

采用残差网络与卷积注意力机制的设备剩余使用寿命预测方法

莫仁鹏, 李天梅, 司小胜, 朱旭
(火箭军工程大学导弹工程学院, 710025, 西安)

摘要: 针对以往剩余使用寿命(RUL)预测方法均等对待深层特征中具有不同重要性的空间域和通道域特征,导致大量的计算资源浪费在不重要的特征上,进而造成 RUL 预测值偏差过大、不能对失效设备进行及时的维护以避免潜在的安全隐患的问题,提出一种采用残差网络与卷积注意力机制的端到端的 RUL 预测方法。该方法以卷积层和池化层对原始监测信号进行浅层特征提取与压缩;利用堆叠残差模块在学习深层特征的同时,缓解梯度弥散以及网络退化现象的发生;由卷积注意力模块对设备的深层退化特征进行加权赋值,分别在其空间维度和通道维度上强化更重要的特征并抑制相对不重要的特征,使网络的注意力集中在对 RUL 预测任务更关键的信息上;将加权后的特征输入到全连接网络中映射得到 RUL 预测值。通过 PHM2012 轴承数据集进行了实验验证,实验结果表明,卷积注意力和残差结构皆对改善模型的预测性能有着积极的作用,所提方法在测试轴承上的均方根误差和平均绝对误差分别为0.107 9和0.083 1,远低于其他对比方法。

关键词: 剩余使用寿命;残差网络;卷积注意力;端到端

中图分类号: TH17 **文献标志码:** A

DOI: 10.7652/xjtuxb202204021 **文章编号:** 0253-987X(2022)04-0194-09



OSID 码

Remaining Useful Life Prediction for Equipment Using Residual Network and Convolutional Attention Mechanism

MO Renpeng, LI Tianmei, SI Xiaosheng, ZHU Xu
(Department of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: Regarding the previous remaining life prediction methods, the spatial domain and channel domain features of different importance in the deep features are often treated equally, causing a lot of computing resources to be wasted on unimportant features, which leads to excessive deviations in RUL predictions and failure to perform timely maintenance on failed equipment to avoid potential safety hazards, an end-to-end RUL prediction method using residual network and convolutional attention mechanism is proposed. This method first uses the convolutional layer and the pooling layer to extract and compress the shallow features of the original monitoring signal. Then uses the stacked residual module to learn deep features while mitigating the occurrence of gradient dispersion and network degradation. Next, the convolutional attention module weights the deep degradation features of the device to strengthen the more important features and suppress the relatively unimportant features in its spatial dimension and channel dimension, so that the network's attention is focused on information that is more critical

to RUL prediction task. Finally, the weighted features are input into the fully connected network and mapped to obtain the RUL prediction value. The experimental verification is carried out through the PHM2012 bearing data set. The experimental results show that both the convolutional attention and the residual structure have a positive effect on improving the prediction performance of the model. In addition, the root mean square error and average absolute error of the proposed method on the test bearing are 0.107 9 and 0.083 1, which are much lower than other comparison methods.

Keywords: remaining useful life; residual network; convolutional attention; end-to-end

设备或零部件的性能一旦退化至失效状态,轻者会阻碍正常的运行过程,更为严重的是将有可能造成难以承受的经济损失或重大安全事故^[1-2]。根据监测数据预测设备的剩余使用寿命(RUL),并由此实施及时精准的预测性维护、提前规划维护资源配置,对于保障设备的运行安全性、可靠性与经济性具有重要意义^[3]。设备的RUL预测技术已成为过去十多年的研究热点,其应用也已拓展到高速列车^[4]、航空航天装备^[5]、风电设备^[6]等领域。

传统上对设备进行RUL预测的方法主要根据其失效机理来建立物理退化模型^[7-8],但是这种方法需要对设备的物理结构和机理特性有着较深入的了解,且在实际应用中难度较大^[9]。随着计算机科学和传感器技术的发展,数据驱动的RUL预测方法得到了广大研究者的关注,借助于传感器对设备的运行过程进行监测,可以得到大量反映设备健康状态的数据,而数据驱动的方法利用这些数据之间的隐藏退化信息来建立模型,进而预测设备的RUL。

近年来,以深度学习算法为基础的数据驱动方法以其强大的学习和拟合能力在语音识别、图像处理、故障诊断等领域取得了很好的效果^[10],这也为设备的RUL预测提供了一种新的思路。各种深度学习算法中,卷积神经网络(CNN)的特征提取能力较强且具有一定的降噪功能^[11],因此CNN在RUL预测领域得到广泛应用。例如,张钢等将轴承原始振动信号的时频图输入到多尺度 AlexNet 网络中构建健康因子,所得的健康因子具有较好的单调性与趋势性,但是并未进一步验证该方法在RUL预测时的效果^[12]。Zhang 等采用时间窗技术和改进的差分方法得到若干个退化特征,并将这些退化特征和原始数据一并输入到CNN中进行RUL预测,虽然具有良好的预测精度,但是其预处理步骤较为繁琐^[13]。Zhu 等将小波变换得到的轴承时频表示输入到CNN中进行自动特征学习,并在网络深层采取多尺度结构以同步学习全局信息和局部信息,

