

DOI:10.19651/j.cnki.emt.2108383

基于 SincNet 网络结合注意力机制齿轮箱故障诊断*

杨永灿^{1,2} 刘 韬^{1,2} 王振亚^{1,2} 张益勤^{1,2} 阮 强^{1,2}

(1.昆明理工大学机电工程学院 昆明 650500; 2.云南省先进装备智能维护工程研究中心 昆明 650500)

摘要: 针对传统卷积神经网络(CNN)在齿轮箱中故障诊断准确率不高、特征提取方面表现欠佳的问题,提出了 SincNet 网络结合注意力机制齿轮箱故障诊断方法。首先,采用参数化的 Sinc 函数设计滤波器作为卷积层来代替传统 CNN 的第 1 个卷积层,得到 SincNet 网络结构,提取输入数据的特征信息;其次,结合具有 Softmax 的注意力机制(Att)增强特征信息。最后,采用齿轮箱故障数据集对所提出的方法进行实验验证,结果表明,所提方法平均诊断准确率达到 99.68%,均高于对比方法。此外,通过特征图可视化分析,该方法能够准确定位输入数据中的识别信息,能更好地理解神经网络的特征提取过程,为机械振动信号的特征提取过程提供了参考。

关键词: 齿轮箱;故障诊断;Sinc 函数;SincNet 网络;注意力机制

中图分类号: TN98;TN06;TH165.3 **文献标识码:** A **国家标准学科分类代码:** 510.4030;460.1520

Fault diagnosis of gearbox based on SincNet and attention mechanism

Yang Yongcan^{1,2} Liu Tao^{1,2} Wang Zhenya^{1,2} Zhang Yiqin^{1,2} Ruan Qiang^{1,2}(1. Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, China;
2. Engineering Research Center for Intelligent Maintenance of Advanced Equipment of Yunnan Province, Kunming 650500, China)

Abstract: Aiming at the problem that the traditional convolutional neural network has low accuracy and poor performance in feature extraction in gearbox fault diagnosis, a SincNet combined with attention mechanism method for gearbox fault diagnosis was proposed. First, use the parameterized Sinc function to design the filter and obtain the Sinc convolutional layer, Sinc convolutional layer replace the first convolutional layer of traditional CNN to construct the SincNet network structure, Extract the characteristic information of the input data. Then, combined attention mechanism with Softmax enhances characteristic information. Finally, the gearbox fault data set was used to verify the proposed method. The results show that the average diagnostic accuracy of the proposed method is 99.68%, which is higher than that of the comparison method. In addition, the method can accurately locate the recognition information in the input data and better understand the feature extraction process of neural network through the visual analysis of the feature map, which provides a reference for the feature extraction process of mechanical vibration signals.

Keywords: gearbox; fault diagnosis; Sinc function; SincNet; attentional mechanism

0 引言

齿轮箱作为机械设备中不可或缺的通用部件,在工业领域得到了广泛的应用^[1]。因结构复杂、工作环境恶劣等因素,齿轮箱中的齿轮等核心部件易发生故障。严重时会引起整台设备的损坏,因此,有必要对齿轮箱进行监测和状态评估,保证机械传动系统正常运行^[2]。

随着人工智能和机器学习迅速发展,使故障诊断逐渐智能化,基于数据驱动的智能故障诊断算法日益受到关注。以卷积神经网络为代表深度学习算法^[3],由于其特征提取能力强,故在机械故障诊断领域得到了广泛的应用。这类

方法取代了传统算法繁琐的特征提取,能够自动挖掘输入数据的深层特征,减少了对专家知识的依赖。例如,Zhao 等^[4]建立了批量归一化的卷积神经网络(convolutional neural network, CNN)并结合指数移动平均的故障诊断模型,完成了变工况的轴承故障诊断;杨春柳^[5]提出了 CNN 的多层域自适应的诊断方法,避免了轴承故障诊断中存在的协方差变移位问题;孔子宇等^[6]用 Inception 结构优化残差网络,可直接对滚动轴承振动信号进行诊断并提高了网络的诊断效率;孔子迁等^[7]将时频融合与注意力机制巧妙结合起来,能够准确诊断齿轮箱的各种故障状态;陈星^[8]针对飞

