

改进卷积胶囊网络的滚动轴承故障诊断方法

赵小强^{1,2,3}, 柴靖轩¹

(1. 兰州理工大学电气工程与信息工程学院, 甘肃 兰州 730050;

2. 甘肃省工业过程先进控制重点实验室, 甘肃 兰州 730050;

3. 兰州理工大学国家级电气与控制工程实验室教学中心, 甘肃 兰州 730050)

摘要: 目前许多基于卷积网络的滚动轴承故障诊断方法受噪声信号以及负荷变化的影响, 存在诊断效果不佳、泛化能力差的问题。针对此问题提出一种改进卷积胶囊网络的滚动轴承变工况故障诊断方法。该方法设计了多尺度非对称卷积模块, 其中采用不同尺度的非对称卷积层对输入数据进行特征提取, 在实现最大化提取数据中的特征信息的同时, 还能够有效减少参数量; 在该模块中引入通道注意力机制, 能更好地提取有用的通道特征, 提高该方法特征提取的能力; 通过将网络中的全连接层改进为胶囊全连接层, 使得胶囊在输出向量特征信息时, 避免了特征信息在空间中的丢失。使用凯斯西储大学轴承数据集和东南大学变速箱数据集来验证所提方法的诊断性能, 并与其他深度学习学习方法进行了比较。实验结果表明, 与其他深度学习学习方法相比, 具有较好的泛化性, 效果更佳。

关键词: 故障诊断; 滚动轴承; 胶囊网络; 非对称卷积; 特征提取

中图分类号: TH165⁺.3; TH133.33⁺1 **文献标志码:** A **文章编号:** 1004-4523(2024)05-0885-11

DOI: 10.16385/j.cnki.issn.1004-4523.2024.05.017

引言

旋转机械中的滚动轴承是关键部件, 其运行状态直接影响到整个旋转机械的工作过程^[1], 发生故障可能会损坏整个设备, 造成重大损失。因此, 对滚动轴承进行更准确、更智能的故障诊断是减少经济损失的重要保障^[2]。

故障诊断技术向智能化阶段发展, 其中人工智能的算法对各种故障类型的识别和分类尤为重要^[3-4]。作为人工智能的主要方法, 机器学习可以有效地学习数据信息, 模拟并实现人类学习行为, 不断改进和提高自身的性能, 取得更好的学习效果。贝叶斯分类器^[5]、支持向量机^[6]和神经网络^[7]等机器学习方法虽然能够提取、识别与分类所需的故障特征, 但是在提取轴承的故障信号特征时, 往往使用复杂的数学工具, 针对不同类型的故障所采用的特征提取方法也不同。传统的故障特征提取方法过于依赖信号处理的相关知识以及故障诊断专家的经验^[8-9], 这对故障检测的效率和成本都是很大的考验。

与传统的人工智能故障诊断方法相比, 深度学习技术不仅直接处理工业生产系统的检测信号, 而且不过多依赖信号处理和专家经验。近年来, 深度

学习技术在轴承故障诊断领域取得了快速发展^[10-11], 例如卷积神经网络^[12]、深度置信网络^[13]、稀疏自编码器^[14]和卷积自编码网络^[15]等方法通过多层非线性特征提取结构, 将振动信号进行多层非线性变换, 可以比较充分地提取故障特征, 诊断生产过程中的设备故障类别。然而在实际的工业应用中, 传感器采集到的振动信号会无法避免地被现场各种噪声所污染^[16-17]。此外, 轴承的转速会因负荷的变化而变化。鉴于各种复杂的系统组件及其内部工作原理, 通常存在错综复杂且强相关的耦合关系。因此, 传统的神经网络在诊断受噪声干扰和负荷变化影响的轴承故障时, 会存在诊断准确率低、泛化能力差的问题。

胶囊网络由 Sabour 等^[18]提出。与传统的神经网络不同, 胶囊网络中的每个神经元都不是由传统的标量组成, 而是由向量组成。因此, 胶囊网络可以从输入数据中提取并存储更详细的特征, 同时有效地减少特征信息的损失。初始胶囊网络提取特征时仅使用了单卷积; 杨平等^[19]提出双卷积层胶囊网络的诊断方法, 但双卷积层胶囊网络只增加了一个卷积层和池化层^[18], 使得原始图像数据特征未被进一步提取, 导致细节特征提取不够完整, 其特征提取能力还有待改善。

收稿日期: 2022-05-18; **修订日期:** 2022-08-19

基金项目: 国家自然科学基金资助项目(61763029, 62163023); 甘肃省科技计划资助项目(21YF5GA072); 甘肃省教育厅产业支撑计划项目(2021CYZC-02)。

针对上述胶囊网络存在的问题,本文提出一种基于卷积胶囊网络的故障诊断方法。首先使用不同尺度非对称卷积层和注意力机制构建多尺度非对称卷积结构,然后使用胶囊全连接层改进了胶囊网络,最后结合多尺度非对称卷积和改进的胶囊网络进行特征分类,可以在强噪声影响或变负荷的情况下诊断轴承的故障,提高故障诊断的准确率。

1 基础理论

1.1 Inception 卷积结构

在深度学习中,卷积神经网络(CNN)^[20]是一种前馈神经网络,具有自动提取特征的能力,它通过多个卷积核逐层提取输入数据中的深层特征,卷积核的每个元素都包含一个权重系数和偏差,并对输入图像或数据执行局部卷积运算。

Inception 结构使用不同大小的卷积核,通过堆叠来增加网络宽度,从而可以提取丰富的特征信息,同时利用 1×1 尺度的卷积核对输入的特征图进行降维,减少参数,加快网络计算及训练速度。由于激活函数的增加,多层卷积核的非线性表达能力得到提升;卷积层的深度增加,避免了一定的梯度弥散。目前,Inception 模块已经更新到 Inception v4 和 Inception-ResNet 模块^[21]。

1.2 通道注意力机制

注意力机制模仿人观察事物、关注重点部位的特点,在自然语言处理、统计学习及语音识别等领域中^[22]被广泛应用。通道注意力机制能更好地提取有用的通道特征,增强模型特征提取的能力。通道注意力模块由三部分组成:压缩模块、激励模块和注意力模块^[23]。压缩模块由池化层将各通道内的全局空间特征信息进行求和压缩,形成各自的通道特征,该特征能反映全局的通道特征信息,等同扩大了网络的感受野;激励模块在增强模块迁移能力的同时降低模块的参数数量;注意力模块在每个通道域上将

