发, 提出了一种基于改进傅里叶模态分解(MFMD)和 Transformer convolutional neural network(Transformer-CNN)的故障诊断方法。设计了振动数据预处理模块, 利用 MFMD 和位置编码对数据样本进行预处理并标记序列位置关系, 随后设计了基于注意力机制的 Transformer-CNN 序列建模单元, 利用最大值池化优化了缩放点积注意力机制的循环堆叠结构, 减少了网络的待训练参数并提升了网络序列建模能力。采用预训练-微调的迁移学习方法, 将预训练模型参数迁移至目标域并进行模型微调, 可以避免数据不足导致的过拟合现象。实验结果表明: 相较于基准算法, Transformer-CNN 可以降低 50% 以上的故障诊断错误率。在变工况和跨型号的小样本迁移学习实验中, 该算法可以提升 8.75% 的诊断准确率, 同时可以提升收敛速度。

关 键 词: 滚动轴承; 故障诊断; Transformer-CNN; 小样本; 迁移学习

中图分类号: V263.6; TH133.3

文献标志码: A

Fault diagnosis method of rolling bearing using MFMD and Transformer-CNN

LIU Junfeng¹, YU Xiang², WAN Haibo², LIU Xiao¹

(1. College of Power Engineering, Naval University of Engineering, Wuhan 430033, China;

2. College of Naval Architecture and Ocean Engineering,

Naval University of Engineering, Wuhan 430033, China)

Abstract: Considering the worse effect and generalization ability of rolling bearing fault diagnosis under variable working conditions and cross-type conditions as well as the shortage of serious samples in practice, a fault diagnosis method using modified Fourier mode decomposition (MFMD) and Transformer convolutional neural network (Transformer-CNN) was proposed based on the sequence characteristics of vibration signal. The vibration data preprocessing module was designed, in which MFMD and position encoding were adopted to preprocess the samples and mark the sequence position relationships. The Transformer-CNN sequence modeling unit with the scaled dot-product attention mechanism was then designed, and the cyclic stack structure was optimized by the max-pooling, which reduced the network parameters and improved the sequence modeling capability. The pre-training-fine-tuning transfer learning method was adopted to transfer the trained mode parameters to the target domain and fine-tune, which avoided the over-fitting caused by insufficient data. The results showed that Transformer-CNN can reduce

收稿日期: 2021-12-15

基金项目: 国家自然科学基金(51679245)

作者简介: 刘俊锋(1997—), 男, 博士生, 主要从事旋转机械故障诊断与智能运维研究。

引用格式: 刘俊锋, 俞翔, 万海波, 等. 基于 MFMD 和 Transformer-CNN 的滚动轴承故障诊断方法[J]. 航空动力学报, 2023, 38(6): 1446-1456.

LIU Junfeng, YU Xiang, WAN Haibo, et al. Fault diagnosis method of rolling bearing using MFMD and Transformer-CNN[J]. Journal of Aerospace Power, 2023, 38(6): 1446-1456.

the fault diagnosis error by more than 50% compared with the benchmark algorithms. In the case of cross-working and cross-type conditions with small samples, the algorithm enables to achieve 8.75% diagnosis accuracy improvement and faster convergence.

Keywords: rolling bearing; fault diagnosis; Transformer-CNN; small sample; transfer learning

滚动轴承是旋转机械设备中不可或缺的关键零部件,对滚动轴承的运行状态进行监测并进行早期故障诊断具有重要意义^[1]。

随着大数据时代来临,以傅里叶分解^[2-3]、小波变换^[4]和变分模态分解^[5]为代表的传统信号处理方法难以满足大数据的需求。一方面,复杂的计算过程导致数据处理效率低下;另一方面,人工选取故障特征难以适应复杂的变工况和跨型号情况。与此同时,以人工神经网络为代表的机器学习算法因其强大的数据处理能力在故障诊断领域得到了广泛应用^[6-7]。但是由于滚动轴承型号繁多,实际运行工况也复杂多变,基于人工神经网络的故障诊断方法在实际应用中常常面临训练样本不足或样本标签缺失等问题。经典监督学习方法在小样本情况下,由于数据特征不足和模型参数冗余常常无法学习到有效数据特征或学习到偶然特征,从而出现过拟合问题^[8],导致在训练中损失值足够小,而测试时损失值很大,并且随着训练次数增长,诊断准确率反而下降^[9]。同时,将深度学习模型直接应用到变工况或跨型号条件下进行故障诊断,其前提为源域和目标域具有相同的数据分布,否则将会导致算法诊断性能显著下降^[10]。

为解决上述问题,目前主要通过改进网络结构、扩充数据和迁移学习的方式。张西宁等^[11]提出了一种基于全局均值优化的卷积神经网络迁移学习方法并成功应用到变工况和跨型号的轴承故障诊断中,达到了 92.25% 的平均分类准确率。Zhang 等^[12]利用基于多尺度特征提取单元的改进神经网络进行变工况轴承故障诊断,一定程度上缓解了泛化能力较差的问题。Zhu 等^[13]利用多层多核的最大均值差异分析框架有效实现了变工况条件下的轴承故障诊断。何强等^[14]设计了一种基于梯度惩罚的距离生成对抗网络,通过对抗训练生成与训练样本分布相似的新样本。这些方法在一定程度上缓解了变工况或跨型号下轴承故障诊断效果不佳的问题,但诊断精度方面仍存在差距。同时,传统神经网络的运行过程中忽略了轴承振动信号的时序特性,难以挖掘输入之间关键的相互关联信息。以长短期记忆神经网络^[15]和

门控循环神经网络^[16]为代表的循环神经网络(recurrent neural network, RNN)通过隐藏层循环单元之间的权重使得序列数据的信息可以向后传递,具有独特的序列特征挖掘优势^[17]。但 RNN 固有的顺序特性阻碍了样本训练的并行化,降低了网络的运行效率。同时, RNN 关联任意两个位置的信号所需的操作次数会随着位置之间的距离增加而激增,导致严重的长依赖问题。2017 年, Ashish 等^[18]提出了一种完全基于注意力机制的 Transformer 神经网络,该网络可以有效缓解了 RNN 的长依赖问题,已在自然语言处理和计算机视觉等领域取得了优异效果。

基于此,本文针对变工况和跨型号滚动轴承故障诊断样本严重不足和振动信号序列信息提取困难的问题,首先提出一种基于改进傅里叶模态分解(modified Fourier mode decomposition, MFMD)的样本预处理方法,利用频带熵指标优化频率边界并确定敏感频带,有效降低了大数据处理中的巨大运算量并避免了人为挑选敏感分量造成的效率低下问题。同时,本文提出一种金字塔结构的 Transformer convolutional neural network (Transformer-CNN)迁移学习轴承故障诊断模型,改进了传统 Transformer 模型的堆叠结构,结合预训练-微调的迁移学习方法,可以有效避免小样本数据导致的过拟合问题。实验结果表明,本文所提算法与基准算法相比较,在准确率和效率方面具有一定的优越性,可以有效解决变工况和跨型号的小样本轴承故障诊断问题。

