

DOI:10.19651/j.cnki.emt.2517972

基于 STA 机制的 AE-BiLSTM 的滚动轴承剩余寿命预测^{*}

施 彤¹ 张自豪¹ 邱晓惠¹ 张 菘^{1,2}

(1.南京信息工程大学自动化学院 南京 210044;2.江苏省大气环境与装备技术协同创新中心 南京 210044)

摘要: 轴承作为机械零部件中不可或缺的组成部分,长时间工作容易导致轴承磨损和疲劳失效,进而影响机械设备的正常运转。因此,对轴承的剩余使用寿命(RUL)的预测可以有效避免意外发生,确保设备安全可靠地运行。为了提高滚动轴承的 RUL 预测精度,本文提出了一种基于空间时间注意力(STA)机制的自编码(AE)和双向长短记忆(BiLSTM)的滚动轴承寿命预测方法,有效地整合轴承数据中的多种模态信息,从而捕捉轴承运行状态的变化。首先,将原始振动信号输入到自编码模型自动提取故障特征;然后,将提取的特征输入到 STA 模型中,对特征数据在特征维度上的空间信息和运行时间步长信息进行深度加权融合,以更全面地捕捉特征维度和时间维度的信息;再结合 BiLSTM 模型对轴承的剩余使用寿命进行预测;最后,通过 PHM2012 挑战赛数据集和 ABLT-1A 轴承全寿命周期数据进行试验验证,实验结果表明所提出的模型的 RMSE 平均降低了约 22.76%,MAE 平均降低了约 26.57%,而 R^2 平均提升了约 12.47%,可以看出所提出方法对 RUL 预测准确度有明显的提升。

关键词: 滚动轴承;空间时间注意力模型;BiLSTM 网络;剩余寿命预测

中图分类号: TH17; TN98 文献标识码: A 国家标准学科分类代码: 460. 99

Rolling bearing remaining life prediction using AE-BiLSTM with STA mechanism

Shi Tong¹ Zhang Zihao¹ Qiu Xiaohui¹ Zhang Wan^{1,2}

(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Jiangsu Collaborative Innovation Center for Atmospheric Environment and Equipment Technology, Nanjing 210044, China)

Abstract: Bearing as an indispensable part of mechanical parts, long-term work is easy to lead to bearing wear and fatigue failure, and then affect the normal operation of mechanical equipment. Therefore, the prediction of the remaining useful life (RUL) of the bearing can effectively avoid accidents and ensure the safe and reliable operation of the equipment. To enhance the prediction accuracy of the RUL for rolling bearings, this paper proposes a rolling bearing life prediction method based on spatio-temporal attention (STA) and bidirectional long short-term memory (BiLSTM). Effectively integrate multiple modal information in bearing data to capture changes in bearing operating state. Firstly, the original vibration signal is input to auto-encoder (AE) model to extract fault features automatically. Then, the extracted features are input into the STA model, and the spatial information and running time step information of the feature data in the feature dimension are deeply weighted to capture the information of the feature dimension and time dimension more comprehensively. Combined with BiLSTM model, the remaining service life of bearing is predicted. Finally, experimental validation is conducted using datasets from the PHM2012 Challenge and ABLT-1A bearing full-life cycle data. The experimental results indicate that the proposed model has achieved an average reduction of approximately 22.76% in RMSE and 26.57% in MAE, while the R^2 has improved by an average of about 12.47%. It can be concluded that the proposed method significantly enhances the accuracy of RUL prediction.

Keywords: rolling bearing; spatio-temporal attention model; BiLSTM network; remaining useful life prediction

0 引言

轴承作为机械零部件中不可或缺的组成部分,被广泛

用于电机、变速箱等机械设备中,是工业生产中广泛使用的配套件和基础件。由于轴承长时间工作可能引起磨损和疲劳失效,滚动轴承一旦发生故障,不仅会造成经济损失,甚

收稿日期:2025-01-20

* 基金项目:国家自然科学基金(62105160)项目资助

至会导致安全事故发生^[1]。对滚动轴承的健康监测和剩余寿命预测,对确保机械设备的安全性和可靠性至关重要。

为了实现轴承的状态监测和健康管理,人工智能方法已逐渐被引入到轴承故障诊断领域中。深度学习作为人工智能领域中非常重要的技术之一,被广泛应用于滚动轴承剩余寿命预测,实现对滚动轴承数据分析、特征提取、和剩余寿命预测。陈曦等^[2]提出了基于个性化联邦学习的轴承剩余寿命预测方法,实现不同工况下轴承的剩余使用寿命(remaining useful life, RUL)预测。袁倩等^[3]利用 D-S 信息融合,以轴心轨迹为初步判断,结合集合经验模态分解和反向传播神经网络进行故障识别,并最终通过 D-S 证据理论做决策融合,提高了磁轴承转子故障诊断结果的可靠性。此外,冯兆熙等^[4]针对高速高精加工过程滚动轴承的剩余寿命预测建模中存在样本少、标注难度大等问题,引入深度置信网络,同时融合无监督与有监督微调学习方法进行滚动轴承剩余使用寿命预测的研究。杨小东等^[5]提出了融合卷积神经网络模型,将单一的输入转换为多维度的输入,使模型能充分提取到轴承振动信号特征,从而有效地实现对轴承进行剩余寿命预测。在剩余寿命预测领域中,深度学习方法经常面临一个问题:存在大量冗余或次要信息。这些信息可能妨碍有效地识别和突出关键特征,导致模型过度泛化或易受噪声干扰,进而影响预测的准确性。