进而提高RUL预测精度,然而该方法得到的预测曲线中蕴含噪声较大、抖动剧烈^[14]。

此外,以上基于CNN的RUL预测方法还存在以下不足:首先,CNN网络利用多个卷积层从设备监测信号中提取得到的深层特征含有通道和空间两个维度,但是不同通道特征或不同空间特征之间所含的退化信息量是不一样的,因此对RUL预测任务的贡献度也是不一样的,而以往的方法往往忽视了这一点,这将对设备的RUL预测带来不利影响;其次,传统的CNN网络需要有足够的深度才能有较好的学习能力,但是当网络堆叠到一定深度时会存在梯度弥散以及网络退化现象,不利于预测任务的进行。

为了克服以上缺陷,本文提出一种结合残差网络与卷积注意力的端到端的设备RUL预测方法,所提方法直接将设备的原始监测信号输入到网络中进行RUL预测,并且利用卷积注意力对网络中的深层特征进行自适应的加权处理以强化重要的通道特征和空间特征,以及抑制相对不重要的特征。此外,通过嵌入残差模块来增加CNN网络的深度,可以在一定程度上缓解梯度弥散,避免网络退化现象的发生。最后,通过PHM2012轴承数据集验证本文方法的有效性与优越性,并对轴承的RUL预测曲线进行了平滑降噪处理。

1 残差网络

为了解决CNN网络过深导致的梯度弥散以及网络退化问题,He 等将残差神经网络(ResNet)引入到CNN中^[15],ResNet在多个非线性的卷积层之间增加了直连边,这种特殊结构具有缓解网络退化、避免梯度消失的功效^[16]。

此外,由于非线性激活函数的存在,每层网络的特征学习都是不可逆的,因此退化信息在网络中传播时或多或少会存在一些丢失^[17]。当网络加深时,过多的信息丢失甚至会影响预测模型的准确性。

残差模块的跨层跳跃连接直接将输入信息绕道传递给输出,有效缓解了退化信息在网络中的传播损耗,从而提高了网络对退化信息的利用效率。图1给出了本文所用残差模块的结构。

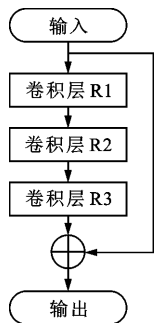


图1 残差模块的结构

Fig.1 Structure of Residual module

在图1残差模块中,卷积层R1、R2、R3的核数分别为 $K/4$ 、 $K/4$ 、 K ,核尺寸分别为 1×1 、 $S \times 1$ 、 1×1 ,这种先缩后扩的结构和嵌入的 1×1 卷积核,将有助于大幅减少网络参数,提高模型的预测速度。

2 卷积注意力

深度学习中的注意力机制借鉴了人类视觉注意力的思想,将注意力集中在对目标任务更重要的信息上,同时抑制无用信息的干扰,以此来提高神经网络的效率。目前已有多个神经网络模型与注意力机制结合在一起使用,并且在各自的任务中都取得了较好的效果^[18-20]。

鉴于此,本文亦在CNN中引入与设备RUL预测任务相适应的卷积注意力模块(CBAM)来进行特征重标定,以增强对预测任务贡献度更大的网络特征。CBAM主要由通道注意力和空间注意力组成^[21],分别在深层特征的通道维度和空间维度上强化重要信息并抑制无效信息。

2.1 通道注意力

卷积层的每个输出通道对预测任务的重要程度是不一致的,因此有必要采用通道注意力机制对不同通道的重要性程度进行评估,并为更重要的通道特征赋予更大的权重,以加强这些通道的影响。通道注意力的原理如图2所示。

设输入特征的维度为 $H \times C$,其中 H 为空间维度, C 为通道维度。通道注意力首先同时使用全局最大池化(GAP)和全局平均池化(GMP)来聚合输入特征的空间信息,将每一个通道内的空间信息压缩为一个表示符,得到两个 $1 \times C$ 的特征向量,再将

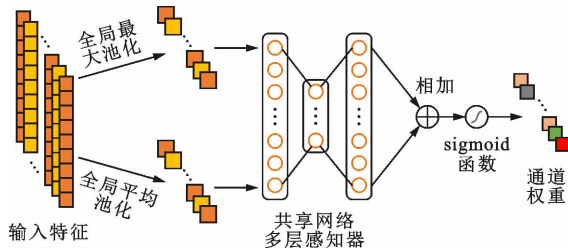


图2 通道注意力的原理

Fig.2 Principle of Channel attention

这两个特征向量分别输入到共享网络中,然后使共享网络输出的两个 $1 \times C$ 的一维向量相加合并,最后利用sigmoid函数进行归一化得到各通道的注意力权重。其中共享网络为一个多层感知器(MLP),共含有两个隐藏层,为了减少参数量,第1个隐藏层的神经元数设置为 C/r , r 为下降率,第2个隐藏层的神经元数再恢复为 C 。通道注意力的具体计算过程如下

$$M_c(F) = \delta[\text{MLP}(\text{GAP}(F)) + \text{MLP}(\text{GMP}(F))] = \delta[W_1(W_0(F_{\text{GAP}}^c)) + W_1(W_0(F_{\text{GMP}}^c))] \quad (1)$$

