

DOI: 10.13382/j.jemi.B2205226

基于 RA-LSTM 的轴承故障诊断方法*

谢锦阳¹ 姜媛媛^{2,3} 王力²

(1. 安徽理工大学人工智能学院 淮南 232000; 2. 安徽理工大学电气与信息工程学院 淮南 232000;

3. 安徽理工大学环境友好材料与职业健康研究院(芜湖) 芜湖 241003)

摘要:为解决轴承故障诊断时网络模型不能对一维振动信号中的特征按照重要程度分配不同的权重,导致无法提取具有代表性意义的特征,进而影响诊断模型的精确度与鲁棒性的问题。提出基于反向注意力机制(reverse attention mechanism, RA)的特征突出处理方法,通过将特征进行注意力反向与剪枝,降低非重要特征占比,从而对重要特征进行凸显。并通过长短期记忆网络(LSTM)进一步学习特征之间的时间信息后通过全连接层进行故障类型分类。通过实验选取了最优数据截取长度、剪枝超参数并对信号添加噪声后模型的稳定性进行了验证。实验结果表明所提出的 RA-LSTM 轴承故障诊断方法具有优异的故障诊断性能,故障诊断精度能达到 100%,且在添加噪声后模型的诊断能力仍具有优异的鲁棒性。

关键词: 轴承故障诊断;反向注意力机制;LSTM

中图分类号: TH17 **文献标识码:** A **国家标准学科分类代码:** D460.4

RA-LSTM based bearing fault diagnosis method

Xie Jinyang¹ Jiang Yuanyuan^{2,3} Wang Li²

(1. School of Institute of Artificial Intelligence, Anhui University of Science and Technology, Huainan 232000, China;

2. School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232000;

3. Institute of Environment-Friendly Materials and Occupational Health, Anhui University of

Science and Technology, Wuhu 241003, China)

Abstract: In order to solve the problem that the diagnostic model cannot assign different weights to the features in the one-dimensional vibration signal according to their importance, which leads to the failure to extract representative features and thus affects the accuracy and robustness of the diagnostic model. A feature highlighting method based on reverse attention amplification mechanism (RA) is proposed to reduce the proportion of non-important features by reversing attention and pruning the features, so as to highlight the important features. The long short term memory (LSTM) network is used to learn the temporal information between the features and to classify the fault types through the fully connected layer. The optimal data interception length, pruning hyperparameters and the stability of the model after adding noise to the signal are selected experimentally. The optimal data interception length, pruning hyperparameters are experimentally selected and verified the stability of the model after adding noise to the signal. The experimental results show that the proposed RA-LSTM bearing fault diagnosis method has excellent fault diagnosis performance, with fault diagnosis accuracy reaching 100% and excellent robustness even after the addition of noise.

Keywords: bearing fault diagnosis; reverse attention mechanism; LSTM

收稿日期: 2022-06-02 Received Date: 2022-06-02

* 基金项目:安徽省重点研究与开发计划(202104g01020012)、安徽理工大学环境友好材料与职业健康研究院研发专项基金(ALW2020YF18)项目资助

0 引言

滚动轴承作为旋转机械的核心部件,其运行状态决定着整个旋转机械的可靠性。由于旋转轴承往往长期工作在复杂的工况环境下导致其极易出现损伤,是导致旋转机械发生故障的常见原因。及时发现轴承故障就能够通过维修和较低的成本避免更大事故的发生。因此,对滚动轴承运行状况的监测与诊断是十分必要的。

