

基于 ARCN 模型的轴承故障诊断

梁海涛¹, 王立纲², 王亮³, 张庆峰³

(1. 四川三星通用航空有限责任公司 机务工程部, 成都 610051;

2. 中国民用航空飞行学院 广汉分院, 四川 广汉 618307;

3. 中国民用航空飞行学院 飞行技术学院, 四川 广汉 618307)

摘 要: 提出注意力循环机制与胶囊网络融合的注意力循环胶囊网络(ARC�)的诊断模型。提取时序特征信息构建初级胶囊;自适应融合路由机制、注意力循环机制构建数字胶囊特征;基于西储大学轴承实验数据,验证了 ARC� 模型的准确率、鲁棒性、稳定性、收敛误差,其准确率相比 Caps 模型识别准确率提高 1.2%、收敛误差达到 0.2。基于实验仿真平台,采集正常、内环故障、外环故障和滚动体故障的振动信号,并通过小波基变换获取的时频图构建 ARC� 模型的数据集。仿真实验结果表明:ARC� 模型下,每类故障被误诊的概率不超过总样本的 1%。

关 键 词: 轴承故障诊断;注意力循环机制;胶囊网络;小波变换;准确率

中图分类号: V234; TH17

文献标志码: A

Bearing fault diagnosis based on ARC� model

LIANG Haitao¹, WANG Ligang², WANG Liang³, ZHANG Qingfeng³

(1. Mechanical Engineering Department,

Sichuan Tri-star General Aviation Company Limited, Chengdu 610051, China;

2. Guanghan Flight College,

Civil Aviation Flight University of China, Guanghan 618307, China;

3. Flight Technology College,

Civil Aviation Flight University of China, Guanghan 618307, China)

Abstract: The attention recurrent and capsule network (ARC�) diagnosis model was proposed by integrating the attention cycle mechanism and capsule network. Firstly, the bidirectional LSTM network was used to extract the time-series characteristic information to construct the primary capsule. Secondly, routing mechanism and attention cycle mechanism were used to construct adaptively digital capsule. The accuracy, robustness, stability and convergence error of ARC� model in bearing fault identification were verified by bearing experiment data of Western Reserve University. The accuracy of ARC� model was 1.2% higher than that of Caps model. The convergence error of the ARC� model reached 0.2. Based on the experimental simulation platform, the vibration signals of normal, inner ring fault, outer ring fault and rolling element fault were collected. The results showed that the misdiagnosis probability of each kind of fault

收稿日期:2021-04-21

基金项目:青年科学基金项目(12002368)

作者简介:梁海涛(1976—),男,高级工程师,研究方向为航空发动机故障诊断。E-mail:lianghaitao7603@163.com

通信作者:张庆峰(1982—),男,副教授,硕士生,研究方向为工程力学与飞机结构健康监测。E-mail:zqcfac@163.com

引用格式:梁海涛,王立纲,王亮,等. 基于 ARC� 模型的轴承故障诊断[J]. 航空动力学报,2021,36(9):1793-1803. LIANG Haitao, WANG Ligang, WANG Liang, et al. Bearing fault diagnosis based on ARC� model[J]. Journal of Aerospace Power, 2021, 36(9):1793-1803.

was less than 1% of the total samples under ARCN model.

Key words: bearing fault diagnosis; attention recurrent; capsule network; wavelet transform; accuracy

滚动轴承是航空发动机等旋转机械系统中支撑转子系统的核心部件,起着承受及传递载荷的作用。在长期使用过程中,由于变载荷等工况的反复,使得轴承容易出现不同程度的磨损或者故障。滚动轴承发生故障时,会导致发动机转静子碰磨、振动过大等故障现象,甚至造成严重破坏事故。由于轴承故障程度不同时,其表现出的振动特征也存在差异,尤其是难以识别的早期微弱故障是国内外研究的重点^[1]。

在国内外研究中,一方面:利用振动信号对轴承健康状态实施监控和诊断是一种有效地手段^[2],但由于振动信号故障特征较为微弱且噪声信号对其干扰性较强,因此部分信号处理方式,如快速傅里叶变换(FFT)等不能够准确有效地反映出有效信息的振动特征^[3-4];另一方面:基于智能故障诊断技术的轴承故障诊断方法,常采用信号特征提取、筛选与分类识别算法相结合的方式^[5],但相关资料证明^[6]其浅层网络结构导致特征提取能力较差,无法实现故障微弱特征的提取。因此,针对以上两方面原因,宫文峰等^[7]将更具特征提取的卷积神经网络运用于轴承故障诊断中,在其研究中为解决传统卷积神经网络(convolutional neural networks, CNN)模型参数过多的问题,引入了全局均值池化技术对CNN进行改进,改进的CNN在轴承故障诊断模型中取得不错的效果。卷积神经网络模型中的池化层都是采用的最大池化,此方式可以增强网络对于平移和旋转的时不变性,增强网络的泛化能力,但是,却存在以下两方面原因:①容易丢失特征图层中的其他特征^[8];②传统卷积网络将提取好的深层特征通过全连接层映射到目标空间时,由于全连接层的参数过多,容易发生过拟合现象^[9]。

为了克服卷积神经网络的缺陷,在2017年Sabour等^[10]提出了胶囊网络。胶囊网络使用向量进行网络运算,每一个向量都构成一个胶囊,向量的每个维度都代表着某个实体的一种特征,所以一个向量即为该实体的抽象表达,而向量的模则表示该实体出现的概率^[11]。Zhu等^[8]利用凯斯西储大学轴承数据库数据以及帕德博恩大学的轴承数据库对他提出的改进胶囊网络模型进行验证,且验证精度达到了97.15%;Chen等^[12]验证了胶

