

DOI:10.19344/j.cnki.issn1671-5276.2024.04.012

# 基于自注意机制胶囊网络的行星齿轮箱故障诊断

聂松雅, 陈则王, 杨林, 王友仁

(南京航空航天大学 自动化学院, 江苏 南京 211106)

**摘要:**针对实际工程中行星齿轮箱故障数据有限、诊断准确率不高的问题,提出一种基于自注意机制胶囊网络的故障诊断方法。直接将采集到的行星齿轮箱振动信号作为输入,用首层宽卷积层提取浅层特征,过滤输入中的高频噪声;引入自注意机制关注信号关键特征;再次将所提特征输入胶囊层,进一步提取特征并实现故障分类;采用行星齿轮箱实验平台数据对所提方法进行实验验证。实验结果表明:在样本数量有限的情况下,所提方法仍能取得不错的诊断准确率。

**关键词:**行星齿轮箱;故障诊断;胶囊网络;自注意机制;小样本

**中图分类号:**TH132.425 **文献标志码:**B **文章编号:**1671-5276(2024)04-0067-04

## Fault Diagnosis of Planetary Gearboxes Based on Self-attentive Mechanism Capsule Network

NIE Songya, CHEN Zewang, YANG Lin, WANG Youren

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** A fault diagnosis method based on self-attentive mechanism capsule network is proposed to solve the problems of limited fault data and low diagnosis accuracy for planetary gearboxes in practical engineering. The acquired planetary gearbox vibration signal is directly used as the input to extract primary features through the first wide convolutional layer and filter the high-frequency noise in the input. The self-attentive mechanism is introduced to focus on the key features of the signal. The proposed features are input into the capsule layer to further extract features and achieve fault classification. The proposed method is verified by the data of planetary gearbox experimental platform. The results show that the proposed method can still achieve good diagnostic accuracy with limited samples.

**Keywords:** planetary gear box; fault diagnosis; capsule network; self-attention mechanism; small sample

### 0 引言

行星齿轮箱因其轻巧高效的特点,普遍应用于各类国计民生行业中<sup>[1]</sup>。在实际应用中,行星齿轮箱因为长期服役极易产生故障,因此,及时的故障诊断是至关重要的。由于齿轮箱内部结构复杂,受力多变,导致所采集振动信号传递路径复杂<sup>[2]</sup>,而传统机器学习方法需叠加信号处理技术,依赖丰富专业知识,仍难以实现智能故障诊断。

深度学习近年来飞速发展,由于其特征学习能力强,特征提取自适应,降低了人为操作带来的不确定性<sup>[3]</sup>,因此广泛应用于智能故障诊断。文献[4]利用1D-CNN直接从原始振动信号中提取特征,实现电机故障智能诊断。文献[5]为提高CNN在复杂多变环境下故障诊断能力,引入VMD算法,实现信号同步滤波和数据增强的功能。文献[6]采用短时傅里叶变换处理时域信号,再将所得时频图输入卷积神经网络进行特征提取,实

现故障诊断。但上述方法皆需大量数据,而由于齿轮箱定期检修维护,无法获取足够多的故障数据。胶囊网络采用胶囊神经元,不同于卷积神经网络采用的标量神经元,捕获信息更加丰富,可以识别输入样本中的微小改变,在所用数据集较小的情况下也能捕获较多特征。因此,在小样本故障诊断领域方面备受关注<sup>[7-8]</sup>,但仅用单卷积层的胶囊网络特征提取并不全面。

针对上述问题,本文提出了一种基于自注意机制胶囊网络的故障诊断方法,以应对模型训练过程中样本不足的情况。该方法将一维振动信号直接输入所提模型中,利用所提的自注意机制胶囊网络自动提取行星齿轮箱振动信号特征。首先使用宽卷积层滤去一维振动信号中混杂的高频噪声;然后将所提特征输入自注意机制层,削弱无关信息,关注重点特征;最后,利用胶囊网络提取精细特征并给出最终分类结果,以实现小样本下行星齿轮箱智能故障诊断。

**第一作者简介:**聂松雅(1998—),女,江西景德镇人,硕士研究生,研究方向为故障诊断,niesongya1412@163.com。

## 1 相关网络理论

### 1.1 胶囊网络概述

卷积神经网络因其优越的学习能力在故障诊断领域被广泛使用。但由于其池化层的存在,限制了网络的空间特性,会造成局部特征信息丢失。为解决上述问题,SABOUR 等<sup>[9]</sup>于 2017 年提出了胶囊网络(capsule network, CapsNet),进一步提升神经网络的上限。区别于传统神经网络使用向量神经元,胶囊网络使用胶囊神经元作为网络的输入与输出,并且引入了动态路由机制将信息从低层胶囊传递到高层胶囊。