原卷积相应的通道特征值与得到的特征权重进行加权相加,使卷积通道特征表现不同的权重,提取出表征目标中的关键信息。

1.3 胶囊网络

胶囊网络使用卷积层来对二维输入信息的各个区域进行卷积计算,并将计算结果堆叠,形成卷积层的输出。传统深度学习方法通常使用最大池化层来实现静态不变性,由于最大池化层持续搜寻二维矩阵的区域并选取区域内最大的计算结果,因此有价值的信息容易损失,并且对编码特征间的相对空间关系缺乏考虑,本文方法使用动态路由算法替代最大池化层,胶囊层提取的重要特征信息以向量形式被胶囊封装,普通标量神经元与胶囊向量神经元如图 1 所示。

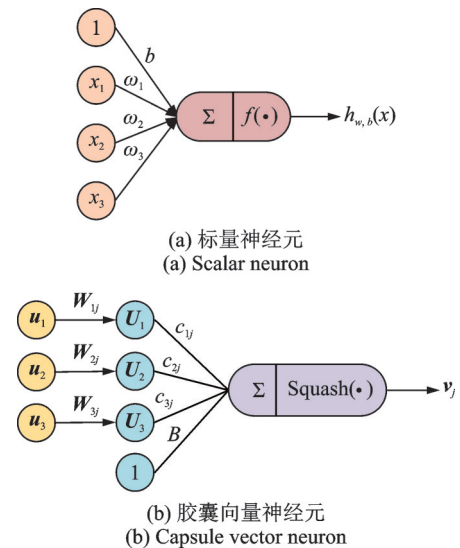


图 1 两种神经元对比

Fig. 1 Comparison of the two neurons

在图 2 胶囊网络结构中,设定上层胶囊为父胶囊,下层胶囊为子胶囊。除第一层胶囊外,其余下层胶囊的总输入 s_j 是来自子胶囊的所有“预测向量” \hat{u}_{ji} 的加权和,记父胶囊层第 i 个神经元为 u_i , W_{ij} 为与神

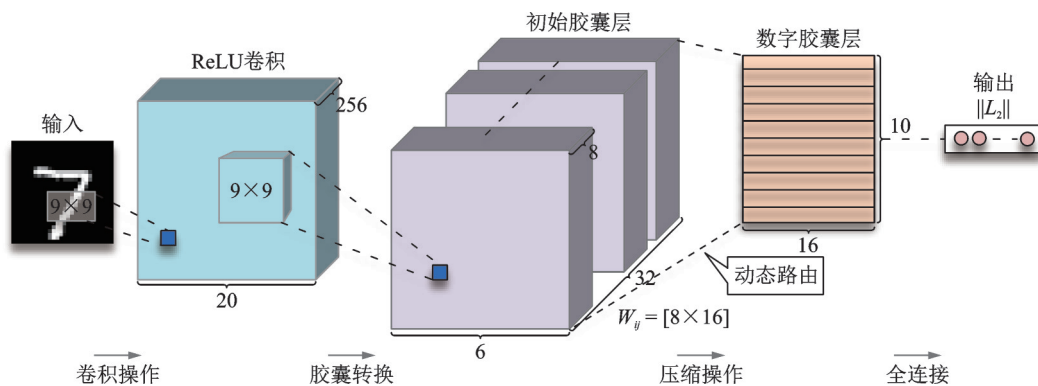


图 2 胶囊网络结构图

Fig. 2 Capsule network structure diagram

经元 u_i 相乘的权重矩阵。用非线性 Squash 函数替代传统激活函数,将总输入 s_j 传递到胶囊网络中,得到最终的输出向量 v_j ,其表达式为:

$$s_j = \sum_i c_{ij} \hat{u}_{ji} \quad (1)$$

$$u_{j|i} = W_{ij} u_i \quad (2)$$

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (3)$$

式中 c_{ij} 是由迭代动态路由过程确定的耦合系数,其目的是使输入的神经元能够自主选择最好的路径传输到下一层神经元。胶囊 i 与上一层,胶囊 $i-1$ 层中所有胶囊之间的耦合系数之和为 1,由“路由 Softmax”确定,其初始 Logits 函数 b_{ij} 为胶囊 i 连接到胶囊 j 的对数先验概率,其表达式为:

$$c_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}} \quad (4)$$

$$b_{ij} = b_{ij} + v_j u_{ji} \quad (5)$$

在每一次动态路由的前向传播中,都先将 b_{ij} 初始化为 0,使用式(5)来更新 b_{ij} 的值,从而更新 c_{ij} 的值,通过前向传播进一步修正 s_j 的值,改变输出向量 v_j 的值,通过动态路由迭代循环,最终得到一组最佳的耦合系数。

2 改进卷积胶囊网络的滚动轴承故障诊断方法

滚动轴承的工作环境通常伴随强噪声、变负荷等情况,导致实际轴承检测的振动信号容易受到外界因素干扰。为了充分利用胶囊网络的特征提取能力,同时改进神经网络需要处理大量的数据和时间来提高学习能力的缺点,并预防梯度损失的问题,本文提出了改进卷积胶囊的滚动轴承变工况故障诊断方法,设计了改进的多尺度非对称卷积模块和改进全连接胶囊层并引入通道注意力机制,实现对强噪声和变工况下故障的诊断。

2.1 多尺度非对称卷积结构

为使胶囊网络从原始振动信号中获取更多的有用信息,并提升胶囊网络提取特征的速度,本文提出了一种改进的多尺度非对称卷积模块。该模块基于 Inception 结构,相比于普通的对称 $k \times k$ 卷积核的卷积,非对称卷积是将一个 $k \times k$ 的卷积拆解成一个 $k \times 1$ 的卷积再串联一个 $1 \times k$ 的卷积,两者的感受野是相同的,但非对称卷积能有效地减少参数量和运算量^[24],因为多个大小兼容的二维核使用相同的步长对同一输入进行操作,这些卷积核会产生具有相

同分辨率的输出。在求和非对称卷积的输出后,相应位置的卷积核会相互叠加,从而形成一个等效的卷积核,其输出与原始输出相同^[25-26],其结构如图 3 所示:在第一层中,使用 1×1 结构, 3×1 与 1×3 的串联,以及 5×1 与 1×5 的串联结构。且并行的卷积层对输入故障数据进行不同尺度的特征提取,通道数设置为 16, 8, 8; 第二层将 5×5 , 3×3 的卷积分别等效拆解成 5×1 与 1×5 , 3×1 与 1×3 的非对称卷积,并与第一层串联,通道数都为 16; 第三层使用两个 1×1 且通道数都为 32 的卷积层。为了增加模型的非线性表达能力,在每个卷积层之后使用批量标准化和 ReLU 激活函数,然后通过 Concat 层对不同分支的特征维度进行堆叠拼接,并使用通道注意力机制获得不同特征信息的重要程度,依据重要程度增强有用特征,抑制干扰。