在轴承故障诊断中,原始信号通常是带有噪声的一维振动信号,如何从信号中提取具有代表意义的特征信息是对轴承信号进行精确故障类型诊断的关键。常用的特征提取方法有变分模态分解(VMD)^[1-3],经验模态分解(EMD)^[4],小波变换(WT)^[5],集成经验模态分解(EEMD)^[6]等。朱少明等^[7]利用人工峭度和相关系数构建加权峭度指标并作为目标函数,通过人工蜂群算法^[8]对变分模态分解过程中的模态数和带宽控制参数进行优化从而对故障信号提取有效特征。丁嘉鑫等^[9]利用广义复合多尺度(GCMWPE)全面表征滚动轴承故障特征信息,构建高维故障特征集,并应用监督等度规映射(S-Isomap)算法进行有效的二次特征提取。上述改进算法能够增强网络对特征的提取能力,但大多采用浅层网络,难以处理复杂工况环境下的诊断问题。随着卷积神经网络在自然语言处理领域的成功,其强大的时序信息的特征提取能力与深层的网络结构使得复杂工况的轴承故障诊断问题有了新的解决方案^[10]。金余丰等^[11]将原始振动信号输入 Inception 模型中提取信号的多尺度抽象特征,并通过 BLSTM 进一步学习特征信息的时间依赖性。Li 等^[12]通过改进双胶囊卷积神经网络^[13]提取信号中的特征,在深入挖掘故障特征的同时充分保留信号间的空间信息,能够在数据量与诊断效果之间达到理想的平衡。Jin 等^[14]通过构建多层自适应卷积神经网络(MACNN)对原始数据进行特征提取,接着使用宽内核去捕获浅层中的多尺度低频特征,再将其通过小卷积核在深层中实现多层非线性映射。上述方法只是通过卷积神经网络提取振动信号中的特征,并未将重要特征与次要特征加以区分,导致网络无法对更需要关注的特征分配更多的关注,进而影响网络对故障类型诊断的精确度及鲁棒性。Transformer 模型中的注意力机制能够将特征中的重要特征进行突出,使神经网络模型能够更好的提取具有代表性的特征。郑直等^[15]将一维时序振动信号输入残差网络进行特征提取,并通过带有注意力机制的双向长短期记忆网络对所提取的特征赋予不同的权重。但是现今的注意力机制往往只是简单地将特征通过占比的形式进行概率换算后进行放大,重要特征在放大的同时次要特征并没有得到有效的抑制,影响模型的诊断精度。

针对上述问题,提出一种基于反向注意力放大机制与 LSTM 网络的轴承故障诊断方法。首先通过一维卷积神经网络初步提取振动信号中的特征后,再由反向注意力放大机制将所提取特征中的重要特征进行放大、次要特征进行剪枝后输入 LSTM 网络进行进一步时间信息特征提取,最后经过全连接层对故障进行分类。并通过实验探究不同数据读取步长、反向注意力放大机制中不同参数以及添加噪声时模型的诊断精确性。

1 相关理论

1.1 注意力机制

人眼在观察事物时会自动过滤次要信息,首先关注重要程度高的信息。通过将注意力集中在重要程度较高的信息上能够减少目标区域的大小,有效提高信息的处理效率。注意力机制通过模仿人眼对信息的处理方式,对输入特征的重点信息进行突出处理增加其在整体上的比重,减少网络对其他信息的关注程度,其流程如图 1 所示。处理方式通常为将特征经 Softmax 函数将其压缩至 0~1 之间,再将原特征矩阵与压缩后的矩阵相乘以达到突出重要特征的目的。

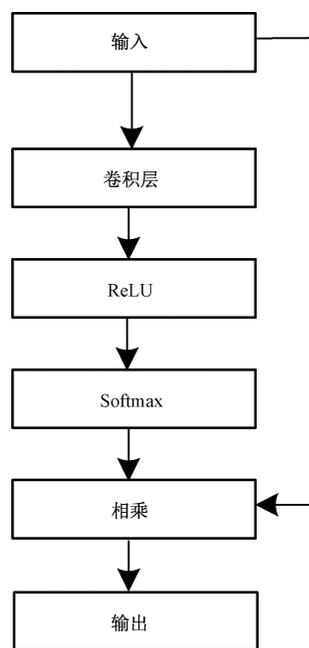


图 1 注意力机制

Fig. 1 Attentional mechanism

1.2 反向注意力放大机制

注意力机制对特征进行突出的方法将重要特征与次要特征进行同等操作,在突出重要特征的同时并没有对次要特征进行进一步的压缩,导致次要特征在特征矩阵

中仍然具有较大占比,影响模型诊断精度。

为对特征按照重要程度进行区分并对其分配不同的权重,提高信息的处理效率。提出基于注意力机制的反向注意力放大机制,首先对输入的特征进行一次卷积操作,卷积核个数为 128,进行初步提取特征;再经过 ReLU 激活函数增加非线性后利用 Softmax 函数将特征转换为 0~1 之间的概率占比。为了对重要特征进行进一步的凸显,提出用 1 减去转换后的特征使得原本突出的特征变为非突出特征,再设置一个合适的超参数使现保留的原突出特征减为负数后再经过 ReLU 激活函数;此时,理想情况下原重要特征变为 0,原次要特征为一较小的正数;然后再用 1 减去现保留的特征矩阵,原重要特征此时得以进一步突出,并再添加一个超参数用以对现特征矩阵进行进一步放大。其操作流程如图 2 所示。

