

DOI:10.19651/j.cnki.emt.2107902

## 基于注意力 TCN 的滚动轴承剩余使用寿命预测方法\*

陈保家 陈正坤 陈学良 郭凯敏

(三峡大学 机械与动力学院 宜昌 443002)

**摘要:** 针对现有的基于数据驱动的滚动轴承剩余使用寿命(RUL)预测方法仍需要大量的先验知识来提取特征、构建健康指标和设定故障阈值的现状,提出了一种基于多头注意力机制的时间卷积网络(TCN)的 RUL 直接预测方法。该方法首先将原始振动信号的短时傅里叶变换(STFT)作为堆栈降噪自动编码器(SDAE)的输入,得到深度特征表示;然后将其输入到注意力 TCN 中进行 RUL 预测。最后,在 PRONOSTIA 的滚动轴承数据集进行验证。结果表明,该方法的预测误差指标 MAE 和 MAPE 分别比其他 4 种方法平均降低了 53.92% 和 46.13%;得分指标也比这些方法平均提高了 52.98%。

**关键词:** 轴承;剩余使用寿命预测;时间卷积网络;多头注意力

**中图分类号:** TH133.3;TP18 **文献标识码:** A **国家标准学科分类代码:** 460.1520

## Prediction method of remaining useful life of rolling bearing based on attentional temporal convolutional network

Chen Baojia Chen Zhengkun Chen Xueliang Guo Kaimin

(College of Mechanical &amp; Power Engineering, China Three Gorges University, Yichang 443002, China)

**Abstract:** Since the existing data-driven remaining useful life (RUL) prediction methods of rolling bearings still need a lot of prior knowledge to extract features, construct health indicators, and set fault thresholds, a direct RUL prediction method based on time convolution network (TCN) with multi-head attention mechanism is proposed. In this method, the short-time Fourier transform (STFT) of the original vibration signal is used as the input of the stack noise reduction automatic encoder (SDAE) to get the depth feature representation, and then input it to the attention TCN for RUL prediction. Finally, an example is verified in the rolling bearing data set of PRONOSTIA. The results show that the prediction error-index MAE and MAPE of this method are 53.92% and 46.13% lower than those of the other four methods, respectively, and the score index is 52.98% higher than that of these methods.

**Keywords:** bearing; remaining useful life prediction; temporal convolutional network; multi-head attention

## 0 引 言

机械设备在现代工业企业中得到了广泛的应用。一方面,由于系统复杂性的提高,另一方面,工业企业通常需要长期保持生产设备的良好状态,以保持竞争力。因此,开展机械设备的工程维修和剩余使用寿命(remaining useful life, RUL)预测,对于保证机械设备高效可靠的正常生产运作具有重要意义<sup>[1-2]</sup>。由于基于深度学习<sup>[3]</sup>的数据驱动 RUL 预测方法能够自动提取历史数据的深度特征表示,有效地克服了未知退化模型的问题,其输入可以是多种不同类型的数据,因此在机械 RUL 预测领域得到了广泛的研

究和应用<sup>[4-7]</sup>。Wang 等<sup>[8]</sup>利用残差块将原始振动信号输入到深度可分离卷积网络(deep separable convolution network, DSCN)中获得敏感特征,并将敏感特征直接映射到机器的剩余寿命。Ren 等<sup>[9]</sup>提出了一种新的基于小波的联合特征来表征滚动轴承的退化过程,并采用基于深度自动编码器的联合特征压缩方法来提取表征滚动轴承退化的深度特征。Guo 等<sup>[10]</sup>提取时域、频域和时频域的相对相似度特征,利用循环神经网络(recurrent neural network, RNN)构建轴承健康指数(health index, HI),并用 HI 预测轴承的 RUL。Zhao 等<sup>[11]</sup>利用卷积网络(convolutional neural network, CNN)从多传感器数据中提取局部稳健特

收稿日期:2021-09-17

\* 基金项目:国家自然科学基金(51975324)、机械传动国家重点实验室开放基金(SKLM-T-MSKFKT-202020)、水电机械设备设计与维护湖北省重点实验室(三峡大学)开放基金(2020KJX02)项目资助

征,并利用双向的长短时记忆网络(long short-term memory, LSTM)预测刀具磨损深度。然而,现有的一些基于深度学习的方法没有考虑到不同特征对滚动轴承 RUL 预测的贡献,难以提高滚动轴承 RUL 预测精度。

注意力机制可以突出信息的关键部分,实现更准确的预测。在故障预测与健康管理(prognostic and health management, PHM)领域已经取得了一些研究和应用<sup>[1,12]</sup>。因此,针对上述问题,提出了一种结合多头注意机制(multi-head attention, MA)与时间卷积网络(temporal convolutional network, TCN)的直接 RUL 预测方法。该方法包括深度特征提取和 RUL 估计两个步骤。首先对轴承原始振动信号进行短时傅里叶变换,接着将短时傅里叶变换取得的数据转换为对数形式,再将预处理后的短时傅里叶时频特征的时间序列输入堆栈降噪自动编码器(stacked denoising autoencoder, SDAE)进行特征降维和深度特征提取。然后,将 SDAE 提取的深度特征输入到所提出的多头注意 TCN 中进行 RUL 预测。该模型能够自动提取与真实 RUL 标签密切相关的特征,并对轴承的 RUL 值进行预测。最后,对 PRONOSTIA 的滚动轴承数据集进行了测试和分析,并将结果与其他方法进行了比较。实验结果表明,所提出的 RUL 预测方法能够比较准确地预测出轴承的 RUL,为旋转机械设备的健康管理提供可靠的指导。