收稿日期:2021-11-17

* 基金项目:国家自然科学基金(52065030,51875272)、云南省重大科技专项计划(202002AC80001)项目资助

机发电机诊断性能下降的问题,利用 CNN 诊断算法提高了飞机发电机故障诊断性能。

虽然深度学习已成功应用于机械故障诊断领域,并且取得很好的效果但仍然存在一些相关的问题。首先,模型的“黑盒性”缺乏工程上的物理解释⁹,无法解释模型学到信号的那些特征,以及如何根据学习的特征来预测结果;其次,由于物理解释性缺乏,使得在工程实际领域无法理解模型的内部机理,从而影响了模型在工程领域的推广。

因此,针对模型学习的特征难以解释性问题,提出了一种基于 SincNet 网络结合注意力机制的故障诊断模型。从模型滤波器的角度设计出发,获得特定的滤波器,以提取嵌入在振动信号中与故障相关的冲击分量,同时结合注意力机制,对提取的冲击分量自适应的加权,提升模型的故障诊断性能。与结合高斯滤波器的一维卷积神经网络(Gaussian-CNN)和一维卷积神经网络(1D-CNN)进行对比,本文所提出的方法在诊断精度、特征图的能量集中程度均优于上述的方法,在工程实际领域有一定的应用价值。

1 理论基础

1.1 SincNet 结构

SincNet 是采用参数化 Sinc 函数实现滤波器,构造网络层,用时域卷积代替频域的乘积,时域卷积的形式对信号频谱进行加窗^[10],从而达到滤波的效果。该方法用 Sinc 函数替代了 CNN 中的卷积核,通过训练帮助 CNN 找到更有意义的过滤器,提取的特征更具有物理意义。为保存机械振动信号的原始结构以及故障相关的信息,因此本文引入的传统 CNN 为一维 CNN(1D-CNN)。

在数字信号处理中,Sinc 滤波器是理想化的带通滤波器。在频域内 Sinc 滤波器相当于一个矩形函数,因此,频域内的 Sinc 滤波器大小可以描述为两个低通滤波器之间的差:

$$G(f, f_l, f_h) = \text{rect}\left(\frac{f}{2f_h}\right) - \text{rect}\left(\frac{f}{2f_l}\right) \quad (1)$$

式中: f 表示频率; f_l 和 f_h 分别对应于带通滤波器的上限截止频率和下限截止频率; $\text{rect}(\cdot)$ 为频域上的矩形函数。

卷积层是 CNN 的主要模块,使用过滤器对输入的信号进行卷积(Conv)操作,并生成相应的特征向量。其卷积过程可描述为:

$$y(t) = x(t) * g(t) = \sum_{l=0}^{L-1} x(l) \cdot g(t-l) \quad (2)$$

式中: $y(t)$ 是卷积后的输出; $x(t)$ 是输入信号; $g(t)$ 表示长度为 L 的过滤器。

为了减少卷积层的参数,使过滤器具有特定的物理意义,将原过滤器替换为只依赖截止频率的过滤器,即在卷积层实现具有上限和下限截止频率的 Sinc 带通滤波器,获得 Sinc 卷积层(SincConv-Layer),对输入的信号进行滤波操作:

$$y(t) = x(t) \cdot g_w(t, f_l, f_h) \quad (3)$$

式中: w 是加窗函数, $g_w(t, f_l, f_h)$ 是时域上的函数, f_l 和 f_h 是在训练过程中需要学习的参数。

时域上的函数 $g_w(t, f_l, f_h)$ 可根据傅里叶逆变换得到,如下所示:

$$g(t, f_l, f_h) = \text{IDFT}(G(f, f_l, f_h)) = 2f_h \text{sinc}(2\pi f_h t) - 2f_l \text{sinc}(2\pi f_l t) \quad (4)$$

