

蜉蝣优化双通道网络在齿轮箱故障诊断中的应用

于宁, 魏沉潜, 田立勇, 赵建军, 于晓涵

(辽宁工程技术大学机械工程学院, 123000, 辽宁阜新)

摘要: 为了有效提取齿轮箱中齿轮和滚动轴承的故障信号特征, 并克服深度学习模型超参数选取依赖人工经验的局限性, 提高故障诊断的准确性和稳定性, 提出了一种基于蜉蝣算法(MA)优化的双通道神经网络故障诊断模型。该模型采用一维时序输入的门控循环单元(GRU)和二维图像输入的卷积神经网络(CNN)构建双通道并行架构, 并引入自适应批标准化(AdaBN)算法。利用 MA 的全局优化能力, 以 CNN-GRU 的诊断精度为优化目标, 自适应调整模型超参数。将蜉蝣算法优化效果与粒子群算法和遗传算法进行了对比验证, 以评估其在模型参数优化方面的有效性。基于东南大学齿轮箱数据集和凯斯西储大学轴承数据集的实验结果表明: 该模型能够有效提取振动信号特征, 其故障识别精度与稳定性均优于典型深度学习模型, 并展现出较强的鲁棒性。在稳态工况下, 优化后的 CNN-GRU(MA-CNN-GRU)模型在各数据集上的识别精度显著提高; 在噪声工况下, MA 优化的 CNN-GRU 模型表现出优异的抗噪性; 在变负载工况下, 结合 AdaBN 算法的 MA-CNN-GRU 模型实现了最高的平均识别精度。所提模型能够高效、准确地检测齿轮箱故障, 为机械设备的维护和稳定运行提供了重要的参考价值。

关键词: 故障诊断; 齿轮箱; 门控循环单元; 卷积神经网络; 蜉蝣算法

中图分类号: TH133.3 **文献标志码:** A

DOI: 10.7652/xjtuxb202505021 **文章编号:** 0253-987X(2025)05-0217-12

Application of Mayfly Algorithm Optimized Dual-Channel Network in Gearbox Fault Diagnosis

YU Ning, WEI Chenqian, TIAN Liyong, ZHAO Jianjun, YU Xiaohan

(College of Mechanical Engineering, Liaoning Technical University, Fuxin, Liaoning 123000, China)

Abstract: In order to effectively extract the fault signal characteristics of gears and rolling bearings in gearboxes, overcome the limitation of deep learning model hyperparameter selection relying on human experience, and improve the accuracy and stability of fault diagnosis, a two-channel neural network fault diagnosis model based on Mayfly algorithm(MA)optimization was proposed. The model uses a one-dimensional timing input gated recurrent unit (GRU) and a two-dimensional image input convolutional neural network (CNN) to construct a dual-channel parallel architecture, and introduces an adaptive batch normalization (AdaBN) algorithm. By using the global optimization ability of MA, the hyperparameters of the model are adaptively adjusted with the diagnostic accuracy of CNN-GRU as the optimization objective. At the same time, MA optimization effort is compared with particle swarm optimization and genetic algorithm to evaluate its effectiveness in model parameter optimization. The experimental results based on the gearbox data set of Southeast

University and the bearing data set of Case Western Reserve University show that the model can effectively extract the characteristics of vibration signals, and its fault recognition accuracy and stability are better than typical deep learning models, and it shows strong robustness. Under steady-state conditions, the recognition accuracy of the optimized CNN-GRU model on each data set is significantly improved. Under noise conditions, the MA-optimized CNN-GRU (MA-CNN-GRU) model exhibits excellent noise immunity. Under variable load conditions, the MA-CNN-GRU model combined with the AdaBN algorithm achieves the highest average recognition accuracy. In summary, the proposed model can detect gearbox faults efficiently and accurately, and it can provide an important reference for the maintenance and stable operation of mechanical equipment.

Keywords: fault diagnosis; gearbox; gated recurrent unit; convolutional neural network; mayfly algorithm

齿轮箱作为机械设备的关键传动部件,其运行的安全稳定性,对工业生产和人身安全至关重要^[1-3]。由于其复杂的工作环境以及长期时变、高负荷运转,齿轮和轴承易出现故障,造成经济损失甚至引发严重的人员伤亡事故,因此准确、有效地识别并排除齿轮箱的早期故障,减少事故发生概率具有重要意义^[4-6]。

随着人工智能迅速发展,机器学习和深度学习算法广泛应用于旋转机械故障诊断中,并且效果显著^[7-9],卷积神经网络是其中性能优越的典型代表^[10-13]。Chen等采用时域统计量对齿轮箱振动信号进行预处理,将256个统计特征转换为尺寸为16的特征矩阵,并作为卷积神经网络(CNN)的输入进行故障诊断,与其他算法相比,该算法优势明显,性能突出^[14]。宫文峰等将全连接层替换为过渡卷积层与全局平均池化层,从而减少网络的参数量,并将改进的CNN结合支持向量机(SVM)应用于电机支撑滚珠轴承的故障诊断中^[15]。Monteiro等以傅里叶变换处理齿轮箱振动信号数据作为CNN的输入数据集进行故障诊断,对齿轮箱断齿故障有着良好的检测效果^[16]。王贡献等提出了一种多分支卷积神经网络,通过多尺度均值化实现振动信号重构,并利用多分支卷积神经网络(MBCNN)的并行分支提取故障特征^[17]。叶壮等提出了多通道一维卷积神经网络模型,相较于传统的一维CNN,故障识别精度有一定的提升^[18]。Zhang等提出了一维宽卷积核模型,在噪声环境下故障诊断效果突出^[19]。

CNN通常采用单通道多层卷积结构,这种结构复杂且超参数众多^[20]。部分超参数对模型的影响较小,但有些关键超参数对分类效果具有决定性的作用,仅依靠经验或传统的启发式算法来设定这些关键参数,容易导致模型陷入局部最优解,从而影响网络输出的确定性,单通道网络结构在提取故障特征时存在局限性,无法全面准确地表征故障信号的

特征,这对识别精度产生了明显的影响^[21]。在实际工况下,工作载荷并非稳定不变,而是会随着生产条件的变化而变化,导致数据集的分布出现差异,在此情形下,如何保持故障诊断模型的高性能和高效率,成为一个亟待解决的重要问题。

基于以上研究,提出了一种基于蜉蝣算法(MA)优化的双通道并行故障诊断模型,用于提取齿轮箱振动信号并识别故障类型,提升故障诊断的准确性和稳定性。针对实际工作过程中,工作载荷随生产条件变化导致数据集分布出现差异的情况,引入自适应批标准化(AdaBN)算法以代替传统的批量标准化层(BN),使得故障信号在不同负载下也能得到有效识别,极大地提高模型的故障诊断精度,具备重要的实际应用价值^[22]。