1 理论分析

1.1 MFMD

MFMD 从共振解调思想出发,引入频带熵分析确定敏感频带,改进了傅里叶分解^[2]的边界频率搜索方式,分解具有完备性、局部性、自适应性和正交性^[19]。对任意满足 Dirichlet 条件的有限长的零均值实信号 $x(t)$ ($t \in [t_0, t_0 + T]$, 其中 t_0 为起始时间, T 为周期), 可将其自适应分解为一系列傅里叶本征模态函数(Fourier intrinsic mode function, FIMF), 具体流程定义如下。

步骤 1 构造信号 $x(t)$ 的周期延拓并进行广义傅里叶级数展开: $x_T(t) = \sum_{m=-\infty}^{\infty} x(t-mT)$, 使得 $x(t) = x_T(t)w(t)$, 其中 $m \in [-\infty, \infty]$, m 为周期数, $w(t) = 1, (t_0 \leq t \leq t_0 + T)$, 其他则 $w(t) = 0$ 。可以得到 $x_T(t)$ 的展开后的复数形式为

$$x_T(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t)] \quad (1)$$

$$\begin{cases} a_0 = \frac{1}{T} \int_{t_0}^{t_0+T} x_T(t) dt \\ a_k = \frac{2}{T} \int_{t_0}^{t_0+T} x_T(t) \cos(k\omega_0 t) dt \\ b_k = \frac{2}{T} \int_{t_0}^{t_0+T} x_T(t) \sin(k\omega_0 t) dt \end{cases} \quad (2)$$

式中角频率 $\omega_0 = 2\pi/T$, k 为波数, 式(1)可改写为 $x_T(t) = a_0 + \frac{1}{2} \sum_{k=1}^{\infty} [c_k \exp(jk\omega_0 t) + c_k^* \exp(-jk\omega_0 t)]$, 其中 $c_k = a_k - jb_k$, $c_k^* = a_k + jb_k$, 可以得到 $x_T(t) = a_0 + \text{Re}[z_T(t)]$, 其中

$$z_T(t) = \sum_{k=1}^{\infty} [c_k \exp(jk\omega_0 t)] \quad (3)$$

对信号 $x_T(t)$ 进行快速傅里叶变换

$$F(f) = \int_{-\infty}^{+\infty} x(t) e^{-jft} dt \quad (4)$$

式中 $F(f)$ 为复系数, f 为频率。

步骤 2 频率边界的准确性对传统傅里叶分解结果有着决定性影响, 而边界频率搜索极易受背景噪声干扰。同时, 傅里叶分解从高频到低频或者从低频到高频的边界频率搜索方法在处理多分量信号时计算量较大, 严重影响效率。因此, 本文提出一种基于频带熵分析^[20]的边界频率搜索方法, 按照如下方法对 $\text{Re}[F(f)]$ 进行自适应分割, 确定频率边界集:

1) 选取频带熵区域熵值极小值点为中心频率, 选取离中心频率最近的频带熵极大值点为敏感频带边界。然后依据敏感频带划分初始边界频率集 $\{A_s\} = \bigcup_{s=1}^S [f_{s-1}, f_s) = [0, F_s/2)$, 其中 $f_0 = 0$, $f_s = F_s/2$, F_s 为频率上限, $s=1, 2, \dots, S$, S 为初始区间个数。

2) 根据初始频率边界在各敏感区间内进行边界频率二次搜索, 在满足瞬时幅值 $a_i(t) \geq 0$ 和瞬时频率 $f_i(t) \geq 0$ 的情况下获取最小数量的解析 FIMFs, 即可得到最终优化的频率边界集。

$$\{B_i\} = \bigcup_{i=1}^I [f_{i-1}, f_i) = [0, F_s/2) \quad (5)$$

其中 $f_0 = 0$, $f_I = F_s/2$, I 为优化后区间个数。然后, 依据边界频率对其复系数的实部 $\text{Re}[F(f)]$ 进行自适应分割。

步骤 3 对信号在区间 $B_i = [f_{i-1}, f_i)$ 内进行逆快速傅里叶变换, 得到每个区间 B_i 内的解析分量 $I_i(t) \equiv x_i(t) = a_i(t) \exp[j\phi_i(t)]$, 其中 $a_i(t)$ 为瞬时幅值, $\phi_i(t)$ 为瞬时相位, $i=1, 2, \dots, I$ 。由此, 原始信号可表示为

$$x_T(t) = \sum_{i=1}^I I_i(t) = \sum_{i=1}^I [a_i(t) \exp[j\phi_i(t)]] \quad (6)$$

MFMD 无需在整个频带上进行边界频率搜索, 因此具有较高效率。同时, 利用频带熵分析搜索敏感频带, 无需人为挑选敏感分量, 极大提高了大数据样本处理的效率。

1.2 基于 Transformer-CNN 的迁移学习框架

Transformer 中的编码器 (encoder) 部分可用于解决分类问题, 但其循环计算结构在处理长序列数据时, 存在严重的参数冗余问题和过拟合风险。针对该问题, 本文引入池化层降采样改进了编码器的循环堆叠结构, 设计了一种完全基于自注意力机制的 Transformer-CNN 轴承故障诊断网络模型, 其完整结构如图 1 所示。该网络模型完全利用缩放点积自注意力机制完成序列建模, 其定义如下^[21]。

步骤 1 定义 3 个参数矩阵 W^Q 、 W^K 和 W^V , 利用这 3 个矩阵对所有的输入向量进行线性变换, 可以得到相应的查询向量、键向量和值向量, 将所有输入向量变换得到的对应向量进行拼接得到查询矩阵 Q 、键矩阵 K 和值矩阵 V 。