囊网络在不同载荷,不同背景噪声下对轴承故障诊断有着良好的诊断效果以及较强的鲁棒性;袁路生等^[13]利用凯斯西储大学的轴承数据库数据对胶囊网络模型的诊断能力进行验证,并且验证精度达到了99.93%;Ru等^[14]在五速汽车变速器上采集了由轴承故障(内圈缺陷)、齿轮故障(缺齿缺陷)和复合故障(轴承和齿轮故障)组成的齿轮箱数据集,验证了胶囊网络在旋转机械的故障诊断中有着良好的表现。

但部分研究中,在构建数字胶囊时,要么采用动态路由机制,要么采用注意力循环机制构建数字胶囊,动态路由机制能够保留所有的特征并通过耦合系数准确反应整体与局部的关系,但却无法有效地抑制无用特征;而注意力机制可从大量信息中筛选出对当前任务更关键的信息,并且抑制无用的信息^[10]。所以,采用其中某一种方式实现数字胶囊的构建,将不能够最大限度的提取故障有效特征信息。在沈炜域^[15]撰写的基于自注意力与动态路由的文本建模方法一文中,利用多层多尺度卷积核并辅以动态路由机制保证图片转换为文本时的特征不变性,并结合注意力机制能保证文本特征间相互性的特点,构建了注意力与动态路由机制的CapSA模型,从而实现文本特征的捕捉,这也本文后续工作的开展提供了一种思路。

为此,本文提出注意力循环机制与胶囊网络融合的注意力循环胶囊网络(attention recurrent and capsule network, ARCN)的诊断模型。与沈炜域提出的采用卷积核构建初始胶囊不同的是:ARCN模型中考虑到胶囊网络的独立性、振动信号的时序性,采用双向长短记忆神经(long short term memory network, LSTM)网络提取轴承故障振动信号的时序信息,从而获得输入样本的全局特征,并作为初级胶囊层的输入;同时,分别采用动态路由机制、注意力循环机制构建数字胶囊,并将路由机制的先验耦合系数与注意力机制的后验耦合系数相融合,从而提高实例预测胶囊与数字胶囊之间的相似度的同时,抑制不相关特征对模型的干扰,提高诊断效果;并利用凯斯西储大学轴承数据检验ARCN模型的有效性,最后将ARCN模型运用到轴承故障实例分析中。

1 基本理论

1.1 经典胶囊网络

随着机器视觉的飞速发展,卷积神经网络被运用到旋转机械故障诊断中^[7]。在卷积网络中,其池化层常采用下采样的方法提取故障特征。虽然下采样法能够增强网络的泛化能力,但却容易破坏图像的姿和空间等信息^[14]。且,随着卷积神经网络通过增加结构宽度和深度来改善网络性能时,容易出现梯度弥散。为了改善卷积网络的缺陷问题,面对卷积网络存在的缺陷问题,Sabour等^[10]在2017年提出了胶囊网络。

胶囊网络主要由初级胶囊层和数字胶囊层组成,初级胶囊层又包括卷积层和初级胶囊构建层,其卷积层计算公式如式(1):

$$a_k^l = f\left(\sum_{g \in h_k} a_k^{l-1} \otimes w_{gk}^l + b_k^l\right) \quad (1)$$

式中 a_k^l 表示第 l 层第 k 个元素; h_k 表示 $l-1$ 层的第 g 个卷积面 ($l > 1$); a_k^{l-1} 表示卷积面中的数值; w_{gk}^l 表示对应卷积面的卷积核; b_k^l 表示阈值; $\otimes$ 为卷积运算符; $f(\cdot)$ 表示 ReLU 激活函数。

胶囊网络在构建初级胶囊时,将卷积层提取出的标量故障特征排列为向量故障特征,并生成初级胶囊 u_i ,其中 i 表示第 i 个故障特征胶囊。数字胶囊层与初级胶囊层之间通过动态路由机制实现的信息通讯。动态路由机制的传递过程如下:

步骤1 基于初级胶囊网络,根据式(2)生成实例预测胶囊

$$\hat{u}_{ji} = w_{ij} u_i \quad (2)$$

式中 i 表示初级胶囊标号; j 为数字胶囊标号; u_i 为初级胶囊层输出的第 i 个初级胶囊; $\hat{u}_{ji}$ 为实例预测胶囊; w_{ij} 为权值矩阵。

步骤2 基于式(3)计算实例预测胶囊与数字胶囊之间的耦合系数,并对所有实例预测胶囊进行加权求和得到数字胶囊

$$\begin{cases} c_{ij} = \frac{\exp(b_{ij})}{\sum_j \exp(b_{ij})} \\ s_j = \sum_i c_{ij} \hat{u}_{ji} \end{cases} \quad (3)$$

式中 b_{ij} 表示实例预测胶囊 $\hat{u}_{ji}$ 与数字胶囊 s_j 之间的先验链接权重,且 b_{ij} 的初始值都为0,并计算 c_{ij} ,且 $\sum c_{ij} = 1$ 。

步骤3 基于式(4)的squash函数将数字胶囊 s_j 的模长压缩至 $[0, 1]$,并得到数字胶囊层最终输出数字胶囊 v_j ;并更新先验连接权重 b_{ij}

$$\begin{cases} v_j = \text{squash}(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} \\ b_{ij} = b_{ij} + \hat{u}_{ji} \cdot v_j \end{cases} \quad (4)$$

1.2 注意力循环机制

注意力机制是基于人观察事物关注重点部位的特点而提出的,其广泛应用于图像识别、自然语言处理等领域。由于采集的轴承振动信号难免会引入部分噪声信号,而噪声信号的存在会在一定程度上对特征向量产生干扰。为此,可采用循环注意力机制,提取振动特征信号中具有更高的敏感度的信息,而忽略不敏感因素,使得经过循环注意力机制后的特征更具表征性。注意力机制算法将在文中第2.2节数字胶囊中详细描述。

