

# 基于深度学习的导航装备轴承剩余使用寿命预测

党慧莹<sup>1,2</sup>, 李海林<sup>1</sup>, 吴北苹<sup>1,3</sup>, 余佳宇<sup>1,4</sup>, 庄银传<sup>5</sup>

(1. 空军工程大学信息与导航学院, 西安, 710077; 2. 31401 部队, 吉林通化, 134000;  
3. 空军通信士官学校, 辽宁大连, 116000; 4. 95486 部队, 成都, 610041; 5. 93127 部队, 北京, 100834)

**摘要** 作为导航装备的重要部件, 轴承影响着导航装备的定位精度和保障能力。在装备剩余使用寿命(RUL)预测中, 传统的机器学习算法在处理复杂非线性传感信号问题上存在局限性, 为此提出了一种基于注意力机制(AM)和双向长短期记忆网络(Bi-LSTM)的轴承 RUL 预测框架(Bi-LSTM-A)。该框架在前端加入一维卷积神经网络(CNN)从原始传感器信号中提取局部特征, 然后利用双向长短期记忆网络与注意力机制相结合的方式对信号进行分析预测, 最后经过网络末端的全连接层输出预测结果。通过与同类算法的对比实验表明, 该方法能够准确地预测装备剩余使用寿命, 具有较好的预测效率和预测精度。

**关键词** 轴承; 深度学习; 长短期记忆网络; 注意力机制; 剩余使用寿命

**DOI** 10.3969/j.issn.2097-1915.2025.02.010

**中图分类号** TH17 **文献标志码** A **文章编号** 2097-1915(2025)02-0081-08

## A Navigational Equipment Bearing Remaining Useful Life Prediction Based on Deep Learning

DANG Huiying<sup>1,2</sup>, LI Hailin<sup>1</sup>, WU Beiping<sup>1,3</sup>, YU Jiayu<sup>1,4</sup>, ZHUANG Yinchuan<sup>5</sup>

(1. Information and Navigation School, Air Force Engineering University, Xi'an 710077, China;  
2. Unit 31401, Tonghua 134000, Jilin, China; 3. Air Force Communication Noncommissioned Officer School, Dalian 116000, Liaoning, China; 4. Unit 95486, Chengdu 610041, China;  
5. Unit 93127, Beijing 100834, China)

**Abstract** As a crucial component of navigation equipment, bearings affect the positioning accuracy and safeguarding capability of the navigation equipment. In predicting the remaining useful life (RUL) of equipment, traditional machine learning algorithms are limited to dealing with the problems of complex nonlinear characteristic signals. For the above-mentioned reasons, a new prediction framework for RUL of bearing based on attention mechanism(AM) and bidirectional long short-term memory (Bi-LSTM) is proposed (Bi-LSTM-A). First, a one-dimensional convolution neural network (CNN) is added to the front of the structure to extract local features from the original signal sequence, and then, the signals are analyzed and predicted by combining bidirectional long short-term memory network with attention mechanism. Finally, the predicted results are output through the fully connected layers at the end of the network. In comparison with the similar algorithms, the results show that the proposed method can accurately predict the equipment remaining useful life, and is good in predicting efficiency and accuracy.

**收稿日期:** 2024-04-01

**作者简介:** 党慧莹(1993—), 女, 陕西渭南人, 硕士生, 研究方向为装备可靠性。E-mail: danghuiyingying@163.com

**引用格式:** 党慧莹, 李海林, 吴北苹, 等. 基于深度学习的导航装备轴承剩余使用寿命预测[J]. 空军工程大学学报, 2025, 26(2): 81-88.  
DANG Huiying, LI Hailin, WU Beiping, et al. A Navigational Equipment Bearing Remaining Useful Life Prediction Based on Deep Learning [J]. Journal of Air Force Engineering University, 2025, 26(2): 81-88.

**Key words** bearing; deep learning; long short-term memory; attention mechanism; remaining useful life

由于航空技术的不断发展,机动导航系统的稳定性和精确性要求不断加强,设备的定期维护以及预测其发生故障的时间对于保证系统的正常运行至关重要。导航平台轴承是导航装备的核心使用部件,其性能、质量对导航平台系统框架的稳定性、可靠性有着重要作用。为了满足航空装备发展的新要求,故障预测与健康管理(prognostics and health management, PHM)技术应运而生,并迅速得到高度重视。PHM 实质是通过数据处理、特征提取,运用智能的推理算法评估系统的运行状态,预测系统发生故障的部位、时间及使用寿命,并给出合理的维修保障建议<sup>[1]</sup>。对轴承的剩余使用寿命预测(remaining useful life, RUL)可以估计其未来的健康状态,及时提醒可能发生的故障,降低事故率,提高生产效率,防止突发性事故的发生,对装备生产具有重要意义<sup>[2]</sup>。

近年来,随着人工智能、信号处理分析等技术的飞速发展,基于数据驱动的方法逐渐成为 RUL 预测的主流手段<sup>[3-4]</sup>,尤其是神经网络在 RUL 的预测上体现出其强大的能力,使得基于数据驱动的方法具有巨大潜力。Zhao 等<sup>[5]</sup>提出了一种基于改进 Transformer 的多头双向系数自注意力网络来预测 RUL,该结构包含了一个多头 ProbSparse 和 LogSparse 自注意力网络,前者旨在滤除主要的点积运算,后者采用了一种基于对数的综合稀疏策略从而减低计算量。Zhang 等<sup>[6]</sup>针对 RUL 预测过程数据存在的非同分布问题,提出了一种局部加权的深度自适应网络,将子域之间的细粒度特征进行对齐。Chen 等<sup>[7]</sup>提出了一种自适应动态聚类方法,将有用的多模态数据选择到不同的聚类中,再构建多个回归网络对航空发动机的 RUL 值进行预测。