1 齿轮箱故障诊断算法

1.1 卷积神经网络

CNN是一种以仿生视觉机制构建的深度学习算法,主要用于图像识别和数据处理领域^[23]。在分类任务中,卷积和池化运算交替进行,有效提取高维非线性特征,捕捉并处理局部信息相关性,通过全连接层的非线性变换,将这些特征整合并最终实现精确的分类。

卷积层的主要作用是特征提取,输入图像或特征图通过卷积核处理对应区域,并将特征经过激活函数映射形成新的卷积特征图,表达式如下

$$\mathbf{M}_j^l = f^l \left(\sum_{i=1}^P \mathbf{M}_i^{l-1} \otimes \mathbf{W}_{ij}^l + \mathbf{b}_i^l \right) \quad (1)$$

式中: l 为网络深度; $\mathbf{M}_j^l$ 为 l 层第 j 个神经元特征值; $\mathbf{W}_i^{l-1}$ 为 $l-1$ 层第 i 个神经元的权值矩阵; P 为神经元数; $\mathbf{W}_{ij}^l$ 为 l 层第 i 个神经元和第 j 个神经元之间的权值矩阵; $\mathbf{b}_i^l$ 为第 l 层第 i 个神经元的偏置; $f^l(\cdot)$ 为 l 层激活函数。

池化层一般位于卷积层之后,通过池化操作,在保留重要信息的同时,能够极大地压缩数据,降低计算复杂度并加快神经网络计算速度。本文采用最大池化操作,表达式如下

$$\mathbf{M}_j^t = \text{poolmax}(\sum_{j=1}^p \mathbf{M}_j^{t-1}) \quad (2)$$

式中:poolmax(·)为最大池化函数。

全连接层作用是将输入特征映射到输出预测结果,表达式如下

$$\mathbf{M}_j^t = f'(W_{ij}^t \mathbf{M}_i^t + \mathbf{b}_i^t) \quad (3)$$

1.2 门控循环单元神经网络

对于时间序列处理,长短期记忆网络(LSTM)能够解决传统循环神经网络(RNN)存在的问题,但是LSTM存在复杂的门控机制和多个权值参数,导致计算量较大。相比于LSTM网络,门控循环单元(GRU)结构更加简单,采用更少的门控单元,能够大幅减少训练参数,提高求解速度^[24]。GRU有两个门,一个是更新门,另一个是重置门,可以很好地解决长间隔和长延迟时间序列的建模问题,故采用GRU网络作为时序数据特征提取通道。

更新门控制了上一个时间步的信息对当前状态的贡献程度,决定了在新的隐藏状态中保留哪些过去的信息,更新门输出为

$$\mathbf{z}_t = \sigma(\mathbf{w}_{xz} \mathbf{p}_t + \mathbf{w}_{hz} \mathbf{h}_{t-1} + \mathbf{b}_z) \quad (4)$$

式中:σ(·)表示激活函数;w_{xz}表示更新门输入层与隐藏层之间的权值矩阵;w_{hz}表示更新门上一隐藏层与当前隐藏层之间的权值矩阵;b_z表示更新门偏置项;p_t表示当前时刻输入;h_{t-1}表示上一时刻隐藏层状态。

重置门控制了是否忽略前一个时间步的信息,本质上,重置门被模型用来决定忘记多少过去的信息。重置门输出为

$$\mathbf{r}_t = \sigma(\mathbf{w}_{xr} \mathbf{p}_t + \mathbf{w}_{hr} \mathbf{h}_{t-1} + \mathbf{b}_r) \quad (5)$$

式中:w_{xr}表示重置门输入层与隐藏层之间的权值矩阵;w_{hr}表示重置门上一个隐藏层与当前隐藏层之间的权值矩阵;b_r表示重置门偏置项。

1.3 蜉蝣算法

蜉蝣算法是Konstantinos等提出的仿生优化算法,用于解决复杂函数优化问题^[25]。该算法模拟了蜉蝣飞行及繁衍后代等社会行为,通过多个并行子空间搜索,搜索效率高,并且根据现有位置判断移动方向,避免过早收敛并陷入局部最优,同时平衡局部搜索和全局搜索能力,其工作原理如下。

以两组蜉蝣代表雄性和雌性种群,定义蜉蝣位置为 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D)$,速度为 $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ 。

(1)雄性蜉蝣寻优。雄性蜉蝣通常在固定区域内并产生群聚行为,会成群结队向更优方向进化,以自身情况及种群行为变化来调整位置。设 $\mathbf{x}_i^t$ 表示 t 时刻雄性蜉蝣 i 的空间位置,其位置更新定义为

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1} \quad (6)$$

式中,速度更新定义为

$$\mathbf{v}_{ij}^{t+1} = \begin{cases} g\mathbf{v}_{ij}^t + a_1 e^{-\beta r_p^2} (\mathbf{b}_{ij} - \mathbf{x}_{ij}^t) + a_2 e^{-\beta r_g^2} (\mathbf{q}_j - \mathbf{x}_{ij}^t), & f(\mathbf{x}_i) > f(\mathbf{q}) \\ g\mathbf{v}_{ij}^t + d\mathbf{r}, & f(\mathbf{x}_i) \leq f(\mathbf{q}) \end{cases} \quad (7)$$

其中, $\mathbf{x}_{ij}^t, \mathbf{v}_{ij}^t$ 为 j 维 t 时刻蜉蝣 i 的位置和速度, g 为重力系数, a_1, a_2 分别为用于缩放认知和社会部分的贡献正的吸引系数; β 为固定可见度系数, d 为婚舞距离系数, $\mathbf{r}$ 为 $[-1, 1]$ 之间的随机矩阵, $\mathbf{b}_{ij}$ 为 j 维蜉蝣 i 的最优位置, $\mathbf{q}_j$ 为 j 维蜉蝣种群的历史最优位置, r_p, r_g 分别为 $\mathbf{x}_i$ 与蜉蝣个体历史最优位置 $\mathbf{b}$ 、 $\mathbf{x}_i$ 与蜉蝣种群的全局最优位置 $\mathbf{q}$ 之间的笛卡尔距离,表达式为

$$\|\mathbf{x}_i - \mathbf{q}_i\| = \sqrt{\sum_{j=1}^n (\mathbf{x}_{ij} - \mathbf{q}_{ij})^2} \quad (8)$$