1.3 双向长短记忆网络

由于轴承振动信号的小波时频图中包含时序信息,所以当前区域的相邻区域的特征都可以用来对当前区域信息进行评价。由于经典胶囊网络中的卷积层无法捕捉时域信息,而LSTM网络通常将前序时刻信息和当前时刻信息进行融合,无法考虑后续时刻特征对当前时刻信息的影响。因此采用双向LSTM网络提取轴承的故障特征,作为胶囊网络的输入。

双向LSTM网络结构如图1所示(图中 X 表示LSTM网络的输入; C 表示权中矩阵; a 表前序

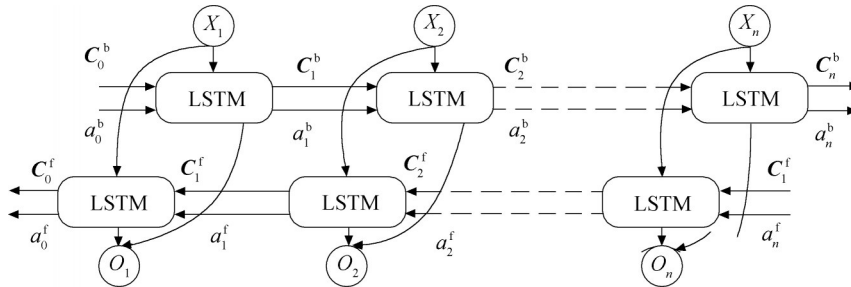


图1 Bi-LSTM 结构图

Fig.1 Structure diagram of Bi-LSTM

LSTM单元的输出; O 表示是网络输出;上标 b 表示正向,上标 f 表示反向;图中下标数值表示LSTM网络单元个数),包含两个反向的LSTM网络,即一个前向LSTM网络和一个后向LSTM网络,分别从输入样本的开始端和结束端工作,并将两个网络的输出结果进行融合获得最终输出,其计算过程见文献[16]。

2 基于ARCN模型的轴承故障诊断

由于本文中采用小波时频域分析方法处理振动信号,因此得到的样本中包含时序信息,而胶囊网络中的卷积层无法捕捉样本的时序信息。因此,采用双向长短型记忆网络提取时序特征信息作为初级胶囊的输入。同时,考虑采用单一构建数字胶囊的方式,无法解释样本之间的全局特征问题。因此,文中分别采用动态路由机制、注意力循环机制构建数字胶囊,并将路由机制的先验耦合系数与注意力机制的后验耦合系数相融合,从而提高实例预测胶囊与数字胶囊之间的相似度的同时,抑制不相关特征对模型的干扰,提高诊断效果。

因此,本文提出的ARCN模型主要包括:初级胶囊构建、数字胶囊构建和自适应特征融合。

2.1 初级胶囊

首先,采用两层二维卷积网络对轴承振动信号的时频图进行特征提取,两层网络的卷积核均采用大小 9×9 的卷积核,激活函数均为ReLU,其数量分别为64、128,其步长分别为1、2;其次,利用双向LSTM网络对卷积层得到的特征向量进行二次特征提取;最后,将双向LSTM网络输出的特征向量输入到胶囊层,从而完成初级胶囊的构建。

2.2 数字胶囊

考虑到动态路由机制虽然能够保留所有的特征,并通过耦合系数准确反应整体与局部的关系,但是却无法有效抑制无用特征;而注意力机制可以从大量信息中筛选出对当前任务更关键的信息,并且抑制无用的信息。因此,在构建数字胶囊时,将路由机制的先验耦合系数与注意力机制的后验耦合系数相融合,提出基于注意力循环机制的胶囊网络。通过先验耦合系数,建立时频图中高能区与高能区之间的关系,从而提高实例预测胶囊与数字胶囊之间的相似度;并利用注意力机制,提高时频图中不同时间和频带的能量强度变化的

影响力,抑制不相关特征对模型的干扰,提高诊断效果。

在提出的ARCN模型中,数字胶囊分别由动态路由机制和注意力循环机制构成。由于在文中第1.1节中已经对动态路由机制进描述,此处不再累述,仅对ARCN模型中的注意力循环机制进描述。其算法描述如下:

步骤1 基于初级胶囊的输入,利用式(3)计算先验耦合系数 c_{ij} 。

步骤2 基于式(2)计算实例预测胶囊 $\hat{u}_{ji}$ 。

步骤3 利用先验耦合系数 c_{ij} 和实例预测胶囊 $\hat{u}_{ji}$,根据式(5)计算输入胶囊 H_{ji}

$$H_{ji} = \tanh [w_{ij} \cdot (c_{ij} \otimes \hat{u}_{ji}) + b_{ij}] \quad (5)$$

步骤4 基于式(6),计算注意力机制中的后验耦合系数 a_i

$$a_i = \frac{\exp(H_{ji})}{\sum_i \exp(H_{ji})}, \sum_i a_i = 1 \quad (6)$$

步骤5 基于式(7),计算注意力循环机制中第 j 层的总输入向量

$$s_j = \sum_i a_i H_{ji} \quad (7)$$

步骤6 基于式(4)得到数字胶囊层最终输出数字胶囊 v_j ,并更新后验权重 b_{ij} 。

2.3 自适应特征融合

对于动态路由机制与注意力循环机制产生的两个独立数字胶囊,采进行自适应融合。特征融合之后再通过softmax分类器得到网络输出,其数学描述如式(8):

$$\begin{cases} C_{aps}^l = \sum_j \sum_i a_{ij}^l \times w_{ij} \\ y = \text{soft max}(C_{aps}) \end{cases} \quad (8)$$

式中 C_{aps}^l 表示第 l 个拼接胶囊; a_{ij}^l 表示第 l 个拼接胶囊的第 i 行第 j 列的特征量; w_{ij} 为特征融合时的权值; y 表示网络输出;softmax是深度神经网络中的一个归一化函数。