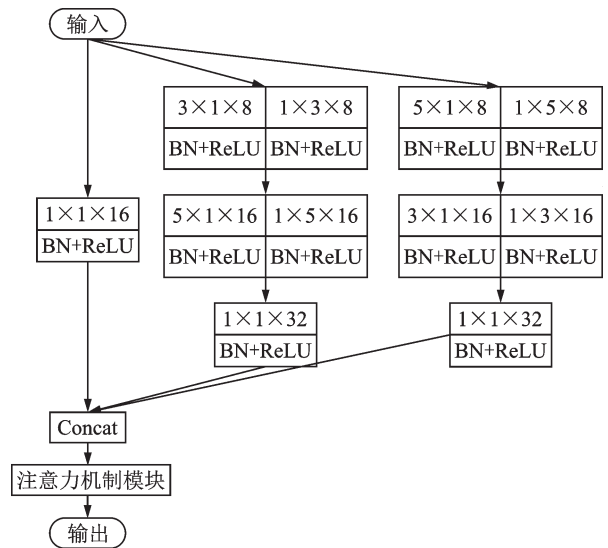


图 3 多尺度非对称卷积模块

Fig. 3 Multi-scale asymmetric convolution module

2.2 胶囊全连接层

全连接层(Fully Connected Layers)在整个神经网络中起到将学习到的“分布式特征表示”映射到样本标记空间的作用^[27]。因为全连接层的特性参数在网络中也是最多的,所以全连接层在综合前边提取的特征时,参数相对冗余,在模型进行最后故障分类时,存在特征信息丢失的问题。为使模型达到更好的故障诊断效果,本文改进了一种胶囊全连接层^[28]来替代传统全连接层。胶囊全连接层是由子胶囊构成,子胶囊保留上层父胶囊提取的特征信息,如图 4 所示。当模型在最后压缩输出时,下层的胶囊被展平成胶囊列表,送入胶囊全连接层,胶囊层中每个子胶囊乘以转换矩阵 W^d ,通过协议路由,以产生适用于每个类别的最终胶囊 v_j 及其概率 a_j 。因为协议路由内部会使用 Softmax 函数产生胶囊间的对数先验

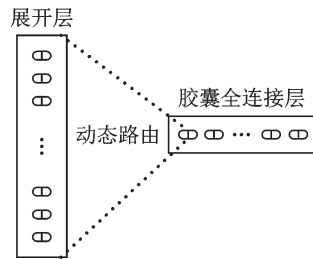


图4 胶囊全连接层

Fig. 4 Capsule fully connected layer

概率 b_{ij} , 所以本文使用 ReLU 激活函数在全连接层内部进行一次再压缩, 针对每个最终胶囊 v_j 进行解码。改进因胶囊全连接层内部的 Softmax 函数被过多使用, 导致数值溢出的问题; 所以在其内部使用 ReLU 函数进行单侧抑制, 增强运算效率的同时, 并没有像传统前连接层一样丢失过多的特征信息。因为各个胶囊尽可能地保留了特征向量, 所以在将各个数字胶囊层提取的特征信息相连时, 依旧保持较

高的信息量, 在最后使用 ReLU 函数来抑制过拟合, 可以提高模型的分类准确度。

2.3 网络结构

为提高卷积胶囊网络的特征提取效率, 利用改进的多尺度特征提取模块进行前端特征提取, 该模块采用多通道不同尺度的非对称卷积核, 有效地减少特征提取时的参数量。采用通道注意力机制获得更多更有效的故障数据信息, 并减少特征提取时的模型计算量, 与主胶囊层相结合, 将前端提取的标量特征信息转换为向量特征信息, 在空间上存储了更多的有效特征信息。以动态路由算法替代了传统最大池化实现特征传递, 并在后端数字胶囊层与输出层之间使用全连接胶囊层进行下一步的特征信息全连接和特征分类的输出传递, 保证特征信息尽可能地存储在胶囊中, 经过交叉熵损失函数输出故障诊断结果, 其结构如图 5 所示。

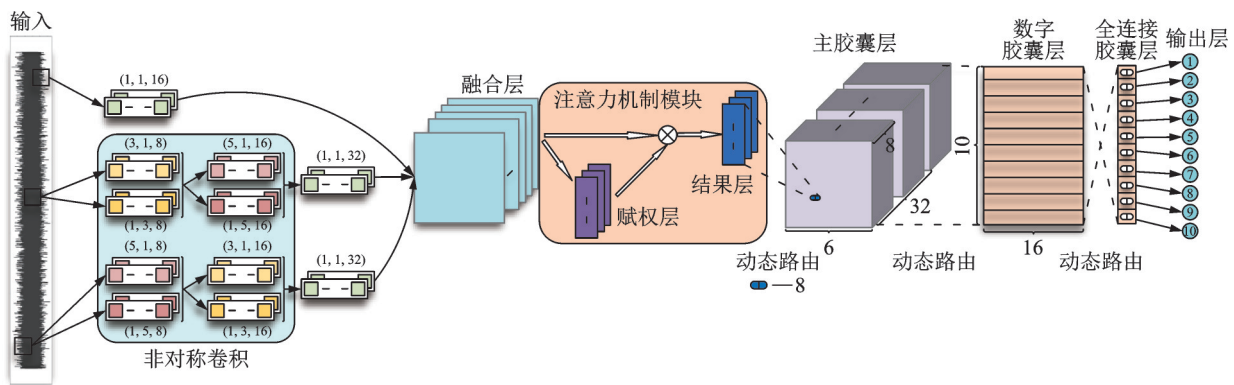


图5 改进的卷积胶囊神经网络结构

Fig. 5 Improved convolutional capsule neural network structure

3 实验验证及结果分析

为了评估本文提出方法对故障诊断的有效性和准确性, 实验在 Windows 10 系统下进行, 处理器为 i9-9900K, GPU 为 RTX 2070 SUPER, 利用 Pycharm 平台, 编程语言为 Python, 深度学习框架为 TensorFlow^[29]。本文实验以滚动轴承为对象, 采用来自美国凯斯西储大学 (CWRU) 的轴承数据^[30]和东南大学变速箱数据集中的轴承数据^[31]进行实验验证与分析。