## 1 方法理论

### 1.1 短时傅里叶变换

短时傅里叶变换(short-time Fourier transform, STFT)被广泛应用于非平稳信号的研究。它可以对信号进行时域和频域分析。其基本思想是将非平稳信号分成多个小的时间间隔,这些小的时间间隔被认为是静态的,然后对不同时间间隔的信号进行快速傅里叶变换,得到不同时间段的频谱。连续 STFT 的输出可以提供原始信号<sup>[13]</sup>的时频表示。假设原始非平稳信号为  $m(t)$ ,则时间  $t$  的频带  $k$  的离散 STFT 定义如式(1)所示。

$$M_l(e^{j\omega_k}) = \sum_{\tau=-\infty}^{+\infty} m(\tau)w(t-\tau)e^{-j\omega_k\tau} \quad (1)$$

式中:  $\omega_k = 2\pi k/N$  是以弧度为单位的频率;  $N$  是频带数;  $\tau$  是所选窗口大小为  $l$  的对称窗口;如果需要重建信号  $l \leq N$ 。

将上述的  $\omega_k = 2\pi k/N$  代入式(1)中可得:

$$M(t, k) = M_l(e^{j\omega_k}) = \sum_{\tau=-\infty}^{+\infty} m(\tau)w(t-\tau)e^{-j2\pi k\tau/N} \quad (2)$$

式中:  $|M(t, k)|$  是对短时傅里叶变换谱的幅度的估计。

则其功率谱密度(PSD)的函数的定义如式(3)所示。

$$P(t, k) = |M(t, k)|^2 = (m(t, k)m(\text{conj}(m(t, k)))) \quad (3)$$

式中:  $\text{conj}(m(t, k))$  表示  $m(t, k)$  的共轭。

为了方便直观地观察频带能量随时间的变化,将 PSD 进一步转化为对数形式:

$$L_{P(t,k)} = 10 \lg \frac{\langle P(t,k) + \epsilon \rangle}{10^{-6}} \quad (4)$$

式中:  $\epsilon$  为正实数取值为  $10^{-6}$ ,以防止非负实值函数  $P(t, k)$  取 0 时而造成对数变换公式中的分子为 0。

### 1.2 时间卷积网络

时间卷积网络(TCN)是在残差网络<sup>[14]</sup>的基础上,由 Bai 等<sup>[15]</sup>于 2018 年为处理时间序列问题而设计的一种卷积网络。其基本残差块结构如图 1 所示,对于一维序列输入  $\mathbf{g} = \{s_1, s_2, \dots, s_{n-1}\} \in \mathbf{R}^n$  和卷积核  $f: \{0, \dots, z-1\} \rightarrow \mathbf{R}$ , 序列元素  $s$  的扩展卷积运算  $F$  定义如式(5)所示。

$$F(s) = (\mathbf{g} *_{d} f)(s) = \sum_{i=0}^{z-1} f(i) \cdot \mathbf{g}_{s-d+i} \quad (5)$$

TCN 的基本结构如图 2 所示,其中 TCN 的每一层都是一个残差块,每一层残差块中卷积神经元的扩张率从浅到深以  $2^{N-1}$  的速度递增,这使得它具有更好的记忆历史信息的能力。与 RNN 相比,TCN 最大的优势在于权重共享和卷积层的局部感知。权值参数的共享可以有效地减少网络中需要训练的参数数量,局部感知特征能够有效地反映当前卷积核覆盖的输入数据的局部结构信息。因此,TCN 不仅可以学习输入时间序列的长期历史时间相关性,而且可以像 CNN 一样进行并行计算。尽管 RNN 理论上可以捕捉无限长的历史,但 TCN 已被证明更适合需要长期历史依赖的领域<sup>[15]</sup>。

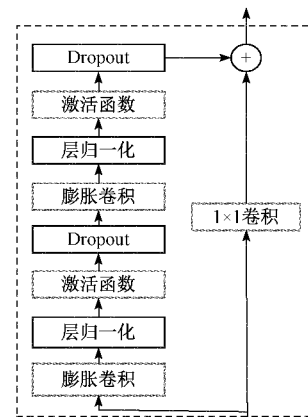


图 1 TCN 中残差块结构

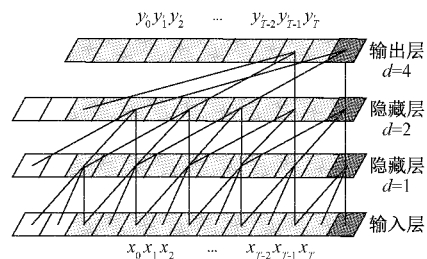


图 2 TCN 的结构

### 1.3 降噪自动编码器

降噪自动编码器(DAE)是自动编码器的一种改进方法。它接受被破坏的输入数据,并通过训练重建原始的未被破坏的数据作为输出,以增强网络的抗噪声能力<sup>[9,13]</sup>。而堆栈降噪自动编码器(SDAE)是由多个 DAE 堆栈而成。DAE 的基本结构如图 3 所示。假设原始输入数据被定义为  $\mathbf{X} = \{x_0, x_1, \dots, x_{n-1}\} \in \mathbf{R}^n$ , 其中  $n$  是输入维度。被破坏的编码器输入为  $\tilde{\mathbf{X}} = \{\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_{n-1}\} \in \mathbf{R}^n$ ,  $p$  为 DAE 的隐层神经元个数,则输入层到隐层的编码器映射关系如式(6)所示。