2.4 基于ARCN模型的轴承故障诊断流程

基于ARCN模型的轴承故障诊断流程如图2所示,可分如下6个部分:①振动信号采集,从实验平台上采集不同工况下的振动信号;②数据集构建,对振动信号进行小波变换,得到二维时频图;③初级胶囊构建,通过卷积核提取特征,并采用双向LSTM网络对卷积核提取的特征进行二次提取;④数字胶囊构建与融合,通过路由机制与注意力循环机制构建数字胶囊,并实现特征自适应融合;⑤softmax层,用于对胶囊结果进行处理;⑥输出判断,用于判断ARCN模型是否终止。

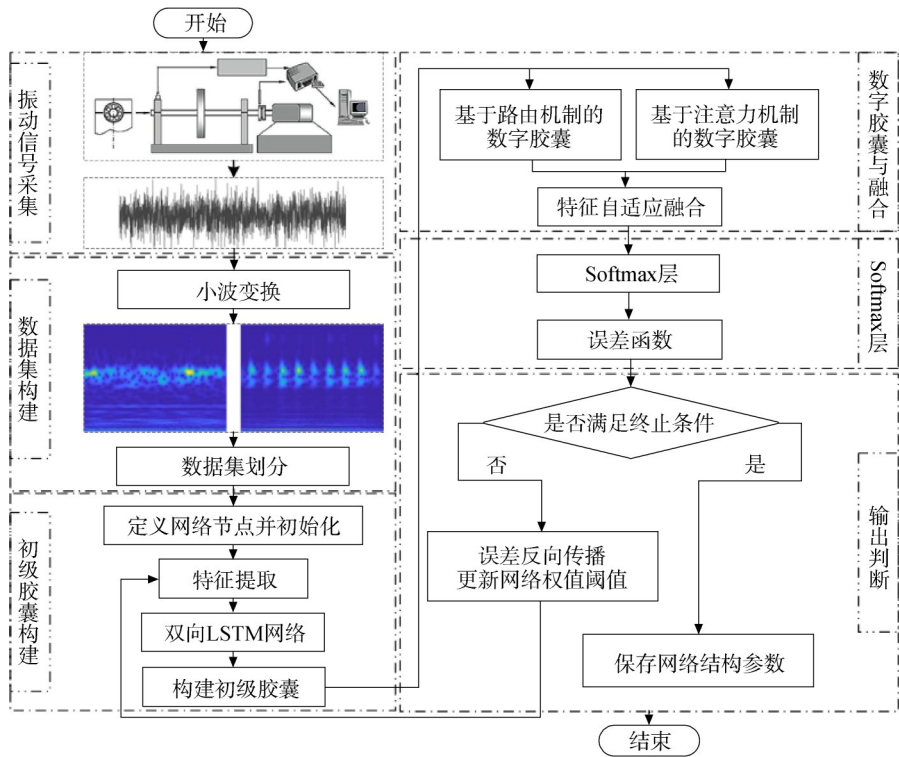


图 2 ARCN 模型诊断流程

Fig.2 ARCN model diagnosis process

基于 ARCN 模型的轴承诊断流程描述如下：

步骤 1 通过振动试验平台获取不同工况下的轴承振动信号。

步骤 2 采用尺度为 256 的 cmor3-3 小波基对不同工况下的故障振动信号进行连续小波变换,生成小波时频图。

步骤 3 采用两层二维卷积网络对时频图进行特征提取,第一层卷积层采用 64 个尺度为 9×9、步长为 1 的卷积核进行特征提取;第二层卷积核采用 128 个尺度为 9×9、步长为 2 的卷积核进行特征提取;并构建出 8×1 的初级胶囊向量雏形,利用双向 LSTM 网络对初级胶囊雏形进行二次特征提取,从而得到初级胶囊向量。

步骤 4 分别采用路由机制、注意力循环机制构建数字胶囊,并将路由机制的先验耦合系数与注意力机制的后验耦合系数相融合,实现特征自适应融合。

步骤 5 将融合的特征通过 softmax 层实现故障模式识别,并通过计算交叉熵误差判断网络是否满足网络误差精度要求,若满足则终止训练并保存网络结构参数,否则重复步骤 3~步骤 5。

3 ARCN 模型性能验证

为验证提出的 ARCN 模型在轴承故障诊断

中的有效性,采用凯斯西储大学的轴承实验数据验证作为验证数据集^[17]。性能验证实验中,选择正常、内环、外环以及滚动体 4 种状况的振动信号,振动数据是在负载为 0.746 kW、采样频率为 12 kHz 下获得。试验中,每次使用 1 024 个点进行小波变换,生成小波时频图,每类故障生成 1 800 个样本,随机不重复的选择 1 300、400、100 分别作为训练集、校验集和测试集。试验中,所使用的轴承类型、故障尺寸如表 1 所示,样本数据信息如表 2 所示。

表 1 轴承参数

Table 1 Parameters of bearing

参数	数值及详情
轴承型号	SKF6205
滚子直径/mm	7.5
节圆直径/mm	39
滚子数量	9
接触角/(°)	0

3.1 ARCN 模型训练分析

为了更好地描述 ARCN 模型在实验数据下的训练效果,采用双纵坐标轴描述 ARCN 模型训

表 2 试验数据组成
Table 2 Composition of the test data set

标签	损伤位置	损伤尺寸/ mm	训练样 本数	校验样 本数	测试样 本数
1	无	0	1 300	400	100
2	滚动体	0.355 6	1 300	400	100
3	内环	0.177 8	1 300	400	100
4	外环	0.533 4	1 300	400	100

练过程中,训练集、测试集的准确率和模型收敛趋势。考虑到每次仿真结果存在波动性,因此将试验重复 10 次,10 次试验的 ARCN 模型平均结果如图 3 所示,实验中,每 100 步输出一次计算结果。