式中: $\mathbf{q}_i$ 为蜉游 i 的最优位置; $\mathbf{q}_{ij}$ 为蜉游 i 在 j 维上的最优位置。

(2)雌性蜉蝣寻优。雌性蜉蝣逐渐靠近适应度值更高的雄性蜉蝣,设 $\mathbf{y}_i^t$ 表示 t 时刻雌性蜉蝣 i 的空间位置,其位置更新定义为

$$\mathbf{y}_i^{t+1} = \mathbf{y}_i^t + \mathbf{v}_i^{t+1} \quad (9)$$

$$\mathbf{v}_{ij}^{t+1} = \begin{cases} g\mathbf{v}_{ij}^t + a_2 e^{-\beta r_{mf}^2} (\mathbf{x}_{ij}^t - \mathbf{y}_{ij}^t), & f(\mathbf{y}_i) > f(\mathbf{x}_i) \\ g\mathbf{v}_{ij}^t + F_L \mathbf{r}, & f(\mathbf{y}_i) \leq f(\mathbf{x}_i) \end{cases} \quad (10)$$

式中: $\mathbf{y}_{ij}^t$ 为 j 维 t 时刻蜉蝣 i 的位置; r_{mf} 为蜉蝣对的笛卡尔距离; F_L 为飞行系数。

(3)蜉蝣交配。蜉蝣以最佳匹配机制交配产生后代,即最优雄性与最优雌性吸引,次优雄性与次优雌性吸引,依次类推,交配产生两个子代为

$$\begin{cases} \mathbf{S}_1 = L\mathbf{F}_m + (1-L)\mathbf{F}_f \\ \mathbf{S}_2 = L\mathbf{F}_f + (1-L)\mathbf{F}_m \end{cases} \quad (11)$$

式中: $\mathbf{F}_m$ 表示父本; $\mathbf{F}_f$ 表示母本; L 为 $[0, 1]$ 内的随机数。

2 MA优化的CNN-GRU(MA-CNN-GRU)模型

2.1 信号预处理

采用变分模态分解(VMD)和连续小波变换(CWT)^[26]进行信号预处理工作,其中将VMD降噪

重构后的时序信号作为时序输入,将 CWT 后的时频图作为图像输入,VMD 能够有效应对模态混叠现象,减少噪声对信号的影响,保证信号在噪声干扰下依旧能提取到故障信号特征,CWT 糅合了时域和频域分析的优势,可把信号分解为不同尺度(频率)的小波进行多频带划分,能够捕捉到信号中的局部变化,分析信号的瞬时特性和多尺度特性,能够兼顾时间特性和频率特性,对复杂多变信号分析效果较好。

2.2 CNN-GRU 结构优化

2.2.1 模型设计

首先通过探究不同结构模型对故障诊断精度的影响,来选择最合适的神经网络模型进行下一步研

究,本实验通过凯斯西储大学轴承数据集进行验证,图像通道每层卷积池化步长均相同,除了卷积层数外,其他训练参数保持一致,时序通道仅改变循环层数,末层神经元数为 128,其他层神经元数为 64,每层丢弃率均保持一致,并按照常规经验设定,其中设定模型学习率为 0.000 1,最大迭代次数为 35,每组进行 5 次实验,取其中训练最优结果作为最终结果。表 1 设计了不同网络层数模型结构,以 $C\times 3+G\times 2$ 为例,表示模型由两个通道组成,其中 2D(图像通道)有 3 个卷积层,1D(时序通道)有两个 GRU 层, K_s 表示 2D 通道卷积核尺寸。 K_c 表示 2D 通道卷积核数, K_g 表示 1D 通道神经元数, K_d 表示神经元丢弃率。

表 1 不同网络层数模型结构参数
Table 1 Structural parameters of different network layer models

模型结构	通道参数	
	2D(CNN)	1D(GRU)
$C\times 1+G\times 1$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=128,K_d=0.5$
$C\times 2+G\times 2$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=128,K_d=0.5$
$C\times 3+G\times 2$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=128,K_d=0.5$
	第 3 层: $K_s=(3,3),K_c=64$	
$C\times 2+G\times 3$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=64,K_d=0.5$
		第 3 层: $K_g=128,K_d=0.5$
$C\times 3+G\times 3$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=64,K_d=0.5$
	第 3 层: $K_s=(3,3),K_c=64$	第 3 层: $K_g=128,K_d=0.5$
$C\times 3+G\times 4$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=64,K_d=0.5$
	第 3 层: $K_s=(3,3),K_c=64$	第 3 层: $K_g=64,K_d=0.5$
		第 4 层: $K_g=128,K_d=0.5$
$C\times 3+G\times 5$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=64,K_d=0.5$
	第 3 层: $K_s=(3,3),K_c=64$	第 3 层: $K_g=64,K_d=0.5$
		第 4 层: $K_g=64,K_d=0.5$
		第 5 层: $K_g=128,K_d=0.5$
$C\times 4+G\times 3$	第 1 层: $K_s=(3,3),K_c=32$	第 1 层: $K_g=64,K_d=0.5$
	第 2 层: $K_s=(3,3),K_c=64$	第 2 层: $K_g=64,K_d=0.5$
	第 3 层: $K_s=(3,3),K_c=64$	第 3 层: $K_g=128,K_d=0.5$
	第 4 层: $K_s=(3,3),K_c=64$	

不同模型结构识别精度对比实验结果如图 1 所示,当 2D 通道为 1 层、1D 通道为 1 层时,模型识别精度最低,训练过程存在较大波动,随着网络深度增加,识别精度也在不断增加。2D 通道为 3 层、1D 通道为 3 层的模型结构时,识别精度达到最大。与 2D 通道为 3 层、1D 通道为 4 层模型结构时,识别效果相当,由于模型深度增加,计算量增大,因此采用一个 3 层 CNN 通道和一个 3 层 GRU 通道组成双通道模型,并在此基础上优化。

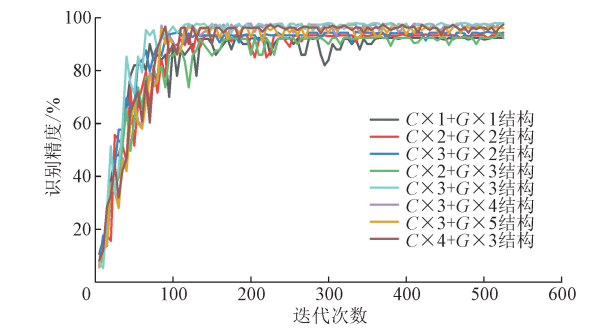


图 1 不同模型结构识别精度的对比结果

Fig.1 Comparison of accurate accuracy of different model structure recognition

采用 AdaBN 算法对 BN 层进一步优化,当训练样本和测试样本分布不一致时,采用 BN 算法会使网络模型性能下降,AdaBN 算法则会将测试样本的均值和方差替换为训练阶段中通过训练样本得到的均值和方差,即模型在测试集中只进行正向传播,而不进行反向传播。