传统的人工神经网络由于其局限性,不能很好地处理长时间序列数据,循环神经网络(recurrent neural network, RNN)作为最原始的递归神经网络,通常用于建立时间序列数据的关系,但是 RNN 网络自身存在缺陷,在训练过程中由于序列的长度问题,很容易造成梯度消失或梯度爆炸<sup>[8]</sup>。RNN 的变体长短期记忆网络(long short-term memory network, LSTM)提出于 1997 年,它有效解决了 RNN 存在的问题,并在长时间序列预测领域取得巨大的成就<sup>[9]</sup>。随着科技的进步,信号数据变得越来越复杂,单一的 LSTM 网络很难处理好复杂的时间信号序列,基于改进 LSTM 的模型不断被提出。Dong 等<sup>[10]</sup>提出一种新的基于迁移学习的 RUL 预

测方法,先利用卷积自编码器(convolutional autoencoder, CAE)判断失效行为,再通过嵌入域自适应结构的多通道传输网络提取轴承退化的多尺度特征,使得模型能够学习域不变性。Tian 等<sup>[11]</sup>提出一种基于空间相关性和时间注意力机制的长短时记忆框架,有效提取了时间维度的信息,衡量了不同变量在不同时间的重要性。

不同于先前方法对 LSTM 内部结构的改变,本文将注意力机制和双向长短期记忆网络(bidirectional LSTM, Bi-LSTM)结合用于预测轴承剩余使用寿命。近些年,注意力机制(attention mechanism, AM)成为深度学习领域最受欢迎的概念之一,被运用到对应的模型处理相应的任务<sup>[12]</sup>。它可以被视为对数据信息的进一步筛选,使网络注意到信息的关键部分,从而实现精准预测分析。基于注意力机制深度神经网络已成功地应用于目标识别<sup>[13]</sup>、机器翻译<sup>[14]</sup>、图像字幕生成<sup>[15]</sup>、视觉对象分类<sup>[16]</sup>等多种任务中。实验表明,将 AM 和 Bi-LSTM 相结合的模型能准确预测轴承的剩余使用寿命。本文的贡献如下:

1)在网络结构方面,与传统 LSTM 单向传递信息的方式不同,本文的 LSTM 采用了双向连接的方式,双向连接使网络能够关注到信息的相互关系,有助于从复杂信号中提取有用的特征信息,使模型处理非线性能力提高。

2)将注意力机制和 Bi-LSTM 结合,AM 通过 Bi-LSTM 保留输入序列的中间输出,并有选择地学习其特征信息,将输入和输出序列关联起来,使模型能够精确捕捉到输入信号关键信息部分。实验证明,本文提出的模型比传统的预测方法具有更高的预测精确度。

## 1 深度学习相关原理

### 1.1 卷积神经网络

卷积神经网络(convolutional neural network, CNN)最早被用于图像处理,目前已经广泛应用在计算机二维视觉方面。CNN 同样在一维序列数据处理中表现突出,包括自然语言处理和语音识别等。一维 CNN 对数据的特征提取主要是采用多个局部滤波器对整个输入序列进行卷积运算,每个局部滤波器对应的每个特征映射都可以通过在序列上滑动滤波器来生成。

在本文提出的网络中,采用的 CNN 由卷积层

和池化层组成来处理时间序列数据,卷积层通过在序列上滑动滤波器提取特征。卷积操作的计算式为:

$$\mathbf{a}_{i:i+m-1} = \mathbf{a}_i \oplus \mathbf{a}_{i+1} \oplus \cdots \oplus \mathbf{a}_{i+m-1} \quad (1)$$

$$\mathbf{x}_i = f(\mathbf{u}^\top \mathbf{a}_{i:i+m-1} + b) \quad (2)$$

式中: $\mathbf{a}_{i:i+m-1}$ 为从第*i*个时间步长开始的*m*个连续时间窗口;*b*为偏置;*f*为线性修正(rectified linear unit, ReLu)激活函数;*u*为过滤器; $\mathbf{x}_i$ 为卷积运算后的结果。式(2)为最终运算。

池化层输入数据经过卷积层提取后的向量为:

$$\mathbf{x}_j = [x_1, x_2, \cdots, x_{l-m+1}] \quad (3)$$

假设池化层的步长为*s*,提取后的向量为 $\mathbf{x}_j$ 中最大的值。压缩后的向量如式(4)和式(5)所示:

$$\mathbf{y} = [y_1, y_2, \cdots, y_{\frac{l-m}{s}+1}] \quad (4)$$

$$\mathbf{y}_j = \max[x_{(j-1)s}, x_{(j-1)s+1}, \cdots, x_{js-1}] \quad (5)$$

一维 CNN 能够有效分析具有固定长度周期的信号数据,并从中提取重要的特征。在提取与数据不具有高度相关性的特征时, CNN 的效果优于传统方法<sup>[17]</sup>。加入池化层能够对 CNN 提取的特征进一步凝练,从而解决单个时间步长信息识别性不够的问题。通过采用 CNN 压缩序列长度能够增加后续模型捕捉时间信息的能力。