式中: Sinc 函数定义为 $\text{sinc}(t) = \sin(t)/t$ 。

通过傅里叶逆变换得到的时域函数 $g_w(t, f_l, f_h)$ 有无限的延迟,如果直接对信号做处理会产生频谱能量泄露,为缓解该问题,将对时域函数 $g_w(t, f_l, f_h)$ 进行加窗函数截断。此处用的是汉明窗,得到所需滤波器的函数:

$$g_w(t, f_l, f_h) = g(t, f_l, f_h) \cdot w(t) \quad (5)$$

Sinc 卷积操作后接着采用激活函数对捕获的特征进行非线性映射,进一步增强所提取特征的分度。此处采用修正线性单元(rectified linear unit, ReLU)激活函数:

$$R(n) = f(y(t)) = \max[0, y(t)] \quad (6)$$

式中: $R(n)$ 是被激活的特征映射, $f(\cdot)$ 是激活函数。

由于卷积操作后,获得了大量的特征映射,因此为了有效减轻计算量,从而减少所需的计算资源和时间,利用池化层进行下采样操作。为了更好地保留数据纹理特征,本文采用最大池化(MaxPooling)操作,函数表示如下^[11]:

$$p_j^l = \max_{(j-1)W+1 \leq t \leq jW} \{a_j^{l-1}(t)\} \quad (7)$$

式中: $a_j^{l-1}(t)$ 为第 $l-1$ 个特征面的第 j 个特征矢量中第 t 个激活值, $t \in [(j-1)W+1, jW]$, W 表示池化区域的宽度。

在一系列堆叠卷积和池化之后,最终特征映射被扁平,输入全连接(FC)层,通常连接 Softmax 函数实现类别分类。为了减小损失率,利用交叉熵^[12]作为模型训练中的损失函数,可表示为:

$$H(r, p) = - \sum_i r_i \log(p_i) \quad (8)$$

式中: i 是样本类别数目, $r \in [0, 1]$ 是样本标签, p 是从 Softmax 函数中得到的概率。

1.2 注意力机制

为了获取不同信号片段对结果的影响程度,通过引入注意力机制(attention mechanism, Att)对不同信号片段自适应的生成权重进行信息筛选,抑制不符合条件的信息,让它的权重为 0,使得网络关注于在输入数据中发现与输出有关的有用信息,从而使网络能够获取更多有判别性的信息。

注意力机制的过程分为注意力权重生成和权重分配两个阶段。首先获取网络输出的特征序列 $\mathbf{H} = [h_1, h_2, h_3, \dots, h_N]$, N 表示长度;接着生成特征序列的注意力权重向量 $\mathbf{S} = [s_1, s_2, s_3, \dots, s_N]$ 。最后,将注意力权重 \mathbf{S} 与原始特征序列相乘得到注意力筛选后的特征序列,其过程可描述如下:

$$\omega_f = \delta(\mathbf{H}) \tag{9}$$

$$\mathbf{S} = \text{softmax}(\alpha^T \omega_f) = [s_1, s_2, s_3, \dots, s_N] \tag{10}$$

$$\mathbf{H}' = \mathbf{H}\mathbf{S}^T \tag{11}$$

式中: $\delta(\cdot)$ 是注意力机制的激活函数, α^T 是可学习的参数, \mathbf{H}' 为注意力筛选后的特征。

1.3 基于 SincNet 结合 Att 故障诊断方法

1) 模型构建

齿轮箱由于受交变工况的影响,不同时刻内采集的信号可能具有不同的频带;另外随机初始化的卷积核很难从一维振动信号中提取与故障信号相关的冲击分量,从而使获得的故障特征失去物理意义。故障诊断模型的第 1 层比较关键,提取的特征会影响整个模型^[13],本文将 SincNet 结合注意力机制技术构建故障诊断模型,即基于 Sinc 函数的卷积(SincConv)层取代 CNN 的第 1 个卷积层,使 CNN 的第 1 层更专注于提取输入信号的冲击分量,从而使结果有明确的物理意义^[14],后面采用常用的 CNN 层;接着采用注意力机制对所提取的特征自适应的生成权重进行信息筛选。其网络结构如图 1 所示。