步骤 2 通过查询矩阵 Q 和键矩阵 K 相乘, 获取每个输入向量的注意力权重, 然后利用比例因子 $\sqrt{d_k}$ 进行放缩, 经过 softmax 后乘以值矩阵 V 得到输出

$$A(Q, K, V) = G(QK^T / \sqrt{d_k}) V \quad (7)$$

其中 A 代表输入的注意力机制运算函数, G 为归一化指数函数, d_k 为键矩阵维度。

步骤 3 为从多个维度挖掘序列信息, 采用多头注意力机制, 其定义为

$$M(Q, K, V) = D(H_1, \dots, H_h) \quad (8)$$

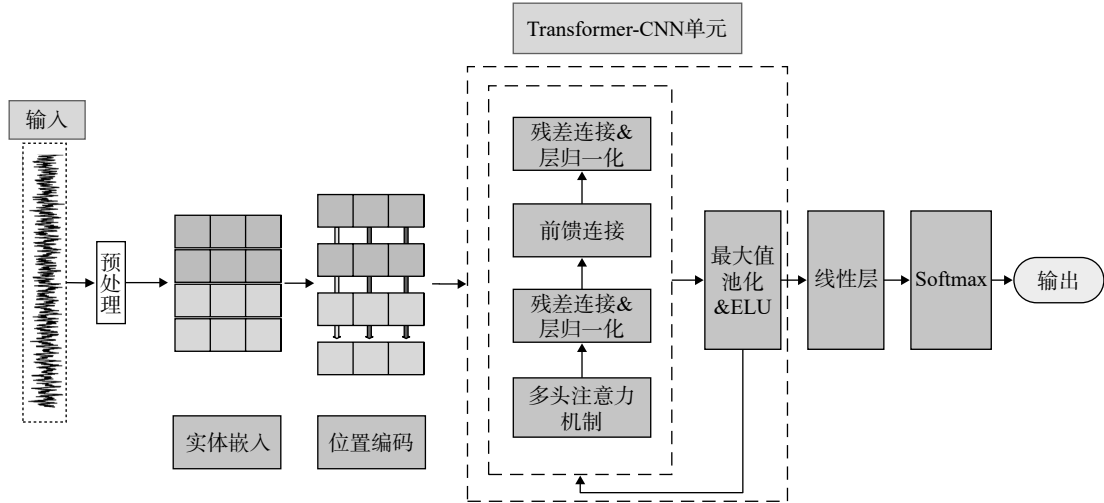


图1 Transformer-CNN 结构图

Fig. 1 Structure diagram of Transformer-CNN

$$H_z = A(OW_z^Q, OW_z^K, OW_z^V) \quad (9)$$

$$W_z^Q \in \mathbf{R}^{d_m \times d_k}, W_z^K \in \mathbf{R}^{d_m \times d_k}, W_z^V \in \mathbf{R}^{d_m \times d_v} \quad (10)$$

式中 M 为多头注意力机制, D 表示数据拼接函数, H_z 代表注意力机制第 z 个头, h 为注意力机制头数, O 为输入矩阵, d_v 为值矩阵维度, d_m 为向量长度, R 表示二维矩阵空间。

为保证分类性能,编码器一般采用循环堆叠的方式进行布置,但其运算量 $C(\cdot)$ 和序列维度 L 的平方成正比,因此处理长序列数据时计算压力较大,存在严重的参数冗余问题。因此,本文引入池化层降采样功能,在循环计算中调整序列长度,形成新的Transformer-CNN单元,如图1所示,其中池化层采用最大值池化,其定义为

$$y_b = \max[r_b^{n \times n} u(n, n)] \quad (11)$$

式中 y_b 为区域 b 的输出特征, $r_b^{n \times n}$ 代表第 b 个大小为 $n \times n$ 的区域, $u(n, n)$ 代表大小为 $n \times n$ 的窗函数。池化操作能够减少特征向量的尺寸,降低过拟合风险。激活函数采用指数线性单元(exponential linear unit, ELU)。Transformer-CNN单元和编码器的结构区别如图2所示。

Transformer-CNN可以有效缓解网络参数冗余情况,相较于Transformer,参数总量减少了45%。完整网络参数如表1所示。

本文输入样本为MFMD获取的频谱信号,由于频谱信号特征主要表现在低频部分,本文截取[0,800]Hz区间数据作为输入样本。其中,区间边界和故障特征频率密切相关,可设置为最大故障特征频率 $f_{\max}$ 的3~5倍。对于不同频率分辨率数据,采用三次样条插值重采样,固定输入长度为

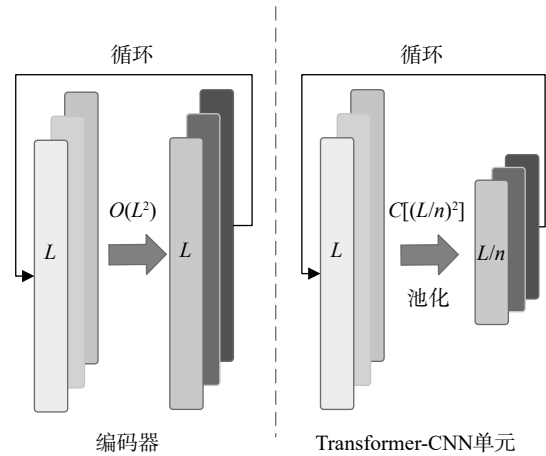


图2 Transformer-CNN 单元

Fig. 2 Transformer-CNN unit

表1 Transformer-CNN 网络参数

Table 1 Network parameters of Transformer-CNN

顺序	网络层	输出尺寸	核大小	标准化
1	输入层	64 × 800 × 1		
2	嵌入层	64 × 800 × 64		
3	位置编码层	64 × 800 × 64		
4	编码器层	64 × 800 × 64		是
5	池化层	64 × 800 × 8	8 × 1	
6	编码器层	64 × 800 × 8		是
7	编码器层	64 × 800 × 8		是
8	编码器层	64 × 800 × 8		是
9	池化层	64 × 800 × 1	8 × 1	
10	线性层	64 × 10 × 1		
11	Softmax	64 × 10		

800。然后利用实体嵌入,将输入样本映射到高维空间,以便进行位置标记和数据特征挖掘,嵌入

维度为 64。采用固定位置编码,可使网络识别数据位置信息,其定义为

$$\begin{aligned} P(p, 2q) &= \sin\left(p/10\,000^{2q/d_m}\right) \\ P(p, 2q+1) &= \cos\left(p/10\,000^{2q/d_m}\right) \end{aligned} \quad (12)$$

其中 P 是编码函数, p 是位置, q 是位置信息编码特征向量的第 q 个元素。通过引入位置编码,可让算法更容易学习到相对位置的依赖关系,因为对任意确定的偏移 g , P_{p+g} 可以表示 P_p 的线性函数。

