

基于孪生网络的行星齿轮箱故障诊断方法

钱心筠,王友仁,赵亚磊

(南京航空航天大学 自动化学院,江苏 南京 211106)

摘要:针对行星齿轮箱在实际工作中故障训练样本有限的问题,设计一种基于孪生卷积神经网络的故障诊断方法。采集行星齿轮箱多方向振动信号,使用卷积神经网络对多路振动信号进行特征提取,利用孪生网络架构对所提取特征样本进行距离度量,确定分类准则,完成故障诊断模型的训练。通过行星齿轮箱实验平台的数据实验,证明孪生卷积神经网络在故障训练样本有限的情况下相比传统方法具有更好的故障诊断性能。

关键词:故障诊断;孪生网络;行星齿轮箱;深度学习;有限数据

中图分类号:TH132.425; TP277.3 **文献标志码:**B **文章编号:**1671-5276(2022)04-0116-04

Method of Planetary Gearbox Fault Diagnosis Based on Siamese Networks

QIAN Xinyun, WANG Youren, ZHAO Yalei

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 21106, China)

Abstract: A fault diagnosis method based on Siamese convolutional neural networks is adopted to address the problem of limited fault training samples for planetary gearboxes in practice. The multi-directional vibration signals of the planetary gearbox are collected and the convolution neural network is used for feature extraction of multiple vibration signals. With Siamese network architecture, the distance of the extracted feature samples is measured and the classification criteria is determined to accomplish the training of the fault diagnosis model. The data gained from the experiments on the planetary gearbox experimental platform demonstrates that the Siamese convolutional neural network has better fault diagnosis performance than the conventional methods in the condition of limited fault training samples.

Keywords: fault diagnosis; Siamese network; planetary gearbox; deep learning; limited data

0 引言

行星齿轮箱作为一种传动效率高、承载能力强的机械装置,被广泛用于新能源发电、军事、航空航天等涉及国计民生的领域。由于行星齿轮箱长期工作在变转速重载的恶劣环境下,太阳轮、行星轮等关键部件极易发生故障,易引发传动系统的失效。由于行星齿轮箱本身受到一定维护加之部分故障发展较慢,因而无法收集到足够的故障样本用于故障诊断^[1]。

随着机器学习尤其是深度学习理论的发展,在训练样本有限的情况下行星齿轮箱的故障诊断已成为当前的研究热点。针对这类问题,在小样本学习领域一般可从数据、算法等角度进行处理,目前,算法层面上主要是通过提取故障敏感特征,并对故障分类上选用支持向量机(support vector machine, SVM)这类有小样本分类能力算法进行改进来解决此类问题,这类方法所提取特征质量密切相关,需要大量专家知识^[2]。因而,最为常用的方法仍然是在数据层面对训练数据集进行增强,效果比较好的是通过生成对抗网络(generative adversarial nets, GAN)实现样本扩充。文献[3]将堆栈自动编码器作为GAN的判别器,生成了与原始分布相似的样本,实现了在故障样本量

较小情况下的故障诊断。文献[4]通过有限元仿真生成的轴承仿真本来弥补故障样本缺失的不足,将仿真样本和真实样本送入GAN实现样本扩充。文献[5]利用条件变分自动编码器获取了原始故障样本分布,通过生成器为解码网络生成大量故障样本,实现了行星齿轮箱故障的有效诊断。然而这类基于GAN的样本增强方法往往只学习到了对象在分布上的特点,当面对某些类别的样本量极少时,所训练出的模型并不能达到令人满意的效果^[6]。

BROMLEY J等提出的孪生网络(Siamese networks)^[7]目前已广泛应用于人像识别、指纹识别、目标追踪等模式识别领域。孪生网络通过两个结构相同的网络对特征进行提取,而后基于距离度量给出分类准则。该网络通过训练集样本组成样本对的方式,增加了对单个样本的利用次数,实现了在有限的训练样本中获取到多类别样本间的共同特征,为在样本量有限情况下的故障诊断提供了一种新的思路。