## 1.2 长短期记忆网络

LSTM 是 RNN 的一种变体,与传统 RNN 的循环神经元相比, LSTM 通过门控状态来控制传输状态,如图 1 所示。

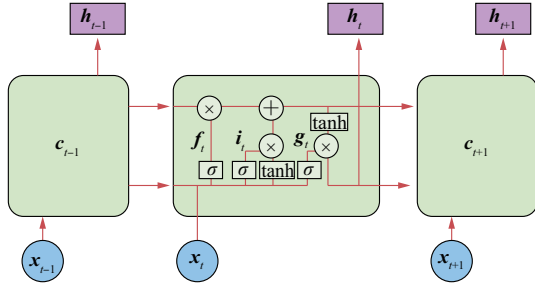


图 1 LSTM 结构

Fig. 1 Structure of LSTM

图 1 中, $f_t$ 为遗忘门控制信号, $i_t$ 为输入门控制信号, $g_t$ 为信息候选状态, $o_t$ 为输出门控制信号。遗忘门 $f_t$ 的功能是控制输入是否移除上一个经过长期状态 $c_{t-1}$ 的信息,输入门 $i_t$ 控制当前输出 $g_t$ 是否加入当前长期状态 $c_t$ 中。输出门 $o_t$ 使用长期状态 $c_t$ 控制当前短期状态的 $h_t$ 信息。长短期状态公式为:

$$f_t = \sigma(\mathbf{W}_{xf}^\top x_t + \mathbf{W}_{hf}^\top \mathbf{h}_{t-1} + b_f) \quad (6)$$

$$i_t = \sigma(\mathbf{W}_{xi}^\top x_t + \mathbf{W}_{hi}^\top \mathbf{h}_{t-1} + b_i) \quad (7)$$

$$g_t = \varphi(\mathbf{W}_{xg}^\top x_t + \mathbf{W}_{hg}^\top \mathbf{h}_{t-1} + b_g) \quad (8)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (9)$$

$$\mathbf{h}_t = \mathbf{z}_t = \mathbf{o}_t \otimes \varphi(\mathbf{c}_t) \quad (10)$$

式中: $\varphi$ 为 tanh 激活函数; $\sigma$ 为 sigmoid 激活函数; $\mathbf{W}$ 为权重矩阵;*b*为偏置值; $\otimes$ 表示对应数组元素相乘。

## 1.3 双向长短期记忆网络

尽管传统 LSTM 单元能够根据以前的信息预测当前状态,但是在输入数据量大且复杂的情况下,仍难以捕捉到多个变量之间的依赖关系。Bi-LSTM 在传统 LSTM 基础上进行改进,通过向前向后传播的方式,使得每个时间序列样本在每个传播方向上都能获得完整的信息,提高了模型处理长序列信息的能力。Bi-LSTM 结构如图 2 所示,它由 2 个不同方向的 LSTM 层组成,网络的输出由正向传播和反向传播的结果共同决定。Bi-LSTM 的传播方式使模型具有稳定的动态学习能力以及从复杂序列数据中提取有用特征的分析能力。

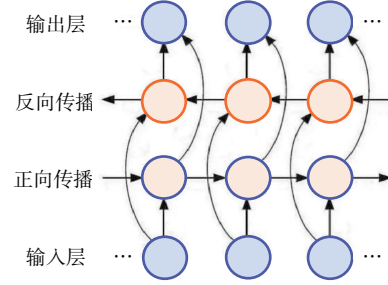


图 2 Bi-LSTM 结构

Fig. 2 Structure of Bi-LSTM

## 1.4 注意力机制

AM 是人类从视觉注意力中获得的灵感,其模型要从序列中学习每一个元素的关键程度,通过分配权重系数的方式将元素进行合并,得到更为优秀的模型<sup>[18]</sup>。神经网络中的注意力机制是一种资源分配方案,它将计算资源分配给更重要的任务<sup>[19]</sup>,解决了计算能力有限情况下的信息过载。在神经网络中,模型的参数越多,其表达能力越强,模型存储的信息越多,但这也可能会导致信息超载。通过引入注意力机制,使网络能够在众多的输入信息中专注于对当前任务更关键的信息,减少对其他无用信息的注意。

$$\mathbf{h}_t = \text{RNN}(x_t, \mathbf{h}_{t-1}) \quad (11)$$

$$\mathbf{c}_i = \sum_{j=1}^n \alpha_{ij} \mathbf{h}_j \quad (12)$$

$$\alpha_{ij} = \frac{\exp(s(\mathbf{h}_i, \overline{\mathbf{h}}_s))}{\sum_{k=1}^n \exp(s(\mathbf{h}_k, \overline{\mathbf{h}}_s))} \quad (13)$$

$$s(\mathbf{h}_i, \overline{\mathbf{h}}_s) = \mathbf{h}_i \cdot \overline{\mathbf{h}}_s \quad (14)$$

式中: $x_t$ 为输入特征;RNN 网络处理输入序列并为每个输入元素 $x_t$ 产生一个状态隐藏变量 $\mathbf{h}_t$ , $\mathbf{h}_t$ 为

时间步  $t$  的隐藏状态;  $\mathbf{h}_0$  为初始隐藏状态;  $\mathbf{c}_i$  为加权平均值, 即输入序列中所有隐藏状态  $\mathbf{h}_j$  的加权表示;  $a_{ij}$  为 2 个隐层单元之间对应的权重比例, 也就是注意力分布, 由注意力机制决定;  $s(x)$  为权重计算函数, 衡量当前隐藏状态  $\mathbf{h}_i$  和时间步  $k$  隐藏状态  $\mathbf{h}_k$  之间的相似度。