针对变工况和跨型号中样本不足导致的过拟合或欠拟合问题,搭建基于 Transformer-CNN 的迁移学习框架,如图 3 所示。

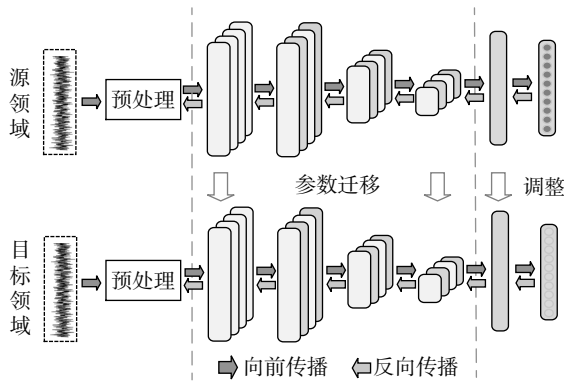


图 3 迁移学习框架

Fig. 3 Transfer learning framework

考虑有一个源域 D_s 和一个目标域 D_g 的迁移学习,其定义为:已知源域 D_s 和源任务 M_s ,目标域 D_g 和目标任务 M_g ,迁移学习的目标是利用从 D_s 和 M_s 中学习到的知识提升或实现 D_g 中目标预测函数 $f_g(\cdot)$ 的学习。本文开展变工况和跨型号迁移故障诊断流程如图 4 所示,该算法采用了

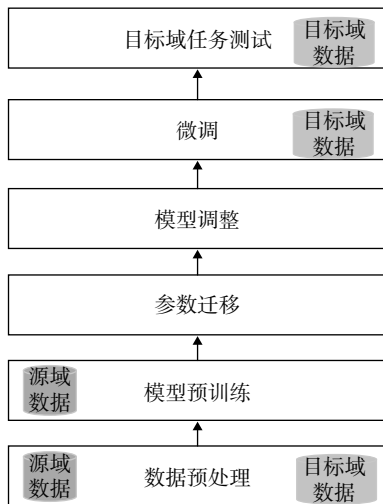


图 4 迁移学习算法流程

Fig. 4 Transfer learning algorithm flow

预训练-微调的迁移学习方式,首先利用源域数据开展故障诊断模型预训练。接着将预训练模型参数迁移到变工况和跨型号故障诊断任务中,固定较浅层并微调最后一层编码器层和线性层的参数,使网络适应目标域样本的数据分布,最后利用训练好的模型进行目标域数据识别。

实验在 Windows10 系统、Intel CORE i7-7700 HQ 处理器、16 GB 内存的计算机上进行,编程语言为 Python、平台为 Pycharm、使用 Pytorch 深度学习框架实现。为保证实验可信度,本文所有算法采用相同的输入数据,网络基本参数设置如表 2 所示。

表 2 网络基本参数设置

Table 2 Basic parameter setting of networks

项目	设置
优化器	Adam 学习率为 0.000 1
损失函数	交叉熵损失
批处理大小	64

2 实验分析

2.1 数据简介及预处理

采用凯斯西储大学(CWRU)轴承数据集进行实验分析,该数据集数据充足、应用广泛,适用于验证算法性能。测试轴承为驱动端轴承,型号为 SKF6205,振动数据通过加速度计采集,采样频率为 48 kHz。数据包含 3 种工况,后文简称工况 1、工况 2 和工况 3,内圈、外圈和滚动体故障都包含 3 种严重程度的故障,因此每种工况都包含 10 类健康状态。为保证样本包含一个完整周期并避免数据冗余,以滚动轴承旋转一圈的采样点数作为数据输入长度,对应 3 种工况分别为 1 625、1 645 和 1 665。为提升模型适应性,采用重叠采样的方式对原始信号进行截取,如图 5 所示。

从如图 6 所示的轴承加速寿命试验台采集轴

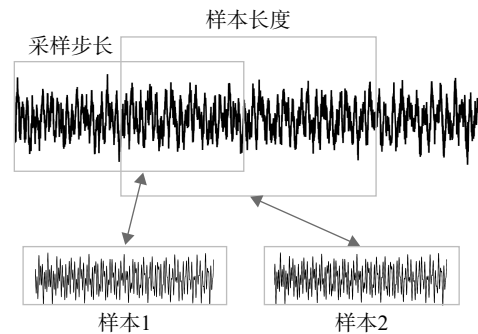


图 5 重叠采样示意图

Fig. 5 Schematic diagram of overlapping sampling

承振动数据^[22]。平台由交流电动机、电动机转速控制器、转轴、支撑轴承、液压加载系统和测试轴承等组成,测试对象为 LDK UER204 滚动轴承,采样频率为 25.6 kHz,故障类型选取外圈故障和内圈故障,转速为 2 250 r/min,施加的径向力为 11 kN。将凯斯西储大学数据集作为数据集 A,轴承寿命加速试验台数据作为数据集 B,样本统计情况如表 3 所示。

利用 MFMD 对截取得到的样本进行故障特征提取,其结果如图 7 所示。然后利用实体嵌入和位置编码使输入样本具备网络可识别的位置标签,如图 8 所示。

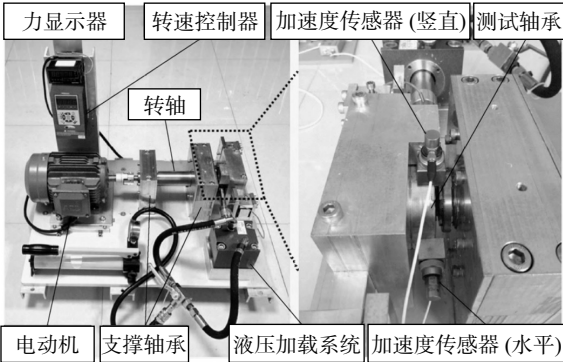


图 6 轴承加速寿命试验台
Fig. 6 Bearing accelerated life test bench

表 3 轴承样本统计
Table 3 Bearing sample statistics

数据集	负载	转速/(r/min)	故障尺寸/mm	故障类型	训练样本数量	测试样本数量
A	0.7 kW	1 772	0.177 8,	正常/内圈/ 外圈/滚动体	800	200
	1.4 kW	1 750	0.355 6,			
	2.1 kW	1 730	0.533 4			
B	11 kN	2 250		正常 内圈 外圈	400	100