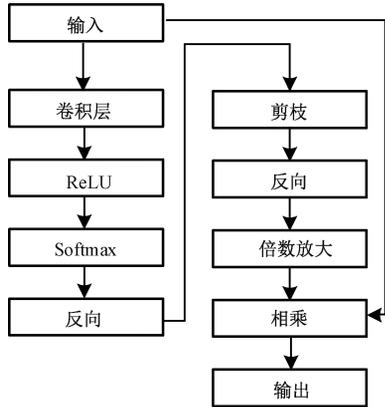


图 2 反向注意力放大机制流程

Fig. 2 Flow chart of reverse attention amplification mechanism

1.3 LSTM

LSTM 结构能够克服模型中梯度的消失和爆炸的缺点,其通过引入“门”结构对细胞状态进行删减或者增添信息 σ 层的输出是 0~1 的值,代表 σ 层所能通过的信息占比。0 表示都不能通过,1 表示都能通过。LSTM 通过 3 个门结构对细胞状态进行控制,称为忘记门、输入门和输出门。忘记门 f_t 通过查看 h_{t-1} 和 x_t 的信息来输出一个 0~1 之间的向量,该向量里面的 0~1 值表示细胞状态中 C_{t-1} 的信息保留或丢弃程度。0 表示不保留,1 表示都保留。

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$c_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t c_{t-1} + i_t c_t \quad (4)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(c_t) \quad (6)$$

其中, x_t 表示 t 时刻的输入, ω 表示权重矩阵, b 表示

偏置矩阵, \hat{c}_t 表示 t 时刻的候选向量, c_t 表示 t 时刻的更新值, h_t, h_{t-1} 表示 $t, t-1$ 时间点模型的所有输出。输入门 i_t 决定 h_{t-1} 和 x_t 的信息更新,然后利用 h_{t-1} 和 x_t 通过一个 tanh 层得到新的候选细胞信息 C_t 。旧细胞中的一部分信息通过忘记门 f_t 选择忘记,输入门选择添加候选细胞信息 C_t 的一部分得到新的细胞信息 C_t 。输出门 o_t 控制着有多少记忆信息将被用于下一阶段的更新中。更新规则如图 3 所示。

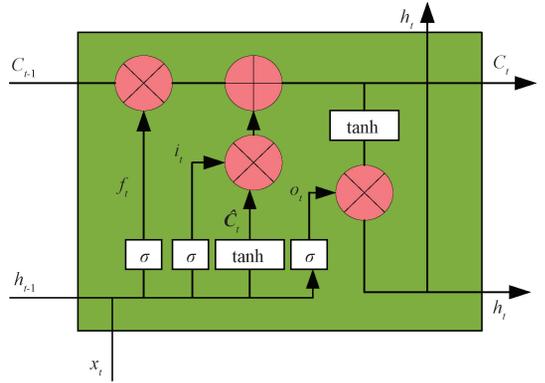


图 3 LSTM 网络结构

Fig. 3 LSTM network structure

2 滚动轴承故障诊断模型设计

LSTM 网络能够学习时间序列中的时序特征,反向注意力放大机制能够对信号中的重要特征进行进一步突出,故本文将两者优势结合,设计了一种基于反向注意力放大机制和 LSTM 网络的轴承故障诊断方法。首先,将轴承一维时序数据按照指定窗口大小以及步幅进行截取作为输入,通过一维卷积层对输入数据进行初步特征提取,再将提取后的特征输入反向注意力放大机制中对重要特征进行突出;为提升网络的非线性逼近能力以及鲁棒性加入批量归一化层,再通过池化层减少参数量后输入第 2 层卷积对突出后的特征进行进一步提取,再由 LSTM 网络对其中的时序信息进行学习,最后经由全连接层、随机失活层后输出诊断结果。模型结构如图 4 所示,其中两层卷积层的卷积核个数均为 128,卷积核大小为 3;池化层中 pooling 大小为 2;LSTM 的隐含层个数为 128;Dropout 层的随机丢弃值为 0.3。