$$\mathbf{h} = A_g(\mathbf{W}\tilde{\mathbf{X}} + \mathbf{b}) \quad (6)$$

式中:  $\mathbf{W}$  是权重,  $\mathbf{b}$  是偏置。  $A_g$  是一个非线性激活函数。在解码器中,隐藏表示  $\mathbf{h}$  通过映射函数映射到输出层  $\hat{\mathbf{X}} \in \mathbf{R}^n$  如式(7)所示。

$$\hat{\mathbf{X}} = A_g(\mathbf{W}'\mathbf{h} + \mathbf{b}') \quad (7)$$

式中:  $\mathbf{W}'$  是权重,  $\mathbf{b}'$  偏置。在模型训练中,采用均方重构误差来最小化预训练过程中的重构损失函数,其计算公式如式(8)所示。

$$L(\theta)_{DAE} = \frac{1}{2n} \sum_{i=1}^n \|\mathbf{X} - \hat{\mathbf{X}}\|^2 \quad (8)$$

式中:  $\mathbf{X}$  为原始输入,  $\hat{\mathbf{X}}$  为重构数据,  $n$  表示信息的维度。

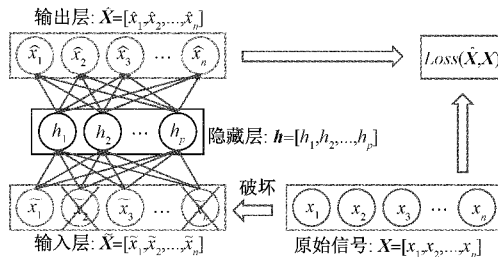


图3 降噪自编码器的结构

### 1.4 多头注意力

与TCN一样,自注意机制能够捕捉退化特征的长期趋势,实现并行计算。相比于TCN,自注意机制可以学习不同退化特征对RUL预测的贡献。一般来说,自注意机制有查询矩阵  $\mathbf{Q}$ 、关键字矩阵  $\mathbf{K}$  和关键字的数值矩阵  $\mathbf{V}$ 。自我注意的定义如式(9)所示。

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (9)$$

式中:  $\mathbf{Q}$ 、 $\mathbf{K}$  和  $\mathbf{V}$  分别是线性映射格式,定义为  $\mathbf{Q} = \mathbf{W}^q\mathbf{e}_i$ ,  $\mathbf{K} = \mathbf{W}^k\mathbf{e}_i$  和  $\mathbf{V} = \mathbf{W}^v\mathbf{e}_i$ ;  $d_k$  表示要投影的尺寸。

多头注意<sup>[16]</sup>通过式(9)的多重自注意力操作将输入特征投射到不同的子空间,获取多个子空间的注意向量,然后将这些注意向量集合在一起。这可以从多个角度探讨不同嵌入方式之间的相关性,提高自注意模型的性能。多头注意的定义如式(10)~(11)所示。

$$Multihead(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(head_1, \dots, head_h) \quad (10)$$

$$head_i = Attention(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (11)$$

本文多头注意的输出与单层前馈神经网络(FFNN)进行级联。注意力载体  $\mathbf{a}_i$  被输入到单层具有残差连接额前馈神经网络中得到输出  $\mathbf{h}^{attention}$ , 其定义如式(12)所示。

$$\mathbf{h}^{attention} = sigmoid(\mathbf{W}_{ffnn}\mathbf{a}_i + \mathbf{b}_{ffnn} + \mathbf{a}_i) \quad (12)$$

## 2 基于多头注意力和TCN的RUL预测方法

### 2.1 深度特征提取

首先利用STFT将原始振动信号从时域转换到时频域,并对结果进行对数变换得到原始特征集。然后将原始特征集输入到由4个DAE堆栈构成的SDAE中进行深度特征提取。

深度特征提取的基本过程如图4所示,DAE由全连接层搭建而成,每一层的激活函数均为Sigmoid,因此可以将特征值的范围限制在(0,1)之间。训练时采用Adam优化器,学习率为0.001,并且从训练集中每次随机划分出25%的样本,使其不参与训练而作为验证集。第1个DAE的输入层、隐藏层和输出层的维度分别为1281-256-1281;第2个DAE的维度分别为256-128-256;第3个DAE的维度分别为128-256-128;最后一个DAE的维度为256-1281-256。通过对SDAE网络的无监督训练,可以将含有噪声或者被破坏的输入恢复到原始输入,获得表征原始输入的强鲁棒性深度特征。具体训练过程如下:首先依次进行单个DAE的无监督预训练,将前一个DAE隐藏层的输出作为后续DAE的输入。预训练阶段完成后,每个DAE都有其重量和偏置参数,然后将每一个DAE的中间隐藏层抽出构造出一个新的网络,再次将原始数据输入到新构建的网络中进行参数微调。最终我们选择SDAE的神经元数目为128的中间隐含层的特征向量作为本文的深层特征集。