如图 3 所示, AM 对训练模型的中间结果进行选择性学习, 通过分配权重的方式将输出序列与之关联。

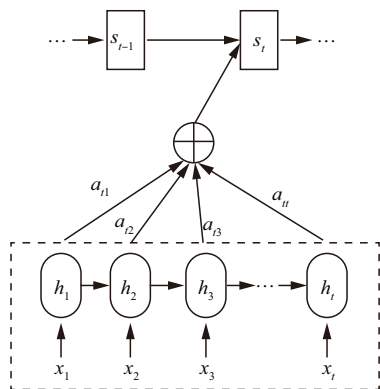


图 3 AM 学习过程

Fig. 3 Learning process of AM

通过注意机制, 将元素与注意分配系数相结合, 得到了较好的模型。注意机制保留训练模型的中间结果, 有选择地学习这些中间结果, 并通过权值将输出序列与它们进行关联。此外, 与传统的 RNN 相比, 注意机制可以并行计算, 每一步的计算不再依赖于前一步的计算结果。与传统方法相比, 该注意机制参数少、计算效率高。

### 1.5 AM 算法与 Bi-LSTM 算法的融合

为了增强 Bi-LSTM 网络处理非线性复杂数据的能力, 建立具有较高预测精度的模型, 本文将 AM 与 Bi-LSTM 网络相结合, 提出轴承 RUL 预测框架, 带有注意力机制的 Bi-LSTM (Bi-LSTM with attention, Bi-LSTM-A)。AM 通过权重系数的分配, 使隐层之间的联系更加紧密, 强调重要信息, 提高预测精度。

如图 4 所示, 中间隐藏层保证了在 Bi-LSTM 后每个时间步的输出都被保留, 这些输出包含了当前时间步的前后上下文信息。分数值计算指 AM 通过 Bi-LSTM 隐藏层的中间变量, 通过一个权重矩阵计算出每个变量对应的分数值。权重计算指权重通过一个前馈网络计算, 确保所有权重加起来等于 1, 按权重分配显示不同信息的重要程度, 代表每个时间步的相对重要性。结合层指将 Bi-LSTM 的输出与其对应的权重相乘, 然后进行求和, 得到一个加权表示, 统一整合信息, 捕捉输入序列中较为关键的特征信息部分。输出关联指这个加权表示随后被用于生成最终的输出序列, 用于预测下一个输出元素, 为后续模块的准确预测打下基础。

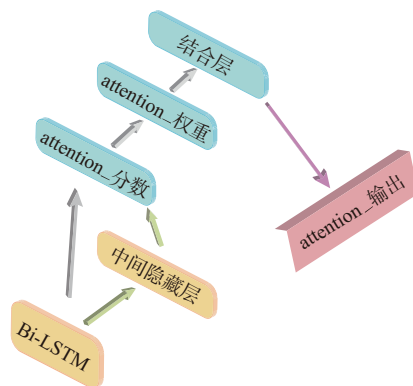


图 4 Bi-LSTM 结合 AM 机制原理

Fig. 4 Schematic diagram of Bi-LSTM with attention mechanism

## 2 模型结构

### 2.1 RUL 预测模型结构

本节提出了一个 RUL 预测模型 Bi-LSTM-A, 模型结构如图 5 所示, 主要由 3 个部分组成: 特征提取器一维卷积神经网络 (one dimensional convolutional neural network, 1DCNN)、具有注意力机制的 Bi-LSTM 和全连接层。

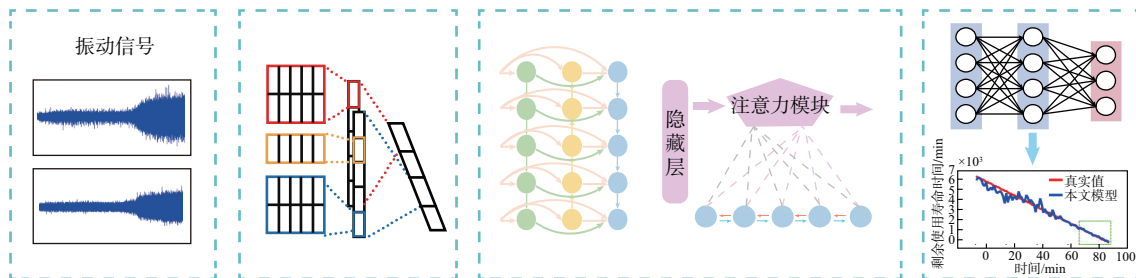


图 5 RUL 预测模型框架

Fig. 5 Structure of RUL prediction model

首先使用 1DCNN 提取原始信号序列, CNN 具有很强的特征提取能力, 可以通过特定的结构从样

本中自动提取特征, 减少了人为参与的资源消耗。1DCNN 通过卷积层和池化层可以有效地从信号数

据中提取重要的特征,所提取的信号通过具有注意机制的 Bi-LSTM 学习更高层次的特征表示。Bi-LSTM 通过双向传播的方式使网络能够在每个时间点了解原始数据的特征。通过增加 AM 机制使模型可以选择性地学习输入,将输出序列与输入联系起来,提高预测精度。最后通过全连接层对 RUL 预测结果进行预测。