3 实验验证

3.1 实验数据

实验数据采用美国凯斯西储大学的 SKF 型轴承的 DE 端加速度数据,转速为 1 730 r/min,样本采用频率为 48 kHz。实验所选轴承单点直径损伤分别为 0.007、

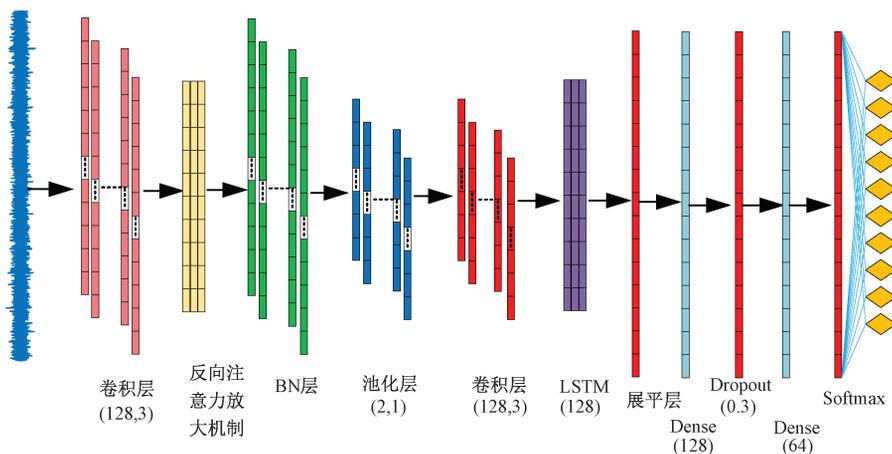


图 4 诊断模型结构

Fig. 4 Diagnostic model structure diagram

0.014、0.021 mm 的故障,每种故障直径内含有滚动体故障、内圈故障以及外圈故障共 3 种故障类型。实验数据集由 9 种故障数据以及一种正常数据组成,生成数据集的截取方法如图 5 所示,通过设置一个定值的窗口对数据进行截取,并对窗口按照指定步幅移动,进行下一次截取。将截取后的数据集的 70% 作为训练集,30% 作为测试集。

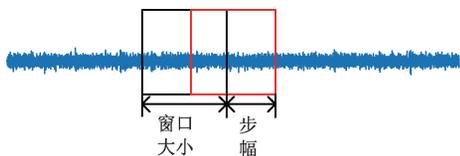


图 5 数据截取示意图

Fig. 5 Schematic diagram of data interception

制层后,并在网络的展平层之前添加一个 LSTM 层。

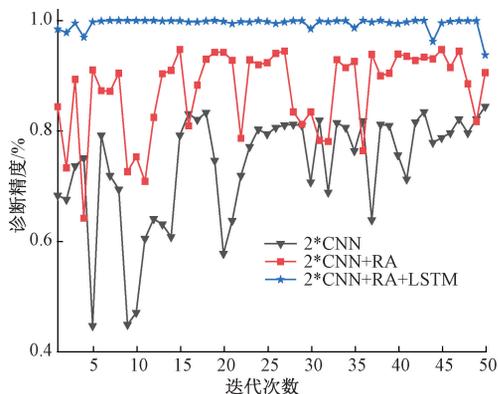


图 6 模型诊断实验结果

Fig. 6 Experimental results of model diagnosis

3.2 实验结果分析

1) 模型诊断效果

轴承故障类型诊断的准确率以及模型的稳定性是判别滚动轴承故障诊断模型优异的重要评判指标,本文以上述指标作为综合评价指标。实验首先验证了网络结构为卷积层、BN 层、池化层以及卷积层的初始模型,后续在模型中加入反向注意力放大机制,实验表明反向注意力放大机制提高了模型的诊断精确率以及稳定性,为进一步提升模型的诊断精度,在模型尾部加入了 LSTM 网络。实验结果表明,所提出的基于反向注意力放大机制与 LSTM 网络的轴承故障诊断方法具有优秀的诊断精度以及稳定性。图 6 为实验结果,其中 2 * CNN 代表模型结构为一层卷积加上 BN 层再加上池化层,并在最后连接一层卷积层;2 * CNN+RA 代表在 2 * CNN 模型的第 1 层卷积后加上一层反向注意力放大机制层;2 * CNN+RA + LSTM 代表在第 1 层卷积后加上一层反向注意力放大机