本文选用一种基于孪生卷积神经网络(Siamese convolution neural networks, Siamese-CNN)的行星齿轮箱故障诊断方法,以应对故障训练样本有限的问题。该方法利用CNN自动提取行星齿轮箱振动信号特征,将特征输入至孪生网络架构,基于样本间的距离度量来优化网络参数。测试阶段通过将测试样本和训练集中部分样本组成样本对

输入至 Siamese-CNN 网络中进行故障诊断,最后以行星齿轮箱实验平台的数据为例验证了该方法的有效性。

1 孪生卷积神经网络理论

1.1 孪生网络概述

孪生网络的基本结构如图 1 所示,网络 1 和网络 2 是两个相同的神经网络,用于将输入样本对映射到同一个高维空间,通过特征向量距离的相似度量,输出一个在 $[0,1]$ 之间的数作为该样本对样本间的相似度。通过这种基于距离度量学习的方式,可以极大地减少网络参数的训练负担。对于一个样本量 M 的训练集而言,可以组成 $M(M-1)/2$ 个不同的训练样本对,这样增加了样本的利用率,又借此减少对训练样本量的依赖。在测试时,一条网络的输入是训练集的样本,用于提供先验知识;另一条支路则是测试集样本。通过训练好的网络获取样本间的相似度,确定测试样本状态。

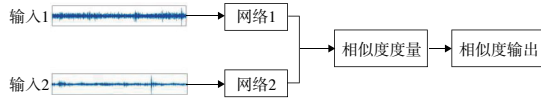


图 1 孪生网络基本结构

1.2 孪生网络训练与测试策略

图 2(a) 所示为孪生网络训练过程,输入网络的是一对相同或者不同类别的样本数据,输出则是经相似度量后这个样本对样本间相似的概率,通过得到的相似概率与实际的差距来调整模型参数。

测试策略上借鉴元学习理论^[8],从训练集中选取每类 N 个样本作为支持集,如图 2(c) 所示,采用 N -shot K -way 测试准则,确定测试样本的分类。

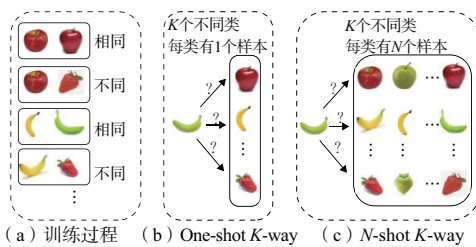


图 2 孪生网络训练与测试策略

在 one-shot K -way 测试过程中,给定一个待分类的测试样本 $\hat{x}$ 以及一个如式(1)的支持集 S 。

$$S = \{(x_1, y_1), \dots, (x_k, y_k)\} \quad (1)$$

式中 $y_1, y_2, \dots, y_k$ 为样本标签。然后将 $\hat{x}$ 与支持集样本一一比对,得到样本对两个样本间的相似概率,如式(2)选取相似概率数值最大的作为当前测试样本的类别。

$$C(\hat{x}, S) = \arg\max_c (P(\hat{x}, x_c)), x_c \in S \quad (2)$$

对于 N -shot K -way 测试过程,则是进行了 N 次 one-shot K -way 的测试过程,支持集 S_n 是 N 个形如式(1)

的集合组成。将每一类别所得到的 N 个相似度求和,选取各个类别求和所得的最大值作为当前测试样本的最终类别,其表示式如(3)所示。

$$C(\hat{x}, (S_1, \dots, S_N)) = \arg\max_c (\sum_{n=1}^N P(\hat{x}, x_{cn})), x_{cn} \in S_n \quad (3)$$

1.3 孪生卷积神经网络训练机制

1) 相似度量

孪生卷积神经网络是指图 1 中的网络 1、网络 2 均为卷积神经网络,在这里选用包含 5 个卷积层、5 个最大值池化交替组成的卷积网络结构。该网络使用 $L1$ 距离度量 CNN 网络,其表达式如(4)所示。

$$d_f(x_1, x_2) = \|f(x_1) - f(x_2)\| \quad (4)$$

式中 $f(\cdot)$ 是所使用的卷积神经网络。最后,如式(5)所示,通过全连接操作生成最终的相似度,

$$P(x_1, x_2) = \text{sigm}\{\text{FC}[d_f(x_1, x_2)]\} \quad (5)$$

式中 FC 为全连接层,这里采用 Dropout 进行全连接降维,增强了模型泛化性;sigm 是 sigmoid 激活函数,用于生成两个向量之间的最终相似度。当值越接近则代表两者越相似。

