



DOI:10.12404/j.issn.1671-1815.2402620

引用格式:王雪纯,李想,杨随先.基于MCNN-MSA-BiGRU的轴承故障诊断[J].科学技术与工程,2025,25(11):4534-4542.

Wang Xuechun, Li Xiang, Yang Suixian. Bearing fault diagnosis based on MCNN-MSA-BiGRU[J]. Science Technology and Engineering, 2025, 25(11): 4534-4542.

机械、仪表工业

基于 MCNN-MSA-BiGRU 的轴承故障诊断

王雪纯,李想,杨随先*

(四川大学机械工程学院,成都 610065)

摘要 针对传统故障诊断模型对特征提取不全面,单一模型稳定性和泛化性差的问题,提出了一种基于多头自注意力机制的多尺度卷积神经网络和双向门控循环单元模型,从空间和时序层面实现特征提取。该模型采用原始一维振动信号作为输入,使用不同尺寸卷积核的卷积网络捕获多尺度信息。引入多头自注意力机制,根据输入的不同部分动态调整输出权重,忽略冗余信息并对所提取特征进行加权融合,将融合后的特征输入至BiGRU(bidirectional gated recurrent units)网络,通过双向信息融合机制,对来自过去和未来两个方向的信息进行挖掘,捕捉输入序列不同部分间的依赖关系。最后,通过Softmax分类实现轴承故障诊断。在3种轴承数据集上进行实验验证,结果表明,所提模型性能指标表现优异,具有良好的泛化性和可行性。

关键词 故障诊断;卷积神经网络;双向门控循环单元;注意力机制;轴承

中图分类号 TH133.3;

文献标志码 A

Bearing Fault Diagnosis Based on MCNN-MSA-BiGRU

WANG Xue-chun, LI Xiang, YANG Sui-xian*

(School of Mechanical Engineering, Sichuan University, Chengdu 610065, China)

[Abstract] To address the issues of incomplete feature extraction, poor stability, and limited generalization in traditional fault diagnosis models, a model based on a multi-scale convolutional neural networks (MCNN), bidirectional gated recurrent units (BiGRU), and multi-head self-attention mechanism (MSA) was proposed. The model was designed to achieve comprehensive feature extraction from both spatial and temporal perspectives. It took raw vibration signals as input, and multi-scale features were extracted through convolution kernels of different sizes. A multi-head self-attention mechanism was used to dynamically adjust output weights, disregarding redundant information and weighting the extracted features for fusion. Then the fused features were input into a BiGRU network, which utilized a bidirectional information fusion mechanism to explore information from both past and future directions, capturing dependencies between different parts of the input sequence. Finally, Softmax was employed for classification. Experimental validation was conducted using three bearing fault datasets, and the results show that the proposed model has excellent performance metrics on different datasets and showcases good generalization and feasibility.

[Keywords] fault diagnosis; convolutional neural network; bidirectional gated recurrent unit; attention mechanism; bearing

随着现代化工业的不断发展,旋转机械广泛应用于航空航天、建筑、交通运输等领域^[1]。作为旋转机械的关键部件,滚动轴承通常在高负载、高速的恶劣环境下长时间工作,在固有退化过程和时变工况下可能出现磨损、变形、断裂等故障^[2]。轴承故障未被及时检测可能导致整个机械装置失效,造成经济损失甚至伤亡^[3]。多项研究的统计分析结果表明,40%以上的设备故障与轴承有关^[4]。开展滚动轴承故障诊断方法研究对于提高设备可靠性、

保证工业生产安全具有重要意义。

传统的基于数据的故障诊断方法一般包括4个步骤:数据采集、特征提取、特征选择与分类^[5]。传统方法通过对原始数据信号进行时域、频域或时频域的分析处理实现故障特征的提取,所提取的特征质量依赖于专家经验与专业知识,泛化能力差。对于一些具有外部环境干扰和非线性内部耦合的复杂系统,传统方法的浅层结构不足以挖掘所有类型故障的特征^[6]。

收稿日期:2024-04-11 修订日期:2024-07-26

基金项目:国家自然科学基金(52275538)

第一作者:王雪纯(1999—),女,汉族,河南开封人,硕士研究生。研究方向:故障诊断。E-mail:wangxc@stu.scu.edu.cn。

*通信作者:杨随先(1965—),男,汉族,四川纳溪人,博士,教授。研究方向:故障诊断、无损检测。E-mail:yangsx@163.com。

投稿网址:www.stae.com.cn

随着大数据分析 with 人工智能的发展,现代智能故障诊断在齿轮、变速箱、轴承、滚动轴承、泵、风力涡轮机和核电站等领域得到广泛应用^[7]。中国在轴承智能故障诊断领域,尤其是数据驱动的故障诊断领域,处于世界领先地位^[8]。作为一种数据驱动方法,深度学习可以捕捉数据与设备状态间复杂的映射关系,其诊断过程无需人工干预,克服了传统故障诊断方法的局限性。卷积神经网络(convolutional neural network, CNN)作为最常用的深度学习模型,已经成功应用于故障诊断任务中^[9]。Song 等^[10]提出了一种基于宽卷积核卷积神经网络(convolutional neural network with wide convolution kernels, WKCNN)的轴承智能故障诊断新方法,解决了传统算法耗时长,通用性差的问题。赵志宏等^[11]将马尔可夫跃迁场(Markov transition field, MTF)与 CNN 相结合,应用于滚动轴承的故障诊断。Zeng^[12]设计了集特征识别和特征分类于一体的 CNN 模型,实现对带式输送机传动托辊滚动轴承的故障诊断。李中等^[13]采用双分支卷积网络将故障位置与损伤程度进行双标签处理,实现故障诊断。在上述研究中, CNN 更多地是对数据进行局部空间特征提取,而无法针对时间序列信息进行挖掘,导致数据的在时间序列的动态性能捕获不佳。RNN(recurrent neural networks)及其变体在解决捕获时序数据中的时间依赖性问题方面具有显著优势^[14]。RNN 可以保留隐藏层在某一时刻的状态信息,建立方向环路中单元之间的连接,并通过表示隐藏层数据环路更新的链式规则实现时间的记忆,在复杂动态系统建模领域具有强大的优势。但单独使用 RNN 可能会导致梯度消失或梯度爆炸,降低预测准确性,常与其他方法结合使用。Ahsan 等^[15]将具有 SoftMax 分类器的高精度深度卷积神经网络(deep convolution neural network, DCNN)与长短期记忆(long short term memory, LSTM)结合,并应用于时变转速下的轴承故障诊断。Sun 等^[16]提出一种基于 CNN 和 LSTM 的混合故障诊断模型,实现混合负载下的轴承故障诊断。刘万宇等^[17]将宽卷积核深度卷积神经网络与深度 LSTM 相结合,实现了端到端的轴承故障诊断。而 LSTM 模型训练时间较长,相较于 LSTM,门控递归单元(gate recurrent unit, GRU)的网络结构更加简洁,参数量更少,训练速度快,并降低了过拟合的风险。而双向门控循环单元(bidirectional gated recurrent unit, BiGRU)结合了前向与后向的双向信息流^[18],可同时学习过去与未来的特征表示,有效捕捉到序列中的双向依赖关系。