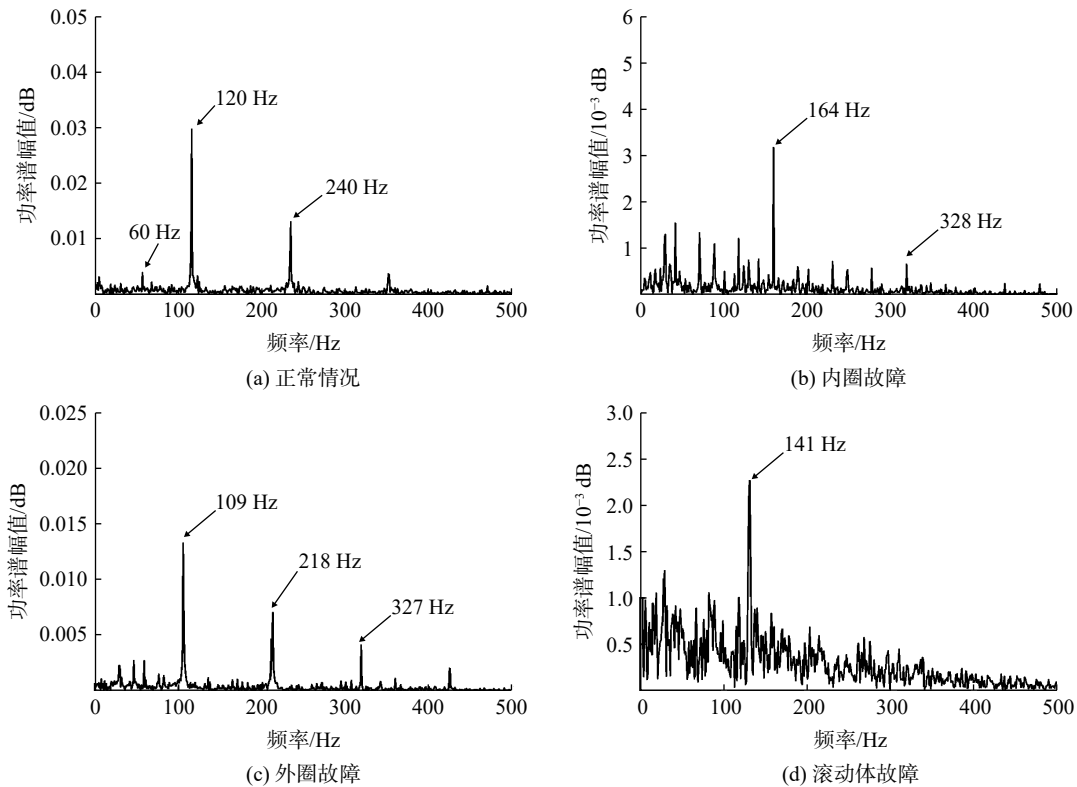


图 7 MFMD 提取的故障特征
Fig. 7 Fault feature extraction results based on MFMD

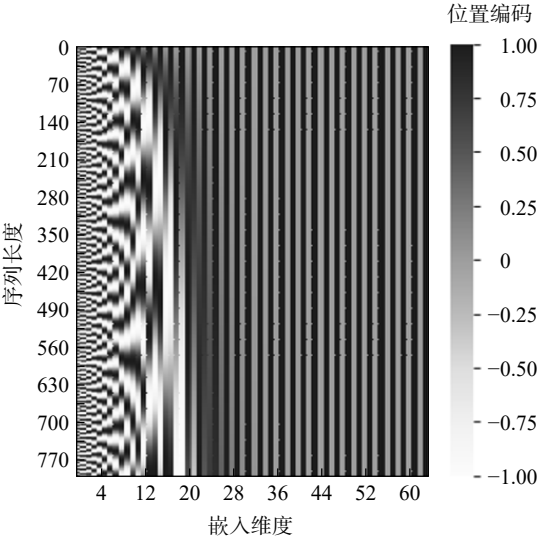


图 8 位置编码示意图
Fig. 8 Schematic diagram of position encoding

2.2 实验设置与结果

2.2.1 实验 1: Transformer-CNN 结构优化实验

编码器数量对 Transformer-CNN 算法的故障诊断效果具有最直接影响,编码器数量过少将降低网络序列建模能力,编码器过多将影响计算效率并存在过拟合风险。为避免网络层数过多导致计算效率降低和过拟合问题,本文以 2 个循环为例,第 1 个循环 1 个编码器,第 2 个循环 3 个编码器用 1+3 表示,同理设置 2+2 和 3+1。采用上述 3 种编码器数量配置进行实验,其他网络结构及参数保持一致。经过迭代训练,3 种方案的诊断准确率变化曲线如图 9 所示。为进一步分析编码器布置对 Transformer-CNN 算法性能的影响,将上述 3 个方案的准确率、平均耗时和参数总量进行统计,结果如表 4 所示。

2.2.2 实验 2: 算法对比实验

将 Transformer^[18]、CNN^[23] 和长短期记忆神经网络 (LSTM)^[24] 作为基准算法。基准算法中包含卷积神经网络、Transformer 和其他经典 RNN,可以较为直观地对比出网络效果。对比算法的结构参数设置如表 5 所示。

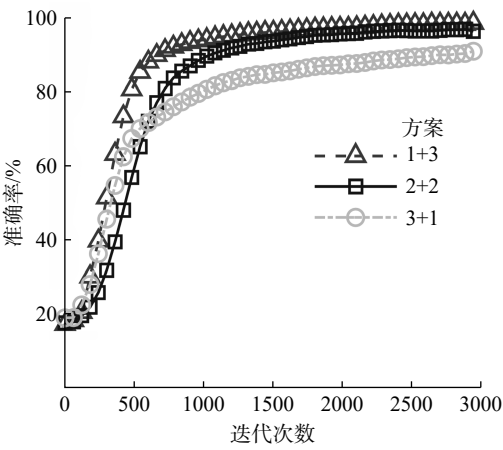


图 9 不同编码器设置的效果对比
Fig. 9 Comparison of the effects of different encoder settings

表 4 不同编码器设置结果
Table 4 Results of different encoder settings

参数	方案		
	1+3	2+2	3+1
准确率/%	99.15±2.15	97.15±2.60	90.55±2.10
平均耗时/s	14.73	15.56	16.16
参数总量	132 259	17 7 563	212 867

表 5 对比算法结构参数设置
Table 5 Contrast algorithm structure parameter settings

算法	设置
Transformer	实体嵌入(800, 64)→ 位置编码(800, 64)→ 编码器×4(800, 64)→线性层(10, 1)→Softmax
CNN	实体嵌入(800, 64)→ 卷积块(800, 64)→ 最大值池化(400, 64)→ 卷积&池化→线性层(10, 1)→Softmax
LSTM	实体嵌入(800, 64)→ LSTM(隐层节点个数为 1 024, 层数为 2)→ 池化层(2 114, 1)→ 线性层(10, 1)→Softmax

经过迭代训练,将最终诊断准确率、平均耗时、网络参数量进行统计,结果如表 6 所示。

表 6 不同算法的效果对比
Table 6 Effect comparison of different algorithms

网络类型	Transformer-CNN	Transformer	CNN	LSTM
准确率/%	99.15±2.15	98.15±2.20	97.45±3.10	97.20±3.15
平均耗时/s	14.73	16.77	21.04	16.23
参数量	132 259	241 415	229 927	255 687