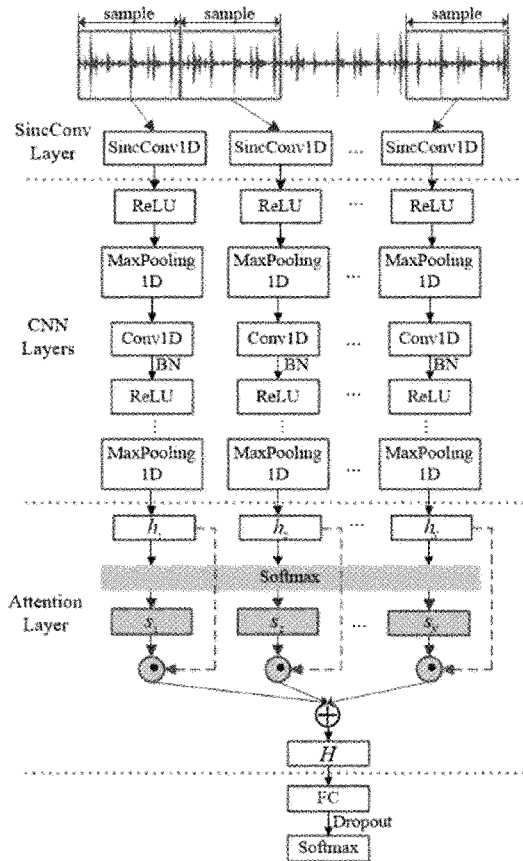


图 1 故障诊断模型

Sinc 函数的卷积形式主要是对输入的故障信号频率分辨率进行识别,提高频域的判别信息,获得表征齿轮箱故障相关的冲击分量。常用的 CNN 层是以交替方式出现卷积和池化层,进一步对信号做深度特征提取;另外随着网络结

构的深化,其特征分布会发生偏移或变动,故在卷积层之后使用批量归一化(batch normalization, BN)^[15]消除特征分布差异,确保模型的性能。提取深度特征后,引入注意力机制技术对特征序列自适应的生成权重进行信息筛选,抑制不符合条件的特征,同时设置 Dropout^[16]机制,在训练过程中随机的丢弃部分权重,提高网络的泛化能力;最后将筛选好的特征序列送到 Softmax 分类器,实现分类诊断。