为解决上述研究中特征提取不全面,时序信息利用不充分,单一模型泛化能力差的问题,提升模型在复杂工况下的性能表现,现将 CNN 的空间特征提取能力与 BiGRU 在时间序列的特征挖掘能力相结合,为增强模型捕捉全局信息与细节特征的能力,引入多头自注意力机制,提出一种基于多头注意力机制的多尺度卷积神经网络-双向门控循环单元故障诊断模型,将原始一维振动信号作为输入,保留信号的完整性,采用不同尺寸卷积核对特征进行多尺度捕获,实现端到端的故障诊断。

1 基础理论

1.1 卷积神经网络

卷积神经网络是一种前馈神经网络,可以实现分层特征的自动抽象,具有权值共享和稀疏连接的特性,即同一个滤波器在不同位置共享权重,每个神经元只与输入数据的局部区域相连,使 CNN 能够更有效地学习和表示数据的特征。CNN 通常由输入层、隐藏层和输出层三层组成,隐藏层一般包括卷积层、池化层和全连接层。卷积层将前一层的局部信号映射至下一层,计算过程表达式为

$$y_j^l = f\left(\sum_{i \in R_j} x_i^{l-1} * k_j^l + b_j^l\right) \quad (1)$$

式(1)中: y_j^l 为第 l 层的输出; $*$ 为卷积运算; x_i^{l-1} 为第 $l-1$ 层的第 i 个输入; k_j^l 和 b_j^l 为第 l 层中第 k 个卷积核所对应的权值与偏置; R_j 为输入的特征集合; $f(\cdot)$ 为激活函数。

池化层又称下采样层,常添加于卷积层之后,进行子采样操作,去除数据中的冗余信息,减少特征数据量,提高计算效率。常见的池化方式有最大池化与平均池化,即选择池化区域内的平均值或最大值,保留突出的特征分别传播到下一层,减少过拟合。全连接层由密集连接的神经元组成,接收从池化层提取得到的特征并映射至输出空间,通过非线性激活函数增加网络的表达能力,实现分类或回归。

1.2 BiGRU 网络

GRU 由 LSTM 网络优化而来,是对 LSTM 模型的简化与调整,并保留了 LSTM 网络遗忘无关信息的特点,具有更少的网络参数与更简单的结构,收敛性更好。当训练数据较多时,可以显著减少训练时间^[19]。LSTM 分别使用遗忘门、输入门和输出门来控制输入输出与状态信息。GRU 由重置门 r_t 和更新门 z_t 组成。更新门 z_t 用于控制上一时刻传递到当前结构的信息,代替 LSTM 中的遗忘

门和输入门,重置门 r_t 用于确定控制当前 GRU 单元忽略上一时刻多少信息。GRU 通过上一时刻的状态 h_{t-1} 和当前时刻的输入 x_t 确定门控状态,运算公式为

$$z_t = [W_z(h_{t-1}, x_t)] \quad (2)$$

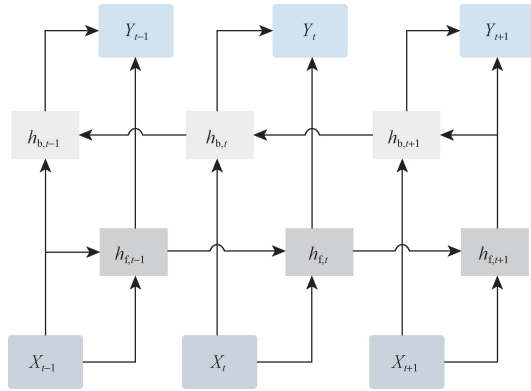
$$r_t = \sigma[W_r(h_{t-1}, x_t)] \quad (3)$$

$$\tilde{h}_t = \tanh[W_{\tilde{h}}(r_t * h_{t-1}, x_t)] \quad (4)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (5)$$

式中: σ 为 Sigmoid 激活函数; W_z 、 W_r 、 $W_{\tilde{h}}$ 分别为权重; \tilde{h}_t 为当前时刻节点; h_t 为当前隐藏状态的输出; $*$ 为逐元素乘法运算。更新门 z_t 的范围为 $(0, 1)$, 当 z_t 趋近于 0 时, h_{t-1} 的状态信息保留越多,通过 $z_t * \tilde{h}_t$ 对 \tilde{h}_t 进行选择记忆^[20]。

BiGRU 由正向和反向的两个 GRU 构成,同时考虑输入的前向与后向信息,并将两个方向的信息进行融合,其结构图如图 1 所示。



$h_{f,t}$ 为前向 GRU 当前时刻状态, $h_{b,t}$ 为后向 GRU 当前时刻状态

图 1 BiGRU 结构图

Fig. 1 The structure of BiGRU

通过双向信息融合机制,对来自过去和未来两个方向的信息进行挖掘,捕捉输入序列不同部分间的依赖关系。BiGRU 的计算公式为

$$h_{f,t} = \text{GRU_forward}(X_t, h_{f,t-1}) \quad (6)$$

$$h_{b,t} = \text{GRU_backward}(X_t, h_{b,t+1}) \quad (7)$$

$$Y_t = [h_{f,t}, h_{b,t}] \quad (8)$$

1.3 注意力机制

注意力机制通过关注输入序列中的关键部分来提高模型的性能,被广泛应用于自然语言处理中^[21]。通过引入注意力机制,模型可以根据输入序列中的不同部分动态调整输出的权值,更好地捕捉序列中的长程依赖关系。自注意力是一种通过计算序列中不同元素之间的相关性来提取信息的注意力机制^[22]。该机制通过对输入特征图进行卷积映射为 3 个相同的特征图,每个像素都用作一个查询