图 3 中虚线表示每次迭代后训练样本的准确率(实际为校验集);点划线表示每次迭代训练完成后使用训练好的 ARCN 模型对测试集的诊断准确率;实线为 ARCN 模型随着训练步数增加,其误差收敛效果。由图 3 可知,ARCN 模型在 40 次迭代之后网络的诊断准确率可以达到 90% 以上,且模型的训练误差降低到 0.2 左右;当迭代步数达到 60 次以后,训练误差及诊断结果均趋于稳

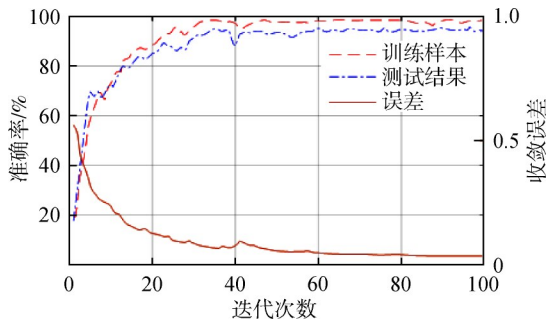


图 3 ARCN 模型训练结果
Fig.3 ARCN model training results

表 3 不同算法的数据集组成
Table 3 Data set composition of different algorithms

模型	数据集组成	标记	训练集样本数	校验集样本数	测试集样本数
ARCN	小波变换	CWT-ARCN	1 300	400	100
胶囊网络	小波变换	CWT-Caps	1 300	400	100
卷积网络	小波变换	CWT-CNN	1 300	400	100
	傅里叶变换	FFT-CNN	1 300	400	100
深度神经网络	小波变换	CWT-DNN	1 300	400	100

定,在迭代步数为 60 次时,计算得测试集的 10 次实验结果平均值为 94.1%。由分析可知,ARCN 模型具有良好的诊断效果。

3.2 算法对比分析

进一步分析本文提出的 ARCN 模型的诊断性能,将胶囊网络、经典卷积网络、深度神经网络作为对比算法,各算法的数据集组成如表 3 所示。

5 种模型 10 次诊断结果如图 4 所示。10 次诊断结果中,5 种算法中位线变化趋势如图 4 中的黑色虚线,其中 CWT-ARCN 的中位数、上四分位数、下四分位数均明显高于其他算法,CWT-Caps、CWT-CNN、FFT-CNN 三种算法的上四分位数基本持平;而采用小波变换的方式处理数据的效果优于短时傅里叶变换,如图 4 中 CWT-CNN 与 FFT-CNN 结果所示,且 FFT-CNN 效果相对最差;通过引入双向 LSTM 网络和循环注意力机制后能够对胶囊的性能得到提升,如 CWT-ARCN 与 CWT-Caps 结果所示。

5 种模型的测试集诊断准确率随着迭代步数变化趋势分别如图 5 所示,5 种算法的平均结果如表 4 所示(10 次结果均取自迭代步骤为 150 时的结果)。由图可知:CWT-DNN 收敛数最慢,且在迭代步数 150 步时,训练集平均准确率为 96.26%、测试集平均准确率仅 36.50%,其可能是由于深度神经网络(deep neural networks,DNN)出现过拟合现象导致;CNN 网络收敛速度较快,能够较快的趋于稳定,但其容易陷入局部最优解,导致对测试集诊断停留在 95% 左右,这说明 CNN 提取特征能力有限;胶囊网络 Caps 虽然收敛速度相对较慢,但其表现出的特征提取能力优于 CNN 网络,随着其迭代步数的增加,其测试集诊断准确率逐渐增加;本文提出的 ARCN 网络收敛速度与 Caps 相当,但其波动相对平稳。

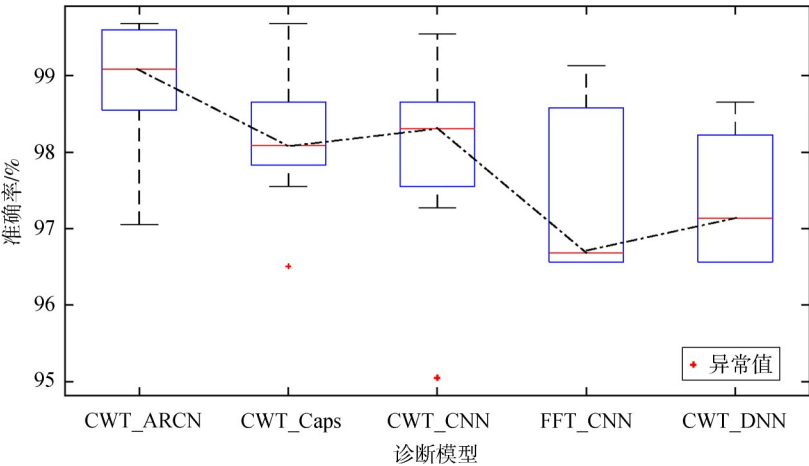


图 4 5 种模型 10 次诊断结果

Fig.4 Five different models 10 times diagnosis results

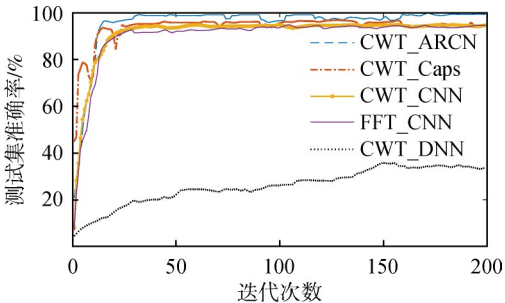


图 5 5 种模型测试集诊断结果

Fig.5 Five different models diagnosis results base on test set

表 4 中 η_{test} 表示测试集准确率; η_{train} 表示训练集准确率。由表 4 结果可以发现: CWT-CNN 与 FFT-CNN 的结果表明, CWT 数据集与 FFT 数据对 CNN 模型的影响较小; 在 CWT 数据集下, 由于过拟合的存在, 使得 DNN 模型拥有较高的训练精度, 却拥有较低的识别精度, 因此应该从机器学习指标进行综合评价模型的优劣; 在 CWT 数据集下, ARCNN、Caps 诊断模型的准确率和召回率的调和平均数(F1 值)得分大于 96%, CNN