式中: F 为输入特征向量; $M_c(F)$ 为相应的通道注意力权重; $\delta(\cdot)$ 为sigmoid函数; W_0 和 W_1 为MLP两个隐藏层的参数; F_{GAP}^c 和 F_{GMP}^c 为两种池化在每个通道上聚合空间信息得到的特征表示符。

2.2 空间注意力

空间注意力关注输入特征空间维度上的重要信息,其原理如图3所示。

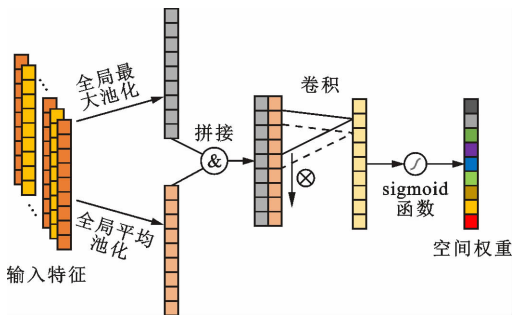


图3 空间注意力的原理

Fig.3 Principle of Spatial attention

空间注意力在输入特征的通道轴上同时使用全局最大池化和全局平均池化,分别得到两个维度为 $H \times 1$ 的特征向量,再将这两个特征向量沿通道方向拼接为 $H \times 2$ 的特征向量,然后通过单核卷积层将其映射为 $H \times 1$ 的特征向量,最后利用sigmoid函数进行归一化得到空间注意力权重,计算公式如下

$$M_h(F) = \delta(\text{conv}((\text{GAP}(F); \text{GMP}(F))) = \delta(\text{conv}((F_{\text{GAP}}^h; F_{\text{GMP}}^h))) \quad (2)$$

式中: $M_h(F)$ 为相应的空间注意力权重; conv 为单核卷积; F_{GAP}^h 和 F_{GMP}^h 为两种池化在每个空间位置上聚合通道信息得到的特征表示符。

2.3 混合注意力

通道注意力与空间注意力分别在输入特征的通道维度和空间维度上关注更有效的信息,两个注意力模块之间有着互补的作用,因此经常结合使用。通道注意力与空间注意力主要有 3 种组合方式^[22],第 1 是先使用通道注意力对输入特征 F 进行加权得到 F_1 ,然后再针对加权后的特征 F_1 使用空间注意力进行再次加权得到最终的输出 F_2 ,本文将这种组合方式简称为 CBAM0,其计算过程如下

$$\begin{cases} F_1 = M_c(F) \odot F \\ F_2 = M_h(F_1) \odot F_1 \end{cases} \quad (3)$$

式中: $\odot$ 表示逐元素相乘。

第 2 组合方式为 CBAM1,将空间注意力模块放在通道注意力模块之前使用,计算过程如下

$$\begin{cases} F_1 = M_h(F) \odot F \\ F_2 = M_c(F_1) \odot F_1 \end{cases} \quad (4)$$

第 3 组合方式为 CBAM2,对输入特征 F 同时使用通道注意力和空间注意力,分别得到通道注意力权重 M_c 和空间注意力权重 M_h ,再将这两种注意力权重一并赋予原输入特征,表达式如下

$$F_2 = M_c(F) \odot M_h(F) \odot F \quad (5)$$

本文使用一维卷积来提取一维时序监测数据的深层特征,以识别设备的退化程度并进行 RUL 预测。与二维卷积不同的是,一维卷积中的“空间”维度只含有 1 个维度,对一维时序数据而言该维度亦可看作为“时间”维度。以设备的单个采样点数据序列作为网络的输入,在经过多个卷积层的映射后,所得到的深层特征仍保留了一定的时序性,即单个采

样点内的短期时序性。

使用空间(时间)注意力模块,可以为更重要的时间点分配更大的权重,有利于模型预测性能的提高。如果采用 CBAM0 组合方式,先使用通道注意力进行加权,无疑会破坏输入特征的时序性,影响后面空间注意力的使用效果。此外,网络本身也具有一定的随机性,每次计算得到注意力权重也会有些许偏差,而 CBAM2 同时使用两种注意力权重与原输入相乘将放大这种偏差,具体表现为网络预测效果的稳定性欠佳。因此,本文最终采用了 CBAM1 的组合方式,先使用空间注意力模块来强调更重要的空间(时间)特征,再以通道注意力作为补充,强化重要通道并抑制相对不重要的通道,CBAM1 的结构如图 4 所示。

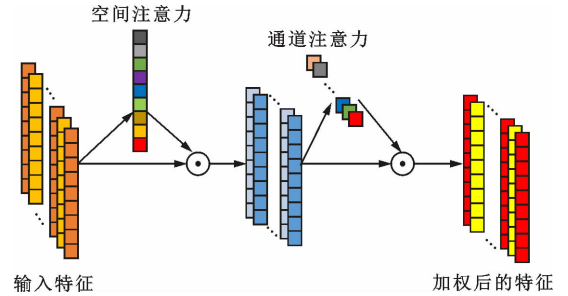


图 4 CBAM1 模块的结构

Fig. 4 Structure of CBAM1 module

3 采用残差网络与卷积注意力的设备剩余使用寿命预测方法

3.1 整体网络模型

本文方法结合了残差网络与卷积注意力的优势,在缓解退化信息在网络传播过程中的丢失与损耗的同时,将网络资源集中在对设备 RUL 预测任务更重要的特征上。图 5 给出了所提方法的整体网络结构,主要包含降维网络、残差网络、注意力网络以及预测网络 4 个子网络。