胶囊网络核心在于动态路由机制,具体操作如图 1 所示。该操作可以分为 3 个步骤:首先,每个输入向量乘以权重得到预测向量,表达式如下:

$$\mathbf{u}_{j|i} = \mathbf{W}_{ij} \mathbf{u}_i \quad (1)$$

式中: $\mathbf{u}_{j|i}$ 为预测向量; $\mathbf{W}_{ij}$ 为权值矩阵; $\mathbf{u}_i$ 为第  $i$  个输出的初级胶囊。

其次,由  $\mathbf{u}_{j|i}$  加权求和获得输出向量  $\mathbf{S}_j$ ,具体表达式如下:

$$\begin{cases} \mathbf{S}_j = \sum_i c_{ij} \mathbf{u}_{j|i} \\ c_{ij} = \frac{e^{b_{ij}}}{\sum_j e^{b_{ij}}} \end{cases} \quad (2)$$

式中: $c_{ij}$ 为耦合系数; $b_{ij}$ 为  $\mathbf{u}_{j|i}$  与  $\mathbf{S}_j$  之间的对数先验概率。

最后,引入向量压缩函数 squash 函数来代替传统神经网络中的激活函数,使向量保持原始方向不变,且长度在 (0,1) 之间。具体表达式如下:

$$\mathbf{v}_j = \frac{\|\mathbf{S}_j\|^2}{1 + \|\mathbf{S}_j\|^2} \cdot \frac{\mathbf{S}_j}{\|\mathbf{S}_j\|} \quad (3)$$

式中: $\mathbf{v}_j$ 为最终输出数字胶囊; $\|\mathbf{S}_j\|$ 表示获得原向量的模长。

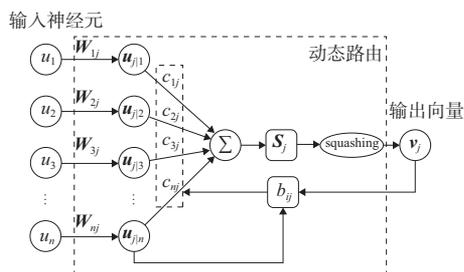


图 1 动态路由机制

### 1.2 自注意机制

在利用卷积层进行浅层特征提取的时候,感

受野具有局部性,无法获取输入信号中的全局特征。而为了更好地提取整个旋转周期中的故障数据,全局特征至关重要。自注意力机制借鉴了生物观察行为的过程,结合了内部经验与外部感知,增强了局部精细程度观察。它能够让模型更着眼于重要特征信息,提高任务处理的效率和准确性。其函数实质是一个查询向量在一组键值对上的投影,计算步骤如下:

- 1) 将输入数据转换为嵌入向量;
- 2) 将嵌入向量分别乘上 3 个权重矩阵,得到查询向量、键向量和值向量;
- 3) 计算查询向量和每个键向量的相似度,其结果作为每个输入特征在自注意力机制中的权重;
- 4) 使用 softmax 函数将这些权重和归一化至 0~1 之间;
- 5) 最后将计算出的权重和对应的值向量进行加权求和,即得到自注意力结果。

## 2 自注意机制胶囊网络模型及故障诊断流程

### 2.1 自注意机制胶囊网络模型

胶囊网络由于其利用向量获取特征,具有保留空间信息的特性,在旋转机械故障诊断领域中被广泛研究。但在故障样本数量极小的情况下,现有的单卷积层胶囊网络模型不足以提取足够的故障敏感特征对行星齿轮箱故障进行诊断。因此,本文提出一种带自注意机制的胶囊网络模型,进一步挖掘胶囊网络潜力,以更适用于实际应用中。本文所提模型由宽卷积层、卷积层、自注意机制层、初始胶囊层、数字胶囊层组成。网络的输入为一维振动信号。该网络旨在进行故障诊断,网络的具体结构如图 2 所示。