诊断模型 F1 值都低于 96%, DNN 模型的 F1 值都低于 60%, 这说明在小样本下胶囊网络的综合性能更好, 能够从轴承振动信号中提取出更多的特征; ARCN 模型相比 Caps 模型中的测试集诊断准确率高出 1.2%, 其召回率、F1 值均最高, 说明注意力循环机制能够为时频图中含有故障信息的高能频带赋予更高的权重, 并且能够有效识别不同故障类型之间的高能频带之间的分布区别。

3.3 鲁棒性分析

考虑到不同负载下得到的振动信号对模型性能的影响, 因此通过测试负载为 0、1.492 kW 和 2.238 kW 验证模型的鲁棒性。数据集、样本、对比模型均按照文中第 4.2 节方式处理, 实验重复 10 次取其平均值(每次取迭代步数为 150 步计算结果), 其诊断结果如图 6 所示。

由图 6 可知, 当实验中负载发生变化时, ARCN、Caps、CNN 诊断模型均能够表现出相对稳定的状态, 这也说明利用小波变换的时频图, 结合图像识别算法能够在轴承故障诊断中获取较佳结果。

表 4 5 种模型的平均结果

Table 4 Average results of five models

诊断模型	训练准确率/%	测试准确率/%	召回率/%	F1 值/%	$(\eta_{\text{test}}/\eta_{\text{train}})/\%$
CWT-ARCN	99.92	98.5	98.65	97.33	98.57
CWT-Caps	99.61	97.30	97.04	96.99	97.68
CWT-CNN	99.75	95.83	95.88	95.82	96.07
FFT-CNN	99.5	95.17	95.29	95.3	95.65
CWT-DNN	96.26	36.50	46.5	45.73	37.92

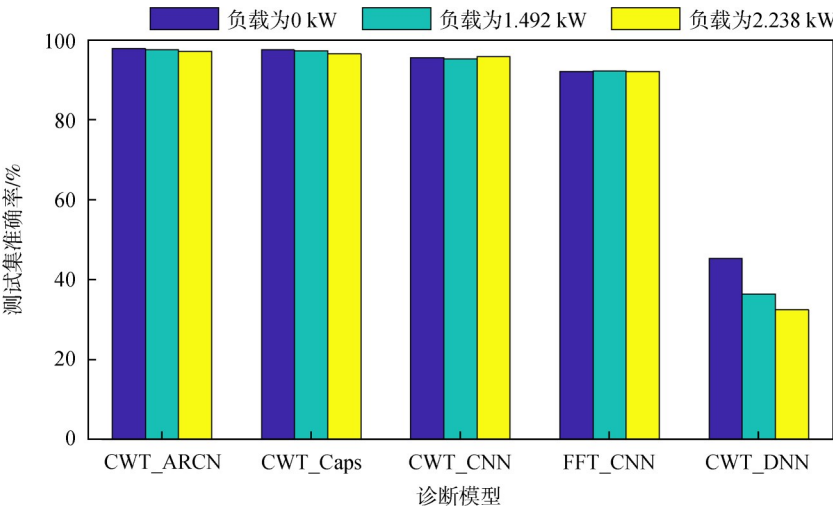


图 6 载荷变化情况下的准确率对比

Fig.6 Comparison of recognition accuracy under load variation

4 基于 ARCN 的轴承故障诊断实例

4.1 轴承实验简述

实验平台简图如图 7 所示,并从轴承试验平台上采集不同工况下的轴承振动信号数据,用于验证 ARCN 模型的有效性。

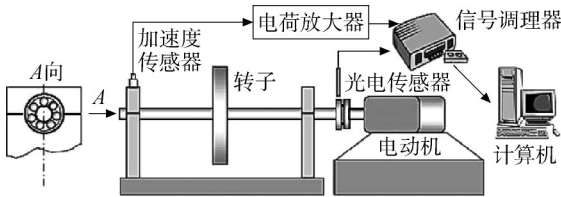


图 7 滚动轴承数据采集平台

Fig.7 Rolling bearing fault data acquisition bench

试验中,主要模拟正常、内环故障、外环故障、滚动体 4 种故障类。试验中所有故障轴承均在轴承未淬火之前完成制备,需要注意的是这些故障深度均为 0.1 mm、故障轴向长度均为 0.6 mm,用以模拟微弱故障类型。试验所用轴承信息如表 5 所示,故障信息表 6 所示,4 种故障类型的时域信号如图 8 所示。

4.2 数据集准备

由于时域信号不能够很好地表征信号特征,为能够获取信号更深层特征,采用尺度为 256 的 cmor3-3 小波基对 4 种故障类型振动信号进行连续小波变换,生成小波时频图^[18]。同时,通过向原始振动信号中增加-10 dB 信噪比的噪声来检测 ACRN 模型的鲁棒性。轴承故障振动信号经小波变换后的时频图如图 9 所示,模型样本比例分

表 5 轴承参数

Table 5 Parameters of bearing

参数	数值
轴承型号	307
滚子直径/mm	14.5
轴承内径/mm	35
轴承外径/mm	80
滚子数量	7
接触角/(°)	0

表 6 滚动轴承故障数据集

Table 6 Fault dataset of rolling bearing

损伤位置	标签	损伤宽度/mm	训练集样本数	校验集样本数	测试集样本数
无	1	0	1 000	400	100
滚动体	2	0.5	1 000	400	100
内环	3	0.25	1 000	400	100
外环	4	0.25	1 000	400	100