诊断模型的诊断结果用混淆矩阵表示如图 7 所示。

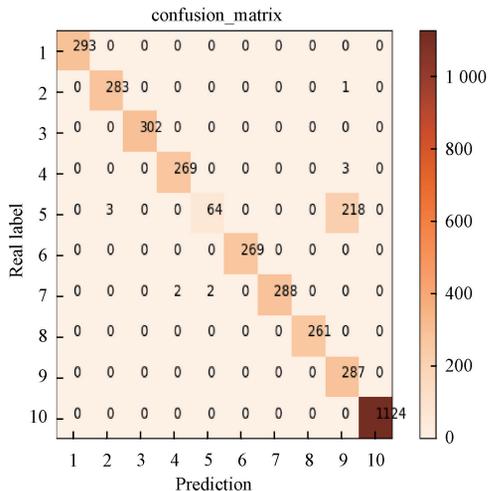


图 7 诊断结果混淆矩阵

Fig. 7 Confusion matrix of diagnostic results

2) 剪枝参数优化

为优化所提出的反向注意力放大机制的性能,通过对算法中的剪枝操作进行超参数对比实验以选取最优参数,由实验结果图 8 以及局部放大图 9 可知,相对于剪枝参数为 0 和 0.7 时,参数设为 0.5 时网络能够在拥有较高诊断精度的同时保持模型的稳定性。在阈值设置为 0.992 5,剪枝参数为 0 时,有 9 次实验结果低于阈值,剪枝参数为 0.5 时有 5 次实验结果低于阈值,剪枝参数为 0.7 时有 8 次实验结果低于阈值。综上,当剪枝参数取 0.5 时能使模型在故障诊断方面达到更好的效果。

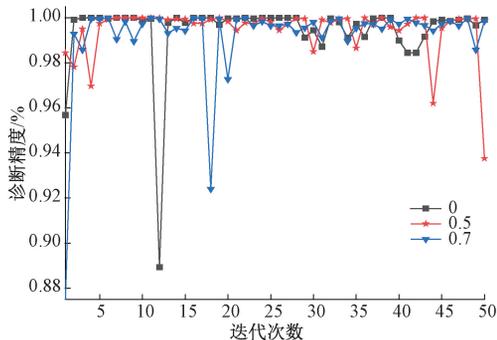


图 8 剪枝参数对比实验结果

Fig. 8 Experimental results of pruning parameter comparison

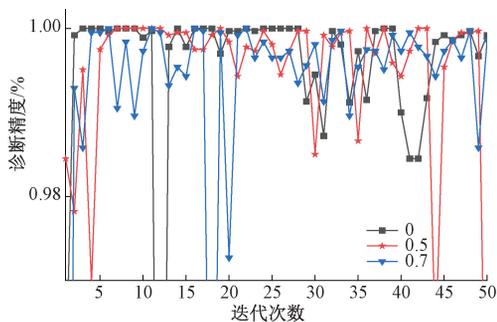


图 9 局部实验结果放大图

Fig. 9 Enlarged view of local experimental results

3) 数据截取长度试验

由于轴承故障数据具有周期性,其截取方式的不同会直接影响所截片段中所包含的故障种类信息,因此设计了在同种诊断模型、不同数据截取长度的对比实验,实验结果如图 10 所示,在诊断精度都能够达到较高水平时,截取长度为 512 的模型具有更优异的稳定性。

4) 特征可视化

为更加清楚的展现 RA-LSTM 模型的诊断效果及特征在网络模型中的分类效果,通过 t-SNE 技术对原始数据、1DCNN、1DCNN+RA、1DCNN+RA+LSTM 模型在二维空间的输出分布分别进行可视化,结果如图 11 所示。相较于其他几种网络模型,在 1DCNN 网络添加了 RA 层之