3.1 CWRU 轴承故障数据集

3.1.1 数据集介绍

本文选取了凯斯西储大学 (CWRU) 滚动轴承数据中心的公开数据集验证本文方法的可行性,

CWRU 实验台由电机、扭矩传感器和测功器组成。测试的故障轴承是型号为 SKF 6205 的电机驱动端轴承, 采用电火花方法在内圈、滚动体和外圈的表面上加工出损伤直径为 0.1778, 0.3556, 0.5334 和 0.7112 mm 的单点凹槽, 以模拟滚动轴承在实际运行中的磨损情况。本实验采样频率为 12 kHz, 分别采集转速为 1797, 1772, 1750 和 1730 r/min 时, 对应负载为 0, 1, 2 和 3 hp (约为 0, 0.75, 1.49 和 2.24 kW) 状态下的加速度数据集, 并将其标记为数据集 A、数据集 B、数据集 C 和数据集 D。将采集到的数据按照不同位置和不同损伤度划分为 16 种状态标签, 且每种状态标签中的样本数量大致相同。在每个数据集中单次选取 8000 个样本, 按照 3:1 的比例划分训练样本和测试样本, 每段的采样点数设置为 784 个, 每一训练批次大小为 100。训练集和测试集如表 1 所示。

表 1 变工况数据集

Tab. 1 Variable working condition dataset

训练数据集	训练样本数	测试数据集	测试样本数	
			变负荷	变噪声
数据集 A	6000	数据集 B 数据集 C 数据集 D	2000	2000
数据集 B	6000	数据集 A 数据集 C 数据集 D	2000	2000
数据集 C	6000	数据集 A 数据集 B 数据集 D	2000	2000
数据集 D	6000	数据集 A 数据集 B 数据集 D	2000	2000

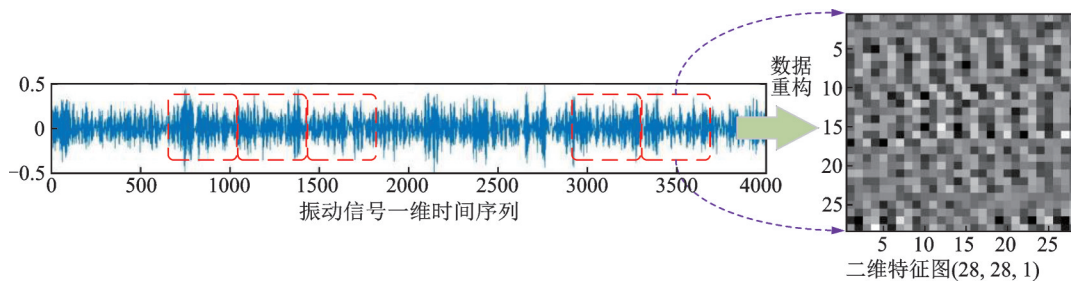


图 6 输入数据格式重构

Fig. 6 Input data reformatting

3.2 网络结构参数

网络结构越深,特征提取能力就越强,但是网络层数越多,越容易产生梯度爆炸等问题。滚动轴承的故障数据为一维时间序列,在转化成二维特征图像输入时,有效特征不是足够多,所以提出方法时要考虑提升网络的计算效率。在使用多尺度特征提取时,用非对称卷积核替代传统对称卷积核,这样可以在同样的感受野下,减少模型计算的参数量,加快模型的计算效率。在通道注意力机制后,需要构建胶囊单元,表 2 中胶囊单元的输出尺寸表示为 $6 \times (8)$,即特征层的宽度为 6 且每一个向量的维度为 8。在胶囊层中通过动态路由将之前卷积输出的特征标量变为特征矢量,在胶囊层之间进行运算。同理, $16 \times (10)$ 表示 16 个维度为 10 的向量, $10 \times (8)$ 表示 10 个维度为 8 的向量。

因为胶囊网络的拟合能力较强,所以训练时在胶囊层使用 Dropout 操作,即对神经元在每次迭代时随机失活,并且失活神经元的权值不再更新,从而降低网络复杂度,防止网络过拟合。本文方法参数设计如表 2 所示。

表 2 结构参数设计

Tab. 2 Structural parameter design

结构名称	结构参数	通道数量	输出大小
输入	(28, 28)	1	28×28
卷积层	(1, 1, 1)	16	28×28
卷积层	(3, 1, 1), (1, 3, 1)/(5, 1, 1), (1, 5, 1)/(1, 1, 1)	32	28×28
卷积层	(5, 1, 1), (1, 5, 1)/(3, 1, 1), (1, 3, 1)/(1, 1, 1)	32	28×28
通道注意力模块	—	—	—
主胶囊层	(9, 9, 2)	32	$6 \times (8)$
数字胶囊层	(10, 16, 1)	256	$16 \times (10)$
全连接胶囊层	(256/1024)	—	$10 \times (8)$
胶囊输出层	(1024/10)	—	—

3.3 对比方法

为了验证本文方法在变噪声、变负荷和变工况实验中,是否能够获得较高的故障诊断准确率和较好的泛化能力,将本文方法和常见的深度学习方法(CNN, IRB, CapsNet 和 FD-CCN)进行对比。对比模型的网络参数如表 3 所示,表 3 中的参数尽量使用与原文献中相同的网络参数。

3.1.2 数据集处理

实验台采集的轴承数据为一维时间序列,为了适应改进卷积胶囊网络模型输入数据格式,有效地进行卷积和下采样操作,根据如图 6 所示的重构方式,本文对长度为 784 的一维数据进行提取。首先,将时间序列样本等分为 28 段,每段包含 28 个数据点;然后,将这 28 段数据堆叠在一起,获得一个 $(28, 28)$ 的二维特征灰度图。

为了加快网络模型训练速度,让数据便于计算、获得更加泛化的结果,对输入数据做标准化处理,有效消除变量量纲和变异范围的影响,其表达式为:

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6)$$

式中 X 为输入数据, X_{\min} 为数据中的最小值, X_{\max} 为数据中的最大值。

表3 对比方法的网络参数

Tab. 3 Network parameters of the contrast methods

方法名称	网络结构
CNN	卷积层(3,32)→卷积层(4×3,64)→卷积层(3,128)→卷积层(3,128)→卷积层(3,64)→全连接层(256/1024)→全连接层(1024/200)→全连接层(200/10)
IRB ^[32]	数据池化(3,32)→残差块1(3,64)→残差块2(3,64)→残差块3(3,64)→残差块4(3,128)→残差块5(3,128)→全连接层(256/1024)→全连接层(1024/10)
FD-CCN ^[19]	卷积层(127×1,32)→池化层(2,32)→卷积层(7,32)→池化层(2,32)→主胶囊层(3,8,32)→数字胶囊层(10×16)→全连接层(1024/10)
CapsNet ^[18]	卷积层(9×9,256)→主胶囊层(9,2,32)→数字胶囊层(10×16)→全连接层(1024/10)

CNN(Convolutional Neural Network)使用传统的全连接层,为适应时频变换后的数据结构,通过卷积和池化的方式,逐层提取和压缩特征。卷积层的参数构建是为了更好地对比IRB^[30]网络,因为IRB网络同样使用卷积层,有着不错的诊断效果。