2) 齿轮箱故障诊断流程

基于 SincNet 结合 Att 的齿轮箱故障诊断方法流程图如图 2 所示,主要分为训练和测试两个阶段,具体步骤信息如下。

(1) 齿轮箱振动信号的采集与截断:采集齿轮箱故障信号数据,将所获取信号数据按等长度截断得到样本,并将其随机划分为训练样本和测试样本。

(2) 故障诊断模型的构建与训练:构建故障诊断模型,将模型参数初始化,利用训练样本数据训练模型参数,根据准确率等因素不断调整模型参数,直至满足要求,并保存模型。

(3) 诊断测试:在训练结束后,使用测试样本数据对模型进行测试,验证其诊断性能。

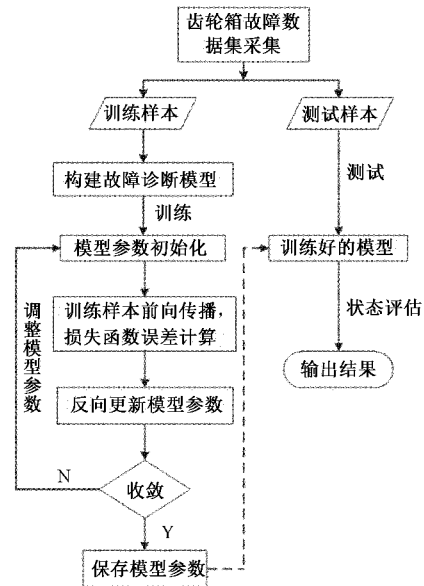


图 2 模型故障诊断流程

2 实验验证

2.1 实测 RV 减速器齿轮箱故障实验

RV(rotate vector reducer)减速器齿轮箱故障试验平台如图 3 所示,试验台主要包括伺服电机、RV 减速器和摆臂。其中,RV 减速器是两级减速传动机构,由行星齿轮减速机(含 1 个 12 齿的太阳轮和 2 个 42 齿的行星轮)和摆线针轮减速机(含 1 个 93 齿的摆线轮和 40 齿的针轮)组成的两级减速传动机构。摆臂主要是为试验台设置负载,实现

实验台的不同负载状态。实验在太阳轮和行星轮分别模拟设置了 3 种故障模式,故障类型为:齿根裂纹(齿根处切深 0.3 mm)、单齿齿面磨损(磨损深度为 0.5 mm)、多齿齿面磨损(3 个齿磨损,磨损深度分别为 0.5、0.3、0.1 mm),如图 4 所示。加上完全正常的减速器状态,故共有 7 种 RV 减速器健康状态。在 RV 减速器表面安装加速度传感器(型号为 PCB353B01),利用 National Instrument 9234 采集卡对信号进行采集,通过 NI_SignalExpress2015 软件平台进行数据记录,采样频率设置为 25.6 kHz。减速器稳定运行时,采集振动信号,排除随机因素的影响,对 RV 减速器的每种健康状态的不同负载进行数据采集。

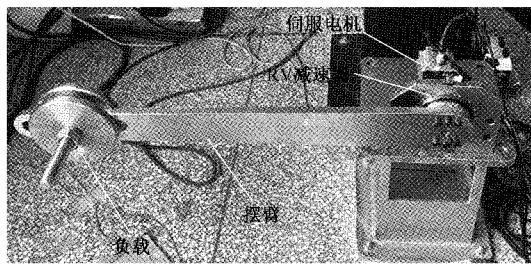


图 3 模拟试验台

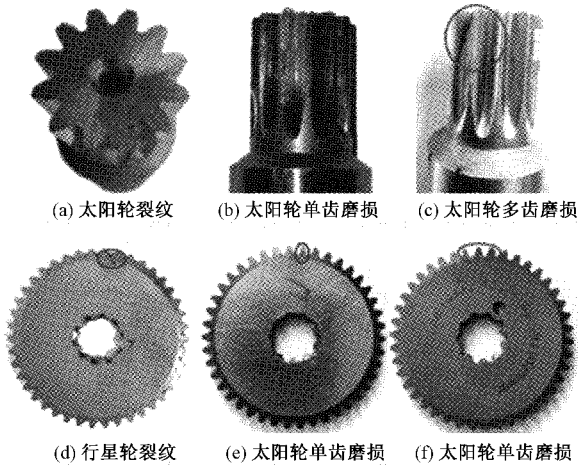


图 4 RV 减速器齿轮箱故障模式

实验开始前先对正常情况下的 RV 减速器数据进行采集。本文实验选取 8 kg 负载状态下的 RV 减速器振动信号进行分析。对获得的实验信号采用滑动窗口截断,其中窗口的长度为 22 800,最终得到 RV 减速器 7 类故障状态各自均有 300 组数据样本,随机划分为训练样本和测试样本,具体信息如表 1 所示。

为降低结果的随机性,每类故障状态的训练和测试循环迭代 30 次。采用网格搜索算法结合深度学习模型参数选择的一般规律来选择模型的主要参数,经过多次选择调试,最终得到的模型主要参数如表 2 所示,其中批处理量为 64。