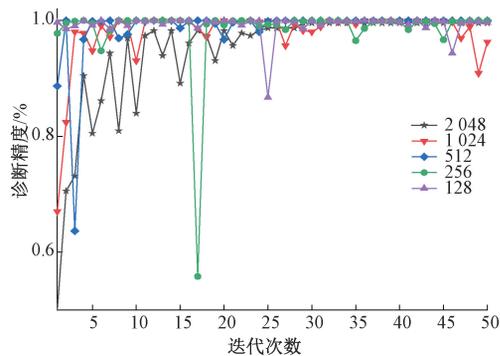


图 10 数据截取长度对比实验结果

Fig. 10 Experimental results of data interception length comparison

后,模型对特征的分类能力得到大幅度提升。

5) 抗噪性实验

为测试本文所提出的轴承故障诊断模型的抗噪性,通过在 10 种状态的轴承数据中分别加入 8、14、20 dB 的白噪声后输入模型中对其进行故障诊断。由图 12 可知,本文所提出的基于 RA-LSTM 的轴承故障诊断方法在添加噪声后仍能达到 100% 的故障诊断精度。

6) 不同故障诊断方法性能对比

最后使用了支持向量机 (support vector machine, SVM), BP 神经网络, 极限学习机 (extreme learning machine, ELM)、概率神经网络 (probabilistic neural network, PNN) 与 RA-LSTM 轴承故障诊断方法进行对比,以体现其诊断性能与鲁棒性。实验结果如表 1 所示,表中第 1 列代表不同的轴承故障诊断方法,第 1 行为所用数据添加的噪声。以诊断方法在不同噪声情况下对故障的最高诊断精度作为评判标准。

表 1 故障诊断方法性能对比实验

Table 1 Performance comparison experiment of fault diagnosis method

精度/%	0 dB	8 dB	14 dB	20 dB
RA-LSTM	100	100	100	100
BP	99.0	70.6	93.0	98.3
SVM	100	68.6	92.0	98.0
ELM	84.0	15.0	19.5	55.0
PNN	87.3	55.0	70.0	83.3

从表 1 可以看出,除 RA-LSTM 以外,其他方法在对添加了噪声的轴承数据进行故障诊断时准确率较低,鲁棒性较低。相比之下,RA-LSTM 在噪声情况下最高准确率仍能达到 100%,鲁棒性较高。