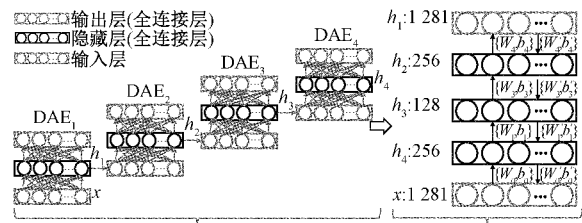


图4 本文SDAE提取深度特征的流程

### 2.2 剩余使用寿命预测

本文采用直接剩余使用寿命(RUL)预测方法,即通过将深度特征直接输入到本文的模型中得到RUL。虽然主流的RUL预测方法都是基于健康指数(HI),但深度神经网络预测的HI大多与传统的具有物理意义的健康指数(PHI)并不相同。因为它是一个虚拟健康指数(VHI),没有实际意义,属于RUL预测的中间变量。传统的均方根指数、峰度指数、熵指数具有较好的单调性和趋势性,能较好地反映机械设备的性能退化情况。因此,机械设备的RUL

可以通过预测其超过设定的阈值来获得。

在日前的深度学习中,由于在提取的特征中没有噪声干扰时,真实的 RUL 标签是线性的,所以从神经网络中提取的 VHI 在理论上是一条直线。它代表了 RUL 的预测过程,即 VHI 是预测的寿命值。在构造 VHI 时,训练集的 VHI 是一条接近真实线性标签值的直线。然而,由于干扰信息的存在,测试集的 VHI 总是非线性的。然后将故障阈值设置为 1,并使用相关向量机(RVM)等模型对测试集的 VHI 进行预测,得到测试集的 RUL。虽然该方法可以得到 RUL 的不确定性表达式,但预测效果与初始预测点的选取、RVM 等预测模型参数的选择密切相关。并且测试集的 VHI 不一定超过故障阈值。该方法引入了人为主观因素和大量先验知识,这与基于深度学习的数据驱动方法的初衷背道而驰。因此,本文通过深层神经网络将深层特征直接映射到轴承 RUL 上,方法简单,且避免了这一问题。所提出的 RUL 预测流程如图 5 所示。

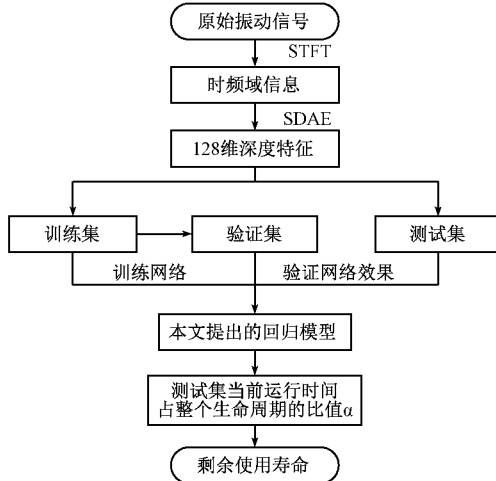


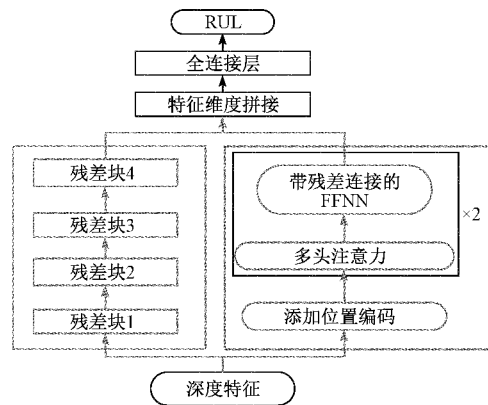
图 5 提出的 RUL 流程

为了探究多头注意力与 TCN 的混合结构模型的有效。本文分别构建了:1)多头注意力模型(multi-head attention, MA);2)并行多头注意 TCN(parallel multi-head attention temporal convolutional network, PMA-TCN)混合结构模型;3)串联多头注意 TCN(tandem multi-head attention temporal convolutional network, TMA-TCN)混合结构模型。其中,PMA-TCN 和 TMA-TCN 的结构分别如图 6(a)和(b)所示。

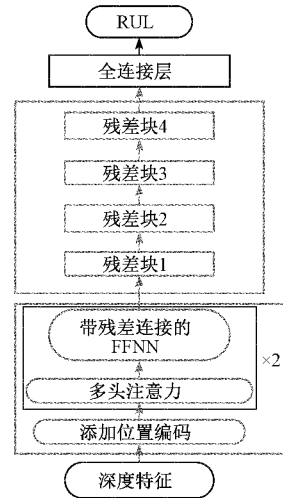
### 2.3 预测模型参数

本文 3 种网络模型的残差块、多头注意力块和全连接层的参数相同,具体参数如下。

1)TCN 残差块的激活函数为泄露修正线性单元(Leaky rectified linear units, LeakyReLU),随机丢失率(Dropout)取 0.25,并采用层规范化方式对卷积操作后的数据进行规范化,使得 LeakyReLU 层的输入数据的均值为 0,方差为 1。残差块 1~4 的神经元数目依次为 128、64、