表 1 齿轮箱的 7 种健康状态

工况状态	训练样本	测试样本	类别
正常	210	90	0
行星轮多齿磨损	210	90	1
行星轮单齿磨损	210	90	2
行星轮裂纹	210	90	3
太阳轮多齿磨损	210	90	4
太阳轮单齿磨损	210	90	5
太阳轮裂纹	210	90	6

表 2 模型主要参数

结构名称	结构参数	激活函数
SincConv	滤波器数目/尺寸 100/64×1	ReLU
池化层 1	池化尺寸 4×1	
卷积层	滤波器数目/尺寸 32/3×1	ReLU
池化层 2	池化尺寸 4×1	
注意力层	节点数 96	
Dropout	丢弃率 0.25	
输出层	—	Softmax

2.2 诊断结果与分析

为展示本文模型的性能,设计了用高斯滤波卷积层替换传统 CNN 第 1 层构建网络(Gaussian-CNN)和传统 1D-CNN 对同一实验数据进行分析并比较。激活函数、优化器、损失函数、Dropout 比例和迭代次数等均和本文构建的模型设置一样。利用交叉熵损失函数计算真实标签与预测标签概率之间的误差得到准确率,为了验证网络的稳定性以及降低随机性的影响,循环迭代 30 次平均实验结果为训练样本数据和测试样本数据的最终结果,如表 3 所示。

表 3 实验结果

准确率/%	SincNet-Att	Gaussian-CNN	1D-CNN
训练	100	94.43	93.24
测试	99.68	92.22	91.1

从表 3 中可以看出,1D-CNN 模型因其结构简单,故诊断准确率较差,为 91.1%。Gaussian-CNN 模型诊断准确率虽提高了 1.12%,但也无法全面地提取到信号所包含的特征,导致诊断准确率都偏低。无论是从故障特征的体现还是诊断准确率的结果来说,笔者所提方法可以充分提取到信号的特征,再引入注意力机制进行自适应加权,模型的平均诊断准确率可达到 99.68%。为进一步验证本文模型的识别能力,引入多分类混淆矩阵^[11],详细分析了本文模型和对比模型齿轮箱的识别结果,如图 5 所示。

图 5 为不同模型的多分类混淆矩阵。分析可知,图 5(a)较图 5(b)与(c)的预测标签与实际标签具有更好地对应度,可以说明本文所提模型具有更好地识别效果。

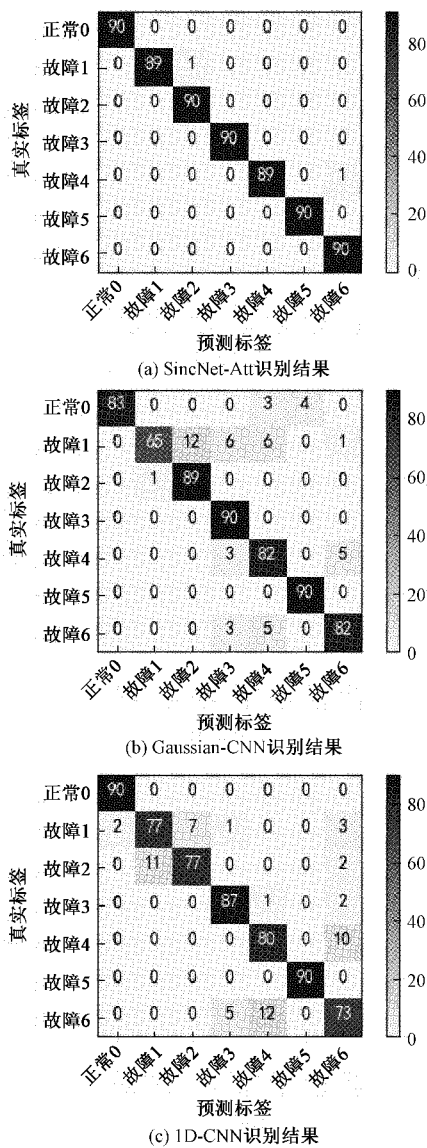


图 5 不同模型的识别结果

为进一步直观展示本文模型对齿轮箱不同状态振动信号数据的诊断能力,采用 t-SNE(t-distribution stochastic neighbor embedding)降维可视化分析,对本文模型和对比模型提取的特征进行展示,分别将各个方法所提取的特征分布情况在二维空间中表示,t-SNE 捕获的特征分布情况如图 6 所示。