4 结 论

本文提出了一种基于注意力机制的反向注意力放大机制,并结合 LSTM 网络应用于处理轴承故障诊断问题。

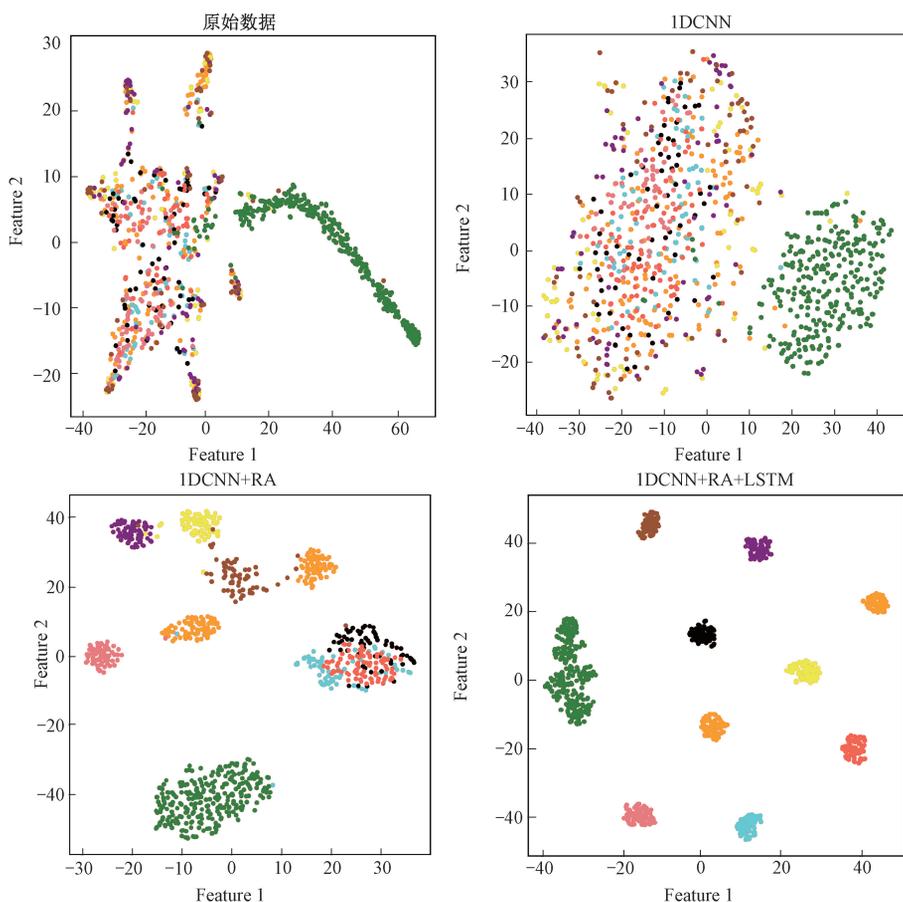


图 11 特征分布图

Fig. 11 Feature distribution map

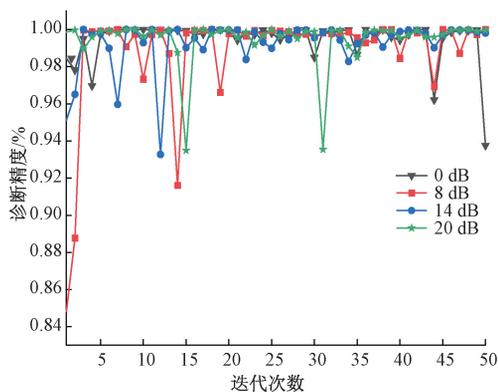


图 12 模型抗噪性实验结果

Fig. 12 Experimental results of model noise immunity

通过实验选取反向注意力放大机制中的剪枝操作超参数以及数据截取长度的最优解,使得模型整体性能达到最优;实验结果表明,在数据截取长度为 512,反向注意力放大机制中剪枝参数为 0.5 时,模型的故障诊断精度能够达到 100%。在模型的抗噪性实验中,本文所提出的模型表现出优异的抗噪性能以及稳定性。

参考文献

[1] 石文杰,黄鑫,温广瑞,等. 基于 DS-VMD 及相关峭度的滚动轴承故障诊断[J]. 振动、测试与诊断,2021, 41(1):133-141,204.
 SHI We J, HUANG X, Wen G R, et al. Fault diagnosis of rolling bearings based on DS-VMD and related cliffness [J]. Journal of Vibration, Measurement and Diagnosis, 2021, 41(1):133-141,204.

[2] 刘泽锐,邢济收,王红军,等. 基于 VMD 与快速谱峭度的滚动轴承故障诊断[J]. 电子测量与仪器学报, 2021,35(2):73-79.
 LIU Z R, XING J SH, WANG H J, et al. Rolling bearing fault diagnosis based on VMD and fast spectral cliffness [J]. Journal of Electronic Measurement and Instrumentation, 2021,35(2):73-79.

[3] 朱哈娜,刘慧明. 基于改进 VMD 与 GS_SVM 的轴承故障诊断[J]. 电子测量技术,2020,43(21):71-76.
 ZHU H N, LIU H M. Bearing fault diagnosis based on improved VMD and GS_SVM [J]. Electronic Measurement Technology,2020,43(21):71-76.