### 2.2 预测流程

为了实验能够有效进行,需要对传感信号数据进行处理。首先将数据按时间顺序重构为矩阵  $\mathbf{X}$ , 矩阵的每一列代表每一阶段的信号序列,分别对应相应的 RUL 值。故输入可表示为  $\mathbf{X}=[X_t, X_{t-1}, \dots, X_{t-a+1}]$ , 其中  $a$  为每一阶段的信号序列长度。输出表示为  $\mathbf{RUL}=[RUL_t, RUL_{t-1}, \dots, RUL_{t-a+1}]$ 。与振动信号序列的幅值相比,真实的 RUL 值相对较大,使得神经网络难以建立输入与输出之间映射关系的模型。为了实现模型的快速收敛和准确的 RUL 预测,对实际 RUL 值进行归一化:

$$t = [t - \text{Min}(t)] / [\text{Max}(t) - \text{Min}(t)] \quad (15)$$

式中:Min 为输入的最小值;Max 为输入的最大振幅。通过归一化,可以得到与真实 RUL 值对应的输出标签。经过模型训练,该框架能够输出预测的标签。最后采用逆归一化方法从预测值得到 RUL。

$$t = t[\text{Max}(t) - \text{Min}(t)] + \text{Min}(t) \quad (16)$$

具体的 RUL 预测流程如下:

**步骤 1** 定义每段时间信号序列的长度为  $a$ , 将数据划分为训练集和测试集,测试集信号不参与训练。

**步骤 2** 对数据进行归一化。

**步骤 3** 初始化网络的权重和偏置。

**步骤 4** 训练模型。通过反向传播的方式最小化损失函数,对权值进行微调。

**步骤 5** 输出预测结果 RUL。

**步骤 6** 更新训练集与对应的真实 RUL 值。

**步骤 7** 重复步骤 4~6,直到训练完毕。

整个网络采用监督训练,所有网络的权重都为随机初始化,通过预测结果与期望的输出进行对比,产生误差。通过给定的代价函数,计算其相对误差,按权值的梯度方向更新初始权重值。不断重复更新权重值,直到代价函数值低于预先给定的阈值或达到预先定义的迭代次数后停止。在实际的训练过程中,本文将完整训练数据集分批输入网络进行训练,使用自适应矩估计(adaptive moment estimation, Adam)优化器对代价函数进行调整。Adam 优化器能够计算每个参数的自适应学习率,并保留历史梯度的指数衰减平均值,将更新步长限制在大致的范

围内,自动调整训练过程中的学习率。

## 3 实验与分析

### 3.1 实验数据集

实验数据使用西安交通大学试验台(见图 6)收集的振动数据,整个数据包含了多个轴承在不同工况下从运行到失效的全过程<sup>[20]</sup>。实验数据如表 1 所示,数据集包含 4 个故障类型,分别为内圈故障、外圈故障、主轴断裂以及外环断裂。

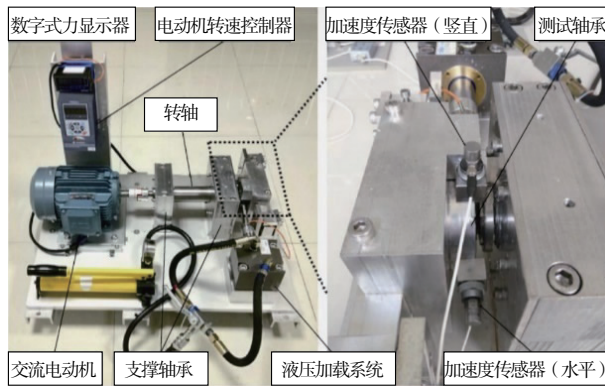


图 6 轴承加速寿命试验台

Fig. 6 Bearing contact fatigue testing machine

表 1 实验数据

Tab. 1 Experimental data

工况	数据集	轴承时间	损坏部位
工况 1 (35 Hz/12 kN)	Bearing1_1	02:03:00	外圈故障
	Bearing1_2	02:41:00	外圈故障
	Bearing1_3	02:38:00	外环断裂
	Bearing1_4	02:02:00	主轴断裂
	Bearing1_5	00:52:00	内圈故障
工况 2 (37.5 Hz/12 kN)	Bearing2_1	08:11:00	内圈故障
	Bearing2_2	02:41:00	外圈故障
	Bearing2_3	08:53:00	主轴断裂
	Bearing2_4	00:42:00	外圈故障
	Bearing2_5	05:39:00	外圈故障
工况 3 (40 Hz/10 kN)	Bearing3_1	42:18:00	外环断裂
	Bearing3_2	41:36:00	主轴断裂
	Bearing3_3	06:11:00	内圈故障
	Bearing3_4	25:15:00	内圈故障
	Bearing3_5	01:54:00	外圈故障

图 7 为数据集 Bearing1\_4 信号各个阶段的振幅值箱线图,从图中可以直观地看出在某个时刻之后,信号的振幅值显著增长。这说明在这个时刻轴承性能出现了明显退化,故障发生的概率远大于正常运行的时刻。本文模型的主要任务是通过这些振动信号的数据准确预测轴承在各个阶段的剩余使用寿命,对工业生产制造有重要意义。