引入故障混淆矩阵对诊断结果进行分析,其中0代表正常情况,1~3代表内圈故障,4~6代

表外圈故障,7~9代表滚动体故障,如图10所示。

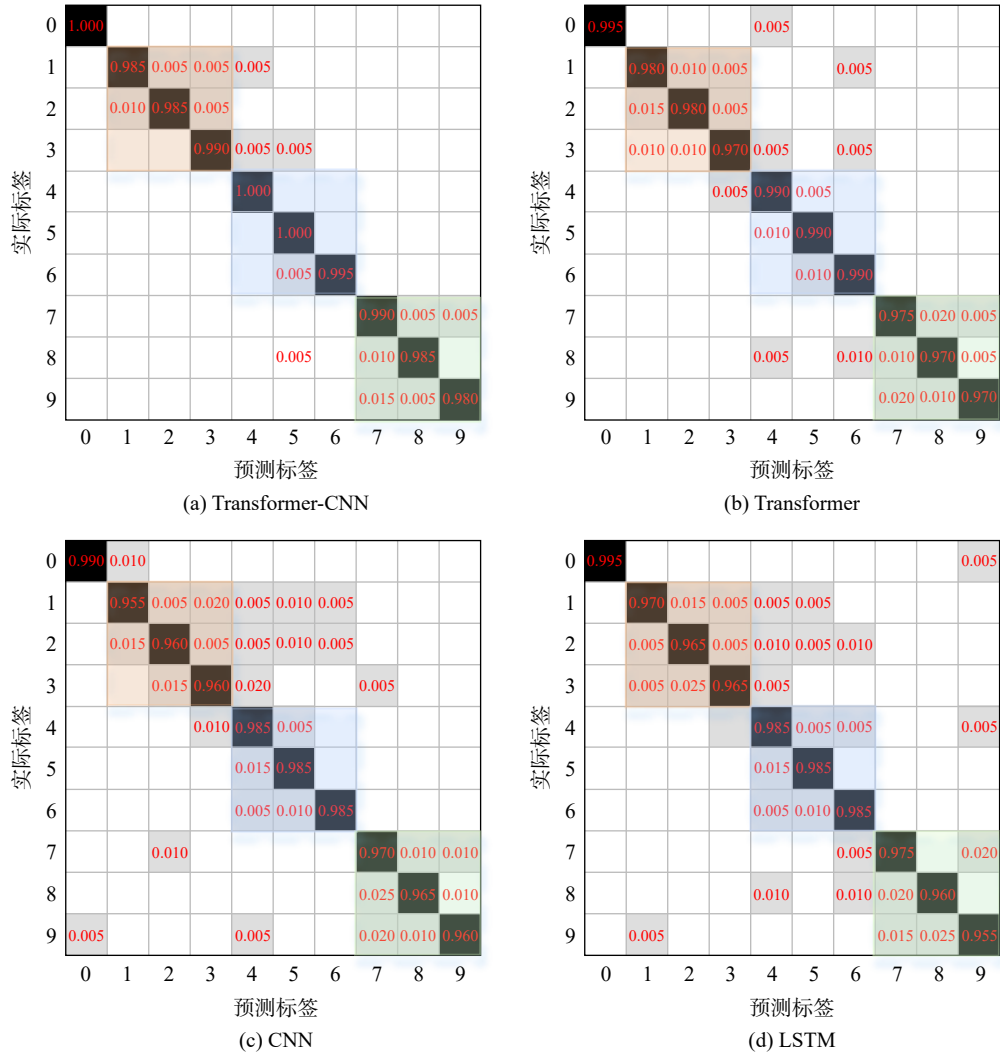


图10 故障分类混淆矩阵

Fig. 10 Fault classification confusion matrix

2.2.3 实验3: 变工况和跨型号小样本故障诊断

针对变工况和跨型号轴承故障诊断中数据样本不足的问题,提出了基于Transformer-CNN的迁移故障诊断算法。设计如表7所示的迁移场景,开展变工况和跨型号的小样本轴承故障诊断研究。为模拟小样本情况,将目标域样本数量设置为总数的5%、10%、15%和20%,样本数量设置符合小样本要求且可以反映出准确率随样本数量的变化趋势,样本选择采用随机抽取的方式。因此,该部分共开展7种场景、28种条件下的轴承故障诊断实验。

在数据集A上开展变工况的迁移学习研究。采用一种工况作为源域,其他工况作为目标域,数据处理流程保持一致。为验证算法效果,利用

重新训练的结果为对比。源域各类样本数量充足,用于诊断模型预训练,最终达到平均准确率为 $(99.15 \pm 2.15)\%$ 。每种情况训练5次然后输出,结

表7 迁移任务的总结

Table 7 Summary of transfer tasks

编号	场景	迁移任务	迁移条件
1	变工况	工况1→工况2	0.7→1.4 kW
2		工况1→工况3	0.7→2.1 kW
3		工况2→工况3	1.4→2.1 kW
4		工况3→工况1	2.1→0.7 kW
5		工况3→工况2	2.1→1.4 kW
6		工况2→工况1	1.4→0.7 kW
7	跨型号	轴承A→轴承B	A→B
8		轴承B→轴承A	B→A

果如表 8 所示。其中,利用 10% 数据样本的迁移学习和重新训练准确率变化如图 11 所示。

利用数据集 A 和数据集 B 开展跨型号轴承故障诊断。每种情况训练 5 次然后输出。为对比迁移学习效果,利用数据集 B 中数据样本重新训练模型,最终故障诊断的结果如图 12 所示。

表 8 变工况迁移学习故障分类结果

Table 8 Fault classification results of transfer learning under variable working conditions

迁移场景	准确率/%			
	5% 样本	10% 样本	15% 样本	20% 样本
1	90.00±1.50	97.25±1.75	98.65±1.60	99.00±1.25
2	90.50±1.50	97.50±1.30	98.30±1.25	99.05±1.15
3	90.25±1.35	97.15±1.50	98.55±1.30	98.95±1.40
4	89.75±1.35	97.10±1.75	98.50±1.40	99.15±1.50
5	91.50±1.50	97.55±1.75	98.05±1.65	99.05±1.20
6	90.20±1.35	97.25±1.65	98.35±1.80	99.10±1.65
场景 1 重新训练	81.25±2.25	92.50±2.35	94.75±1.75	95.65±1.50
迁移学习准确率增幅	8.75	4.75	3.90	3.35