IRB(Inception+Residual Block)全局残差网络^[32]在残差网络中添加了注意力机制,并使用了五个残差块来提取特征信息,两个全连接层来逐层进行特征压缩。网络内部参数参照原文献来复现网络,同时使用Inception模块。

FD-CCN(Fault Diagnosis-Convolution Capsule Network)卷积胶囊网络^[19]用传统卷积层和池化层组合先进行特征提取,并二次使用卷积与池化,再与胶囊层相组合,结合了卷积网络和胶囊网络的各自特点,采用ReLU激活函数,以传统的全连接层来进行特征压缩,也采用原文献中的网络参数,验证本文方法对改进胶囊网络具有明显的提升和优势。

CapsNet(Capsules Network)胶囊网络^[18]先通过一个二维卷积层进行特征提取并作为胶囊层的输入,在初级胶囊层与数字胶囊层之间通过动态路由算法进行特征向量的矢量转化。胶囊网络使用文献[18]中的网络参数,验证了本文方法对原有胶囊网络的改进具有显著效果。

分别使用构建好的数据集B进行实验,样本按3:1的比例随机划分训练集和测试集。4种方法均采用Adam优化方法,学习率为0.001,衰减率为0.9,训练时单次读入数据量批次大小为100,全部样本迭代批次设置50,运行3次取平均结果。表4为在数据集B上的不同方法诊断结果对比。

由表4可以看到,改进卷积胶囊网络对训练集的诊断精度有一定提升。这说明改进的卷积胶囊网络相比于其他深度学习方法在数据量更大时,不仅诊断精度优于其他方法,而且在减少特征损失方面也有着不错的表现。

表4 不同方法训练结果

Tab. 4 Training results with different methods

诊断方法	准确率/%	损失值
CNN	95.66	0.0460
IRB	97.48	0.0249
FD-CCN	99.14	0.0063
CapsNet	98.90	0.0125
本文方法	99.98	0.0034

3.4 变工况条件下的故障诊断结果与分析

3.4.1 变负荷故障诊断

由于在实际中,滚动轴承经常工作在变负荷状态下,所以要求故障诊断方法具有良好的泛化能力。为了验证本文方法在变负荷情况下的诊断性能,网络训练与测试用的数据均为不同负荷下的数据集。以数据集A、数据集B、数据集C和数据集D中的一种依次作为训练样本,另外三种负荷数据集作为测试样本,实验结果如图7所示。

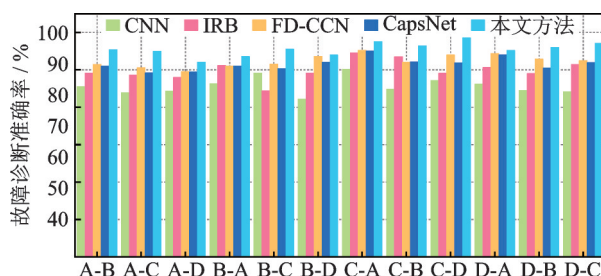


图7 变负荷工况下不同方法的诊断准确度对比

Fig. 7 The comparison of the diagnostic accuracy of different methods under variable load working conditions

由图7可以看到,本文提出的方法在变负荷实验中的诊断准确率均高于其他三种方法,其中以数据集C作为训练样本,数据集A、数据集B和数据集D作为测试样本时,IRB虽然改进了数据池化层,增强了网络特征学习能力,但是诊断准确率不如FD-CCN和本文方法,这是由于胶囊网络能保留卷积丢失的特征信息;CapsNet相较于卷积方法的性能有

提升,但是因为其单一的卷积层,没有充分提取故障特征信息,提升优势并不是很明显;FD-CCN虽然已经有了较明显的提取特征能力的提升,但是在前端卷积层输出特征信息到胶囊层的过程中过度使用池化层,在特征提取过程中过滤了不活跃的特征信息,影响了故障诊断准确率提升。以数据集C为训练集,本文方法对测试集为数据集A、数据集B、数据集D时的故障诊断准确率分别为97.52%,96.41%和98.56%,平均诊断准确率达到97.49%,而FD-CCN方法的诊断准确率平均值为92.38%,本文方法相比于FD-CCN准确率提升5%左右。

3.4.2 变噪声故障诊断

在实际应用中,检测的信号受到强噪声的干扰,为了验证本文方法在故障诊断时的抗强噪声性能,将不同信噪比(SNR)的高斯白噪声添加在测试数据集中。信噪比是评价信号中所含噪声的重要指标,其表达式为:

$$SNR = 10 \cdot \lg \frac{P_{\text{signal}}}{P_{\text{noise}}} \quad (7)$$

式中 P_{signal} 为信号有效功率; P_{noise} 为噪声功率。

本文方法以滚动轴承数据集C作为训练样本,并且在数据集C的测试样本中加入信噪比分别为-3,3,6和9 dB的高斯白噪声。给原始信号加入噪声以后,原始信号的特征就被强噪声所淹没,如图8所示。人眼已无法清晰地区分强噪声信号故障是否与原始信号故障属于相同故障。将最终得到的实验结果与CNN,IRB,CapsNet和FD-CCN方法的诊断结果进行对比分析,实验结果如表5所示。

由表5的数据对比可知,在不同信噪比的实验结果中,FD-CCN和CapsNet相比CNN和IRB有着

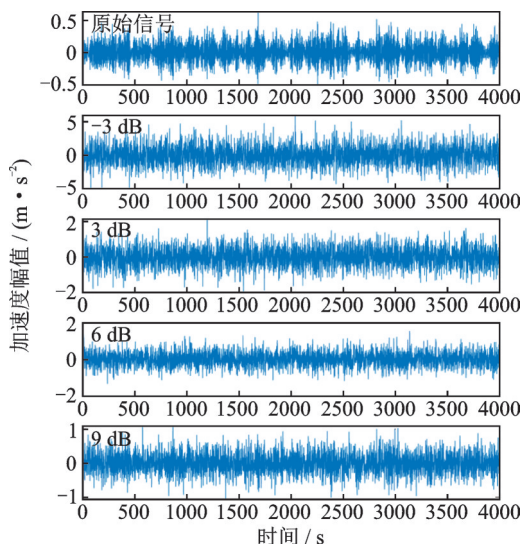


图8 不同信噪比状态(原始,-3,3,6,9 dB)时域图

Fig.8 Time domain diagram of different SNR states (original, -3, 3, 6, 9 dB)

表5 不同方法在不同信噪比下结果对比

Tab.5 The comparison results of different methods under different signal-to-noise ratio

故障诊断方法	信噪比/dB			
	-3	3	6	9
CNN	90.22%	92.69%	94.61%	95.37%
IRB	91.13%	93.24%	96.35%	97.32%
FD-CCN	96.68%	97.03%	98.67%	99.19%
CapsNet	94.97%	96.55%	96.99%	98.16%
本文方法	97.95%	98.47%	99.19%	99.71%