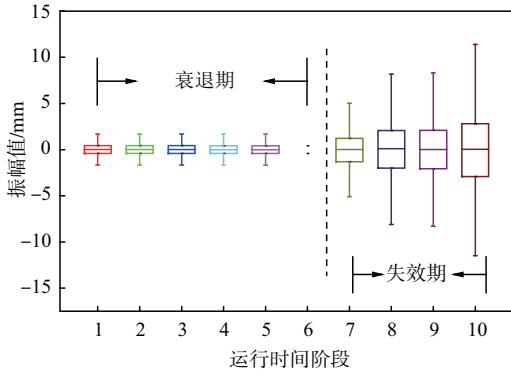


图7 Bearing1\_4 信号振幅变换

Fig. 7 Boxplot of Bearing1\_4 amplitude variation

### 3.2 模型参数设置

在输入网络之前,对输入数据进行归一化处理,以保证训练网络的有效性。一维卷积层包括 16 个卷积核,卷积核大小设为 16,激活函数为 ReLu。最大池化层的步长设为 2。LSTM 的神经元数量为 128,Bi-LSTM 正向传播神经元和反向的数量一致,均选取 64。对应的 AM 神经元与 LSTM 神经元数量一致。网络最后共设有 3 个全连接层,数量分别为 32、16、1。最后一层输出实际预测 RUL。训练过程中采用 Adam 优化器,初始学习率设置为 0.001。模型的评价指标为均方根误差(root mean square error, RMSE)和平均绝对误差(mean absolute error, MAE),即:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y^i - \text{RUL}^i)^2} \quad (17)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\text{RUL}_i - y_i| \quad (18)$$

式中: $y$  为预测值;RUL 为真实值。

为了验证本文模型的性能,将其与以下几种模型进行对比实验:①只带有卷积层和池化层的 CNN;②支持向量机(support vector machine, SVM);③循环神经网络(recurrent neural network, RNN);④单层 LSTM(one-layer LSTM, OLSTM);⑤双向 LSTM (bidirectional LSTM, Bi-LSTM);⑥带有注意力机制的 Bi-LSTM (Bi-LSTM with attention, Bi-

LSTM-A)。

由于 SVM 无法提取信号特征,因此在模型前添加了卷积层。在 SVM 中,正则化参数  $C$  在  $\{0.001, 0.01, 0.1, 1, 10\}$  中取值。

### 3.3 数据集划分

首先,训练数据和测试数据来自同一工况下同一个轴承数据集。例如,对于 Bearing1\_4,从整个振动信号中选取了 2 500 个样本,每个样本对应一个实际的 RUL 值。每个所选样本信号的长度为 1 024 个数据点。在训练过程中,随机打乱这些样本的顺序,从中选取 2 000 个作为训练数据,500 个作为测试数据,以确保训练数据和测试数据之间不存在相同的部分。对每个数据集进行 20 次交叉验证实验,得到平均损失值作为预测精确度的评判指标,避免了预测结果的偶然性。

### 3.4 RUL 预测

对第 3.3 节划分的数据集进行对比实验,表 2~表 4 为 3 个数据集上所有模型 RUL 预测结果的 RMSE 值和 MAE 值。从表中可以看出,CNN 虽然具有较强的特征提取能力,但是对于时间序列的预测能力较差。对于 SVM 模型,即使输入序列经过 CNN 进行了特征提取,其预测精度仍低于递归模型。结果表明,相对于传统线性模型,递归模型在信号序列分析和预测方面有其独特的优势。

在递归模型中,Bi-LSTM 的性能优于传统单向 LSTM,这说明通过双向传播的方式可以学习到更抽象的特征表示,有利于对长期复杂信号序列的预测。在 Bi-LSTM 中加入注意机制,可以更准确地捕获输入信号的关键信息。图 8 为本文提出的 Bi-LSTM-A 模型在各个模块提取的特征结果,从图 8 可以看出,原始振动信号的特征难以区分,大多数特征都堆叠在一起,不利于 RUL 预测。经过 Bi-LSTM 和 AM 模块的处理后,信号特征能够被有效地区分开,为本文模型提供精确的 RUL 预测作出了合理的解释。

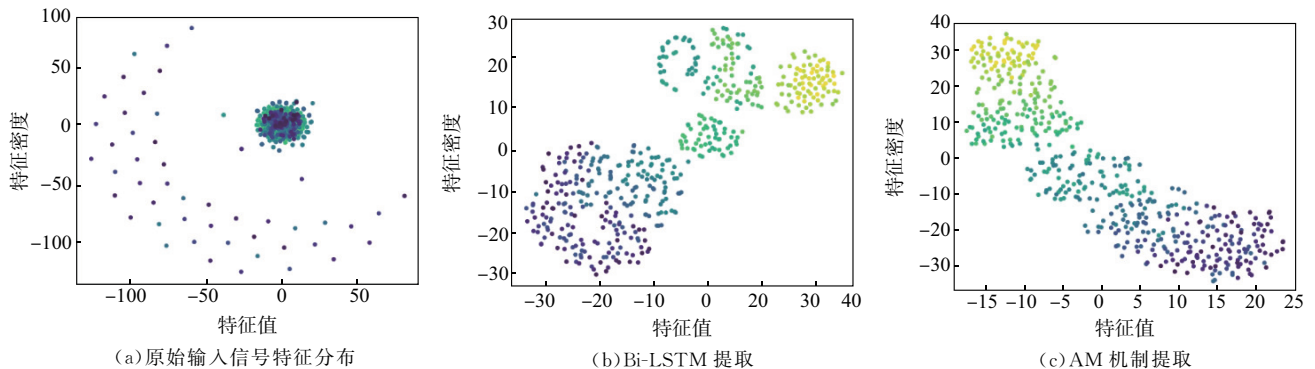


图8 不同模块提取的特征分布

Fig. 8 Feature distribution diagram extracted by different modules

表 2 不同模型预测结果的 RMSE 和 MAE (Bearing1\_1 数据集)

Tab. 2 RMSE and MAE of different model prediction results (Bearing1\_1 dataset)