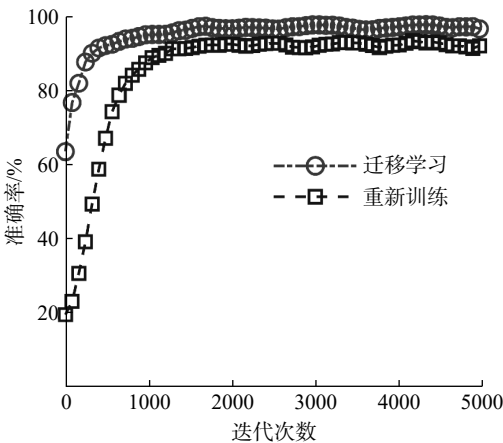


图 11 变工况迁移学习和重新训练对比

Fig. 11 Comparison between cross-working transfer learning and retraining

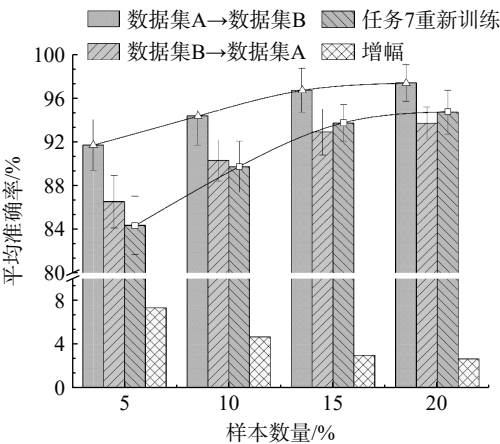


图 12 跨型号迁移学习和重新训练效果对比

Fig. 12 Effect comparison between cross-type transfer learning and retraining

2.3 实验分析

实验 1 结果表明: 方案 3+1 虽然收敛速度较快,但最终准确率较低;方案 2+2 最终准确率比较理想,但收敛速度较慢;方案 1+3 在收敛速度和准确率上都拥有最优效果。由表 4 可知,方案 1+3 相较于其他两种方案在参数总量方面具有较大优势,平均运算时间也由此缩短。从网络结构分析,1+3 方案的 Transformer-CNN 网络结构呈现金字塔型,可以一定程度上避免反向传播时梯度下降过快的问题。由此证明:编码器布置在采用先少后多的方案时具有较好效果。

实验 2 结果表明,Transformer-CNN 相较于基准算法 Transformer、CNN 和 LSTM 分别降低了 54%、66% 和 69% 的识别错误率,具有较高的轴承故障诊断可靠性。由于池化层的降采样功能,Transformer-CNN 结构总体呈金字塔型,其参数总量相较于基准算法具有较大优势。参数总量的变化最终反映到平均运算时间上,相较于 Transformer、CNN 和 LSTM 分别提升了 12%、30% 和 9% 的运算效率。由混淆矩阵结果可知,基于 Transformer-CNN 的轴承故障诊断方法具有较高的准确性和稳定性。同时由统计结果可知外圈和正常情况诊断效果较好,滚动体和内圈故障诊断较差。从传递路径分析,内圈和滚动体故障引起的冲击振动传播至传感器的过程中经历了更多调制和转换过程,因此其信号变化较为复杂。其次从故障机理分析,滚动体与滚道之间可能会存在

滑动现象, 因此其故障特征较为复杂^[25]。

实验 3 在变工况故障诊断实验中, 由表 8 可知, 所提算法利用 5% 目标域数据样本可达到 90% 左右的平均分类准确率, 利用 10% 数据样本即可达到 97.10% 以上的平均诊断准确率, 基本可以满足故障诊断需求。由图 11 可知, 基于 Transformer-CNN 的迁移学习方法与重新训练相比较, 既可以大幅度提升诊断准确率, 也可以大幅度加快神经网络学习的收敛速度。

在跨型号轴承故障诊断实验中, 所提算法同样可以大幅提升平均分类准确率。同时, 场景 7 的识别准确率远远高于场景 8, 这是由于数据集 A 的故障数据更加完备和充足。

2.4 结果可视化

为直观展现本文算法的故障诊断效果, 引入 t 分布随机邻域嵌入 (t -distributed stochastic neighbor embedding, t -SNE) 分析分别对 CWRU 数据集样本的原始数据、预处理数据和算法输出结果进行可视化处理, 结果如图 13 所示。其中算法输出结果采用的是 Transformer-CNN 单元的输出数据, 0~9 表示 10 类故障状态。

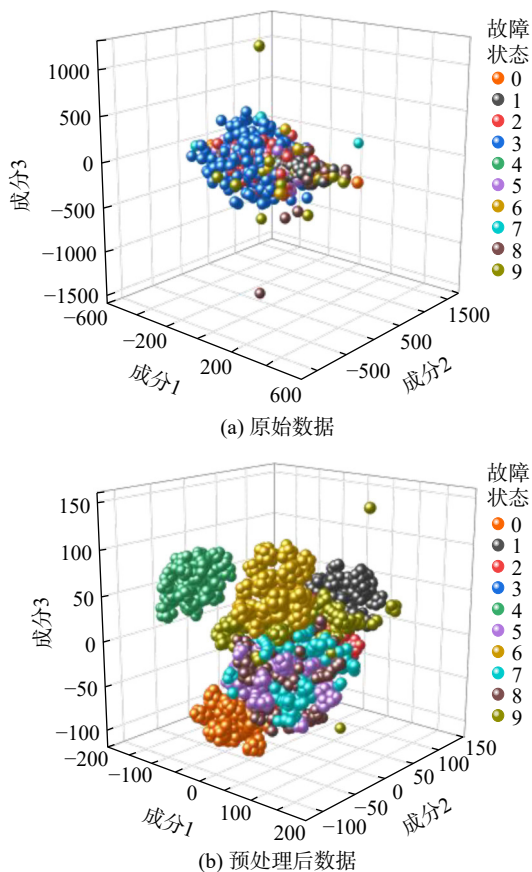


图 13 t -SNE 可视化分析

Fig. 13 Visualization analysis based on t -SNE

结果表明: 经过预处理, 各类故障数据从混乱状态改善为聚集状态, 但各类别数据之间混叠严重, 经过算法处理以后, 各类别故障数据分割距离较大, 分类状态独立清晰。由此证明, 本文算法可以有效实现变工况和跨型号的小样本滚动轴承故障诊断, 并且基于 MFMD 的预处理模块具有较高可行性。

3 结论

本文针对滚动轴承故障样本严重不足和振动信号序列特征提取困难的问题, 利用 MFMD 建立了数据预处理模块, 提出了一种基于 Transformer-CNN 和迁移学习的滚动轴承故障诊断算法, 并开展了变工况和跨型号的迁移学习故障诊断研究, 得到以下结论:

1) 本文所提诊断模型采用金字塔型结构, 可以一定程度上避免反向传播时梯度下降过快的问题。利用 CWRU 数据集开展 Transformer-CNN、Transformer、CNN 和 LSTM 的轴承故障诊断性能对比实验。结果表明: 本文算法可以降低 50% 以上的故障诊断错误率, 同时具有更高的运行效率。通过混淆矩阵分析可知, 该算法也具有较稳定的稳定性和可靠性。

2) 开展了基于 Transformer-CNN 迁移学习的变工况和跨型号轴承故障诊断研究。利用 CWRU 轴承数据集和轴承加速寿命试验数据集进行实验验证, 结果表明: 在变工况和跨型号条件下, 本文迁移学习算法利用小样本数据可以提升诊断准确率, 同时可以提升收敛速度, 具有一定的泛化能力。