明显的诊断优势,在信噪比为-3 dB的噪声下达到95%左右的诊断准确率,相比于IRB和CNN提升了5%以上;在信噪比为-3 dB时,本文方法比CapsNet的故障诊断准确率提升了2.98%,相比FD-CCN的诊断准确率提升了1.27%。本文方法在数据池化层采用多尺度非对称卷积,最大限度地提取了故障数据中的信息,在6 dB以上的噪声下故障诊断准确率达99%以上,相比于其他4种方法其抗噪性能更好。

图9是本文方法在不同信噪比下的故障诊断结果,可以直观得到,在-3 dB的强噪声环境下,测试集的诊断准确率达到95%以上。

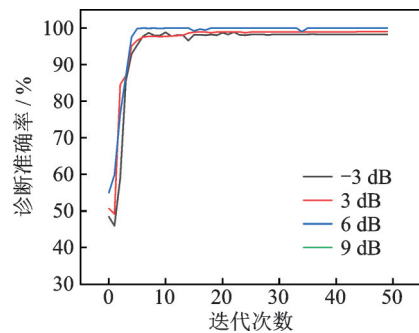


图9 本文方法不同信噪比下的诊断结果

Fig.9 Results of different signal-to-noise ratios of the proposed method

3.4.3 变工况故障诊断

在实际工程应用中,由于轴承的工作环境复杂,振动信号可能同时受到强噪声和负荷变化的影响。在变噪声和变负荷的单工况故障诊断实验中,本文提出的方法均取得了良好的诊断效果,为了验证本文方法在噪声环境下,且负荷在变化时的诊断性能,以数据集A作为训练样本,数据集B、数据集C和数据集D作为测试样本并且向测试样本中分别添加-3,3,6和9 dB的高斯白噪声,与CNN,IRB,CapsNet和FD-CCN模型的诊断结果进行对比分析,结果如表6所示。

表 6 变工况故障诊断实验结果对比

Tab. 6 The comparison of experimental results of fault diagnosis for variable working conditions

数据集名称	故障诊断方法	信噪比/dB			
		-3	3	6	9
数据集 B	IRB	84.91%	93.24%	95.81%	96.89%
	CapsNet	85.87%	94.76%	96.84%	97.02%
	FD-CCN	85.96%	95.88%	97.39%	97.41%
	本文方法	90.37%	96.26%	97.92%	99.11%
数据集 C	IRB	76.21%	85.81%	91.21%	93.37%
	CapsNet	82.99%	89.38%	95.39%	96.99%
	FD-CCN	86.67%	97.21%	97.99%	98.18%
	本文方法	88.55%	98.39%	99.06%	99.32%
数据集 D	IRB	70.99%	82.57%	86.89%	87.16%
	CapsNet	81.83%	89.67%	95.79%	96.31%
	FD-CCN	88.04%	95.96%	97.34%	97.99%
	本文方法	89.41%	96.09%	98.81%	99.02%

从表 6 可知,本文方法在不同信噪比和变负荷的环境下,其故障识别准确率都优于 IRB, CapsNet 和 FD-CCN。以数据集 B 为测试样本时,随着信噪比的减小,FD-CCN 的诊断准确率从 97.41% 降到了 85.96%,而本文方法从 99.11% 降到了 90.37%,诊断准确率仍在 90% 以上,说明本文方法拥有更好的泛化能力。

3.5 变速箱的滚动轴承实验分析

3.5.1 变速箱数据集介绍与处理

为了验证本文所提方法在其他类型轴承应用的可行性,采用来自东南大学变速箱数据集中的轴承数据进行实验验证^[31],该数据从传动系统动力学模拟器(DDS)上获取,DDS 由电机、行星齿轮箱、并联齿轮箱和制动器构成。在实验中,分别在两种运行条件下(20 Hz, 0 V 和 30 Hz, 2 V)对齿轮和轴承的故障进行诊断实验,其中轴承状态包括正常、外圈裂缝故障、内圈裂缝故障、滚动体裂缝故障和内外圈裂缝联合故障。DDS 实验台表面采用 7 个 608A11 振动传感器,其频率范围为 0.5~10 Hz,测量范围为 $\pm 50g$,测量精度为 100 mV/g。测量了齿轮箱 x, y, z 三个方向的振动信号,并采用紧凑型光谱板数据采集仪(最多 20 个通道)进行数据采集,采样频率为 1024 Hz,采样窗口为 512 s^[33]。本文在每种故障中随机选取 800 个样本,按照 3:1 的比例划分训练集和测试集,如表 7 所示。

3.5.2 故障诊断结果与分析

(1) 变负荷故障诊断

该实验转速-负载分别设置为 20 Hz-0 V 和

表 7 变速箱轴承数据集参数

Tab. 7 Gearbox bearing dataset parameters

轴承状态	20 Hz-0 V		30 Hz-2 V	
	训练集	测试集	训练集	测试集
正常状态	600	200	600	200
外圈故障	600	200	600	200
内圈故障	600	200	600	200
滚动体故障	600	200	600	200
内外联合故障	600	200	600	200

30 Hz-2 V,以轴承 5 种故障类型为基础,构建 10 个实验数据集。读入数据共计批次大小为 100,设置迭代批次为 50,将全部样本运行 3 次并取平均结果。将实验结果与 CNN, IRB, CapsNet 和 FD-CCN 方法作比较,结果如表 8 所示。

表 8 各方法的诊断结果对比

Tab. 8 The comparison of diagnostic results of each method

故障诊断方法	训练集	测试集
CNN	97.33%	95.03%
IRB	98.62%	96.91%
CapsNet	99.18%	98.33%
FD-CCN	99.73%	99.07%
本文方法	99.98%	99.83%

由表 8 可知,在变负荷环境下,本文所提方法的故障诊断准确率相比其他 4 种方法有显著提升。这是由于本文方法设计了改进后的多尺度非对称卷积模块进行特征提取,能够充分提取故障数据中的信息,提取的特征进入胶囊网络,胶囊网络可以更充分地提取并保存数据中的特征信息,从而使故障诊断准确率得到显著提升。

为进一步观察本文所提方法对故障误判的情况,对测试结果做了混淆矩阵实验,结果如图 10 所示,图中横坐标为预测标签的诊断状态,纵坐标为真实标签的实际状态。

从图 10(a)可看出,FD-CCN 方法对于轴承滚动体故障的诊断准确率仅为 98.29%,因为 FD-CCN 使用了两次池化操作,在特征提取过程中损失了一些细节特征。这导致两种故障的诊断准确率分别为 0.96 与 0.97,对整个故障诊断造成了影响。而从图 10(b)可看出,本文方法对各个故障状态均有较高的诊断准确率,其整体诊断准确率达 99.81%,这说明本文方法相比于 FD-CCN 拥有更佳的诊断性能,能够更充分地提取故障特征信息。也验证了本文所提出的故障诊断方法应用在其他种类轴承数据集的可行性和泛化能力。