模型	RMSE	MAE
CNN	195.67(±25.47)	124.36(±15.32)
SVM	164.88(±19.58)	104.92(±17.36)
RNN	118.94(±20.86)	83.11(±10.95)
LSTM	55.52(±12.78)	41.80(±8.74)
Bi-LSTM	43.56(±10.51)	34.13(±9.84)
Bi-LSTM-A	36.24(±7.66)	27.02(±6.42)

表 3 不同模型预测结果的 RMSE 和 MAE (Bearing2\_3 数据集)

Tab. 3 RMSE and MAE of different model prediction results (Bearing2\_3 dataset)

模型	RMSE	MAE
CNN	175.68(±28.36)	118.84(±16.15)
SVM	158.05(±26.37)	98.84(±18.52)
RNN	101.64(±22.17)	75.34(±13.24)
LSTM	51.24(±15.29)	38.35(±11.32)
Bi-LSTM	41.99(±16.78)	28.57(±10.52)
Bi-LSTM-A	33.76(±11.38)	23.68(±9.45)

表 4 不同模型预测结果的 RMSE 和 MAE (Bearing3\_5 数据集)

Tab. 4 RMSE and MAE of different model prediction results (Bearing3\_5 dataset)

模型	RMSE	MAE
CNN	180.64(±26.83)	116.25(±19.36)
SVM	165.98(±28.97)	109.17(±25.47)
RNN	130.49(±25.71)	94.68(±16.02)
LSTM	66.82(±17.47)	46.39(±13.59)
Bi-LSTM	53.85(±19.34)	34.57(±15.59)
Bi-LSTM-A	42.17(±10.62)	29.63(±11.93)

图 9 和图 10 分别给出了 20 次交叉验证实验的平均 RMSE 值和 MAE 值,黑线表示其置信区间。以 Bearing1\_1 为例,通过 RUL 预测置信区间的估计结果可以看出,对于 Bi-LSTM-A 模型,有 95% 的可能性预测的 RUL 的 RMSE 值为 28.58~43.90, MAE 值为 20.60~33.44。与其他模型相比,本文提出的模型 Bi-LSTM-A 的预测结果更加准确,相较于其他模型性能更优越。

图 11 为 Bearing1\_1 轴承数据集部分模型的 RUL 预测结果。RUL 预测中,RNN 算法能够处理序列数据,适用于时间序列预测问题,但在处理长序列时容易发生梯度消失或梯度爆炸,不能很好地捕捉长期依赖关系,对于轴承从运行到失效这类时间跨度大的序列数据预测效果不佳。LSTM 算法引入了遗忘门、输入门和输出门,能够有效地解决梯度

消失问题,适合处理长序列数据,但在输入数据量大且复杂的情况下,仍难以捕捉到多个变量之间的依赖关系。Bi-LSTM 算法结合了前向和后向 LSTM,能够捕捉到序列数据中的双向依赖关系,较单向 LSTM 性能更好,但训练和推断的计算成本比单向 LSTM 更高,对于非常长的序列,仍然可能存在梯度消失或爆炸。Bi-LSTM-A 算法可以帮助模型聚焦于序列中最重要的部分,提高预测准确性的同时可以更好地捕捉到序列数据中的长期和短期依赖关系,尤其是处理复杂的序列数据时,通常比单独使用 Bi-LSTM 性能更好。从图 11 可以直观地看出本文提出的 Bi-LSTM-A 模型的 RUL 预测准确度明显优于其他模型,且模型预测 RUL 结果与真实值的拟合效果优于其他模型,验证了 Bi-LSTM 与 AM 机制结合的有效性。

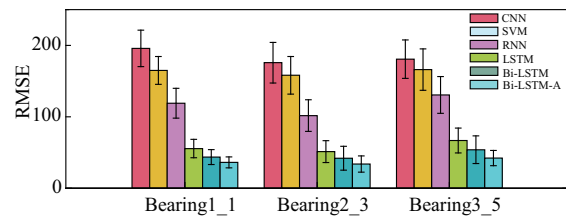


图 9 不同模型 RMSE 置信区间

Fig. 9 Confidence interval of RMSE for different models

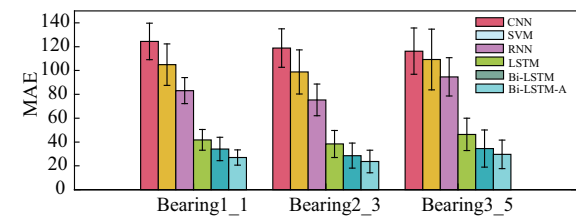


图 10 不同模型 MAE 置信区间

Fig. 10 Confidence interval of MAE for different models

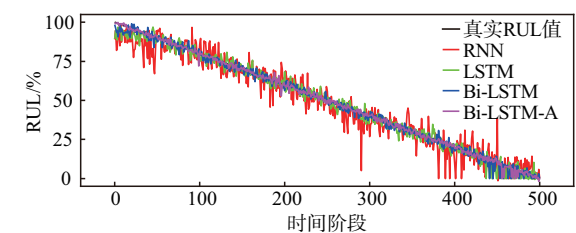


图 11 不同模型在 Bearing1\_1 数据集的 RUL 预测结果

Fig. 11 RUL prediction results of different models from Bearing1\_1

## 4 结语

本文提出了一种基于深度学习的轴承寿命预测模型。设计了一个具有卷积层和池化层的一维 CNN 用于从原始信号序列中提取局部特征,然后利用具有注意力机制的 Bi-LSTM 对时间信息进行编码。双向传播信息的方式可以使模型通过观察相邻