向量 Q , 一个键向量 K 和一个值向量 V ^[23]。公式为

$$Q = W_Q X, K = W_K X, V = W_V X \quad (9)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

通过输入与可学习的参数矩阵 W_Q 、 W_K 、 W_V 线性变换得到 Q 、 K 和 V , $\sqrt{d_k}$ 为查询向量 Q 和键向量 K 的维度,通过计算查询向量 Q 和键向量 K 的点积衡量二者的相似度,将结果通过 $\sqrt{d_k}$ 进行缩放,并使用 Softmax 函数进行归一化得到注意力权重,将注意力权重与对应的值向量 V 相乘并进行加权求和。

多头自注意力机制由多个独立的自注意力机制组成,将输入通过线性变换映射到不同的子空间,其运算公式为

$$\text{MultiHead}(Q, K, V) = \text{concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n)W^O \quad (11)$$

$$\text{head}_i = \text{Attention}(QW_{Qi}, KW_{iK}, VW_{iV}) \quad (12)$$

式中: n 代表注意力头的个数,将每个注意力头计算注意力权重的结果拼接,通过矩阵 W 将结果映射回输入的维度。模型对输入序列的部分给予不同的关注,减少了信息的丢失,提高了模型的表达能力和泛化能力。

2 MCNN-MSA-BiGRU 模型

针对传统模型中时序信息利用不充分,对重点信息关注不充足的问题,本文中结合多尺度卷积不同层次的特征提取与 BiGRU 时序信息的关注,并通过多头自注意力机制的优势,提出了 MCNN-MSA-BiGRU 模型,模型结构如图 2 所示。

模型采用原始一维振动信号作为输入,通过不同尺寸的卷积核对输入进行特征提取,捕获多尺度信息将不同尺度卷积核提取到的特征进行融合,并通过多头自注意力机制对融合后的信息进行权重划分,从多个角度提取特征之间的相关性。引入 dropout 等控制模型的复杂度,减少对数据的过拟合与对特定输入的依赖,提高模型的鲁棒性与泛化能力。通过 BiGRU 从正反两方向实现特征的提取,对全局信息进行进一步整合,利用振动信号的时空特征,强化特征的表达能力。最终,将特征通过 Softmax 函数实现轴承故障的多分类。

3 实验研究与分析

3.1 数据来源与描述

凯斯西储大学(Case Western Reserve University, CWRU)数据集在轴承故障诊断领域应用广泛。轴承实验平台由电机驱动,并在轴承驱动端与

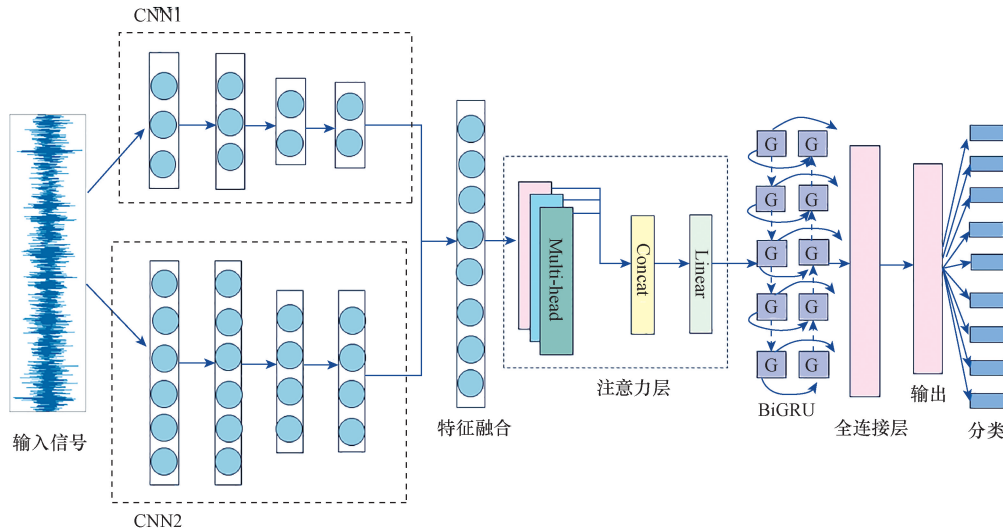


图2 MCNN-MSA-BiGRU 结构图

Fig. 2 The structure of MCNN-MSA-BiGRU

风扇端放置传感器进行信号采集。通过电火花加工技术对轴承植入故障,根据所加工故障的位置不同,可以将其划分为外圈故障、内圈故障与滚动体故障,每种故障包含不同程度的损失,其损伤范围分别为0.007、0.014、0.021 in(1 in = 0.025 4 m)。根据故障位置与损伤程度的不同,可将故障划分为9种类型。此外,实验平台采集了正常状况下的轴承信号,故障信号与正常信号共计10类。

实验选用12 kHz的驱动端采集的故障数据,轴承转速为1 750 r/min,电机负载为2 HP,如表2所示,其中1 HP = 745.7 W。每类状态选取100个样本,按照7:2:1的比例划分为训练集、验证集和测试集,通过长度为1 024的采样点对原始振动数据随机切割实现采样,具体如表1所示。

表1 CWRU数据集样本数据

Table 1 Sample data of the CWRU dataset

标签	故障位置	电机负载/HP	电机转速/(r·min ⁻¹)	故障直径/in	样本数量	样本长度
0	滚动体	2	1 750	0.007	100	1 024
1	滚动体	2	1 750	0.014	100	1 024
2	滚动体	2	1 750	0.021	100	1 024
3	内圈	2	1 750	0.007	100	1 024
4	内圈	2	1 750	0.014	100	1 024
5	内圈	2	1 750	0.021	100	1 024
6	外圈	2	1 750	0.007	100	1 024
7	外圈	2	1 750	0.014	100	1 024
8	外圈	2	1 750	0.021	100	1 024
9	正常	2	1 750	—	100	1 024