2.2.2 实验参数设定

由于模型中超参数众多,综合考虑优化模型时间成本以及优化效果,对网络结构影响较小的超参数按照经验设定。设定 2D 通道首层卷积核数为 32,第 2、3 层为 64,设定卷积步长为 2×2 ,池化窗口大小为 2×2 ,池化步长为 2×2 ,第 3 层卷积无池化操作,设定 1D 通道第 3 层神经元数为 128,1D 通道和 2D 通道提取的信号特征经特征融合后连接,通过 softmax 分类器进行分类。设定优化算法种群数为 20,迭代次数为 10,其中由于蜉蝣算法收敛速度快,故种群数与迭代次数相对较少,为了平衡局部搜索能力和全局搜索能力,设定子代变异率为 0.4,激活函数类型为 Sigmoid、Tanh、Relu、LeakyRelu,优化器类型为 SGD、RMSprop、Adam。需要进行寻优的超参数及寻优范围见表 2。

表 2 寻优的超参数及寻优范围	
Table 2 Hyperparameters and optimization range for optimization	
超参数	取值范围
卷积核大小	$3\times 3,4\times 4,5\times 5$
时序隐藏神经元数	16~128
学习率	0.000 01~0.010 00
最小批量	32~256
丢弃率	0.2~0.6

2.3 MA-CNN-GRU 模型的流程

本文提出的 MA-CNN-GRU 模型包括两部分:通过 MA 优化 CNN-GRU 参数来搜寻 CNN-GRU 最优参数组合;提取训练好超参数配置模型进行测试。构建 MA-CNN-GRU 模型的具体步骤如下。

步骤 1 将 CNN-GRU 的识别精度作为蜉蝣算法适应度,根据表 2 寻优参数集,定义蜉蝣个体向量坐标,初始化设置蜉蝣种群大小、位置和速度等参数,并设定参数取值范围。

步骤 2 计算初始化后适应度,并根据适应度由高到低依次排序,将最高适应度蜉蝣作为此代最优个体和全局最优个体。

步骤 3 按照位置和速度更新公式更新蜉蝣位置坐标及速度,按照最佳匹配机制进行配对,并生成子代蜉蝣。

步骤 4 再次计算更新后蜉蝣以及子代蜉蝣的适应度,若子代蜉蝣适应度高于亲代适应度较低蜉蝣,则用子代取代亲代,若子代蜉蝣优于全局最佳蜉蝣,则将子代取代全局最佳蜉蝣,所有蜉蝣对更新完毕后,以更新后群体作为新一代亲本,重新划定种群,同时,通过加入随机突变机制抑制算法过早收敛,使算法向着其他空间继续探索,提高全局搜索能力,一定程度上防止陷入局部最优。

步骤 5 判断是否达到迭代上限,若满足条件,以全局最优蜉蝣参数重构 CNN-GRU 模型,若不满足条件,循环执行步骤 3~5。

3 实验验证

3.1 数据集描述

为验证本文方法的有效性,采用东南大学(SEU)齿轮箱数据集、凯斯西储大学(CWRU)滚动轴承数据集以及实验室自测振动平台数据集,以 1 024 个数据点为一个样本,采用 VMD 和 CWT 分别生成时序样本和图像样本,验证本文方法的有效性。SEU 齿轮箱实验台包含正常、齿轮故障、轴承

故障及其复合故障共计 10 类齿轮箱健康状态信号。凯斯西储大学选取零负载下故障直径为 0.007 mm 时的 9 种故障数据,构成实验数据集。实验室自测振动平台数据集如图 2 所示,轴承故障预设 在输出轴侧,齿轮故障预设 在从动齿轮上。在行星齿轮箱上预 制了 11 种齿轮箱健康状态,其中轴承故障包括滚动体故障、内圈故障、外圈故障,齿轮故障包括齿面磨损、齿轮断齿,复合故障包括齿轮断齿加轴承滚动体故障、齿轮断齿加轴承内圈故障、齿轮断齿加轴承外圈故障、齿面磨损加轴承滚动体故障、齿面磨损加轴承内圈故障、齿面磨损加轴承外圈故障。



图 2 实验室自测振动平台
Fig. 2 Laboratory self-testing vibration platform

3.2 实验结果与分析

3.2.1 超参数优化结果对比

为展示 蜚蜚算法优化效果,以粒子群算法、遗传

算法作为对比算法,图 3 为 3 种算法迭代过程识别精度变化情况。根据图 3 中曲线分析,3 种优化算法均能通过对超参数优化提高识别精度。粒子群算法以最快速度收敛,但识别精度较低,遗传算法收敛速度较慢,相比于这两种优化算法,蜚蜚算法优化效果更好。

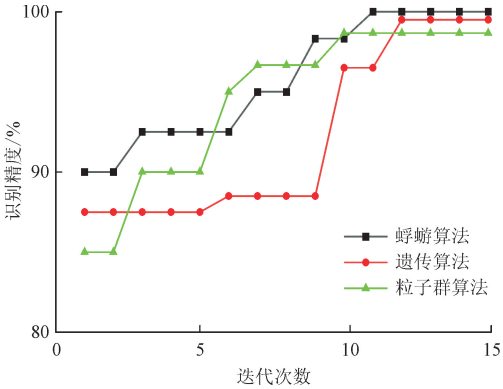


图 3 3 种算法寻优识别精度
Fig. 3 Three algorithms for optimizing recognition accuracy

3 种优化算法得到的超参数组合识别精度结果如表 3~5 所示,3 种参数组合识别精度均超过 98%,并且各超参数最佳组合并非一成不变,不同参数组合之间能够起到较好效果,优化算法对超参数优化结果影响较大,相比于遗传算法和粒子群算法,经过蜚蜚算法优化后模型识别精度最高,优化曲线更加平稳。因此,采用蜚蜚算法作为模型优化算法。

表 3 3 种算法优化的超参数(2D)
Table 3 Superparameters optimized by three algorithms (2D)

优化算法	卷积核尺寸			卷积核数			丢弃率
	卷积层 1	卷积层 2	卷积层 3	卷积层 1	卷积层 2	卷积层 3	
遗传算法	4×4	3×3	3×3	32	64	64	0.600 0
粒子群算法	5×5	4×4	3×3	32	64	64	0.367 2
蜚蜚算法	3×3	5×5	5×5	32	64	64	0.506 0

表 4 3 种算法优化的超参数(1D)
Table 4 Superparameters optimized by three algorithms(1D)