(a) PMA-TCN 模型的基本结构



(b) TMA-TCN 模型的基本结构

图 6 不同多头注意力和 TCN 的混合结构模型

32、32,核大小均为 3,采用因果卷积的方式,膨胀率依次为 1、2、4 和 8,并采用 He 初始化的方式<sup>[14]</sup>。

2)多头注意力块由多头注意力和具有残差连接的前馈层(FFNN)组成,单元类型均为全连接。多头注意力部分的神经元数目为 128;FFNN 神经元的数目也为 128,激活函数为高斯误差线性单元(Gaussian error linear unit, GELU)。

3)全连接层神经元数目为 32,并直接与神经元数目为 1 的输出层相连。输出层的激活函数为 Sigmoid,以保证预测出的不同运行周期时间点对应的当前寿命在全寿命周期的占比处于 0~1 的范围内。

3 种模型均采用 Adam 优化器,学习率为 0.001,训练次数为 100 次。每一次训练过程中,从训练集中随机划分出 25% 的样本数据作为验证集,以增加网络的泛化性能,最后在 5 个测试集轴承上进行测试。

### 3 实验验证与分析

本文的实验部分是在基于 CUDA10.1 和 Tensorflow 2.1 GPU 版框架下,采用 Python 3.7 完成网络模型的构



建;并在配置信息为 Intel(R) Core(TM) i5-9300H CPU@ 2.4 GHz,NVIDIA GeForce GTX 1660Ti GPU,16 G RAM 的计算机上完成模型训练与计算。

3.1 数据描述

本部分采用 PHM 2012 数据挑战赛提供的 PRONOTIA 滚动轴承全寿命数据集<sup>[17]</sup>,来评估提出方法的性能。在后面的部分还对最近的新方法进行了进一步的比较。该数据集通过加速轴承退化试验台收集而来,如图 7 所示。水平和垂直振动信号均由安装在轴承外圈上的加速度计监测。试验台启动时,每次记录被测振动信号持续 0.1 s,采样频率为 25.6 kHz,每 10 s 记录一次。关于该平台和实验的更多细节可以在文献<sup>[17]</sup>中找到。本文仅使用工况 1 下的数据集如表 1 所示。

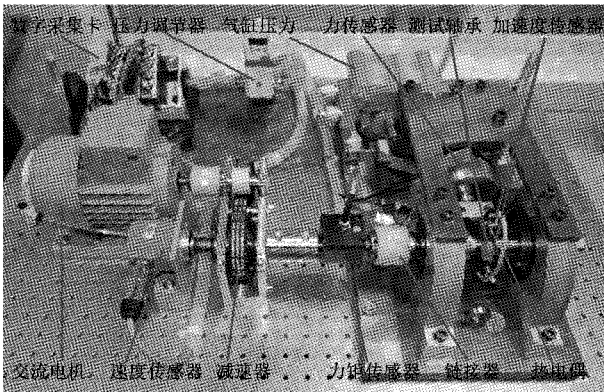


图 7 PRONOTIA 试验台

表 1 本文采用的实验数据及工况

工况	转速/ (r·min <sup>-1</sup> )	载荷/ N	训练集	测试集
1	1 800	4 000	Bearing1-1 Bearing1-2	Bearing1-3 Bearing1-4 Bearing1-5 Bearing1-6 Bearing1-7

3.2 评价指标

为了对本文的 5 个测试集轴承的截断剩余使用寿命进行定量的分析,采用了滚动轴承 RUL 预测中常用的 3 个评价指标,分别为平均绝对误差(mean absolute error, MAE)  $S_{MAE}$ 、平均绝对百分比误差(mean absolute percentage error, MAPE)  $S_{MAPE}$  和评分函数(scoring function, SF)的均值  $S_{SF}$ 。

$$S_{MAE} = \frac{1}{5} \sum_{i=1}^5 | P_{ActRUL}^i - P_{PreRUL}^i | \quad (13)$$

$$E_r^i = \frac{P_{ActRUL}^i - P_{PreRUL}^i}{P_{ActRUL}^i} \times 100 \quad (14)$$

$$C^i = \begin{cases} e^{\ln(0.5) \cdot (E_r^i/5)}, & E_r^i \leq 0 \\ e^{+\ln(0.5) \cdot (E_r^i/20)}, & E_r^i > 0 \end{cases} \quad (15)$$

$$S_{MAPE} = \frac{1}{5} \sum_{i=1}^5 | E_r^i | \quad (16)$$

$$S_{SF} = \frac{1}{5} \sum_{i=1}^5 C^i \quad (17)$$

式中:  $P_{ActRUL}^i$  和  $P_{PreRUL}^i$  分别表示第  $i$  个测试集轴承的 RUL 真实值与 RUL 预测值,而  $P_{PreRUL}^i = T^i/\alpha^i - T^i$ ,  $\alpha^i$  是该轴承当前运行周期占整个全寿命时间  $T^i$  的比值的预测值;  $E_r^i$  表示第  $i$  个轴承的百分比预测误差;  $C^i$  表示第  $i$  个轴承的得分函数值。