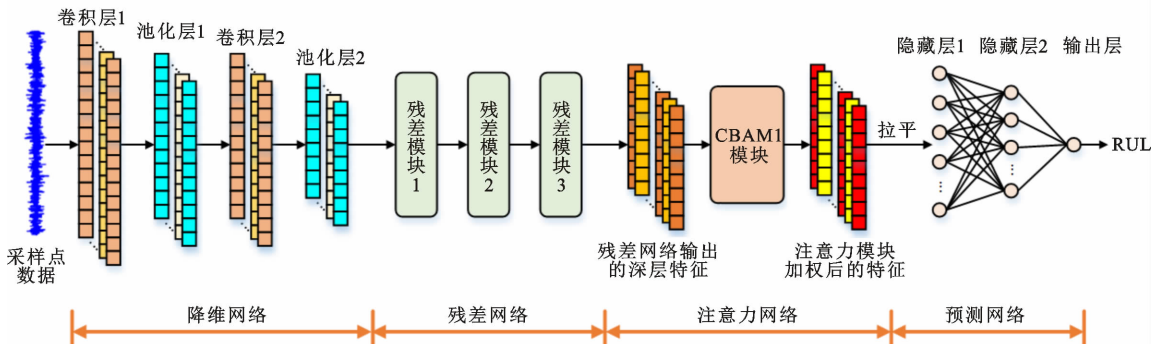


图 5 采用残差网络与卷积注意力的剩余使用寿命预测模型

Fig. 5 Residual life prediction model using residual network and convolutional attention

使用降维网络中的两个卷积层对设备的采样点数据进行初步的特征提取,并且通过两个最大池化层来压缩特征,以减少后续网络的计算量,从而提高模型的预测速度。其中卷积层的计算公式如下

$$x_{i,j}^{l+1} = \sigma(\sum_{i,j} x_i^l \otimes W_j^i + b) \quad (6)$$

式中: $\otimes$ 为卷积运算; W_j^i 为第 j 个卷积核上的第 i 个权值; b 为可学习的偏置; $\sigma(\cdot)$ 为ReLU激活函数; $x_{i,j}^{l+1}$ 表示第 l 层的第 i 个特征 x_i^l 经过卷积核 W_j^i 运算与激活函数激活后得到的第 $l+1$ 层的第 i 个特征。

利用残差网络进行深层特征提取,该网络由3个残差模块叠加组成,残差模块中的跳跃连接可以有效缓解退化信息在网络中传播损耗。

将残差网络学习到的深层特征输入到注意力网络中的CBAM1模块,该模块先后使用空间注意力和通道注意力以强化深层特征中相对重要的空间信息和通道信息。

将经过注意力网络加权后的深层特征拉平,并输入到预测网络中进行RUL预测。预测网络由3个全连接层组成,计算公式如下

$$x_i^{l+1} = \sigma(\sum_{j=1}^D W_{i,j} x_j^l + b) \quad (7)$$

式中: x_i^{l+1} 为第 $l+1$ 层第 i 个神经元的输出; D 为第 l 层的神经元总数; $W_{i,j}$ 为第 $l+1$ 层第 i 个神经元与第 l 层第 j 个神经元之间的权重参数。

限定预测网络输出层的神经元数为1,因此当加权后的深层特征经过3个全连接层的映射后,最终会得到1个单值数据,即设备的RUL预测值。

3.2 参数优化

设预测模型的训练集样本为 $\{(x_i, y_i)\}$,样本数据 x_i 为第 i 个采样点内的监测数据序列,如轴承的振动加速度序列,则样本标签 y_i 为该采样点所对应的RUL,表示第 i 个时刻的轴承还有 y_i 才失效。将样本数据 x_i 输入到预测模型中,经过网络的前向传播后最终可得到RUL预测值 $\hat{y}_i$,以预测值 $\hat{y}_i$ 与真实值 y_i 之间的均方误差作为损失函数 η ,表达式

$$\begin{cases} \hat{y}_i = f(x_i; \theta) \\ \eta(f(x_i; \theta)) = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \end{cases} \quad (8)$$

式中: $f(\cdot)$ 为预测模型整体网络所代表的映射; θ 为网络中可训练的参数; N 为样本总数。

在训练时以损失函数最小化为目标,对网络参

数进行更新与优化,从而降低RUL预测值与真实值之间的偏差。

3.3 RUL预测流程

所提方法对设备进行RUL预测的流程如图6所示,首先将设备的监测数据分为训练集和测试集,再以训练集作为输入对模型进行多次训练,若干次迭代后得到训练好的网络模型,然后就利用训练好的网络模型对测试集设备进行RUL预测,而测试集的RUL预测值曲线与真实值曲线之间的拟合程度则反映了模型的预测效果。