3.2 参数设置

试验在 tensorflow 框架下进行,采用 python 编程语言,在 Linux 系统中进行模型搭建与训练。模型

的主要结构参数如表2所示。使用 Adam 优化器对算法进行优化,并应用交叉熵损失函数作为损失度量函数,模型中的 l2 正则化大小设置为 0.01, dropout 层参数设置为 0.5, batch_size 值设置为 256, epoch 次数设置为 200。所有实验均进行 5 次交叉验证,取平均值作为最终分类结果进行分析。

表2 MCNN-MSA-BiGRU 模型结构参数

Table 2 The structural parameters of the MCNN-MSA-BiGRU

网络层	核大小	核数量	输出	激活函数
Conv1_1	20	64	(None, 256, 64)	Relu
Pool1_1	2	—	(None, 128, 64)	—
Conv1_2	10	32	(None, 64, 32)	Relu
Pool1_2	2	—	(None, 32, 32)	—
Conv1_3	10	32	(None, 32, 32)	Relu
Pool1_3	2	—	(None, 16, 32)	—
Conv2_1	4	64	(None, 256, 64)	Relu
Pool2_1	2	—	(None, 128, 64)	—
Conv2_2	2	32	(None, 64, 32)	Relu
Pool2_2	2	—	(None, 32, 32)	—
Conv2_3	2	32	(None, 32, 32)	Relu
Pool2_3	2	—	(None, 16, 32)	—
BiGRU	64	—	(None, 128)	—
Softmax	—	—	(None, 10)	—

3.3 试验结果及分析

将测试集的准确率、损失率及混淆矩阵作为模型的性能评判指标,并选择 GRU、BiGRU、1D-CNN、CNN-BiGRU、MCNN-BiGRU 共 5 种模型进行对比实验,5 种模型的主要结构参数与本文所提模型保持一致。6 种模型的性能变化如下。

准确率反映了模型预测的正确程度。由图3可知,GRU 的准确率与稳定性均比 BiGRU 差,在时序

信息的提取利用层面,相较于 GRU, BiGRU 更具优势。与 BiGRU 相比, 1D-CNN 准确率更高。CNN-BiGRU 准确率有进一步提升, 但模型稳定性不足, 存在一定波动。MCNN-BiGRU 和 MCNN-MSA-BiGRU 训练样本与验证样本的准确率随迭代次数增