3.3 RUL 预测结果与分析

本文的模型结构探究的消融实验设置如表 2 所示,分为 MA、PMA-TCN 和 TMA-TCN 三种模型。在每一种模型中,根据多头注意力的特性,分别设置了头数为 1、2、4、8、16、32、64 和 128 这 8 种实验条件,每一种条件运行 10 次,得到其预测结果的均值和标准差。

表 2 本文的消融实验

方案	多头注意力的头数							
MA	1	2	4	8	16	32	64	128
PMA-TCN	1	2	4	8	16	32	64	128
TMA-TCN	1	2	4	8	16	32	64	128

通过消融实验,探究不同模型对 RUL 预测结果的影响,并以测试集中 5 个轴承的  $S_{MAE}$ 、 $S_{MAPE}$  和  $S_{SF}$  作为评价指标。其中前两个是主要评价指标,定量表示测试集的 RUL 预测误差,而  $S_{SF}$  起到补充评价的作用。当  $E_r^i$  越接近 0 时,则  $S_{SF}$  越接近 1;而且  $S_{SF}$  会随着  $|E_r^i|$  的增大而呈指数级下降;但是当  $E_r^i < 0$  时,  $S_{SF}$  值会比  $E_r^i > 0$  时的值下降得更快。

从图 8(a)~(c) 可以看出:1) MA、PMA-TCN 和 TMA-TCN 的最佳注意力头数分别为 2、16 和 64 个;2) 并且 TMA-TCN 在注意力头数为 64 时,其预测误差最小、得分函数值最大,预测性能表现最突出;3) 从  $S_{MAE}$  与  $S_{MAPE}$  指标上看,这两个指标的值与注意力的头数并没有线性关系,但是这两个指标具有正相关性;4) 从  $S_{SF}$  指标上看,该指标与  $S_{MAE}$  和  $S_{MAPE}$  没有严格的线性关系,但是可以看出 3 种方案的  $S_{SF}$  大致在注意力的头数为 4~64 之间,与  $S_{MAE}$  和  $S_{MAPE}$  具有一定的相关性;5) 从 8 种注意力头数的 3 种指标的均值与标准差上分析可知,PMA-TCN 方案的标准差最大,MA 方案次之,TMA-TCN 方案的标准差最小。这一现象说明 PMA-TCN 的稳定最佳性差,TMA-TCN 的稳定性最好。

由于 3 种评价指标的均值和标准差只能表示测试集整体的 RUL 预测状况,不能具体地表现出每一个轴承的 RUL 预测情况。因此,为了更加直观地进行每一个轴承的 RUL 预测结果的比较,本文选择 3 种模型在最佳注意力头数时的最好预测效果下进行对比。从表 3 可以得出:1) 从

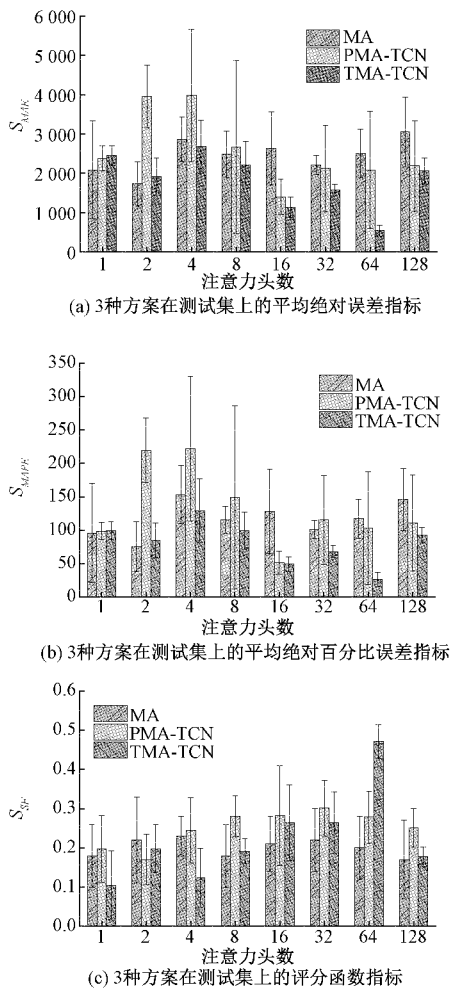


图 8 3 种方案不同头数的 RUL 预测结果比较

单个轴承的预测误差上分析，TMA-TCN 方法在 5 个轴承上的  $|E_r|$  均比 TCN 方法的小，并且也只在轴承 1-4 上的  $|E_r|$  比 MA 方法的大；2) 从测试集轴承整体的预测情况上分析，相较于 TCN、MA 方法，本文的 TMA-TCN 方法在  $S_{MAE}$  和  $S_{MAPE}$  指标上分别下降了 63.25%、59.74%；而在  $S_{SF}$  指标上分别提升了 33.33% 和 45.45%；3) 相比于 PMA-TCN 方法，只有轴承 1-5 的  $|E_r|$  比较差，但是测试集整体的预测误差指标  $S_{MAE}$  和  $S_{MAPE}$  分别下降了 49.71% 和 55.76%；整体的评分函数均值  $S_{SF}$  相等。