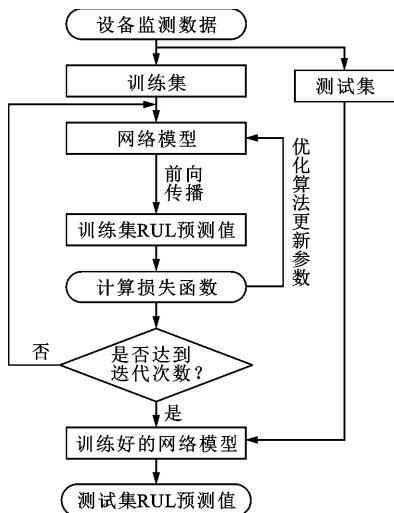


图6 本文方法对设备进行RUL预测的流程

Fig. 6 Process of this method for RUL prediction of equipment

4 实验仿真与结果分析

4.1 实验数据

通过PHM 2012轴承数据集^[23]来验证所提方法的有效性,该数据集包含3个工况下17个轴承的全寿命振动加速度数据。每个轴承内含有由正常阶段运行至失效过程中的若干采样点数据,其采样间隔为10 s,并且每个采样点内采集了2 560个振动加速度。选用轴承1-1、1-2、1-4、1-5、1-6作为训练集,轴承1-3、1-7、2-3、3-1作为测试集。

设训练集或测试集的样本为 $\{(x_i, y_i)\}$,样本序号 i 对应轴承的第 i 个采样点, x_i 为该采样点内的2 560个振动加速度。第 i 个采样点所对应的RUL为 $(N-i) \times 10$,为了降低网络的训练难度,将采样点RUL归一化后再作为样本的标签,表达式如下

$$y_i = \frac{N-i}{N}, 1 \leq i \leq N \quad (9)$$

4.2 实验设置

网络参数包括可训练的参数,例如卷积核的权重和偏置,以及不可训练的超参数,例如卷积核的数量和尺寸。超参数需要提前设置,并且需要综合考虑其对网络预测效果及训练时间的影响,经过反复实验对比,最终确定的超参数设置为:降维网络中两个卷积层的核数分别为 16、32,核尺寸皆为 6×1 ,而两个池化层的长度皆为 4;残差网络中 3 个残差模块的 K 参数分别为 32、64、64, S 参数皆为 5;注意力网络中的 r 参数为 2,而单核卷积层的核尺寸为 5×1 ;预测网络中 3 个全连接层的神经元数分别为 128、32、1。

需要说明的是,本文的网络模型基于 Keras 框架搭建实现,训练时选择 Adam 算法作为优化器,学习率为 0.000 1,一共训练 30 次。

此外,为了量化模型的预测效果,采用测试集轴承的 RUL 预测值与真实值之间的均方根误差 E_{rmse} 以及平均绝对误差 E_{mae} 作为评价指标来评估模型的预测效果,表达式如下

$$E_{\text{rmse}} = \sqrt{\frac{1}{P} \sum_{i=1}^P (y_i - \hat{y}_i)^2}$$

(10)

$$E_{\text{mae}} = \frac{1}{P} \sum_{i=1}^P |y_i - \hat{y}_i|$$

(11)

式中: P 为某测试轴承的样本总数。

4.3 结构分析

在第 2.3 节 3 种组合方式的 CBAM 中,选择了 CBAM1 作为预测模型中的注意力网络部分。为了验证该选择的合理性,在预测模型中分别加入 CBAM0、CBAM1、CBAM2 模块进行对比实验,此外,还额外增加了 3 个参照对象,分别是无注意力、只含通道注意力、只含空间注意力。为了减少实验复杂度,以轴承 1-3 为载体进行结构分析实验,观察含不同注意力模块的预测模型对该测试轴承的 RUL 预测效果,实验结果如表 1 所示。

表 1 不同组合方式的 CBAM 的评价指标
Table 1 Evaluation indicators of CBAM in different combinations

注意力结构	E_{rmse}	E_{mae}	预测时间/s
无注意力	0.059 7	0.050 4	3.35
通道注意力	0.057 2	0.047 6	3.64
空间注意力	0.056 3	0.047 7	3.68
CBAM0	0.053 3	0.045 2	3.88
CBAM1	0.050 9	0.041 2	3.94
CBAM2	0.052 9	0.045 1	3.87

由表 1 可以发现,相较于无注意力模块的网络模型,含注意力模块的网络模型的预测性能都有一定的提高,特别是使用 CBAM1 时,相应的 E_{rmse} 和 E_{mae} 的下降幅度最大,分别为 0.008 8 和 0.009 2。这验证了注意力机制对 RUL 预测任务的有效性,以及先使用空间注意力再使用通道注意力的策略的成功性。此外,CBAM1 为一个轻量模块,带来的额外预测耗时仅为 0.59 s,对 RUL 预测任务的时效性影响较小。

为了进一步观察注意力机制对模型训练的影响,分别对含有和不含有 CBAM1 的预测模型进行训练,并观察如图 7 所示两种模型损失函数(loss)在训练过程中的下降趋势。由图 7 可以看出,含有 CBAM1 的预测模型的损失函数下降得更快,尤其在训练的第一个迭代内,其损失函数就已经达到了一个较低的程度。这是因为在加入注意力机制后,预测模型可以将注意力集中在对 RUL 预测任务贡献度更大的特征上,所以能更快地校正更新网络参数,损失函数也就下降得更快。