2) 损失函数

采用交叉熵损失函数,对于一个训练样本对而言,令 t 表示两个样本是否属于同一类。当 $t=1$ 表示输入样本对的两个样本属于同类, $t=0$ 代表两个样本不属于同一类,同时引入带自适应系数 λ 的 $L2$ 权重衰减系数项 $\lambda \|W\|^2$ 来防止模型过拟合,得到如式(6)所示的误差控制函数。

$$L(x_1, x_2, t) = t \log(P(x_1, x_2)) + (1-t) \log(1-P(x_1, x_2)) + \lambda \|W\|^2 \quad (6)$$

3) 优化

孪生网络使用 mini-batch 与 Adam 相结合的优化训练机制,计算每个参数单独的自适应学习率,网络参数更新遵循以下规则:

$$\begin{aligned} s_w^{(T+1)} &= \rho_1 s_w^{(T)} + (1-\rho_1) \nabla_w L^{(T)} \\ v_w^{(T+1)} &= \rho_2 v_w^{(T)} + (1-\rho_2) (\nabla_w L^{(T)})^2 \\ \hat{s}_w &\leftarrow \frac{s_w^{(T+1)}}{1 - (\rho_1)^{T+1}} \\ \hat{v}_w &\leftarrow \frac{v_w^{(T+1)}}{1 - (\rho_2)^{T+1}} \\ w^{(T+1)} &\leftarrow w^{(T)} - \beta \frac{\hat{m}_w}{\sqrt{\hat{v}_w} + \delta} \end{aligned} \quad (7)$$

式中 $w^{(T)}$ 是在第 T 个 epoch 的参数; $L^{(T)}$ 是如式(6)所示的损失函数; ρ_1 、 ρ_2 分别是 Adam 优化算法的一阶矩、二阶矩遗忘因子; s 和 v 代表移动平均线。

4) 训练

对于第 i 个 mini-batch 的输入样本对 (x_1^i, x_2^i) ,通过式(5)生成两者的相似度,损失计算则通过式(6)完成,模型参数优化由式(7)完成。

2 行星齿轮箱故障诊断算法

本文采用一种基于孪生卷积神经网络的行星齿轮箱故障诊断方法,该方法通过卷积神经网络提取行星齿轮箱

多方向振动信号特征,利用孪生网络的度量学习能力,实现了在故障训练样本有限的情况下的故障特征分类。

基于 Siamese-CNN 的故障诊断算法流程如图 3 所示,具体实施步骤如下:

1) 获取行星齿轮箱多方向的振动信号,划分训练集和测试集,并将训练集两两一组形成样本对,输入 Siamese-CNN 网络中。

2) 训练 Siamese-CNN。首先通过卷积神经网络提取样本对的故障特征,经历 $L1$ 距离度量、全连接生成相似度,根据全连接生成的相似度、样本对实际情况构建的误差损失函数反向优化 CNN 模型参数。反复训练直至完成所设定的迭代次数。