优化算法	GRU 神经元		丢弃率		
	GRU 层 1	GRU 层 2	第 1 层	第 2 层	第 3 层
遗传算法	29	24	0.269 0	0.200 0	0.200 0
粒子群算法	29	21	0.492 1	0.600 0	0.600 0
蜚蜚算法	118	47	0.600 0	0.585 3	0.530 8

表 5 不同超参数组合对识别精度的影响
Table 5 Influence of different hyperparameter combinations on recognition accuracy

优化算法	学习率	优化器类型	最小批次	识别精度/%
遗传算法	0.003 9	Adam	89	99.67
粒子群算法	0.004 3	Adam	40	98.67
蜚蜚算法	0.001 8	Adam	29	100.00

3.2.2 不同实验平台下实验结果与对比分析

将经过 MA 优化初始化参数后的 CNN-GRU 模型与 CNN、GRU、深度适配网络(DAN)^[27] 和 CNN-GRU 模型进行对比,在 CWRU 数据集、SEU 数据集和自测数据集上的对比结果如图 4 所示。从图 4 中可以得出:采用传统 GRU 和 CNN 模型识别效果不如 CNN-GRU 模型,未经过优化的 CNN-GRU 模型在 CWRU 数据集上识别精度为 97.9%,比 GRU 模型高 5.4%,比 CNN 模型高 3.3%,比 DAN 模型高 2.8%;在 SEU 数据集上识别精度为 96.8%,比 GRU 模型高 6%,比 CNN 模型高 3.1%,比 DAN 模型高 2.6%;在自测数据集上识别精度为 97.5%,比 GRU 模型高 6.3%,比 CNN 模型高 3.6%,比 DAN 模型高 2.6%;在自测数据集上识别精度为 97.5%,比 GRU 模型高 6.3%,比 CNN 模型高 3.6%,比 DAN 模型高 2.6%。经优化的 CNN-GRU 模型,在 CWRU 数据集上识别精度高达 99.8%,在 SEU 数据集上识别精度为 99.5%,在自测数据集上识别精度为 99.6%,相比未经优化的 CNN-GRU 模型有明显提升。

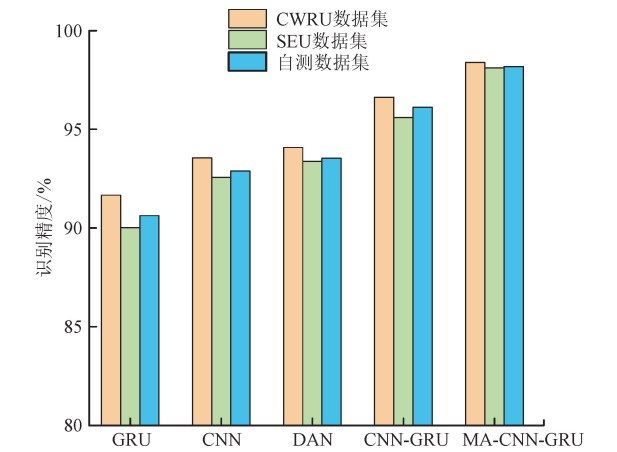
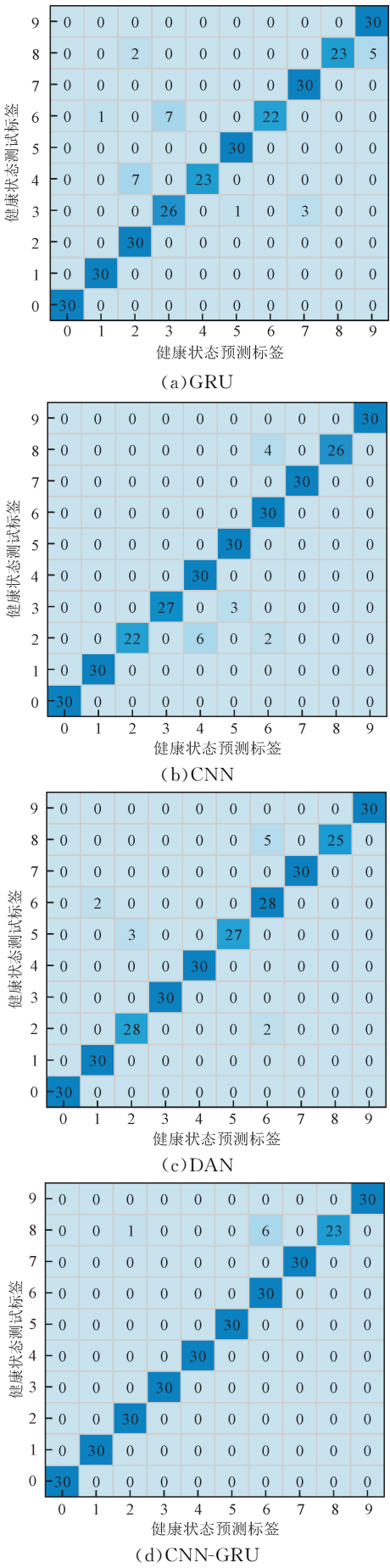


图 4 不同工况下识别精度对比

Fig. 4 Comparison diagram of recognition accuracy under the same working condition

为了直观、清晰地分析模型性能,通过混淆矩阵显示某次测试中故障诊断结果,具体如图 5 所示。GRU 模型共识别正确 274 个样本,识别错误 26 个样本,识别精度为 91.33%;CNN 模型共识别正确 285 个样本,识别错误 15 个样本,识别精度为 95%;DAN 模型共识别正确 288 个样本,识别错误 12 个样本,识别精度为 96%;CNN-GRU 模型共识别正确 293 个样本,识别错误 7 个样本,识别精度为 97.67%;MA-CNN-GRU 模型共识别正确 299 个样本,识别错误 1 个样本,识别精度为 99.67%,由此可见,双通道模型能一定程度上提升故障识别精度,同时也证明双