特征之间的映射关系来发现网络输入与输出之间的关系,提高模型处理非线性的能力。加入注意机制可以使隐含层之间的联系更加紧密,从而提高预测精度。结合 Bi-LSTM 的特征提取能力和注意力机制强大的监督能力,实现了对轴承的精确 RUL 预测。通过比较不同模型的预测结果,验证了该方法的有效性和可行性。

### 参考文献

- [1] ZIO E. Prognostics and Health Management Methods for Reliability Prediction and Predictive Maintenance [J]. *IEEE Transactions on Reliability*, 2024, 73(1): 41.
- [2] KAMAT P, KUMAR S, PATIL S, et al. Anomaly-Informed Remaining Useful Life Estimation (AIRULE) of Bearing Machinery Using Deep Learning Framework [J]. *MethodsX*, 2024, 12: 102555.
- [3] CAESARENDRA W, WIDODO A, THOM P H, et al. Combined Probability Approach and Indirect Data-Driven Method for Bearing Degradation Prognostics [J]. *IEEE Transactions on Reliability*, 2011, 60(1): 14-20.
- [4] ZHAO R, YAN R Q, CHEN Z H, et al. Deep Learning and Its Applications to Machine Health Monitoring [J]. *Mechanical Systems and Signal Processing*, 2019, 115: 213-237.
- [5] ZHAO J H, ZHU Y, ZHANG B, et al. Review of State Estimation and Remaining Useful Life Prediction Methods for Lithium-Ion Batteries [J]. *Sustainability*, 2023, 15(6): 5014.
- [6] ZHANG J S, LI X, TIAN J L, et al. A Variational Local Weighted Deep Sub-Domain Adaptation Network for Remaining Useful Life Prediction Facing Cross-Domain Condition [J]. *Reliability Engineering & System Safety*, 2023, 231: 108986.
- [7] CHEN J X, LI D P, HUANG R Y, et al. Aero-Engine Remaining Useful Life Prediction Method with Self-Adaptive Multimodal Data Fusion and Cluster-Ensemble Transfer Regression [J]. *Reliability Engineering & System Safety*, 2023, 234: 109151.
- [8] LIPTON Z C, BERKOWITZ J, ELKAN C, et al. A Critical Review of Recurrent Neural Networks for Sequence Learning [EB/OL]. (2015-10-17) [2024-03-01]. <https://arxiv.org/abs/1506.00019v4>.
- [9] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [10] DONG S J, XIAO J F, HU X L, et al. Deep Transfer Learning Based on Bi-LSTM and Attention for Remaining Useful Life Prediction of Rolling Bearing [J]. *Reliability Engineering & System Safety*, 2023, 230: 108914.
- [11] TIAN H X, YANG L Z, JU B T. Spatial Correlation and Temporal Attention-Based LSTM for Remaining Useful Life Prediction of Turbofan Engine [J]. *Measurement*, 2023, 214: 112816.
- [12] ZHU H, LUO Y, LIU Q, et al. Multistep Flow Prediction on Car-Sharing Systems: A Multi-Graph Convolutional Neural Network with Attention Mechanism [J]. *International Journal of Software Engineering & Knowledge Engineering*, 2019, 29(11): 1727-1740.
- [13] 徐诚极, 王晓峰, 杨亚东. Attention-YOLO: 引入注意力机制的 YOLO 检测算法 [J]. *计算机工程与应用*, 2019, 55(6): 13-23, 125.
- [14] XU C J, WANG X F, YANG Y D. Attention-YOLO: YOLO Detection Algorithm that Introduces Attention Mechanism [J]. *Computer Engineering and Applications*, 2019, 55(6): 13-23, 125. (in Chinese)
- [15] CHOI H, CHO K, BENGIO Y. Fine-Grained Attention Mechanism for Neural Machine Translation [J]. *Neurocomputing*, 2018, 284: 171-176.
- [16] ANDERSON P, HE X D, BUEHLER C, et al. Bottom-up and Top-down Attention for Image Captioning and Visual Question Answering [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018: 6077-6086.
- [17] WU L, WANG Y, LI X, et al. Deep Attention-Based Spatially Recursive Networks for Fine-Grained Visual Recognition [J]. *IEEE Transactions on Cybernetics*, 2019, 49(5): 1791-1802.
- [18] 王进花, 韩金玉, 曹洁, 等. 基于 AM 和 CNN 的多级特征融合的风力发电机轴承故障诊断方法 [J]. *太阳能学报*, 2024, 45(5): 51-61.
- [19] WANG J H, HAN J Y, CAO J, et al. Fault Diagnosis Method of Wind Turbine Bearing Based on AM and CNN Multi-Level Feature Fusion [J]. *Acta Energetica Solaris Sinica*, 2024, 45(5): 51-61. (in Chinese)
- [20] RAN X D, SHAN Z G, FANG Y F, et al. An LSTM-Based Method with Attention Mechanism for Travel Time Prediction [J]. *Sensors*, 2019, 19(4): 861-870.
- [21] LIN L, LUO H, HUANG R J, et al. Recurrent Models of Visual Co-Attention for Person Re-Identification [J]. *IEEE Access*, 2019, 7: 8865-8875.
- [22] 雷亚国, 韩天宇, 王彪, 等. XJTU-SY 滚动轴承加速寿命试验数据集解读 [J]. *机械工程学报*, 2019, 55(16): 1-6.
- [23] LEI Y G, HAN T Y, WANG B, et al. XJTU-SY Rolling Element Bearing Accelerated Life Test Datasets: A Tutorial [J]. *Journal of Mechanical Engineering*, 2019, 55(16): 1-6. (in Chinese)

(编辑:徐楠楠)