配如表 4 所示。

如图 8 所示,加噪信号的小波时频图高能区由原来的低频和中频区域分散到整个时频区域,使得通过小波时频图已经无法直观的进行故障诊断。

4.3 诊断结果分析

几种诊断模型对轴承故障样本集识别准确率

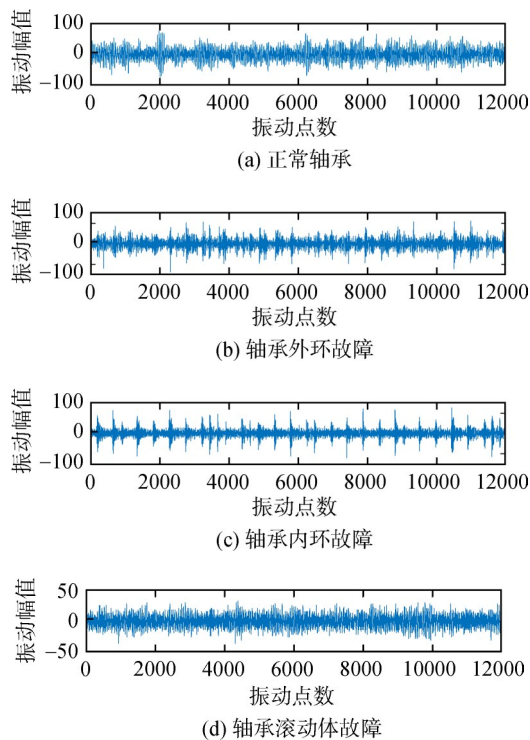


图 8 轴承故障振动信号
Fig.8 Vibration signal of bearing fault data

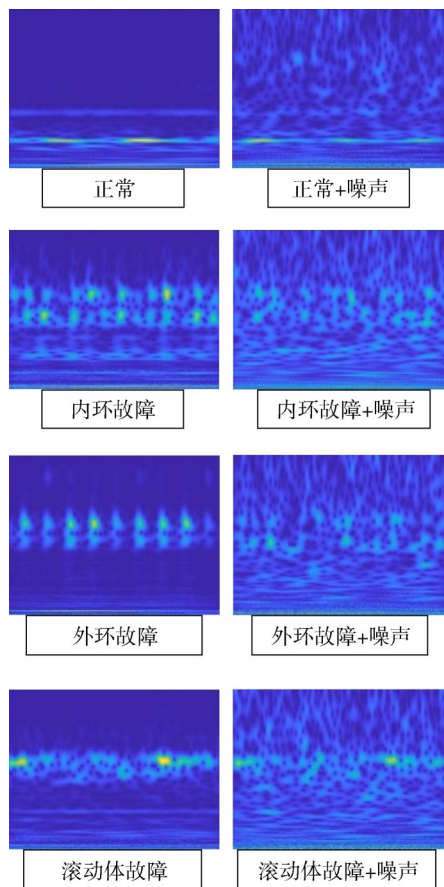


图 9 数据样本集
Fig.9 Data sample set

如表 7 所示,可以发现:ARCN 模型在识别率、鲁棒性比其他几种对比模型更具优势,而胶囊网络次之,从而说明在胶囊网络中引入注意力循环机制后,网络会对时频图中含有故障信息的高能频带赋予更高的权重,能够有效识别不同故障类型之间的分布区别。

表 7 诊断结果			
Table 7 Fault diagnosis results			
诊断模型	迭代次数	测试集	准确率/%
ARCN	150	无噪	98.94
		-10 dB 噪声	93.75
Caps	150	无噪	96.85
		-10 dB 噪声	90.31
CNN	150	无噪	94.58
		-10 dB 噪声	67.21
DNN	150	无噪	91.25
		-10 dB 噪声	43.82

为了更清楚地展示模型在测试集中各个类别的识别结果,引入混淆矩阵对实验结果进行分析,ARCN 模型在无噪声测试集下的故障识别结果如图 10 所示。图中横轴为真实值,纵轴为预测值;每类故障样本 100 个,共 4 种故障类型;主对角线上的数字代表 ARCN 模型对每一类故障状态正确诊断的样本个数,及占总样本的比例;且每一列均量化为精确信息,即正确分类的个数和误分类的个数,以及某故障被误判为哪一类故障。

图 10 中每个小方格上方的数字表示样本被

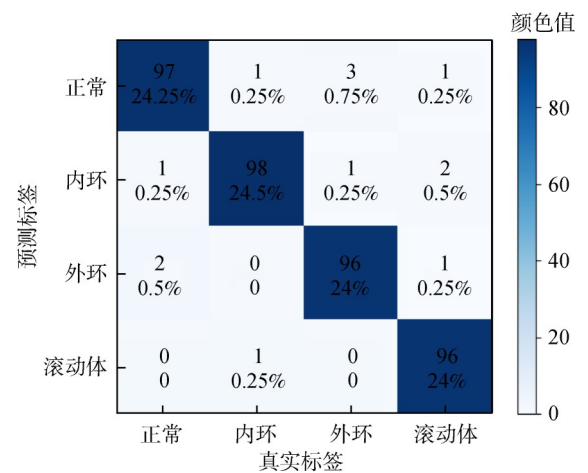


图 10 测试结果的混淆矩阵
Fig.10 Confusion matrix of test results

诊断为该类的个数,下方的百分数表示所占百分比;图中色谱中0~100数值所以对应的为颜色值,小方格上方数值越大,方格颜色越深。由图10可知:ARCN模型能够较好地地区别正常、内环故障、外环故障、滚动体故障4种类型,其每类故障识别准确的样本分别占总样本的24.25%、24.5%、24%、24%;而每类故障被误诊的概率不超过总样本的1%;ARCN模型对外环故障、滚动体故障识别效果相比正常、内环故障效果较差,其误诊样本达到总体样本的1%。通过实验分析,在一定程度上验证了本文提出的ARCN模型具有一定的鲁棒性、诊断准确率,具有一定的工程实用性。