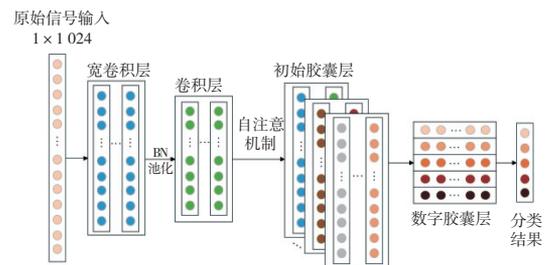


图 2 自注意机制胶囊网络模型结构

### 2.2 行星齿轮箱故障诊断流程

带自注意机制的胶囊网络行星齿轮箱故障诊断方法流程如图 3 所示。该流程的具体操作步骤

阐述如下:

1) 数据采集:通过传感器获取行星齿轮箱在各种健康状态下的振动信号;

2) 数据预处理与数据集划分:将收集到的振动信号分割成不同样本,使每个样本都能包含齿轮转动一圈的信息,并将分割好的样本划分为训练集和测试集;

3) 模型搭建:搭建带自注意力机制的胶囊网络模型,设置模型参数并选取恰当的激活函数;

4) 模型训练:将训练集送入所提模型进行训练,并采用 Adam 优化器优化训练模型;

5) 故障诊断:将测试数据输入训练好的模型中进行故障诊断并输出最终结果。

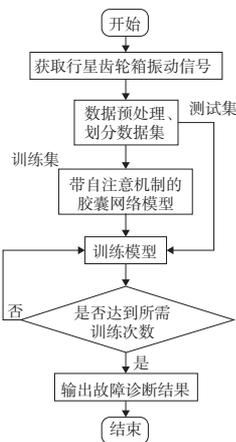


图3 小样本下行星齿轮箱故障诊断方法流程图

### 3 实验验证与结果分析

#### 3.1 实验台搭建及其数据采集

采用的行星齿轮箱故障模拟实验平台如图4所示。该实验平台由变速驱动电机、平行齿轮箱、振动传感器、行星齿轮箱和磁粉制动器组成。其中行星齿轮箱内包含1个太阳轮和3个行星轮,箱内各零部件详细参数如表1所示。

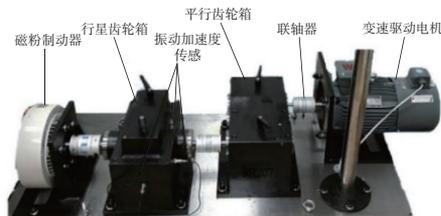


图4 行星齿轮箱故障模拟实验平台

表1 行星齿轮箱参数表

零部件名称	太阳轮	行星轮	齿圈
齿数/个	18	27	72

为了模拟齿轮不同程度的损伤,太阳轮点蚀故障和行星轮裂纹故障的几何参数分别设置如表2、表3所示的3种尺寸,其中裂纹深度为不同斜深度。具体故障位置如图5、图6所示。

表2 太阳轮点蚀故障尺寸

样本类型	深度/mm	数量/个	总点蚀面积/mm <sup>2</sup>
点蚀故障1	0.01	12	37.68
点蚀故障2	0.02	12	37.68
点蚀故障3	0.02	24	75.36

表3 太阳轮裂纹故障尺寸 单位:mm

样本类型	长度	宽度
裂纹故障1	15	0.04
裂纹故障2	15	0.06
裂纹故障3	18	0.06

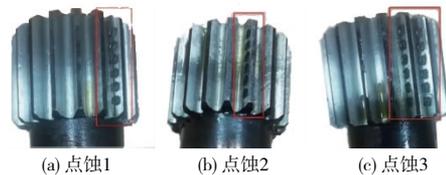


图5 不同故障尺寸的点蚀故障模拟件

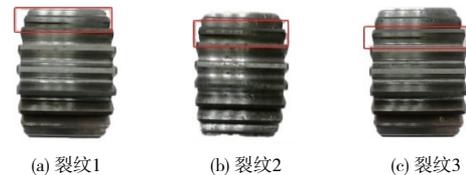


图6 不同故障尺寸的裂纹故障模拟件

为了实现转速与负载的变化,借助控制台调节驱动电机转速与磁粉制动器电流以模拟行星齿轮箱在多种工况下运行。实验采集了共195种工况下的数据:15种转速600~2000r/min,每次递增100r/min;13种负载0~27Nm,每次递增2.25Nm。使用的数据采集器型号为uT3604FRS-ICP,设置的采样率为40960Hz,采样时间为25.6s。