3) 本文所使用神经网络模型的输入尺寸、嵌入维度、网络层数和激活函数类型等均参考相关学者研究成果由经验设定。由于这些模型参数对

辨识准确性和计算复杂度都有一定影响,所以合理的模型参数设计问题仍然是下一步的研究重点。同时,可不断收集实测故障数据对算法进行训练,提升其在实际场景中的应用效果。

参考文献:

- [1] 张向阳,陈果,郝腾飞,等.基于机匣信号的滚动轴承故障卷积神经网络诊断方法[J].航空动力学报,2019,34(12): 2729-2737.
ZHANG Xiangyang, CHEN Guo, HAO Tengfei, et al. Rolling bearing fault convolution neural network diagnosis method based on gearbox signal[J]. Journal of Aerospace Power, 2019, 34 (12): 2729-2737. (in Chinese)
- [2] SINGH P, JOSHI S D, PATNEY R K, et al. The Fourier decomposition method for nonlinear and nonstationary time series analysis[J]. Proceedings Mathematical Physical and Engineering Sciences, 2015, 473(2199): 871-0897.
- [3] 郑近德,潘海洋,程军圣,等.基于自适应经验傅里叶分解的机械故障诊断方法[J].机械工程学报,2020,56(9): 125-136.
ZHENG Jinde, PAN Haiyang, CHENG Junsheng, et al. Mechanical fault diagnosis method based on adaptive empirical Fourier decomposition[J]. Journal of Mechanical Engineering, 2020, 56(9): 125-136. (in Chinese)
- [4] CHENG Yiwei, LIN Manxi, WU Jun, et al. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network[J]. Knowledge-Based Systems, 2021, 216: 106796.1-106796.12.
- [5] LI Jimeng, CHENG Xing, LI Qiang, et al. Adaptive energy-constrained variational mode decomposition based on spectrum segmentation and its application in fault detection of rolling bearing[J]. Signal Processing, 2021, 183: 108025.1-108025.17.
- [6] 雷亚国,贾峰,孔德同,等.大数据下机械智能故障诊断的机遇与挑战[J].机械工程学报,2018,54(5): 94-104.
LEI Yaguo, JIA Feng, KONG Detong, et al. Opportunity and challenge of mechanical intelligent fault diagnosis under big data[J]. Journal of Mechanical Engineering, 2018, 54(5): 94-104. (in Chinese)
- [7] 张西宁,郭清林,刘书语.深度学习技术及其故障诊断应用分析与展望[J].西安交通大学学报,2020,54(12): 1-13.
ZHANG Xining, GUO Qinglin, LIU Shuyu. Deep learning technology and its application analysis and prospect of fault diagnosis[J]. Journal of Xi'an Jiaotong University, 2020, 54(12): 1-13. (in Chinese)
- [8] WEISS K, KHOSHGOFTAAR T M, WANG D D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3(1): 1-40.
- [9] NITISH S, GEOFFREY E. HINTON, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [10] LI Xiang, ZHANG Wei, DING Qian. Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks[J]. IEEE Transactions on Industrial Electronics, 2019, 66(7): 5525-5534.
- [11] 张西宁,余迪,刘书语.基于迁移学习的小样本轴承故障诊断方法研究[J].西安交通大学学报,2021,55(10): 30-37.
ZHANG Xining, YU Di, LIU Shuyu. Research on small sample bearing fault diagnosis method based on transfer learning[J]. Journal of Xi'an Jiaotong University, 2021, 55(10): 30-37. (in Chinese)
- [12] ZHANG Kaiyu, CHEN Jinglong, ZHANG Tianci, et al. A compact convolutional neural network augmented with multiscale feature extraction of acquired monitoring data for mechanical intelligent fault diagnosis[J]. Journal of Manufacturing Systems, 2020, 55: 273-284.
- [13] ZHU J, CHEN N, SHEN C. A new deep transfer learning method for bearing fault diagnosis under different working conditions[J]. IEEE Sensors Journal, 2019, 20(15): 8394-8402.
- [14] 何强,唐向红,李传江,等.负载不平衡下小样本数据的轴承故障诊断[J].中国机械工程,2021,32(10): 1164-1171,1180.
HE Qiang, TANG Xianghong, LI Chuanjiang, et al. Bearing fault diagnosis of small sample data under unbalanced load[J]. China Mechanical Engineering, 2021, 32(10): 1164-1171,1180. (in Chinese)
- [15] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [16] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. [2021-08-20]. <https://arxiv.org/abs/1412.3555>.
- [17] 周奇才,沈鹤鸿,赵炯,等.基于改进堆叠式循环神经网络的轴承故障诊断[J].同济大学学报(自然科学版),2019,47(10): 1500-1507.
ZHOU Qicai, SHEN Hehong, ZHAO Jiong, et al. Bearing fault diagnosis based on improved stackable cyclic neural network[J]. Journal of Tongji University (Natural Science Edition), 2019, 47(10): 1500-1507. (in Chinese)
- [18] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[EB/OL]. [2021-06-20]. <https://arxiv.org/abs/1706.03762>.
- [19] LIU Junfeng, YU Xiang. Rolling element bearing fault diagnosis for complex equipment based on MFMD and BP neural network[J]. Journal of Physics Conference Series, 2021, 1948(1): 012113.1-012113.10.
- [20] LIU Tao, CHEN Jin, DONG Guangming, et al. The fault detection and diagnosis in rolling element bearings using frequency band entropy[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2013, 227(1): 87-99.
- [21] KIM Y, DENTON C, HOANG L, et al. Structured attention networks[EB/OL]. [2021-06-20]. <https://arxiv.org/abs/1702.00887>.
- [22] WANG Biao, LEI Yaguo, LI Naipeng, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. IEEE Transactions on Reliability, 2018, 69(1): 401-412.
- [23] JOHNSON R, ZHANG T. Convolutional neural networks for text categorization: shallow word-level vs. deep character-level[EB/OL]. [2021-06-20]. <https://arxiv.org/abs/1609.00718>.
- [24] LA S W, XU L H, LIU K, et al. Recurrent convolutional neural networks for text classification[C]// Twenty-ninth AAAI conference on Artificial Intelligence. Menlo Park, US: AIAA, 2015: 2267-2273.
- [25] 赵海峰,杨国斌.基于振动信号的深沟球轴承滚动体故障诊断研究[J].化工机械,2014,41(3): 291-295.
ZHAO Haifeng, YANG Guobin. Research on rolling body fault diagnosis of deep groove ball bearing based on vibration signal[J]. Chemical Machinery, 2014, 41(3): 291-295. (in Chinese)

(编辑:陈越)