加不断提高, 在所用数据集中的分类准确率可达 100%, 验证了模型结构设计的可靠性。引入 MSA 机制, 其准确率曲线更稳定。

损失率量化了模型预测与实际数据之间的差异。由图 4 模型的损失率曲线可知, 与空间层面提取

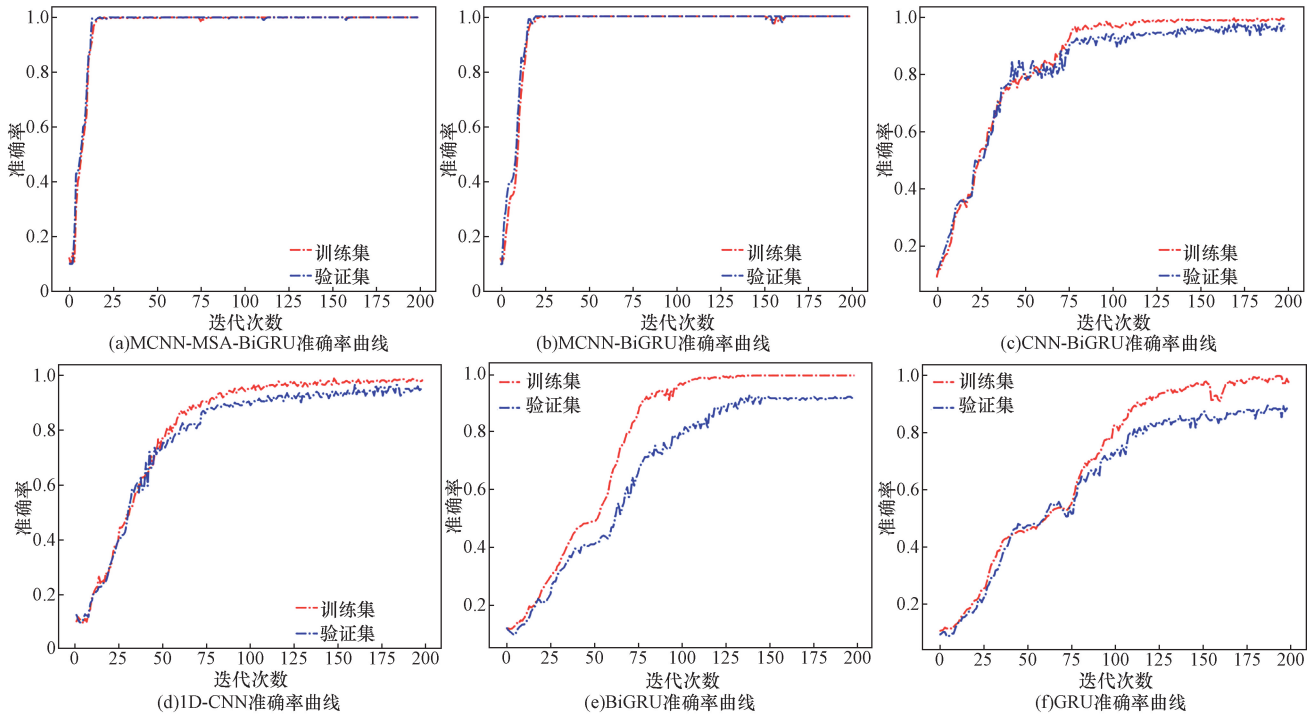


图 3 不同模型识别准确率对比

Fig. 3 Comparison of recognition accuracy of different models

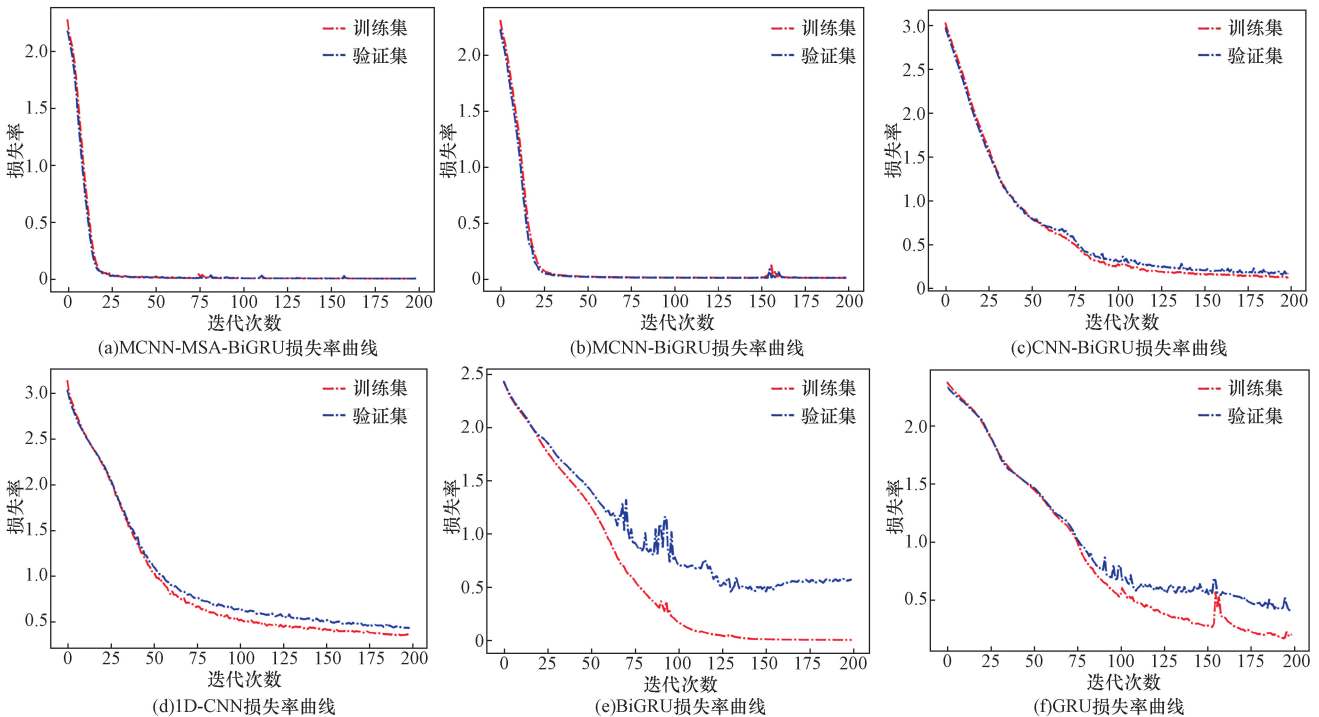


图 4 不同模型损失率对比

Fig. 4 Comparison of loss rates of different models

特征的 1D-CNN 相比,从时序层面提取特征的 BiGRU 和 GRU 的损失率波动较大,拟合效果一般。CNN-BiGRU 拟合程度优于单一模型 1D-CNN。相较于 MCNN-BiGRU,所提模型的损失率迭代曲线斜率绝对值更大且曲线较光滑,收敛速度更快,且伴随迭代次数的增加其损失率曲线斜率接近于 0,模型趋于稳定。

混淆矩阵反映了模型在每个类别上的分类准确度,包括正确分类的样本数与错误分类的样本数。图 5 所示不同模型混淆矩阵对比情况,结果表明,所提模型的每一类故障模型都实现了正确诊断,整体分类性能稳定可靠。

不同负载条件下系统工作状态有所差异,同一故障在不同负载下的故障信号特征有所不同。为进一步验证所提模型在不同工况状态下的性能,选取变负载的工况条件进行实验。其结果如图 6 所示,分别为负载在 0、1、3 HP 的条件下各模型的故障诊断表现结果。由图 6 可知,混合模型的稳定性优于单一诊断模型,在负载变化时,其他模型的性能具有一定程度的波动,MCNN-MSA-BiGRU 模型仍具有较好的性能指标。

3.4 泛化能力验证

为验证本模型在不同轴承数据集的通用性,将模型在 Ottawa 数据集和江南大学 (Jiangnan Univer-

sity, JNU) 数据集上进行泛化能力验证,实验参数设置保持不变。

3.4.1 Ottawa 数据集验证

Ottawa 数据集包含了滚动轴承不同故障类型的振动和转速数据。该数据集包含了 4 种工况状态,即升速,降速,先升速后降速与先降速后升速,每种工况下有 3 种类型状态,包括健康、外圈故障和内圈故障。以 200 kHz 的频率对数据进行采样,采样持续时间为 10 s。实验采用不同转速状况下的轴承振动数据,根据其转速变化类型将其划分为 4 种状态,每种状态选取 300 个样本,按照 7:2:1 的比例划分为训练集、验证集和测试集,并通过长度为 1 024 的采样点随机切割进行采样,具体如表 3 所示。

为验证本模型的鲁棒性与泛化能力,在不同工况条件下进行对故障的诊断,将准确率、损失率和 F_1 作为评价模型指标,诊断结果如表 4 所示。

结果表明,所提出的模型在 4 种工况下均具有较高的准确率。在降速条件下,SEBCNN 和 WDCNN 的准确率低且损失率高,分类效果欠佳,而 MCNN-MSA-BiGRU 在降速数据集上表现优异。在升速和先降速后升速条件下,引入 MSA 机制的模型准确率提升相对未引入 MSA 机制的 MCNN-BiGRU 模型而言效果不显著,但损失率有所降低,模型稳定度高,且准确率均优于 SEBCNN 与 WDCNN,损失率