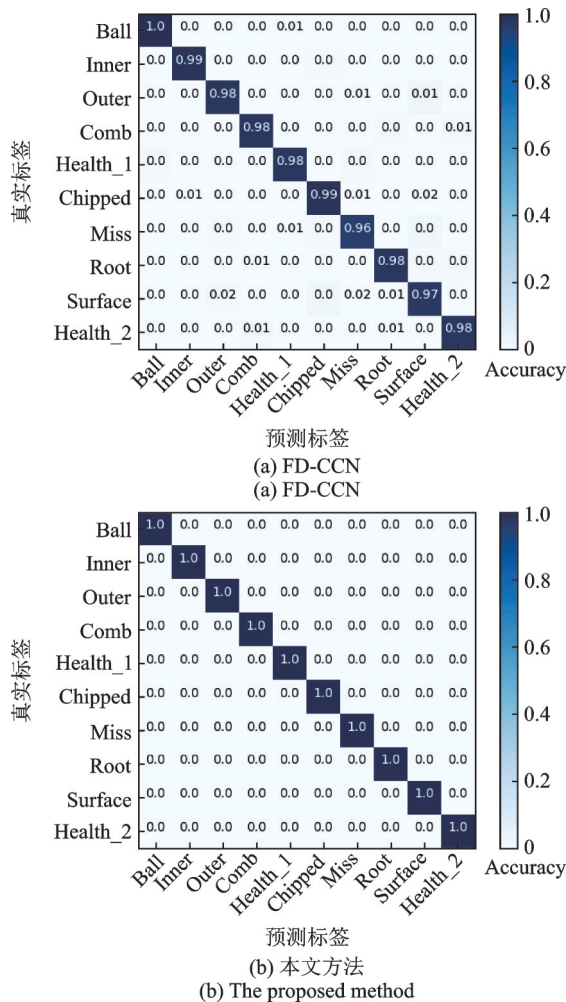


图 10 不同方法的混淆矩阵对比

Fig. 10 Confusion matrix comparisons for different methods

(2) 变噪声故障诊断

在变速箱数据集上依旧采用和 CWRU 轴承数据集一样的处理方式,分别加入信噪比为-3,3,6 和 9 dB 的高斯白噪声,具体结果如表 9 所示。

表 9 不同方法在各信噪比下变速箱轴承数据集结果对比
Tab. 9 The comparison results of gearbox bearing dataset by different methods at each signal-to-noise ratio

故障诊断方法	信噪比/dB			
	-3	3	6	9
CNN	91.09%	92.34%	93.99%	95.40%
IRB	92.88%	93.74%	97.03%	97.94%
FD-CCN	96.98%	97.35%	98.97%	99.31%
CapsNet	95.21%	97.07%	97.69%	98.66%
本文方法	97.63%	98.39%	99.24%	99.83%

由表 9 可知,在-3 dB 的强噪声环境下,本文方法相比传统的 CapsNet 故障诊断准确率提升了 2.42%,相比 FD-CCN 提升了 0.65%。这说明本文方法依旧保持着良好的特征提取性能,且本文方法在轴承故障诊断的抗噪方面有着不错的表现。

3.6 模型参数量

计算成本是评价深度学习方法性能的重要指标。本文方法和各对比方法的参数量如表 10 所示。传统 CNN 和 IRB 方法的参数量虽然明显小于胶囊相关方法,但是它们的故障诊断准确率也是最低的。本文方法相比传统 CapsNet 的参数量显著减少,尽管比 FD-CCN 的参数量要多,但是本文方法在变速箱数据集变负荷实验中每批次平均耗时最短,诊断结果更优。这是因为参数量相同时,非对称卷积比对称卷积计算效率更高。同时由于本文方法在胶囊网络前未使用池化操作过滤部分特征信息,能够充分提取故障特征信息,所以具有良好的鲁棒性。

表 10 各诊断方法的参数量

Tab. 10 Number of parameters for each diagnostic method

故障诊断方法	参数量/kB	耗时/s
CNN	92.6	161
IRB	103.2	159
CapsNet	940.5	185
FD-CCN	206.3	127
本文方法	426.2	118

4 结 论

本文提出了一种改进卷积胶囊网络的滚动轴承变工况故障诊断方法。该方法将一维时域信号转为二维图像数据作为网络输入,通过大量的数据学习,能够自适应地提取轴承故障特征,而无需对原始数据进行过多的人工预处理,通过将多尺度非对称卷积与通道注意力模块结合,并且改进胶囊网络中的胶囊全连接层,起到抗噪和改善特征提取能力的作用,使得本文方法的稳定性和泛化能力更好。实验结果表明,相比于传统卷积神经网络、全局残差网络和双卷积胶囊神经网络,本文方法可以实现更高的诊断准确率和更好的泛化能力。未来在强噪声环境影响下,需要对本文方法进一步优化,改善网络参数选择的稳定性和通用性。

参考文献:

[1] 张妮,车立志,吴小进.基于数据驱动的故障诊断技术研究现状及展望[J].计算机科学,2017,44(6A):47-52.
Zhang N, Che L Z, Wu X J. Present situation and prospects of data-driven based fault diagnosis technique[J].