从图 6 中可以看到,原始振动信号通过 SincNet-Att 模型自适应的学习处理之后,在二维空间中齿轮箱大部分同一故障类型的样本很好地聚集在一起,并且有效地区分出不同故障类型齿轮箱。Gaussian-CNN 和 CNN 模型中同一类别的样本没有很好的聚集在一起,甚至不同类别的样本特征映射点有相互重叠。总之,t-SNE 从特征分类能力的角度也证明了本文模型能有效地自适应提取特征,进而根据提取出的有效特征对设备进行精确故障诊断。

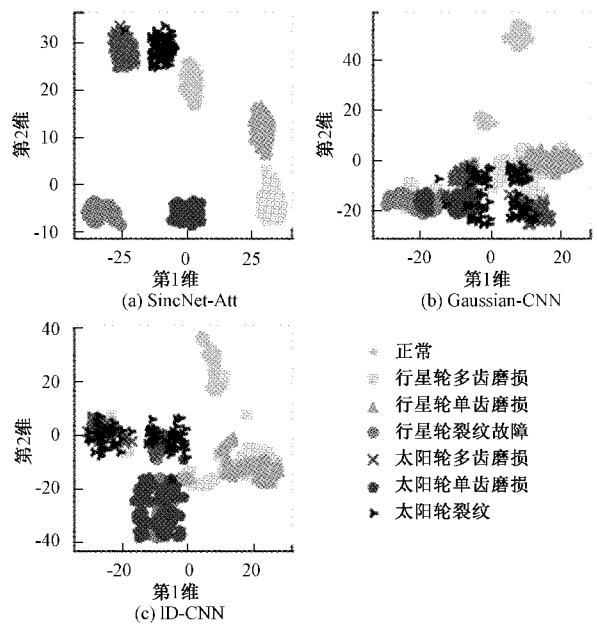


图 6 可视化

为了说明设计的 SincConv 层能够提取具有特定的物理意义的特征,将从测试集样本中随机选取信号样本输入到训练好的网络模型中,对经过 SincConv 层处理之后的特征图与经过其他模型第 1 层的特征图进行比较,如图 7 所示。

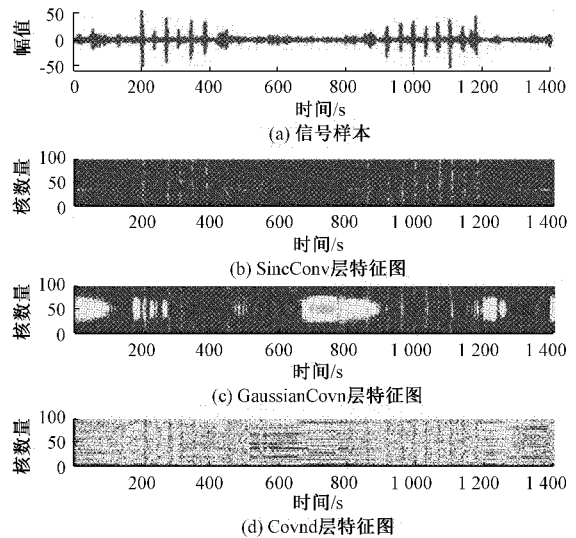


图 7 特征图

从图 7 中可以看出,SincConv 层的特征图能量比较集中,准确显示了输入信号的主要冲击分量的位置。GaussianCovnd 层的特征图中提取了与信号无关的特征分量,难以快速识别输入信号的主要冲击分量;Covnd 层的特征图中的能量分散,冲击分量非常模糊。通过以上特征图对比可以得出,SincConv 层能够有效地滤除与故障无关的信息,重点关注信号的主要冲击分量,即可以学习到信号的识别特征,符合故障诊断的基本规律。