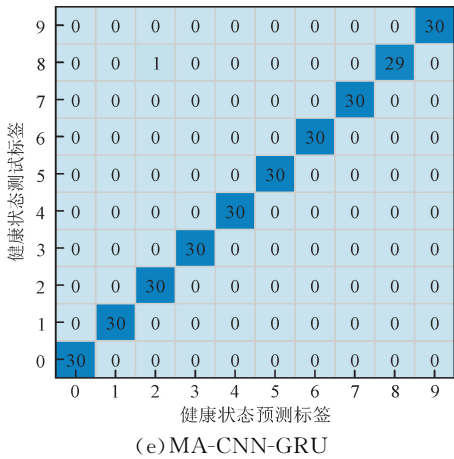


图 5 故障分类混淆矩阵

Fig. 5 Fault classification confusion matrix

通道特征提取具有互补性,相比于单通道,能提取到更多故障特征,故障信息更加完整。

图 6 进一步展示所提模型及对比模型的测试精度,MA-CNN-GRU 模型首次迭代准确率为 22.33%,比 CNN-GRU 模型高 1.67%,比 DAN 模型高 10.04%,比 GRU 模型高 11%,比 CNN 模型高 17.33%。初始训练阶段,MA-CNN-GRU 模型测试精度曲线斜率较大,证明其正在迅速收敛,4 种对比算法测试精度曲线斜率均小于 MA-CNN-GRU 模型。MA-CNN-GRU 模型在第 54 次迭代时准确率为 100%,并在迭代 74 次后达到稳定,收敛于 100%;CNN-GRU 模型在第 156 次迭代时准确率为 97.67%,并在迭代 246 次后达到稳定,收敛于 97.67%;GRU 模型测试精度曲线经过 67 次迭代后达到最大识别精度,221 次迭代后才趋于稳定;CNN 模型测试精度曲线经过 61 次迭代后达到最大识别精度,但后续精度仍有一定波动,较不稳定;DAN 模型测试

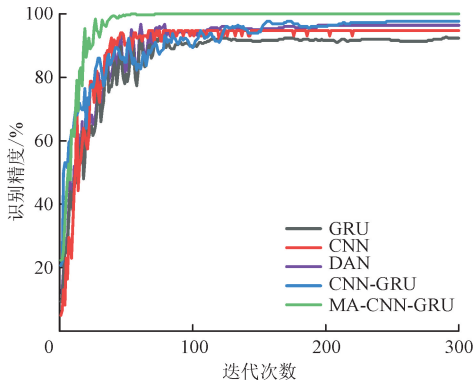


图 6 5 种网络模型训练识别精度的对比

Fig. 6 Comparison of training diagnosis rates for five network models

精度曲线经过 82 次迭代后达到最大识别精度,180 次迭代后才趋于稳定。相比于其他算法,MA-CNN-GRU 模型测试精度曲线均能迅速收敛,训练精度曲线更为平滑,所提方法识别精度明显优于对比方法。

3.2.3 噪声工况下实验结果与对比分析

为了验证模型鲁棒性,使实验更逼近真实环境下采集的信号特征,以 SEU 数据集作为实验数据,在信号中添加高斯白噪声,并与 GRU、CNN、DAN、CNN-GRU 模型进行对比,结果如图 7 所示。由图 7 可知,MA-CNN-GRU 模型具有较强的鲁棒性,能够在高频噪声影响下保持良好的分类能力,相较于 GRU、CNN、DAN、CNN-GRU 模型有明显优势。

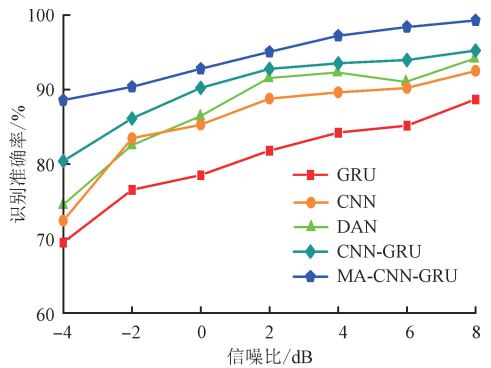


图 7 5 种网络模型噪声干扰下识别精度的对比

Fig. 7 Comparison of recognition accuracy under noise interference of five network models

3.2.4 变负载工况下实验结果与对比分析

以凯斯西储大学数据集作为实验数据,将经过 MA 优化初始化参数后的 CNN-GRU 模型与 CNN、DAN 模型进行对比。故障诊断结果如表 6 所示,表 6 中,任务 A→B 表示以 A 数据集为信号源域,B 数据集为信号目标域,采用迁移学习策略训练故障识别模型。其中,CNN 模型未采用迁移学习策略,其平均识别精度为 83.85%,DAN 模型平均识别精度为 86.51%,使用 BN 层进行归一化的 CNN-GRU 模型的平均识别精度为 88.57%,使用 BN 层进行归一化的 MA-CNN-GRU 模型相对表现较好,平均识别精度为 90.51%。在 12 组变负载工况下的诊断任务中,使用 AdaBN 层进行归一化的 MA-CNN-GRU 模型故障诊断精度最高,平均识别精度为 92.90%,比 CNN、DAN、使用 BN 层进行归一化的

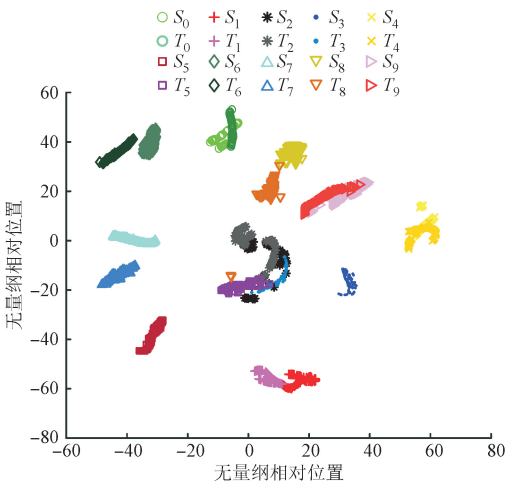
CNN-GRU 和使用 BN 层进行归一化的 MA-CNN-GRU 模型分别高 9.05%,6.56%,4.33% 和 2.39%,结果表明:AdaBN 的引入提高了变负载情况下 MA-CNN-GRU 模型的诊断性能。

为显示训练后源域和目标域信号分布,本文采用 t 分布邻域嵌入(t -SNE)将全连接层可视化。图

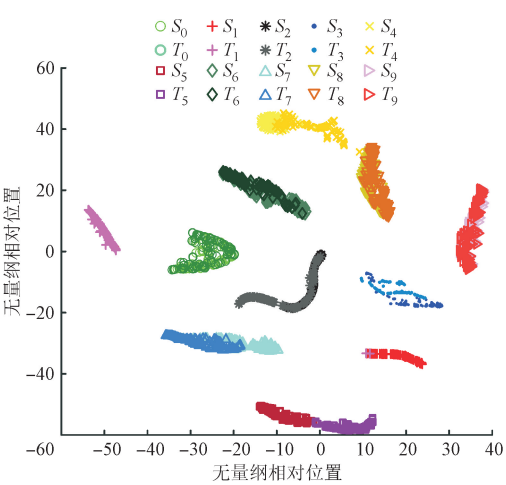
8 以 CNN 和 MA-CNN-GRU(AdaBN)模型为例,显示任务 B→C 可视化特征图,图 8 中横纵坐标表示经过(t -SNE)降维后前两列特征。其中, $S_0 \sim S_9$ 为源域对 10 种轴承健康状态训练得到的网络的特征分布, $T_0 \sim T_9$ 为目标域对 10 种轴承健康状态训练得到的网络的特征分布。