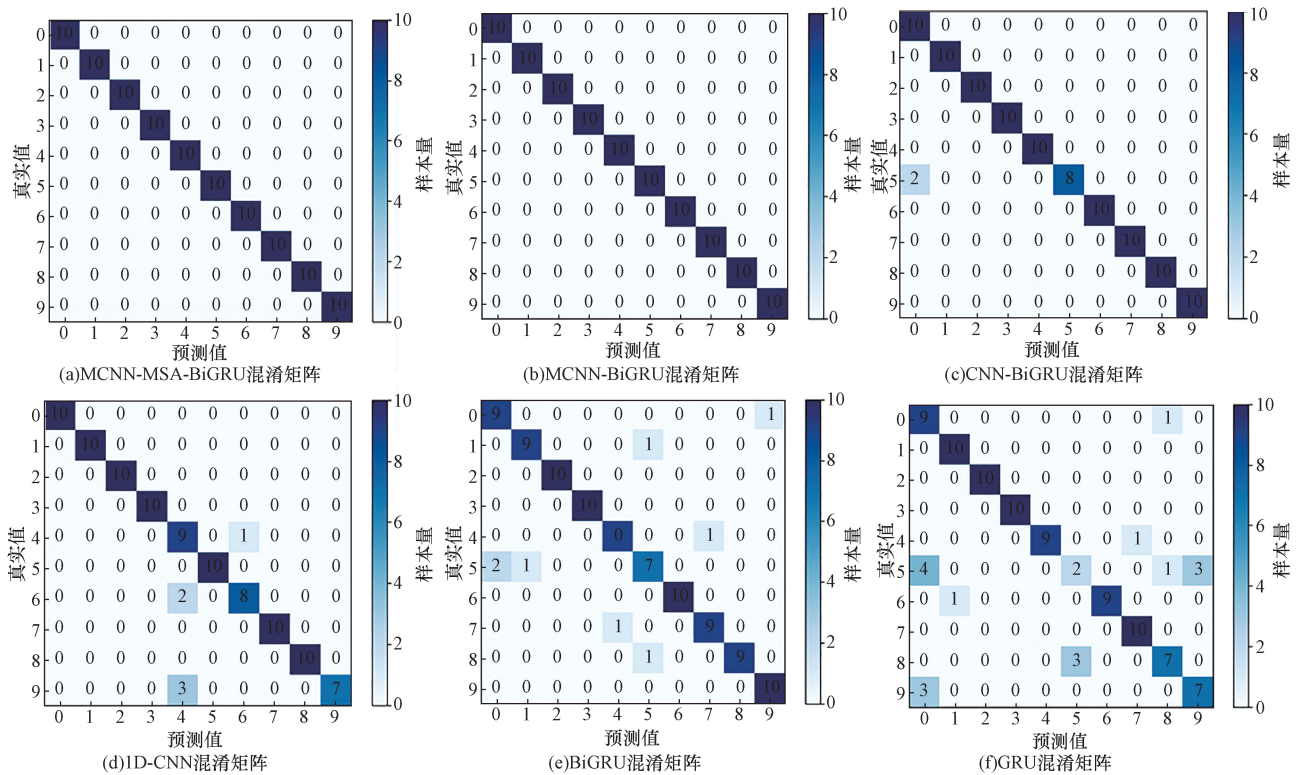


图 5 不同模型混淆矩阵对比
Fig. 5 Comparison of confusion matrix of different models

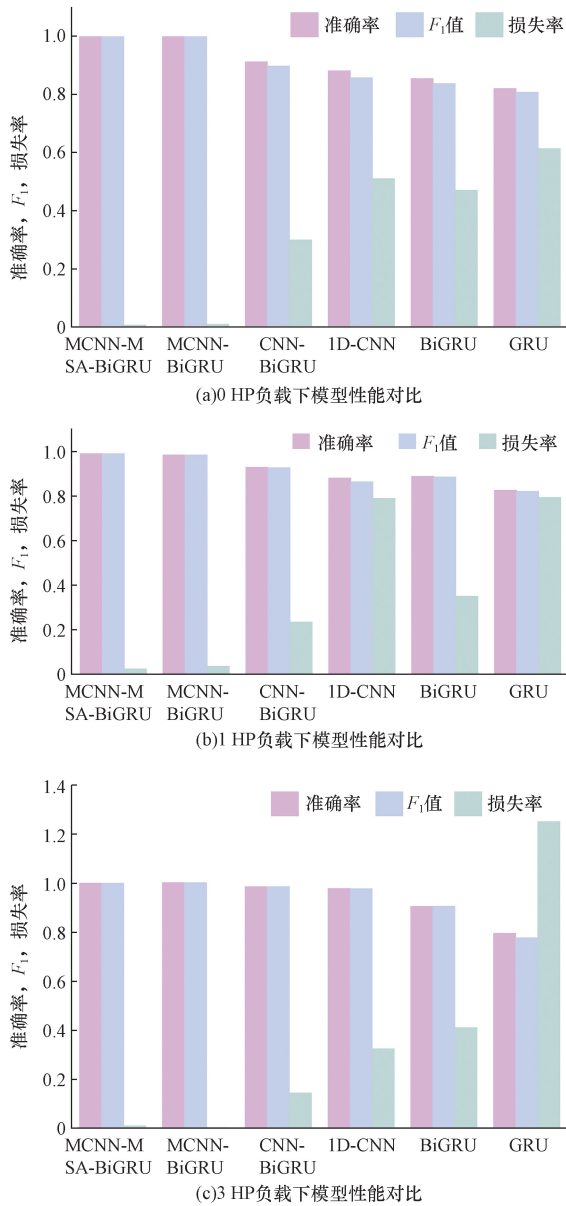


图6 不同负载下各模型性能对比
Fig. 6 Performance comparison of models under different loads

表3 Ottawa数据集样本数据

Table 3 Sample data of the Ottawa dataset

故障标签	故障位置	工况	样本数量	样本长度
0	健康	升速	300	1 024
1	内圈故障	升速	300	1 024
2	外圈故障	升速	300	1 024
0	健康	降速	300	1 024
1	内圈故障	降速	300	1 024
2	外圈故障	降速	300	1 024
0	健康	升速-降速	300	1 024
1	内圈故障	升速-降速	300	1 024
2	外圈故障	升速-降速	300	1 024
0	健康	降速-升速	300	1 024
1	内圈故障	降速-升速	300	1 024
2	外圈故障	降速-升速	300	1 024