图 9(a)~(d) 分别为表 2 中 4 种方法在测试集的轴承 1-3 上的预测结果可视化对比。从图中可知 TMA-TCN 方法兼具了 TCN 与 MA 的优势，其预测拟合效果最佳、波动最小，在截断周期处预测的全寿命周期占比  $\alpha$  与实际的最为接近，说明本文方法的有效性。

表 4 将提出的 TMA-TCN(头数为 64)方法运行 10 次的结果与其他 4 种优秀方法的预测结果进行了比较。结果表明，提出的 TMA-TCN 方法的两个预测误差指标最低，得分指标最高。本文方法在 5 个测试集轴承中的  $S_{MAE}$  和  $S_{MAPE}$  分别比 MSCNN 低 44.56% 和 38.90%；比 RNN-HI 方法低 52.04% 和 21.42%；比 SOM-HI 方法低 61.40% 和 66.34%；与 SVR-HI 法相比，分别下降了 57.67% 和 57.85%。本文方法的得分指标  $S_{SF}$  比上述方法分别提高了 29.88%、68.23%、36.64% 和 77.15%。这些实验结果有力地证明了该方法的优越性能。因此，TMA-TCN 可以为机械设备的健康管理和维护提供可靠的帮助。

表 3 PHM 2012 数据集工况 1 测试集轴承 RUL 预测结果与对比

轴承型号	当前运行时间/s	实际的 RUL/s	预测的 RUL/s	本文方案 $E_r$ /%		其他方案 $E_r$ /%	
				TMA-TCN (头数 64)	TCN	MA (头数 2)	PMA-TCN (头数 16)
轴承 1-3	18 010	5 730	5 510	6.30	17.40	38.37	11.94
轴承 1-4	11 380	2900	2 310	20.32	10.69	-6.14	54.22
轴承 1-5	23 010	1 610	1 870	-16.48	-148.20	22.08	-0.83
轴承 1-6	23 010	1 460	1 610	-10.25	-19.45	-47.41	-61.92
轴承 1-7	15 010	7 570	6 930	8.47	19.55	20.60	-10.78
$S_{MAE}$				<b>401.28</b>	1 091.80	996.73	797.94
$S_{MAPE}$				<b>12.36</b>	43.06	26.92	27.94
$S_{SF}$				<b>0.48</b>	0.36	0.33	0.48

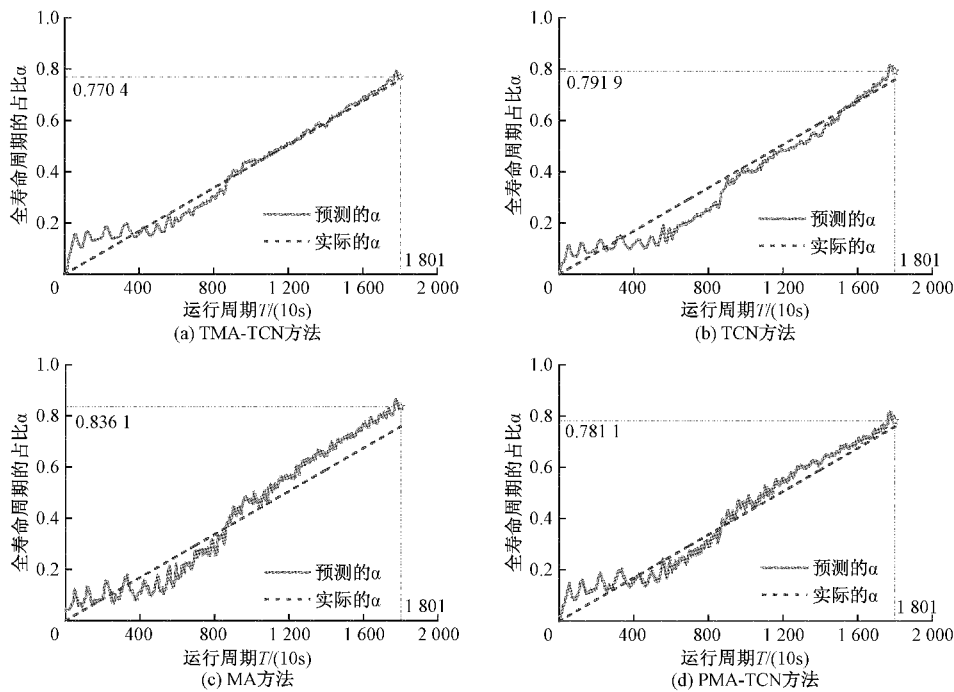


图9 4种方法在轴承1-3上的预测性结果对比

表4 本文方法与其他方法的比较

评价指标	TMA-TCN	MSCNN <sup>[5]</sup>	RNN-HI <sup>[10]</sup>	SOM-HI <sup>[18]</sup>	SVR-HI <sup>[19]</sup>
$S_{MAE}$	605.3±220.5	1 091.8	1 262	1 568	1 430
$S_{MAPE}$	26.31±10.58	43.06	33.48	78.17	62.42
$S_{SF}$	0.470 7±0.043 2	0.362 4	0.279 8	0.360 5	0.265 7