- Computer Science, 2017, 44(6A): 47-52.
- [2] Lu C, Wang Z Y, Qin W L, et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification[J]. Signal Processing, 2017, 130: 377-388.
- [3] 刘强, 柴天佑, 赵立杰. 基于数据和知识的工业过程监视及故障诊断综述[J]. 控制与决策, 2010, 25(6): 801-807.
- Liu Q, Chai T Y, Zhao L J. Progress of data-driven and knowledge-driven process monitoring and fault diagnosis for industry process[J]. Control and Decision, 2010, 25(6): 801-807.
- [4] 卢文涛. 基于胶囊网络模型过程故障识别的应用与研究[D]. 南昌: 华东交通大学, 2020.
- Lu Wentao. Application and research of process fault identification based on CapsNet Model[D]. Nanchang: East China Jiaotong University, 2020.
- [5] Muralidharan V, Sugumaran V. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis[J]. Applied Soft Computing, 2012, 12(8): 2023-2029.
- [6] Yang Y, Yu D, Cheng J. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM[J]. Measurement, 2007, 40(9-10): 943-950.
- [7] 贺岩松, 黄毅, 徐中明, 等. 基于小波奇异熵与SOFM神经网络的电机轴承故障识别[J]. 振动与冲击, 2017, 36(10): 217-223.
- He Y S, Huang Y, Xu Z M, et al. Motor bearing fault identification based on wavelet singular entropy and SOFM neural network[J]. Journal of Vibration and Shock, 2017, 36(10): 217-223.
- [8] Verstraete D, Ferrada A, Drognet E L, et al. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings[J]. Shock & Vibration, 2017, 2017: 1-17.
- [9] He Z, Shao H, Ding Z, et al. Modified deep autoencoder driven by multisource parameters for fault transfer prognosis of aeroengine[J]. IEEE Transactions on Industrial Electronics, 2021, 69(1): 845-855.
- [10] 胡越, 罗东阳, 花奎, 等. 关于深度学习的综述与讨论[J]. 智能系统学报, 2019, 14(1): 1-19.
- Hu Y, Luo D Y, Hua K, et al. Overview on deep learning[J]. Journal of Intelligent Systems, 2019, 14(1): 1-19.
- [11] 邵海东, 张笑阳, 程军圣, 等. 基于提升深度迁移自动编码器的轴承智能故障诊断[J]. 机械工程学报, 2020, 56(9): 84-90.
- Shao H D, Zhang X Y, Cheng J S, et al. Intelligent fault diagnosis of bearings using enhanced deep transfer auto-encoder[J]. Journal of Mechanical Engineering, 2020, 56(9): 84-90.
- [12] 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.
- Qu J L, Yu L, Yuan T, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network[J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134-143.
- [13] Gan M, Wang C, Zhu C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings[J]. Mechanical Systems & Signal Processing, 2016, 72-73: 92-104.
- [14] Sun W, Shao S, Zhao R, et al. A sparse auto-encoder-based deep neural network approach for induction motor faults classification[J]. Measurement, 2016, 89: 171-178.
- [15] 张西宁, 向宙, 唐春华. 一种深度卷积自编码网络及其在滚动轴承故障诊断中的应用[J]. 西安交通大学学报, 2018, 52(7): 1-8.
- Zhang X N, Xiang Z, Tang C H. A deep convolutional auto-encoding neural network and its application in bearing fault diagnosis[J]. Journal of Xi'an Jiaotong University, 2018, 52(7): 1-8.
- [16] Laha S K. Enhancement of fault diagnosis of rolling element bearing using maximum kurtosis fast nonlocal means denoising[J]. Measurement, 2017, 100: 157-163.
- [17] Shao H, Li W, Xia M, et al. Fault diagnosis of a rotor-bearing system under variable rotating speeds using two-stage parameter transfer and infrared thermal images[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3524711.
- [18] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA, 2017: 3859-3869.
- [19] 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.
- Yang P, Su Y C, Zhang Z. A study on rolling bearing fault diagnosis based on convolutional capsule network[J]. Journal of Vibration and Shock, 2020, 39(4): 55-62.
- [20] Chen Y. Convolutional neural network for sentence classification[D]. Waterloo: University of Waterloo, 2015.
- [21] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA, 2017: 4278-4284.
- [22] Bahdanau D, Cho K H, Bengio Y. Neural machine

- translation by jointly learning to align and translate [C]//3rd International Conference on Learning Representations. San Diego, United States, 2015: 1-15.
- [23] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [24] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016: 2818-2826.
- [25] 王贺兵, 张春梅. 基于非对称卷积-压缩激发-次代残差网络的人脸关键点检测[J]. 计算机应用, 2021, 41(9): 2741-2747.
Wang H B, Zhang C M. Facial detection based on ResNet with asymmetric convolution and squeeze excitation[J]. Journal of Computer Applications, 2021, 41(9): 2741-2747.
- [26] 杨宏业. 卷积神经网络的多光谱遥感图像超分辨率重建[D]. 徐州: 中国矿业大学, 2020.
Yang Hongye. Multispectral remote sensing images super-resolution based on convolutional neural network [D]. Xuzhou: China University of Mining and Technology, 2020.
- [27] 王婷. 基于LSTM深度网络的电力负荷预测[D]. 太原: 山西大学, 2020.
Wang Ting. Power load forecasting based on LSTM deep network[D]. Taiyuan: Shanxi University, 2020.
- [28] Zhao W, Ye J, Yang M, et al. Investigating capsule networks with dynamic routing for text classification [C]//2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 3110-3119.
- [29] Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems[J]. arXiv:1603.04467, 2016.
- [30] 袁壮, 董瑞, 张来斌, 等. 深度领域自适应及其在跨工况故障诊断中的应用[J]. 振动与冲击, 2020, 39(12): 281-288.
Yuan Z, Dong R, Zhang L B, et al. Deep domain adaptation and its application in fault diagnosis across working conditions[J]. Journal of Vibration and Shock, 2020, 39(12): 281-288.
- [31] Shao S, McAleer S, Yan R, et al. Highly accurate machine fault diagnosis using deep transfer learning[J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2446-2455.
- [32] 赵小强, 梁浩鹏. 使用改进残差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, 2020, 54(9): 23-31.
Zhao X Q, Liang H P. Fault diagnosis method of rolling bearing under variable condition using improved residual neural network [J]. Journal of Xi'an Jiaotong University, 2020, 54(9): 23-31.
- [33] Zhao R, Wang D, Yan R, et al. Machine health monitoring using local feature-based gated recurrent unit networks[J]. IEEE Transactions on Industrial Electronics, 2018, 65(2): 1539-1548.

Improved convolutional capsule network method for rolling bearing fault diagnosis

ZHAO Xiao-qiang^{1,2,3}, CHAI Jing-xuan¹

(1.College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China; 2.Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou 730050, China; 3.National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: At present, many rolling bearing fault diagnosis methods based on convolutional networks have the disadvantages of poor diagnosis effect and poor generalization ability under the influence of noise signals and load variations. Aiming at these problems, an improved convolutional capsule network fault diagnosis method of rolling bearing under variable operating conditions is proposed. This method designs a multi-scale asymmetric convolution module, in which asymmetric convolution layers of different scales to extract features from the input data to maximize the extraction of feature information in the data and reduce the number of parameters effectively. In this module, the channel attention mechanism is introduced to better extract useful channel features and improve the feature extraction ability of the method in this paper. By improving the fully connected layer in the network to the fully connected layer of the capsule, the capsule can avoid the loss of characteristic information in the space in the process of outputting vector feature information. Case Western Reserve University bearing dataset and Southeast University gearbox dataset are used to verify the diagnostic performance of the proposed method and compare with other deep learning methods. The experimental results show that the proposed method has a better generalization and performance.

Key words: fault diagnosis; rolling bearing; capsule network; asymmetric convolution; feature extraction

作者简介: 赵小强(1969—),男,博士,教授。E-mail: xqzhao@lut.edu.cn。