表4 不同模型的性能对比

Table 4 Performance comparison of different models

工况	模型	准确率	损失率	F ₁
升速	MCNN-MSA-BiGRU	0.982	0.070	0.982
	MCNN-BiGRU	0.980	0.050	0.980
	WDCNN	0.913	0.325	0.913
	SEBCNN	0.944	0.196	0.944
降速	MCNN-MSA-BiGRU	0.996	0.010	0.996
	MCNN-BiGRU	0.989	0.052	0.989
	WDCNN	0.847	0.587	0.843
	SEBCNN	0.862	0.365	0.861
升-降	MCNN-MSA-BiGRU	0.978	0.079	0.978
	MCNN-BiGRU	0.969	0.103	0.969
	WDCNN	0.918	0.477	0.918
	SEBCNN	0.947	0.271	0.947
降-升	MCNN-MSA-BiGRU	0.973	0.106	0.973
	MCNN-BiGRU	0.973	0.110	0.973
	WDCNN	0.920	0.275	0.920
	SEBCNN	0.898	0.349	0.896

有明显降低。在先升速后降速条件下 MCNN-MSA-BiGRU 准确率有所降低,但仍比 WDCNN 高出 6%。引入 MSA 机制可以对重要信息进行优化,增强了模型对不同数据集的特征提取能力。在变工况的状态下所提模型各项性能指标仍表现优异,表明该模型在不同数据集下具有良好的适用性与可行性。

3.4.2 江南大学数据集验证

江南大学数据集的试验平台由电机、滚动轴承、加速度计和数据采集系统组成。该数据集包括 3 种转速下的轴承数据,每种转速下对应 4 种轴承状态,即健康、外圈故障、内圈故障和滚动体故障。以 50 kHz 的频率对数据进行采样,采样持续时间为 10 s。实验选取 3 种转速下的轴承数据,并将其划分为 3 种状态,每种状态选取 500 个样本,按照 7:2:1 的比例划分为训练集、验证集和测试集,并通过长度为 1 024 的采样点随机切割进行采样,具体如表 5 所示。

表5 江南大学数据集样本数据

Table 5 Sample data of the JNU dataset

故障标签	故障位置	电机转速/ (r·min ⁻¹)	样本数量	样本长度
0	健康	600	500	1 024
1	内圈故障	600	500	1 024
2	外圈故障	600	500	1 024
3	滚动体故障	600	500	1 024
0	健康	800	500	1 024
1	内圈故障	800	500	1 024
2	外圈故障	800	500	1 024
3	滚动体故障	800	500	1 024
0	健康	1 000	500	1 024
1	内圈故障	1 000	500	1 024
2	外圈故障	1 000	500	1 024
3	滚动体故障	1 000	500	1 024

在三种工况条件下进行对故障的诊断,模型在 800 r/min 转速下的准确率曲线,损失率曲线,混淆矩阵如图 7 所示。将准确率、损失率和 F_1 作为评价模型指标,模型在三种工况下的诊断结果如表 6 所示。

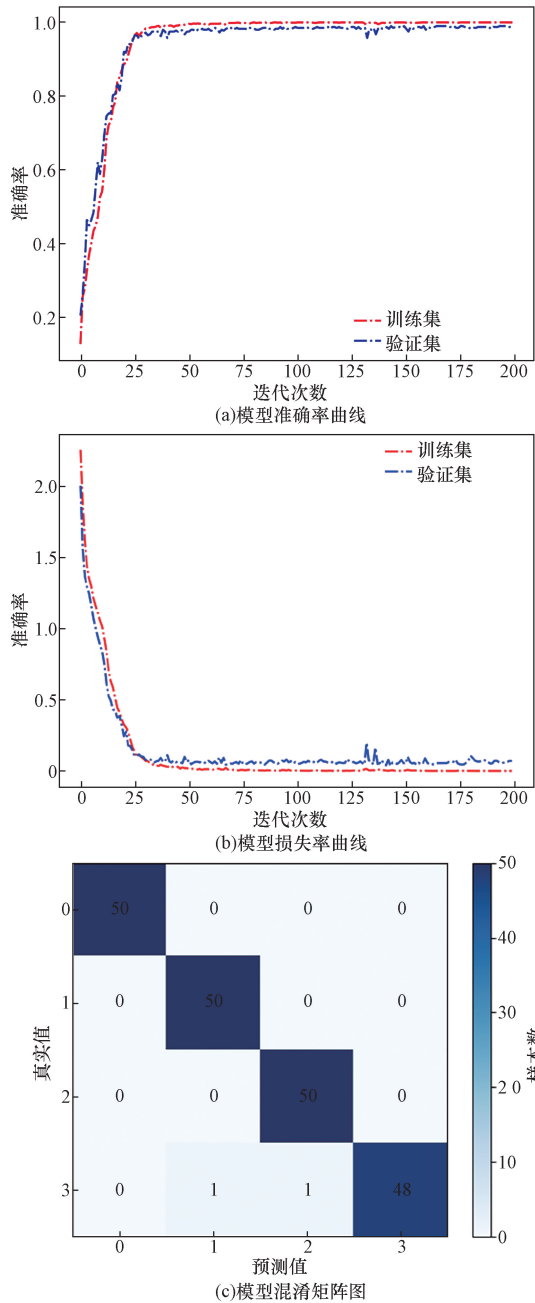


图 7 800 r/min 转速下模型的性能图

Fig. 7 Picture of the model performance at 800 r/min speed

表 6 不同工况下模型的性能测试

Table 6 Performance test under different conditions

转速/(r·min ⁻¹)	准确率	损失率	F_1
600	0.975	0.156	0.976
800	0.990	0.048	0.990
1 000	1.000	0.003	1.000

模型在江南大学数据集中仍具有良好的故障诊断能力,在 800 r/min 转速下,模型收敛速度快,约在第 40 次迭代过程中实现收敛,在收敛后有较小波动,但整体损失率低,模型稳定程度高,在 1 000 r/min 转速下的故障诊断准确率可达 100%,表现出优异的泛化能力。

4 结论

针对现有轴承故障诊断模型特征提取不全面,时序信息利用不充分,复杂工况下故障诊断准确率低的问题,提出了一种基于多头注意力机制的多尺度卷积神经网络-双向门控循环单元故障诊断模型。通过一系列实验验证了该模型的可行性,得到如下结论。