表 6 变负载工况下的故障诊断结果
Table 6 Diagnostic results under variable load conditions

任务	平均识别精度/%				
	CNN	DAN	CNN-GRU(BN)	MA-CNN-GRU(BN)	MA-CNN-GRU(AdaBN)
A→B	91.40	92.20	92.80	93.10	95.10
A→C	83.30	89.47	90.19	92.23	96.33
A→D	75.93	79.33	83.68	85.55	91.27
B→A	92.60	93.37	93.66	93.66	94.43
B→C	92.13	93.40	93.93	95.10	95.27
B→D	80.20	82.13	84.05	87.88	93.20
C→A	88.87	91.33	92.22	95.18	96.10
C→B	89.57	92.57	93.54	95.98	96.73
C→D	84.20	86.30	88.41	89.20	90.37
D→A	73.97	74.87	80.65	83.21	86.53
D→B	69.77	72.40	79.54	82.88	84.80
D→C	84.30	88.73	90.13	92.18	94.70
平均值	83.85	86.34	88.57	90.51	92.90



(a)CNN 特征可视化



(b)MA-CNN-GRU(AdaBN)特征可视化

图 8 全连接层特征可视化

Fig. 8 Fully connected layer feature visualization

从图 8 可以看出,CNN 模型中大部分样本源域和目标域特征距离接近,例如正常状态 S_0 、 T_0 ,但分布差异较大,例如内圈故障损伤尺寸为 0.36 mm

(S_5 、 T_5),并且部分特征之间存在重叠现象,例如滚动体故障损伤尺寸为 0.36 mm(S_2 、 T_2),这是由于 CNN 没有使用迁移策略。

与之相比,MA-CNN-GRU(AdaBN)中源域和目标域特征重合度较高,无混叠现象出现,不同负载下对其识别效果良好,不同类别之前区分度较高,这进一步验证了MA-CNN-GRU模型在处理变负载状态下故障诊断的优越性。综上,所提优化后的MA-CNN-GRU(AdaBN)算法模型对于变负载后的目标域特征具备很好的自适应能力,即使面对特征邻域分布存在差异的情况,依旧能够很好地应对轴承故障诊断任务。

4 结 论

针对齿轮箱故障难以准确提取出微小特征故障,深度学习模型超参数选用对于人工经验的高度依赖性问题,提出一种基于MA算法优化的双通道故障诊断模型CNN-GRU,在SEU齿轮箱数据集以及CWRU滚动轴承数据集进行了对比实验,得到以下结论。

(1)建立双通道故障诊断模型CNN-GRU,在对原始数据进行预处理后,与GRU、CNN和DAN进行对比,实验结果表明,采用时序和图像双通道输入的CNN-GRU模型相比于传统模型具有更高识别精度,证明了双通道特征提取具有一定的互补性。

(2)采用MA算法优化CNN-GRU超参数,以遗传算法和粒子群算法作为对比算法,实验结果表明,蜉蝣算法优化效果明显优于对比算法,优化后的CNN-GRU模型分类特性明显高于未优化的版本,本文所提方法能够对齿轮箱故障进行准确检测,对维修和保持机械设备稳定运行具有一定的实用价值。

(3)针对实际工作过程中工作载荷随生产条件变化导致数据集分布差异的情况,引入AdaBN算法代替传统的BN层,实现了故障信号在不同负载下的有效识别,极大地提高模型的故障诊断精度。

(4)基于MA优化的CNN-GRU诊断模型适用于齿轮箱的故障诊断,具有较高的分类精度,由于其双通道设计,计算参数量较多,模型训练时间较长,对于故障诊断时间要求不高的场合有重要的应用参考价值。

参考文献:

[1] 陈向民,舒文伊,韩梦茹,等.基于URP-ANCNN的变转速齿轮箱智能故障诊断方法[J].噪声与振动控

制,2024,44(2):129-135.

CHEN Xiangmin, SHU Wenyi, HAN Mengru, et al. Intelligence fault diagnosis method for gearboxes under variable rotational speed based on URP-ANCNN [J]. Noise and Vibration Control, 2024, 44(2): 129-135.

[2] 谢锋云,汪淦,王玲岚,等. STFT 结合 2D CNN-SVM 的齿轮箱故障诊断方法 [J]. 噪声与振动控制, 2024, 44(4): 103-109.

XIE Fengyun, WANG Gan, WANG Linglan, et al. Fault diagnosis method of gearbox based on STFT and 2DCNN-SVM [J]. Noise and Vibration Control, 2024, 44(4): 103-109.

[3] 吴胜利,周赓,邢文婷.基于SDP和MCNN-LSTM的齿轮箱故障诊断方法[J].振动与冲击,2024,43(15):126-132.

WU Shengli, ZHOU Yi, XING Wenting. Gearbox fault diagnosis based on SDP and MCNN-LSTM [J]. Journal of Vibration and Shock, 2024, 43(15): 126-132.

[4] 陈子旭,朱振杰,卢国梁.一种新的图谱域滚动轴承早期故障检测与识别方法[J].振动与冲击,2022,41(6):51-59.

CHEN Zixu, ZHU Zhenjie, LU Guoliang. Novel early fault detection and diagnosis for rolling element bearings in graph spectrum domain [J]. Journal of Vibration and Shock, 2022, 41(6): 51-59.

[5] 王岩红,温笑欢,揭永琴,等.基于对比学习的滚动轴承早期故障在线检测方法[J].振动与冲击,2023,42(14):229-236.

WANG Yanhong, WEN Xiaohuan, JIE Yongqin, et al. Online detection method for bearing incipient faults based on contrastive learning [J]. Journal of Vibration and Shock, 2023, 42(14): 229-236.

[6] 李迎秋,周威,郝德成.基于孪生网络的旋转机械故障诊断方法[J].组合机床与自动化加工技术,2023(11):116-120.

LI Yingqiu, ZHOU Wei, HAO Decheng. A fault diagnosis method for rotating machinery based on twin networks [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2023(11): 116-120.

[7] ZHANG Jing, TIAN Jing, WEN Tao, et al. Deep fault diagnosis for rotating machinery with scarce labeled samples [J]. Chinese Journal of Electronics, 2020, 29(4): 693-704.