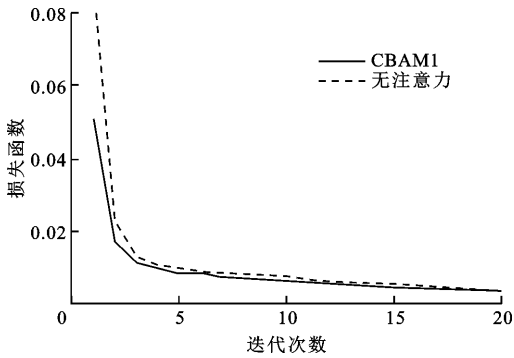


图 7 有无 CBAM1 的模型的损失函数下降曲线
Fig. 7 Loss drop curve of model with or without CBAM1

4.4 对比实验

将本文方法与深度神经网络(DNN)方法^[24]、卷积神经网络(CNN)方法^[25]以及去掉残差结构后的本文方法进行对比,这 4 种方法在 4 个测试轴承上 RUL 预测效果如表 2 所示。

由表 2 可知,去掉残差结构之后所提方法的预测性能有所下降,这从反面验证了具有缓解网络退化、减少信息传播损耗等功效的残差结构,对于改善模型预测性能有着积极的作用。结合残差网络与卷积注意力机制的优势,本文方法不仅可以有效克服特征信息在网络传播过程中的损失,并且将注意力集中在对 RUL 预测任务更重要的特征上,因此本文方法不仅在工况 1 中的轴承 1-3、1-7 上表现出了

较好的预测效果,在工况 2、工况 3 中的两个轴承上亦具有最佳表现,验证了所提方法的稳定性与鲁棒性。此外,虽然所提方法较其他方法结构更复杂,但是耗费的预测时间仅略微高于 CNN 方法,满足实时预测的要求,而 CNN 和 DNN 方法虽然耗时更短,但是预测效果不佳。

图 8 为本文方法、CNN 方法以及 DNN 方法在

轴承 1-3 上的 RUL 预测曲线,且都经过了加权平均方式的平滑降噪处理^[26]。由图 8 可以看出,本文方法的 RUL 预测曲线与真实曲线之间的拟合程度最好,更直观地展示了本文方法在对轴承进行 RUL 预测时的有效性与优越性。此外,由于 CNN 网络本身具有一定的降噪能力,因此本文方法与 CNN 方法的预测曲线的抖动较 DNN 方法小。

表 2 4 种方法在测试轴承上的预测效果

Table 2 Predictive effects of four methods on test bearings

方法	E_{rmse}					E_{mae}					平均预测
	轴承 1-3	轴承 1-7	轴承 2-3	轴承 3-1	平均值	轴承 1-3	轴承 1-7	轴承 2-3	轴承 3-1	平均值	耗时/s
DNN	0.158 6	0.272 7	0.281 4	0.262 5	0.243 8	0.286 4	0.254 3	0.252 1	0.254 7	0.261 9	0.71
CNN	0.066 1	0.166 4	0.135 6	0.239 7	0.152 0	0.055 6	0.126 7	0.115 1	0.207 1	0.126 1	2.39
去残差	0.064 9	0.160 9	0.098 7	0.248 3	0.143 2	0.049 8	0.113 0	0.078 0	0.213 3	0.113 5	2.78
本文	0.050 9	0.154 8	0.083 4	0.142 4	0.107 9	0.041 2	0.109 5	0.064 8	0.116 9	0.083 1	2.90

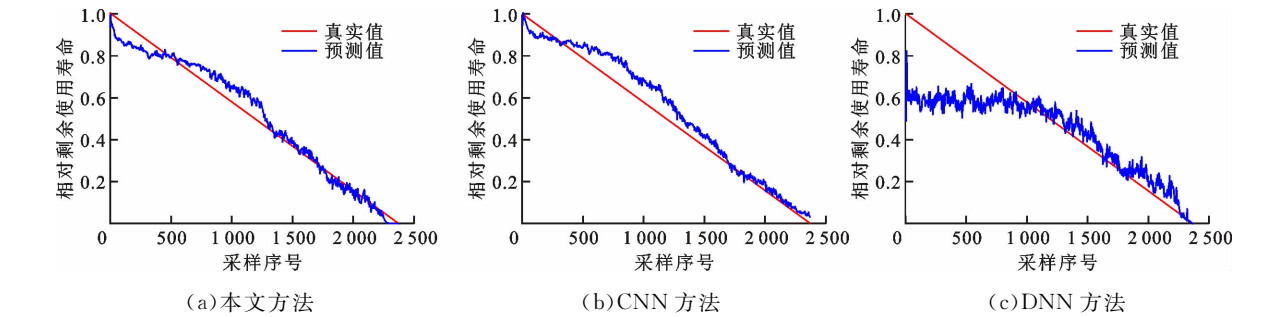


图 8 3 种方法在轴承 1-3 上的 RUL 预测曲线

Fig. 8 Three methods of RUL prediction curve on bearing 1-3

5 结 论

本文结合残差网络与卷积注意力的优势提出一种新的深度学习方法对设备进行 RUL 预测,并通过公开的轴承数据集对所提方法进行了验证,主要结论如下:

- (1)所提方法为端到端的 RUL 预测方法,将原始监测数据直接输入到网络中便可得到设备的 RUL 预测值,无需提前手动计算退化特征。
- (2)在深层网络引入卷积注意力先后在空间维度和通道维度上强化更重要的特征,不仅使模型在测试轴承 1-3 上的均方根误差和平均绝对误差分别下降了 0.008 8 和 0.009 2,有效改善了轴承的 RUL 预测精度,而且可以显著加快模型的收敛速度。
- (3)在不同工况轴承上的测试结果表明,所提方法的均方根误差和平均绝对误差的均值分别为 0.107 9 和 0.083 1,远低于其他对比方法,验证了所提方法的优越性与鲁棒性。其中,去残差模型的均

方根误差和平均绝对误差的均值分别为 0.143 2 和 0.113 5,高于所提方法完整结构的两个评价指标,这从反面证明了残差结构对设备 RUL 预测任务有着积极的作用。此外,所提方法的平均预测耗时为 2.90 s,满足实时预测的要求。最后,所提方法在轴承 1-3 上的 RUL 预测曲线与真实曲线之间的拟合效果最佳,更直观地展示了所提方法的有效性。

参考文献:

[1] 裴洪,胡昌华,司小胜,等. 基于机器学习的设备剩余寿命预测方法综述 [J]. 机械工程学报, 2019, 55(8): 1-13.
PEI Hong, HU Changhua, SI Xiaosheng, et al. Review of machine learning based remaining useful life prediction methods for equipment [J]. Journal of Mechanical Engineering, 2019, 55(8): 1-13.
[2] MOGHADDASS R, ZUO M J. An integrated framework for online diagnostic and prognostic health moni-

- toring using a multistate deterioration process [J]. Reliability Engineering & System Safety, 2014, 124: 92-104.
- [3] 李天梅, 司小胜, 刘翔, 等. 大数据下数模联动的随机退化设备剩余寿命预测技术 [J/OL]. 自动化学报 [2021-06-25]. <https://doi.org/10.16383/j.aas.c201068>.
- LI Tianmei, SI Xiaosheng, LIU Xiang, et al. Data-model interactive remaining useful life prediction technologies for stochastic degrading devices with big data [J/OL]. Acta Automatica Sinica [2021-06-25]. <https://doi.org/10.16383/j.aas.c201068>.
- [4] 董永刚, 仪帅, 黄鑫磊, 等. 重载列车紧急制动过程车轮踏面疲劳裂纹萌生寿命预测 [J]. 中国铁道科学, 2021, 42(5): 123-131.
- DONG Yonggang, YI Shuai, HUANG Xinlei, et al. Prediction of fatigue crack initiation life of wheel tread during emergency braking of heavy haul train [J]. China Railway Science, 2021, 42(5): 123-131.
- [5] HUANG Chenggeng, HUANG Hongzhong, LI Yanfeng. A bidirectional LSTM prognostics method under multiple operational conditions [J]. IEEE Transactions on Industrial Electronics, 2019, 66(11): 8792-8802.
- [6] 孙秋云, 朱才朝, 樊志鑫, 等. 误差对风电齿轮箱轴承疲劳寿命影响 [J]. 太阳能学报, 2021, 42(7): 308-315.
- SUN Qiuyun, ZHU Caichao, FAN Zhixin, et al. Effects of error on fatigue life of bearing for wind turbine gearbox [J]. Acta Energaie Solaris Sinica, 2021, 42(7): 308-315.
- [7] 孙振铎, 侯东勃, 李伟, 等. 基于裂纹萌生和扩展的渗氮钢疲劳寿命预测 [J]. 兵器材料科学与工程, 2021, 44(5): 76-82.
- SUN Zhenduo, HOU Dongbo, LI Wei, et al. Prediction of fatigue life of nitrided steel based on crack initiation and propagation [J]. Ordnance Material Science and Engineering, 2021, 44(5): 76-82.
- [8] 张军, 张犁, 成渝. IGBT 模块寿命评估研究综述 [J]. 电工技术学报, 2021, 36(12): 2560-2575.
- ZHANG Jun, ZHANG Li, CHENG Yu. Review of the lifetime evaluation for the IGBT module [J]. Transactions of China Electrotechnical Society, 2021, 36(12): 2560-2575.
- [9] 苗强, 蒋京, 张恒, 等. 工业大数据背景下的航空智能发动机: 机遇与挑战 [J]. 仪器仪表学报, 2019, 40(7): 1-12.
- MIAO Qiang, JIANG Jing, ZHANG Heng, et al. Development of aviation intelligent engine under industrial big data: chances and challenges [J]. Chinese Journal of Scientific Instrument, 2019, 40(7): 1-12.
- [10] 彭开香, 皮彦婷, 焦瑞华, 等. 航空发动机的健康指标构建与剩余寿命预测 [J]. 控制理论与应用, 2020, 37(4): 713-720.
- PENG Kaixiang, PI Yanting, JIAO Ruihua, et al. Health indicator construction and remaining useful life prediction for aircraft engine [J]. Control Theory & Applications, 2020, 37(4): 713-720.
- [11] 刘惠, 刘振宇, 郑维强, 等. 深度学习在装备剩余使用寿命预测技术中的研究现状与挑战 [J]. 计算机集成制造系统, 2021, 27(1): 34-52.
- LIU Hui, LIU Zhenyu, JIA Weiqlang, et al. Current research and challenges of deep learning for equipment remaining useful life prediction [J]. Computer Integrated Manufacturing Systems, 2021, 27(1): 34-52.
- [12] 张钢, 田福庆, 梁伟阁, 等. 基于多尺度 AlexNet 网络的健康因子构建方法 [J]. 系统工程与电子技术, 2020, 42(1): 245-252.
- ZHANG Gang, TIAN Fuqing, LIANG Weige, et al. Construction method of bearing health indicator based on multi-scale AlexNet network [J]. Systems Engineering and Electronics, 2020, 42(1): 245-252.
- [13] ZHANG Xiaoyong, XIAO Pengcheng, YANG Yingze, et al. Remaining useful life estimation using CNN-XGB with extended time window [J]. IEEE Access, 2019, 7: 154386-154397.
- [14] ZHU Jun, CHEN Nan, PENG Weiwen. Estimation of bearing remaining useful life based on multiscale convolutional neural network [J]. IEEE Transactions on Industrial Electronics, 2019, 66(4): 3208-3216.
- [15] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [16] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Identity mappings in deep residual networks [C]// Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer International Publishing, 2016: 630-645.
- [17] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2021-06-01]. <https://arxiv.org/abs/1704.04861>.
- [18] YANG Zhibo, ZHANG Junpeng, ZHAO Zhibin, et al. Interpreting network knowledge with attention