实验选取转速为1200r/min、载荷为9Nm工况下采集的振动信号作为实验数据。用滑动窗口对时域信号截取样本,设置窗口长度为2048,步长为512,确保每个样本包含足够信息。最终截取各类状态样本分别为400个,具体信息如表4所示。

表 4 具体样本信息

样本类型	样本数		样本标签
	训练集	测试集	
正常	340	60	0
点蚀故障 1	340	60	1
点蚀故障 2	340	60	2
点蚀故障 3	340	60	3
裂纹故障 1	340	60	4
裂纹故障 2	340	60	5
裂纹故障 3	340	60	6

在本实验中,分别从表 4 每类样本中抽取 3、6、9、12、18、30、60 个组成训练集,模拟故障样本有限的情况,另取每种健康状态样本 60 个组成测试集。

### 3.2 故障诊断结果分析

为了验证所提方法的优越性,将所提网络分别与胶囊网络、WDCNN+自注意机制、WDCNN 进行对比。其中,对比网络均与所提网络中对应结构参数一致。所有算法优化均采用 Adam 优化器,其初始学习率设置为 0.001,衰减速率设置为 0.01,并采用 Keras 深度学习框架,计算机配置为 Windows10 系统、intel core i5 处理器。各模型均训练 10 次,取其平均值为最终诊断准确率,所得结果如图 7 所示。

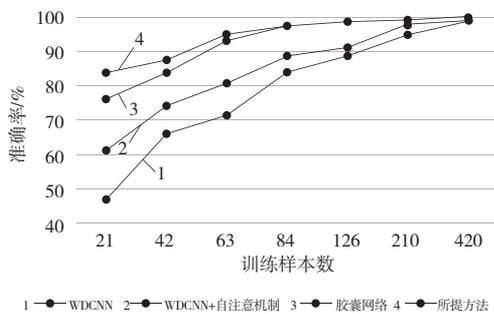


图 7 不同样本数下各模型诊断准确率

由图 7 可知,4 种方法诊断准确率均随样本数增加而提升,在样本数达到 210 的情况下,各模型均能有效提取特征并获得 90% 以上的准确率。但是在样本数量较小 (21、42、63) 的情况下,使用胶囊网络的模型所得结果明显优于另外两种模型。这主要是因为胶囊网络有很强的细节特征提取能力。在训练样本数仅为 21 的情况下,胶囊网络准确率为 75.89%,而所提方法准确率高达

83.66%,表明自注意机制能关注到对故障更加敏感的特征,提高了模型识别能力。综上分析,所提方法集胶囊网络与自注意机制的优势于一体,能更好地处理有限样本数据。

为了更直观地看出各故障模式的分类准确率,图 8 展示了训练集样本数为 126 时对所提方法随机进行一次实验其结果的混淆矩阵。其中,横轴代表测试集预测标签,纵轴代表测试集真实标签,主对角线代表所提模型对各类别预测准确率。

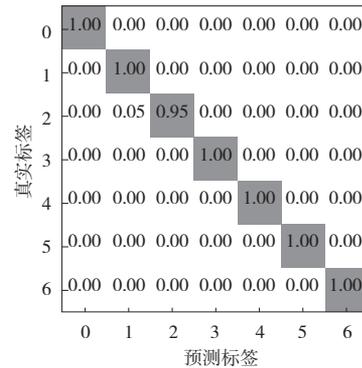


图 8 混淆矩阵结果

该次所获得的诊断准确率结果为 99.29%,从图 8 中可以看出,各类故障模式的识别准确率可达 95% 及以上,均取得了较高的结果,证明了所提模型的有效性与优越性。

由于深度学习“黑箱”的特征,其学习过程难以表述,为了直观显示特征提取结果,对胶囊网络和所提方法最终的输出进行可视化,如图 9 所示。图中 0 为正常,1—6 为不同的故障类型(本刊黑白印刷,相关疑问咨询作者)。可以看出,所提自注意机制胶囊网络模型可以使同类样本聚集更紧凑,各类之间空间分布差异更大,并且所提模型更易将故障 1、故障 2 及故障 5、故障 6 区分开来。这意味着所提模型能更好区分不同程度损伤故障。