3) 将测试样本与支持集样本形成的样本对,输入训练好的 Siamese-CNN 中进行行星齿轮箱故障诊断,分析诊断结果。

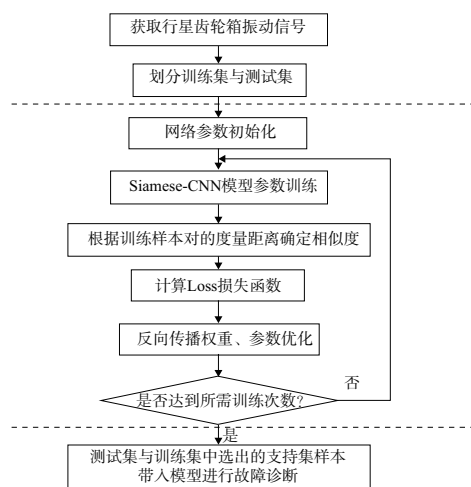


图3 行星齿轮箱故障诊断流程图

3 实验结果与分析

3.1 行星齿轮箱故障诊断实例

1) 实验平台及数据简介

采用的行星齿轮箱故障模拟实验平台如图 4 所示。实验平台主要由变速驱动电机、平行齿轮箱、行星齿轮箱与磁粉制动器组成。在本实验中分别在行星轮和太阳轮上设置了裂纹、磨损、断齿这 3 类故障,6 种故障模式。实验采集了共 121 种工况下的数据:11 种转速 $600 \sim 1\,500$ r/min,11 种负载 $0 \sim 22.5$ N·m。振动加速度采样率为 20 480 Hz,采样时间为 51.2 s。

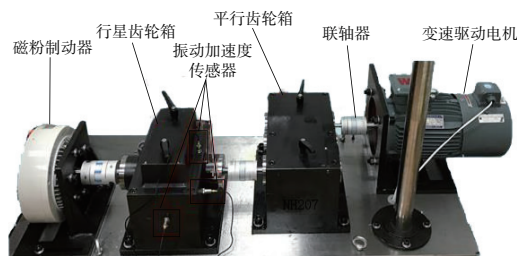


图4 行星齿轮箱实验平台

2) 样本数据集

实验选取工况为 900 r/min、 0 N·m; $1\,050$ r/min、 9 N·m, $1\,200$ r/min、 13.5 N·m 这 3 种工况下的三通道时域数据,将 3 路信号作为整体,选取每种工况下 6 s 数据,前 3 s 用于训练,剩余用于测试。其中,在训练样本的构建上使用时域信号滑窗的方式划分样本,步长为 80,滑窗长度为 2 048,以减少因信号截断对诊断造成的影响。在测试样本的构建上,则以 2 048 为长度截取测试样本。最终获取到每种模式 2 100 个训练集样本、90 个测试样本,具体信息如表 1 所示。

表1 故障训练集

故障类型	样本量		故障标签
	训练集	测试集	
正常	2 100	90	0
行星轮裂纹	2 100	90	1
行星轮磨损	2 100	90	2
太阳轮断齿	2 100	90	3
太阳轮裂纹	2 100	90	4
太阳轮磨损	2 100	90	5
行星轮断齿	2 100	90	6

实验中,分别从表 1 所示的样本集中随机抽取每种模式 9、12、20、30、40、70、150 个样本组成训练集,用于模拟故障训练样本量不足的情况,测试集则总样本量为 630。

3.2 故障诊断结果分析

本文所使用算法均在 Python 环境下使用 Google 研发的 Tensorflow 框架实现,神经网络通过 keras 搭建,程序在搭载有 Intel Core i5-8500 处理器 16G 内存的计算机上运行,操作系统为 Windows 10。

本实验将 SVM、分类器为 Softmax 的 CNN 与 Siamese-CNN 方法进行比较。其中,SVM 方法在核函数上采取高斯核函数,分别输入与上文所述样本集(SVM1)以及该样本集经人工提取特征后所得样本(SVM2)。CNN 方法的特征提取架构与 Siamese-CNN 相同,两种深度学习架构学习率设置为 0.001,优化方法采用 Adam 算法。每种模型训练 5 次,取其平均值作为诊断准确率(图 5)。

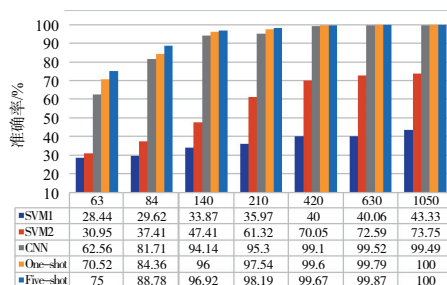


图5 不同诊断模型在不同训练样本量下比较

由图 5 可知,任何一种方法的故障诊断准确率都随样本量的增大而增大。然而,SVM 这类传统方法在样本量

较小的情况下表现不佳。当训练集样本总量为 63 时,使用两种类型数据训练的 SVM 方法仅能达到不足 30% 的故障诊断准确率,远小于深度学习方法。直接使用所截取的原始时域数据训练的 SVM 所得模型较经特征提取后所得模型而言,故障准确率随样本增加上升缓慢。