(1)所提模型通过双向门控循环单元在时间层面挖掘振动信号蕴含的特征信息,学习信号在过去与未来的特征表示,有效捕捉原始振动信号中的序列关系。

(2)所提模型经多尺度卷积神经网络进行不同层次的特征提取与多头自注意力的并行计算,丰富了空间层面的特征表示,增强了模型在变工况下对滚动轴承故障诊断的特征提取能力。

(3)实验结果表明,相较于其他深度学习模型,所提模型准确率和 F_1 值具有更优异的表现;尤其在变工况条件下仍具有较高准确率。其次,所提模型在国内外的轴承故障数据集上展现了良好的泛化性和鲁棒性。综上所述,所提模型对工业生产中复杂工况下的轴承故障诊断具备较好的辅助作用。

参 考 文 献

- [1] Wang H, Guo Y, Liu X, et al. Fault diagnosis method for imbalanced data of rotating machinery based on time domain signal prediction and SC-ResNeSt [J]. IEEE Access, 2023, 11: 38875-38893.
- [2] Cui W, Meng G, Wang A, et al. Application of rotating machinery fault diagnosis based on deep learning[J]. Shock and Vibration, 2021, 23: 1-30.
- [3] Wang M, Li J, Xue Y. A new weight-based dual domain adaptation transfer model for bearing fault diagnosis under noisy and cross-domain conditions[J]. IEEE Access, 2023, 11: 123766-123783.
- [4] George G, Theodore L, Chrysostomos D S, et al. Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition[J]. Mechanical Systems & Signal Processing, 2013, 41(1/2): 510-525.
- [5] Lee S, Kim T. Impact of deep learning optimizers and hyperparameter tuning on the performance of bearing fault diagnosis [J]. IEEE Access, 2023, 11: 55046-55070.
- [6] Chen X, Zhang B, Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model[J]. Intell Manuf, 2021, 32: 971-987.
- [7] Tang S, Yuan S, Zhu Y. Deep learning-based intelligent fault diagnosis methods toward rotating machinery[J]. IEEE Access, 2020,

- 8: 9335-9346.
- [8] 张伟. 基于卷积神经网络的轴承故障诊断算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2018.
Zhang Wei. Bearing fault diagnosis algorithm study based on convolutional neural network[D]. Harbin: Harbin Institute of Technology, 2018.
- [9] Ugli E M, Lee K H, Lee C H. Automatic optimization of one-dimensional cnn architecture for fault diagnosis of a hydraulic piston pump using genetic algorithm [J]. IEEE Access, 2023, 11: 68462-68472.
- [10] Song X, Cong Y, Song Y, et al. A bearing fault diagnosis model based on CNN with wide convolution kernels[J]. Ambient Intell Human Computer, 2022, 13: 4041-4056.
- [11] 赵志宏, 李春秀, 窦广鉴, 等. 基于 MTF-CNN 的轴承故障诊断研究[J]. 振动与冲击, 2023, 42(2): 126-131.
Zhao Zhihong, Li Chunxiu, Dou Guangjian, et al. Bearing fault diagnosis method based on MTF-CNN [J]. Journal of Vibration and Shock, 2023, 42(2): 126-131.
- [12] Zeng Y. Fault diagnosis method of rolling bearing based on Xception-CNN model[C]//2023 4th International Conference on Computer Engineering and Intelligent Control (ICCEIC). Guangzhou: IEEE, 2023: 570-573.
- [13] 李中, 卢春华, 王星, 等. 考虑滚动轴承故障位置与损伤程度的双分支卷积神经网络故障诊断方法[J]. 科学技术与工程, 2022, 22(4): 1441-1448.
Li Zhong, Lu Chunhua, Wang Xing, et al. Double-branch convolutional neural network fault diagnosis method considering the fault location and damage degree of rolling bearings[J]. Science Technology and Engineering, 2022, 22(4): 1441-1448.
- [14] Vo T T, Liu M K, Hsieh C L. Hybrid 1D CNN-RNN network for fault diagnosis in induction motors using electrical signals[C]//2023 Inter-national Conference on System Science and Engineering (ICSSE). New York: IEEE, 2023: 212-217.
- [15] Ahsan M, Salah M M. Efficient DCNN-LSTM model for fault diagnosis of raw vibration signals: applications to variable speed rotating machines and diverse fault depths data-sets[J]. Symmetry, 2023, 15(7): 1413.
- [16] Sun H, Fan Y. Fault diagnosis of rolling bearings based on CNN and LSTM networks under mixed load and noise[J]. Multimed Tools Appl, 2023, 82: 43543-43567.
- [17] 刘万宇, 李宇鹏, 石怀涛, 等. 基于 WDCNN-DLSTM 的滚动轴承故障诊断方法[J]. 科学技术与工程, 2023, 23(13): 5522-5529.
Liu Wanyu, Li Yupeng, Shi Huaitao, et al. Fault diagnosis method for rolling bearings based on WDCNN-DLSTM [J]. Science Technology and Engineering, 2023, 23(13): 5522-5529.
- [18] Zhang B, Song J, Li B, et al. Prognostics of rolling element bearing based on combined SSA-VMD and CNN-BiGRU models[C]//2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS). Yibin: IEEE, 2023: 1-6.
- [19] Choe D E, Kim H C, Kim M H. Sequence-based modeling of deep learning with LSTM and GRU networks for structural damage detection of floating offshore wind turbine blades[J]. Renewable Energy, 2021, 174: 218-235.
- [20] Li Z, Wang P, Li X. Fault diagnosis of asynchronous motors based on 2DCNN-GRU network optimization [C]//2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS). Yibin: IEEE, 2023: 1-6.
- [21] Kim Y, Kim Y K. Physics-informed time-frequency fusion network with attention for noise-robust bearing fault diagnosis[J]. IEEE Access, 2024, 12: 12517-12532.
- [22] Cui L, Tian X C, Wei Q Z, et al. A self-attention based contrastive learning method for bearing fault diagnosis[J]. Expert Systems with Applications, 2024, 238: 121645.
- [23] 祁宣豪, 智敏. 图像处理中注意力机制综述[J]. 计算机科学与探索, 2024, 18(2): 345-362.
Qi Xuanhao, Zhi Min. Review of attention mechanisms in image processing[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(2): 345-362.