- mechanism for bearing fault diagnosis [J]. *Applied Soft Computing*, 2020, 97: 106829.
- [19] 方军, 管业鹏. 基于双编码器的会话型推荐模型 [J]. *西安交通大学学报*, 2021, 55(8): 166-174.
FANG Jun, GUAN Yepeng. A session recommendation model based on dual encoders [J]. *Journal of Xi'an Jiaotong University*, 2021, 55(8): 166-174.
- [20] 曹仰杰, 李昊, 段鹏松, 等. AHNNet: 融合注意力机制的行为识别混合神经网络模型 [J]. *西安交通大学学报*, 2021, 55(5): 123-132.
CAO Yangjie, LI Hao, DUAN Pengsong, et al. AHNNet: human activity recognition based on hybrid neural network combining attention mechanism [J]. *Journal of Xi'an Jiaotong University*, 2021, 55(5): 123-132.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// *Proceedings of the European Conference on Computer Vision*. Berlin, Germany: Springer International Publishing, 2018: 3-19.
- [22] 王美华, 吴振鑫, 周祖光. 基于注意力改进 CBAM 的农作物病虫害细粒度识别研究 [J]. *农业机械学报*, 2021, 52(4): 239-247.
WANG Meihua, WU Zhenxin, ZHOU Zuguang. Fine-grained identification research of crop pests and diseases based on improved CBAM via attention [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(4): 239-247.
- [23] NECTOUX P, GOURIVEAU R, MEDJAH K, et al. PRONOSTIA: an experimental platform for bearings accelerated degradation tests [C]// *Proceedings of the IEEE International Conference on Prognostics and Health Management*. Piscataway, NJ, USA: IEEE, 2012: 1-8.
- [24] REN Lei, CUI Jin, SUN Yaqiang, et al. Multi-bearing remaining useful life collaborative prediction: a deep learning approach [J]. *Journal of Manufacturing Systems*, 2017, 43(2): 248-256.
- [25] GUO Liang, LEI Yaguo, LI Naipeng, et al. Deep convolution feature learning for health indicator construction of bearings [C]// *Proceedings of the 2017 Prognostics and System Health Management Conference*. Piscataway, NJ, USA: IEEE, 2017: 1-6.
- [26] 张继冬, 邹益胜, 邓佳林, 等. 基于全卷积层神经网络的轴承剩余寿命预测 [J]. *中国机械工程*, 2019, 30(18): 2231-2235.
ZHANG Jidong, ZOU Yisheng, DENG Jialin, et al. Bearing remaining life prediction based on full convolutional layer neural networks [J]. *China Mechanical Engineering*, 2019, 30(18): 2231-2235.

(编辑 武红江 刘杨)

(上接第 193 页)

- [22] 李睿, 张小栋, 张黎明, 等. 面向假肢的场景动画稳态视觉诱发脑控方法 [J]. *西安交通大学学报*, 2017, 51(1): 115-121.
LI Rui, ZHANG Xiaodong, ZHANG Liming, et al. Brain-controlled prosthesis manipulation based on scene graph-SSVEP [J]. *Journal of Xi'an Jiaotong University*, 2017, 51(1): 115-121.
- [23] 何柳诗, 谢俊, 于鸿伟, 等. 融合眼动追踪和目标动态可调的稳态视觉诱发电位脑机接口系统设计 [J]. *西安交通大学学报*, 2021, 55(10): 87-95.
HE Liushi, XIE Jun, YU Hongwei, et al. A design of brain-computer interface system for steady-state visual evoked potential integrating eye movement tracking and target dynamical adjustment [J]. *Journal of Xi'an Jiaotong University*, 2021, 55(10): 87-95.
- [24] MANSHOURI N, MALEKI M, KAYIKCIOGLU T. An EEG-based stereoscopic research of the PSD differences in pre and post 2D&3D movies watching [J]. *Biomedical Signal Processing and Control*, 2020, 55: 101642.
- [25] PALM R B. Prediction as a candidate for learning deep hierarchical models of data [D]. *Lyngby: Technical University of Denmark*, 2012: 46-51.

(编辑 武红江 刘杨)