## 4 结 论

为了减少人为因素的干扰,提出了一种基于多头注意力与时间卷积网络相结合的滚动轴承 RUL 直接预测方法。通过网络模型中的多头注意力层从不同角度自动学习不同的输入特征对模型输出的贡献度,从而突出与 RUL 相关的特征并抑制干扰特征。同时,又充分利用 TCN 层的扩张卷积核的局部特征感知与对历史信息记忆功能,更好地捕获输入特征随时间变化的退化趋势。探讨了多头注意机制的最佳头数以及与 TCN 的不同组合模型。最后,通过实验对不同模型进行了探索并找到了最佳的 TMA-TCN 模型。与文献[5]、文献[10]和文献[19]相比,本文方法预测误差指标 MAE 和 MAPE 分别平均降低了 53.92% 和 46.13%;得分函数的均值也比这些方法平均提高了 52.98%。证明了提出的方法能有效地提高了滚动轴承的 RUL 预测精度。当然,本文的研究仍然有一些不足之处。对网络超参数并未进行优化;对于将注意力层放置于模型顶层的结构还有待进一步探究。未来将继续研究机械设备 RUL 的不确定性表达,以提高 RUL 预测的可信度。

## 参考文献

- [1] ZHANG H, ZHANG Q, SHAO S, et al. Attention-based LSTM network for rotatory machine remaining useful life prediction [J]. IEEE Access, 2020, 8: 132188-132199.
- [2] 陈志强,陈旭东,DE OLIVIRA J V,等.深度学习在设备故障预测与健康中的应用[J].仪器仪表学报,2019,40(9):206-226.
- [3] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [4] 杨小东,纪国宜.基于融合 CNN 的滚动轴承剩余寿命预测[J].国外电子测量技术,2021,40(3):62-67.
- [5] ZHU J, CHEN N, PENG W. Estimation of bearing remaining useful life based on multiscale convolutional neural network[J]. IEEE Transactions on Industrial Electronics, 2018, 66(4):3208-3216.
- [6] 车昱娇,陈云霞,崔宇轩. KPCA 和改进 LSTM 在滚动轴承剩余寿命预测中的应用研究[J].电子测量与仪器学报,2021,35(2):109-114.
- [7] LEI Y, LI N, GUO L, et al. Machinery health

- prognostics: A systematic review from data acquisition to RUL prediction[J]. *Mechanical Systems and Signal Processing*, 2018,104:799-834.
- [8] WANG B, LEI Y, LI N, et al. Deep separable convolutional network for remaining useful life prediction of machinery[J]. *Mechanical Systems and Signal Processing*, 2019,134:106330.
- [9] REN L, SUN Y, CUI J, et al. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks[J]. *Journal of Manufacturing Systems*, 2018,48:71-77.
- [10] GUO L, LI N, JIA F, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings[J]. *Neurocomputing*, 2017, 240:98-109.
- [11] ZHAO R, YAN R, WANG J, et al. Learning to monitor machine health with convolutional bi-directional LSTM networks[J]. *Sensors*, 2017, 17(2):273.
- [12] YU W, PI D, XIE L, et al. Multiscale attentional residual neural network framework for remaining useful life prediction of bearings[J]. *Measurement*, 2021,177:109310.
- [13] ZHOU S, XIAO M, BARTOS P, et al. Remaining useful life prediction and fault diagnosis of rolling bearings based on short-time fourier transform and convolutional neural network[J]. *Shock and Vibration*, 2020,DOI:10.1155/2020/8857307.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [15] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. *ArXiv Preprint*, 2018,ArXiv:1803.01271.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. *Advances in Neural Information Processing Systems*, 2017: 5998-6008.
- [17] NECTOUX P, GOURIVEAU R, MEDJAHER K, et al. PRONOSTIA: An experimental platform for bearings accelerated degradation tests[C]. *IEEE International Conference on Prognostics and Health Management, PHM'12*. IEEE Catalog Number: CFP12PHM-CDR, 2012: 1-8.
- [18] HONG S, ZHOU Z, ZIO E, et al. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method[J]. *Digital Signal Processing*, 2014,27:159-166.
- [19] LOUTAS T H, ROULIAS D, GEORGOULAS G. Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic c-support vectors regression[J]. *IEEE Transactions on Reliability*, 2013,62(4):821-832.

#### 作者简介

陈保家(通信作者),工学博士,教授,主要研究方向为机械信号处理与故障诊断、机械可靠性评估与寿命预测、机械装备健康监测与维护决策等。

E-mail:cbjia@163.com

陈正坤,硕士研究生,主要研究方向为基于深度学习的机械设备健康状态评估与寿命预测。

E-mail:chenzhkun@foxmail.com

陈学良,硕士研究生,主要研究方向为机械信号处理与故障诊断。

E-mail:cxueliang1995@163.com

郭凯敏,硕士研究生,主要研究方向为基于深度学习的机械设备可靠性评估与寿命预测。

E-mail:1641703375@qq.com