[8] 王椿晶,王海瑞.深度在线小波极限学习在旋转机械故障诊断中的应用[J].机械科学与技术,2023,42

- (7): 1029-1034.
- WANG Chunjing, WANG Hairui. Application of depth online wavelet extreme learning machine in rotating machinery fault diagnosis [J]. Mechanical Science and Technology for Aerospace Engineering, 2023, 42(7): 1029-1034.
- [9] 吴春志, 冯辅周, 吴守军, 等. 深度学习在旋转机械设备故障诊断中的应用研究综述 [J]. 噪声与振动控制, 2019, 39(5): 1-7.
- WU Chunzhi, FENG Fuzhou, WU Shoujun, et al. Review of application research of deep learning in fault diagnosis of rotating machinery [J]. Noise and Vibration Control, 2019, 39(5): 1-7.
- [10] 李益兵, 马建波, 江丽. 基于 SFLA 改进卷积神经网络的滚动轴承故障诊断 [J]. 振动与冲击, 2020, 39(24): 187-193.
- LI Yibing, MA Jianbo, JIANG Li. Fault diagnosis of rolling bearing based on an improved convolutional neural network using SFLA [J]. Journal of Vibration and Shock, 2020, 39(24): 187-193.
- [11] 陈功胜, 唐向红, 陆见光, 等. 基于 CNN-ETR 的滚动轴承故障诊断研究 [J]. 兵器装备工程学报, 2021, 42(6): 251-255.
- CHEN Gongsheng, TANG Xianghong, LU Jianguang, et al. Research on fault diagnosis of rolling bearing based on CNN-ETR [J]. Journal of Ordnance Equipment Engineering, 2021, 42(6): 251-255.
- [12] 赵小强, 张亚洲. 利用改进卷积神经网络的滚动轴承变工况故障诊断方法 [J]. 西安交通大学学报, 2021, 55(12): 108-118.
- ZHAO Xiaoqiang, ZHANG Yazhou. Improved CNN-based fault diagnosis method for rolling bearings under variable working conditions [J]. Journal of Xi'an Jiaotong University, 2021, 55(12): 108-118.
- [13] 肖雄, 王健翔, 张勇军, 等. 一种用于轴承故障诊断的二维卷积神经网络优化方法 [J]. 中国电机工程学报, 2019, 39(15): 4558-4568.
- XIAO Xiong, WANG Jianxiang, ZHANG Yongjun, et al. A two-dimensional convolutional neural network optimization method for bearing fault diagnosis [J]. Proceedings of the CSEE, 2019, 39(15): 4558-4568.
- [14] CHEN Zhiqiang, LI Chuan, SANCHEZ R V. Gearbox fault identification and classification with convolutional neural networks [J]. Shock and Vibration, 2015, 2015(1): 390134.
- [15] 宫文峰, 陈辉, 张美玲, 等. 基于深度学习的电机轴承微小故障智能诊断方法 [J]. 仪器仪表学报, 2020, 41(1): 195-205.
- GONG Wenfeng, CHEN Hui, ZHANG Meiling, et al. Intelligent diagnosis method for incipient fault of motor bearing based on deep learning [J]. Chinese Journal of Scientific Instrument, 2020, 41(1): 195-205.
- [16] MONTEIRO R, BASTOS-FILHO C, CERRADA M, et al. Convolutional neural networks using Fourier transform spectrogram to classify the severity of gear tooth breakage [C]//2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SD-PC). Piscataway, NJ, USA: IEEE, 2018: 490-496.
- [17] 王贡献, 付泽, 胡志辉, 等. 基于多分支卷积神经网络的轴承变工况故障诊断 [J]. 噪声与振动控制, 2023, 43(6): 135-141.
- WANG Gongxian, FU Ze, HU Zhihui, et al. A multi-branch convolution neural network for fault diagnosis of rolling bearings under variable working conditions [J]. Noise and Vibration Control, 2023, 43(6): 135-141.
- [18] 叶壮, 余建波. 基于多通道一维卷积神经网络特征学习的齿轮箱故障诊断方法 [J]. 振动与冲击, 2020, 39(20): 55-66.
- YE Zhuang, YU Jianbo. Gearbox fault diagnosis based on feature learning of multi-channel one-dimensional convolutional neural network [J]. Journal of Vibration and Shock, 2020, 39(20): 55-66.
- [19] ZHANG Wei, LI Chuanhao, PENG Gaoliang, 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: 439-453.
- [20] 丁丁, 刘文哲, 盛常冲, 等. 神经网络架构搜索研究进展与展望 [J]. 国防科技大学学报, 2023, 45(6): 100-131.
- DING Ding, LIU Wenzhe, SHENG Changchong, et al. State of the art and prospects of neural architecture search [J]. Journal of National University of Defense Technology, 2023, 45(6): 100-131.
- [21] 汪祖民, 张志豪, 秦静, 等. 基于卷积神经网络的机械故障诊断技术综述 [J]. 计算机应用, 2022, 42(4): 1036-1043.
- WANG Zumin, ZHANG Zhihao, QIN Jing, et al. Review of mechanical fault diagnosis technology based on convolutional neural network [J]. Journal of Computer Applications, 2022, 42(4): 1036-1043.

[22] 吴宣勇, 黄忠全, 李琪康, 等. 无源数据约束下多源域自适应的风电齿轮箱故障诊断方法 [J]. 太阳能学报, 2024, 45(4): 238-246.

WU Xuanyong, HUANG Zhongquan, LI Qikang, et al. Multi-source domain adaptive fault diagnosis method of wind turbine gearbox under no-accessing source data constraints [J]. Acta Energiae Solaris Sinica, 2024, 45(4): 238-246.

[23] 陈建蓉. 新型多源特征融合模型及应用研究 [D]. 南京: 南京大学, 2020.

[24] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. Eprint ArXiv, 2014. DOI:10.48550/arXiv.1412.3555.

[25] 陈涛, 李欣. 太赫兹光谱在转基因菜籽油鉴别中的应用: 基于改进蜉蝣算法的支持向量机模型 [J]. 物理学报, 2024, 73(5): 360-368.

CHEN Tao, LI Xin. Application of terahertz spectroscopy in identification of transgenic rapeseed oils: a support vector machine model based on modified mayfly optimization algorithm [J]. Acta Physica Sinica, 2024, 73(5): 360-368.

[26] GAO Zhongke, WANG Zibo, MA Chao, et al. A wavelet time-frequency representation based complex network method for characterizing brain activities underlying motor imagery signals [J]. IEEE Access, 2018, 6: 65796-65802.

[27] 王眺. 基于深度交互适配网络的通用多模态学习方法研究 [D]. 杭州电子科技大学, 2024.

(编辑 武红江)