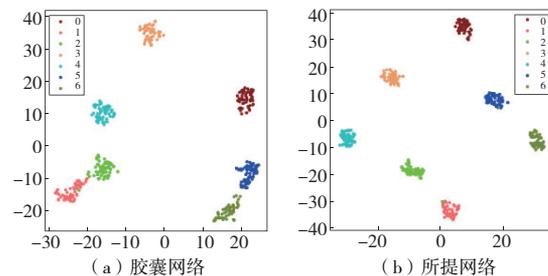


图 9 t-SNE 可视化图

(下转第 105 页)

具头以提升加工效率。

3) 层进给量增大, 会提高板料的成形性能。但随着层进给量增大, 成形力也会随之增大, 需要相应地提高振幅。

4) 较优的敲击式渐进成形工艺参数为: 振幅 15 mm、频率 16 Hz、工具头直径 6 mm、层进给量 1.5 mm。

#### 参考文献:

- [1] 鲁世红, 付婧颐, 丁腾飞, 等. 金属板料数控渐进成形工艺的研究现状[J]. 锻压技术, 2022, 47(10): 1-11.
- [2] 谷岩波, 王辉, 王会廷, 等. AA5754 铝合金板材渐进成形壁厚均匀性研究[J]. 锻压技术, 2018, 43(1): 33-41.
- [3] BEHERA A K, SOUSA R D, INGARAIO G, et al. Single point incremental forming: an assessment of the progress and technology trends from 2005 to 2015[J]. Journal of Manufacturing Processes, 2017, 27: 37-62.
- [4] 王新太, 王进, 顾宁, 等. 渐进成形工艺参数对 6061 铝合金板材成形性能的影响[J]. 锻压技术, 2019,

44(7): 29-33, 46.

- [5] 安治国, 庞鹏辉, 陈科衡, 等. 有模单点渐进成形工艺参数对 1060 铝合金成形性能的影响[J]. 锻压技术, 2017, 42(11): 94-98, 171.
- [6] KUMAR A, GULATI V, KUMAR P, et al. Parametric effects on formability of AA2024 - O aluminum alloy sheets in single point incremental forming[J]. Journal of Materials Research and Technology, 2019, 8(1): 1461-1469.
- [7] WANG J, LI L H, JIANG H S, et al. Incremental sheet punching on the base of sinusoidal tool path[J]. The International Journal of Advanced Manufacturing Technology, 2017, 89(9): 3527-3534.
- [8] ASGARI A, SEDIGHI M, RIAHI M. Investigation of punching parameters effect on mechanical properties of Al - 1100 - O in incremental sheet metal hammering process [J]. Strength of Materials, 2015, 47(6): 882-889.

收稿日期: 2022 - 11 - 25

(上接第 70 页)

#### 4 结语

本文提出了一种基于自注意机制胶囊网络故障诊断方法, 由实验所得结论如下:

1) 在数据样本不足的情况下, 自注意机制胶囊网络分类准确率明显优于其他模型, 证明所提方法故障诊断性能极佳;

2) t-SNE 可视化结果表明, 自注意机制胶囊网络能更好地捕获易于区分的故障特征, 在特征学习方面更加高效智能。

#### 参考文献:

- [1] 邱星辉, 韩勤轲, 褚福磊. 风力机行星齿轮传动系统动力学研究综述[J]. 机械工程学报, 2014, 50(11): 23-36.
- [2] 陈雪峰, 郭艳婕, 许才彬, 等. 风电装备故障诊断与健康监测研究综述[J]. 中国机械工程, 2020, 31(2): 175-189.
- [3] HAN T, LIU C, WU L, et al. An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems[J]. Mechanical Systems and Signal Processing, 2019, 117: 170-187.
- [4] INCE T, KIRANYAZ S, EREN L, et al. Real-time motor

fault detection by 1-D convolutional neural networks[J]. IEEE Transactions on Industrial Electronics, 2016, 63(11): 7067-7075.

- [5] XU Z, LI C, YANG Y. Fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks[J]. Applied Soft Computing, 2020, 95: 106515.
- [6] 余传粮, 梁睿君, 冉文丰, 等. 基于 STFT 和 CNN 的齿轮箱故障诊断[J]. 机械制造与自动化, 2022, 51(3): 152-154, 195.
- [7] HAN T, MA R, ZHENG J. Combination bidirectional long short-term memory and capsule network for rotating machinery fault diagnosis [J]. Measurement, 2021, 176: 109208.
- [8] 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.
- [9] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules [EB/OL]. (2017-11-07) [2022-11-25]. <https://arxiv.org/abs/1710.09829>. pdf.

收稿日期: 2022 - 11 - 25