5 结 论

针对轴承故障识别难度大、且频谱分析方法较为繁琐等问题,提出注意力循环机制与胶囊网络融合的注意力循环胶囊网络诊断模型。

1) 为充分利用振动信号中包含的时序信息,采用双向LSTM网络对卷积层提取的特征向量进行二次提取,双向LSTM提取的特征向量作为胶囊层的输入;同时,分别采用路由机制、注意力循环机制构建数字胶囊,并将路由机制的先验耦合系数与注意力机制的后验耦合系数相融合,实现特征自适应融合。

2) 为验证ARCN模型在轴承故障识别上的有效性,基于利用西储大学轴承实验数据,将胶囊网络、经典卷积网络、深度神经网络作为对比算法,从算法训练、对比、鲁棒性方面验证了ARCN模型在轴承故障诊断上的有效性。

3) 基于实验仿真平台,采集正常、内环故障、外环故障和滚动体故障的振动信号,并通过小波基变换获取的时频图构建ARCN模型的数据集。在相同的数据集下,将胶囊网络、CNN、DNN作为对比算法,进一步验证了ARCN模型轴承故障诊断中的有效性。

参考文献:

- [1] 马圣. 极限学习机在航空轴承故障诊断中的应用研究[D]. 天津: 中国民航大学, 2019.
MA Sheng. Applied research on aeroengine bearing fault diagnosis based on extreme learning machine[D]. Tianjin: Civil Aviation University of China, 2019. (in Chinese)
- [2] 蔡赛男, 宋卫星, 班利明, 等. 基于鲸鱼算法优化LSSVM的滚动轴承故障诊断[EB/OL]. [2021-04-17]. <https://doi.org/10.13195/j.kzyjc.2020.1147>.
- [3] 宋霖, 宿磊, 李可, 等. 基于SSD和1DCNN的滚动轴承故障诊断方法[J]. 华中科技大学学报(自然科学版), 2020, 48(12): 38-43.
SONG Lin, SU Lei, LI Ke, et al. Fault diagnosis method of rolling bearing based on SSD and 1DCNN[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2020, 48(12): 38-43. (in Chinese)
- [4] 李红贤, 韩延, 吴敬涛, 等. 基于ICA包络增强MEMD的滚动轴承故障诊断[J]. 航空动力学报, 2021, 36(2): 405-412.
LI Hongxian, HAN Yan, WU Jingtao, et al. Rolling bearing fault diagnosis based on MEMD with ICA envelop enhancement[J]. Journal of Aerospace Power, 2021, 36(2): 405-412. (in Chinese)
- [5] 皮骏, 马圣, 贺嘉诚, 等. 基于IGA-ELM网络的滚动轴承故障诊断[J]. 航空学报, 2018, 39(9): 233-244.
PI Jun, MA Sheng, HE Jiacheng, et al. Rolling bearing fault diagnosis based on IGA-ELM network[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(9): 233-244. (in Chinese)
- [6] HAI DONG S, HONG KAI J, YING L, et al. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders[J]. Mechanical Systems and Signal Processing, 2018, 102(1): 278-297.
- [7] 宫文峰, 陈辉, 张泽辉, 等. 基于改进卷积神经网络的滚动轴承智能故障诊断研究[J]. 振动工程学报, 2020, 33(2): 400-413.
GONG Wenfeng, CHEN Hui, ZHANG Zehui, et al. Intelligent fault diagnosis for rolling bearing based on improved convolutional neural network[J]. Journal of vibration engineering, 2020, 33(2): 400-413. (in Chinese)
- [8] ZHU Z A, PENG G A, CHEN Y A, et al. A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis[J]. Neurocomputing, 2019, 323: 62-75.
- [9] ZHANG S W, ZHANG S B, ZHANG C L, et al. Cucumber leaf disease identification with global pooling dilated convolutional neural network[J]. Computers and Electronics in Agriculture, 2019, 162: 422-430.
- [10] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[R]. Long Beach, US: the 31st International Conference on Neural Information Processing Systems, 2017.
- [11] MADHU G, GOVARDHAN A, SRINIVAS B S, et al. Imperative dynamic routing between capsules network for malaria classification[J]. Computers, Materials and Continua, 2021, 68(1): 903-919.
- [12] CHEN T Y, WANG Z H, YANG X, et al. A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals[J]. Measurement, 2019, 148: 106857.1-106857.15.
- [13] 衷路生, 刘东东. 多级神经网络的轴承故障诊断研究[J]. 计算机工程与应用, 2020, 56(7): 193-199.
ZHONG Lusheng, LIU Dongdong. Research on bearing fault diagnosis of multi-level neural network[J]. Computer Engi-

- neering and Applications, 2020, 56(7):193-199. (in Chinese)
- [14] RU Y, HUANG, YI X, et al. Deep decoupling convolutional neural network for intelligent compound fault diagnosis[J]. IEEE Access, 2018, 7:1848-1858.
- [15] 沈伟域. 基于自注意力与动态路由的文本建模方法[J]. 软件导刊, 2019, 18(1):56-60, 64.
SHEN Weiyu. A text modeling method based on self-attention and dynamic routing[J]. Software Guide, 2019, 18(1):56-60, 64. (in Chinese)
- [16] 张文涛. 基于 BiLSTM 网络的领域问答系统关键技术研究[D]. 长沙:湖南师范大学, 2020.
ZHANG Wentao. Research on the key technologies of domain question answering system based on BiLSTM networks[D]. Changsha: Hunan Normal University, 2020. (in Chinese)
- [17] Case Western Reserve University Bearing Data Center. Bearing data center fault test data[EB/OL]. [2020-10-01]. <http://www.eecs.case.edu/laboratory/bearing>.
- [18] 朱娜, 母其海, 刘志邦, 等. 基于多尺度小波变换和时频图的微波源故障检测[J]. 电子测量技术, 2021, 44(13):48-53.
ZHU Na, MU Qihai, LIU Zhibang, et al. Fault detection of microwave source based on multi-scale wavelet transform and time-frequency diagram[J]. Electronic measurement technology, 2021, 44(13):48-53. (in Chinese)

(编辑:王碧珺)