相比而言,基于 CNN 的诊断方法有着较高的诊断准确率。为了更好地理解基于孪生网络的故障诊断方法在样本量有限的情况下的效果,绘制出在训练集总样本量为 140 情况下的 CNN、Siamese-CNN 中 one-shot 测试方法提取的特征 TSNE 可视化散点图以及网络分类结果(图 6、图 7)。从图 6 中可以看到,相比传统 CNN 方法,Siamese-CNN 所提取的特征更加充分,尤其在行星轮断齿与太阳轮磨损间的故障特征上更加明显。从图 7 分类结果上看,使用 Siamese-CNN 方法在行星轮断齿、太阳轮磨损故障相比传统 CNN 方法更容易诊断,分类精度明显高于 CNN 方法。

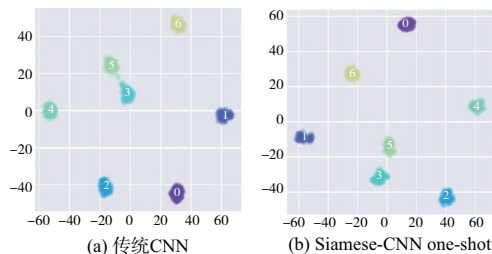


图 6 故障特征可视化

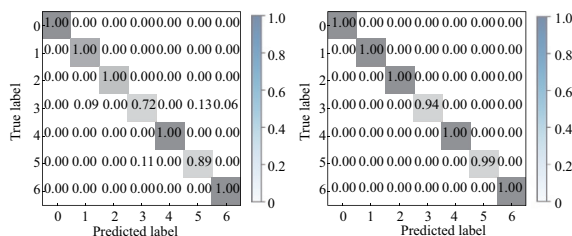


图 7 分类结果混淆矩阵

根据上述分析可知,孪生卷积神经网络能在一定程度上减轻对样本的依赖,在样本量有限的情况下取得较高的诊断准确率以及分类精度。

4 结语

本文采用了一种基于孪生卷积神经网络的故障诊断方法,通过实验得出以下结论:

- 1) 基于孪生卷积神经网络的故障诊断方法通过 CNN 自动提取敏感故障特征,而后基于度量学习方法确定样本分类,在一定程度上减少了对样本的依赖。
- 2) 基于孪生卷积神经网络的故障诊断方法,故障诊断准确率明显优于 SVM 这类传统解决方法。
- 3) 相比传统 CNN 方法,基于孪生卷积神经网络区分度好,在每一种模式上的分类精度都更高。

参考文献:

- [1] 雷亚国,何正嘉,林京,等. 行星齿轮箱故障诊断技术的研究进展[J]. 机械工程学报,2011,47(19):59-67.
- [2] WANG Y Q, YAO Q M, KWOK J T, et al. Generalizing from a few examples[J]. ACM Computing Surveys, 2020, 53(3): 1-34.
- [3] WANG Z R, WANG J, WANG Y R. An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition[J]. Neurocomputing, 2018, 310: 213-222.
- [4] GAO Y, LIU X Y, XIANG J W. FEM simulation-based generative adversarial networks to detect bearing faults [J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4961-4971.
- [5] WANG Y R, SUN G D, JIN Q. Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network [J]. Applied Soft Computing, 2020, 92: 106333.
- [6] SUH S, LEE H, JO J, et al. Generative oversampling method for imbalanced data on bearing fault detection and diagnosis [J]. Applied Sciences, 2019, 9(4): 746.
- [7] BROMLEY J, BENTZ J W, BOTTOU L, et al. Signature verification using a “Siamese” time delay neural network [J]. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4): 669-688.
- [8] 李凡长,刘洋,吴鹏翔,等. 元学习研究综述[J]. 计算机学报, 2021, 44(2): 422-446.

收稿日期: 2021-04-19