[4] 谷玉海,朱腾腾,饶文军,等. 基于 EMD 二值化图像和 CNN 的滚动轴承故障诊断[J]. 振动、测试与诊断,

- 2021,41(1):105-113,203.
- GU Y H, ZHU T T, RAO W J, et al. Rolling bearing fault diagnosis based on EMD binarized image and CNN [J]. Journal of Vibration, Measurement and Diagnosis, 2021, 41(1):105-113,203.
- [5] 韩涛,袁建虎,唐建,等. 基于 MWT 和 CNN 的滚动轴承智能复合故障诊断方法[J]. 机械传动,2016,40(12):139-143.
- HAN T, YUAN J H, TANG J, et al. Intelligent composite fault diagnosis method for rolling bearings based on MWT and CNN [J]. Mechanical Transmission, 2016,40(12):139-143.
- [6] 田晶,王英杰,王志,等. 基于 EEMD 与空域相关降噪的滚动轴承故障诊断方法[J]. 仪器仪表学报,2018,39(7):144-151.
- TIAN J, WANG Y J, WANG ZH, et al. Rolling bearing fault diagnosis method based on EEMD and airspace correlation noise reduction [J]. Chinese Journal of Scientific Instrument,2018,39(7):144-151.
- [7] 朱少民,夏虹,王志超,等. 改进变分模态分解的核电厂轴承故障诊断[J]. 哈尔滨工程大学学报,2021,42(10):1550-1556.
- ZHU SH M, XIA H, WANG ZH CH, et al. Improved variational modal decomposition for bearing fault diagnosis in nuclear power plants [J]. Journal of Harbin Engineering University,2021,42(10):1550-1556.
- [8] 李文霞,刘林忠,代存杰,等. 基于多种群组组合策略的人工蜂群算法[J]. 计算机应用,2021,41(11):3113-3119.
- LI W X, LIU L ZH, DAI C J, et al. Artificial bee colony algorithm based on multiple swarm combination strategy [J]. Computer Applications, 2021, 41(11):3113-3119.
- [9] 丁嘉鑫,王振亚,姚立纲,等. 广义复合多尺度加权排列熵与参数优化支持向量机的滚动轴承故障诊断[J]. 中国机械工程,2021,32(2):147-155.
- DING J X, WANG ZH Y, YAO L G, et al. Generalized composite multiscale weighted permutation entropy and parameter optimization support vector machine for rolling bearing fault diagnosis [J]. China Mechanical Engineering,2021,32(2):147-155.
- [10] 洪腾蛟,丁凤娟,王鹏,等. 深度学习在轴承故障. HONG T J, DING F J, WANG P, et al. Research on the application of deep learning in the field of bearing fault diagnosis [J]. Science, Technology & Engineering, 2021,21(22):9203-9211.
- [11] 金余丰,姚美常,刘晓锋,等. 残差网络和注意力机制相结合的滚动轴承故障诊断模型[J]. 机械科学与技术,2020,39(6):919-925.
- JIN Y F, YAO M CH, LIU X F, et al. Combined residual network and attention mechanism for rolling bearing fault diagnosis model [J]. Mechanical Science and Technology,2020,39(6):919-925.
- [12] LI D C, ZHANG M, KANG T B, et al. Fault diagnosis of rotating machinery based on dual convolutional-capsule network (DC-CN) [J]. Measurement, 2022, 187: 110258.
- [13] 杨平,苏燕辰,张振. 基于卷积胶囊网络的滚动轴承故障诊断研究[J]. 振动与冲击,2020,39(4):55-62,68.
- YANG P, SU Y CH, ZHANG ZH. Research on rolling bearing fault diagnosis based on convolutional capsule network [J]. Journal of Vibration and Shock, 2020, 39(4):55-62,68.
- [14] JIN T, YAN C, CHEN C, et al. New domain adaptation method in shallow and deep layers of the CNN for bearing fault diagnosis under different working conditions [J]. The International Journal of Advanced Manufacturing Technology, 2021: 1-12.
- [15] 郑直,张华钦,潘月. 基于改进鲸鱼算法优化 LSTM 的滚动轴承故障诊断[J]. 振动与冲击,2021,40(7):274-280.
- ZHENG ZH, ZHANG H Q, PAN Y. Fault diagnosis of rolling bearings based on improved whale algorithm optimized LSTM [J]. Journal of Vibration and Shock, 2021,40(7):274-280.

作者简介



谢锦阳,现为安徽理工大学硕士研究生,主要研究方向为轴承故障诊断。

E-mail: 976161176@qq.com

Xie Jinyang is a M. Sc. candidate at Anhui University of Science and Technology. His main research interests include bearing fault diagnosis and artificial intelligence.



姜媛媛(通信作者),2018年于南京航空航天大学获得博士学位,现为安徽理工大学教授,主要研究方向为电力电子系统故障诊断。

E-mail: jyyl672@163.com

Jiang Yuanyuan (Corresponding author) received her Ph. D. from Nanjing University of Aeronautics and Astronautics in 2018. Now she is a professor at Anhui University of Science and Technology. Her main research interest includes fault diagnosis for power electronic systems.



王力,现为安徽理工大学硕士研究生,主要研究方向为电力电子电路系统故障诊断。

E-mail: 1499728325@qq.com

Wang Li is a M. Sc. candidate at Anhui University of Science and Technology. His main research interests include power electronic circuit fault diagnosis and prediction.