3 结 论

为探讨深度学习在齿轮箱故障诊断中的应用,本文对卷积操作进行了优化,设计了一种具有发现物理意义的滤波器,来取代卷积层的滤波器,提出了基于 SincNet 网络结合注意力机制齿轮箱故障诊断方法。利用模拟试验台 RV 减速器齿轮箱故障数据集验证,发现本文模型可以有效地滤除故障无关的信息,并重点关注信号的主要冲击分量,与故障诊断的基本规律有相似性。通过与其他模型对比,本文模型有良好的诊断性能。未来的工作,我们将进一步设计更深入的 SincConv 层的网络结构以及探索通过注意力机制获得的信息片段的高维特征与最终输出之间的关系。

参考文献

- [1] ZHANG M, WANG K S, WEI D, et al. Amplitudes of characteristic frequencies for fault diagnosis of planetary gearbox[J]. *Journal of Sound & Vibration*, 2018, 432:119-132.
- [2] QIU G, GU Y, CAI Q. A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions [J]. *Measurement*, 2019, 145:94-107.
- [3] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553):436-444.
- [4] ZHAO B, ZHANG X, HAI L, et al. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions[J]. *Knowledge-Based Systems*, 2020, 199: 105971.
- [5] 杨春柳. 基于卷积神经网络的多层域自适应滚动轴承故障诊断[J]. *电子测量与仪器学报*, 2021, 35(2):122-129.
- [6] 孔子宇, 王海瑞. 基于 Inception-ResNet 模型的轴承故障分类[J]. *电子测量技术*, 2021, 44(5): 55-62.
- [7] 孔子迁, 邓蕾, 汤宝平, 等. 基于时频融合和注意力机制的深度学习行星齿轮箱故障诊断方法[J]. *仪器仪表学报*, 2019, 40(6): 221-227.
- [8] 陈星. 基于卷积神经网络算法的飞机发电机故障诊断研究[J]. *国外电子测量技术*, 2020, 39(11):57-60.
- [9] YANG Z B, ZHANG J P, ZHAO Z B, et al. Interpreting network knowledge with attention mechanism for bearing fault diagnosis[J]. *Applied Soft Computing*, 2020, 97:106829.
- [10] 刘人源. 端到端语音合成算法实现与改进[D]. 昆明: 云南大学, 2020.
- [11] 曲建岭, 余路, 袁涛, 等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法[J]. *仪器仪表学报*, 2018, 39(7):134-143.
- [12] MIN X, TENG L, LIN X, et al. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks [J]. *IEEE/ASME Transactions on Mechatronics*, 2018, 23(1):101-110.
- [13] LI T, ZHAO Z, SUN C, et al. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis [J]. *ArXiv Preprint*, 2019, ArXiv:1911.07925.
- [14] RAVANELLI M, BENGIO Y. Interpretable convolutional filters with SincNet[J]. *ArXiv Preprint*, 2018, ArXiv:1811.09725.
- [15] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. *Proceeding of the 32nd International Conference on Machine Learning*, JMLR.org, 2015:448-456.
- [16] 彭鹏, 柯梁亮, 汪久根. 噪声干扰下的 RV 减速器故障诊断[J]. *机械工程学报*, 2020, 56(1):30-36.

作者简介

杨永灿, 硕士研究生, 主要研究方向为基于深度学习的设备状态监测等。

E-mail:1772216040@qq.com

刘韪(通信作者), 博士, 副教授, 主要研究方向为设备状态监测与健康评估、机器学习在信号处理中的应用等。

E-mail:kmlitao@aliyun.com