为了克服这一挑战,采用注意力机制已被证明是一种有效的方法。注意力机制在深度学习领域是一种模拟人类视觉和认知系统的方法,能够分配不同的权重或关注度给输入数据的不同部分,以便更好地处理重要信息。特别是在深度学习模型中,通过引入注意力机制来动态调整权重,使模型有针对性地关注轴承运行过程中的关键时刻或特定频率的振动信号,从而显著提高了模型预测的准确性和鲁棒性。近年来,结合深度学习和注意力机制的机械剩余寿命预测方法得到广泛应用,兰杰等^[6]为实现一种端对端的 RUL 预测并提升预测的精度,提出了一种基于注意力机制和残差深度分离卷积的 RUL 预测方法,动态地更新注意力权重,从而使模型达到最好的预测性能。Cao 等^[7]提出了一种新型时间卷积网络与残差自注意机制,用于滚动轴承剩余寿命的预测。该方法可以充分利用自注意力机制与编码器-解码器结构的优势,解决轴承剩余寿命预测中序列过长、特征提取困难等问题,提高了预测准确率。此外, Hu 等^[8]提出了一种新的基于特征注意力的端到端方法,可以实现自适应地加权输入特征,从而提高了 RUL 预测精度。Liu 等^[9]提出了一种新的基于深度神经网络的编码器-解码器模型,用于处理多特征数据的 RUL 预测问题,该模型利用自注意力机制动态地给不同特征赋予不同的权重,并结合长短期记忆网络(long short-term memory, LSTM)和全连接层构建了一个深度神经网络模型。上述方法中通过注意力机制来深层挖掘特征维度或时间维度的重要信息,并没有结合多维特征中空间和时间两个维度的

信息,导致在预测模型中并不能很好地捕捉数据的特征信息,预测也难以达到理想的效果。

为了克服上述注意力机制的局限性,空间时间注意力机制可以根据不同时间步和空间位置的需求,分配不同的权重或关注度。空间时间注意力机制主要用于处理时空数据,如视频、语音和图像等。付孟丹等^[10]提出了一种基于空间和时间注意力机制的视频异常检测方法,引用空间和时间注意力机制,对特征图从空间和时间两个维度进行加权操作,提升了模型的有效性。张志浩等^[11]提出了一种基于空间时间注意力(spatio-temporal attention, STA)卷积循环神经网络的语声情感识别,使用卷积神经网络提取空间特征和长短期记忆网络提取时间特征,加入空间注意力和时间注意力机制,能够有效提升识别的准确率。李大湘等^[12]提出了一种基于 STA-LSTM 的自发微表情识别算法,将微表情视频图像作为输入,构建空间注意力和时间注意力模块,并对不同帧给予不同程度的关注度,能够提高微表情识别精度。

上述研究中,空间和时间注意力机制被广泛应用于二维图像的处理,而在一维序列的预测方面应用还很有限。滚动轴承的原始振动信号经过自动特征提取后构成了多维特征空间,为了更好的理解多维特征空间中的特征维度和时间维度信息,本文提出了一种基于空间时间注意力机制的 AE-BiLSTM 模型的剩余寿命预测方法。首先,滚动轴承全寿命原始振动信号输入到自编码模型进行自动特征提取;然后,通过空间和时间注意力机制将特征空间维度和时间步长信息进行深度融合,再输入到 BiLSTM 模型中进行剩余寿命预测;最后,通过 PHM2012 挑战赛的数据集和 ABLT-1A 轴承全寿命周期数据集对所提出的模型进行试验验证,表明所提出的模型能够更深层次地捕捉到数据的特征信息,进一步提升了预测的准确度。

1 理论分析

1.1 AE 模型

自编码(auto-encoder, AE)模型是一种无监督训练的神经网络,通过编码和解码过程来实现,经过不断压缩和恢复数据,可以学习数据特征的深度表达,被广泛用于特征学习和数据降维。编码阶段可以被认为是通过确定性映射将输入转换为隐藏层表达式,而解码阶段试图将隐藏层表达式重新映射到输入。AE 模型的结构如图 1 所示,其中隐层单元 a 的输出是根据输入样本数据计算得出的。通常情况下,AE 模型的最终目标是学习到一个低维表示,这个表示通常比原始输入 x 提供更优质的输入表征。在 AE 模型中,期望输出 y 与输入 x 尽可能地相似。

在 AE 模型中,从输入层到隐藏层的编码过程可以表示为:

$$g_1(x) = f(\mathbf{W}_1x + \mathbf{b}_1) \quad (1)$$

从隐藏层到输出层的解码过程可以表示为:

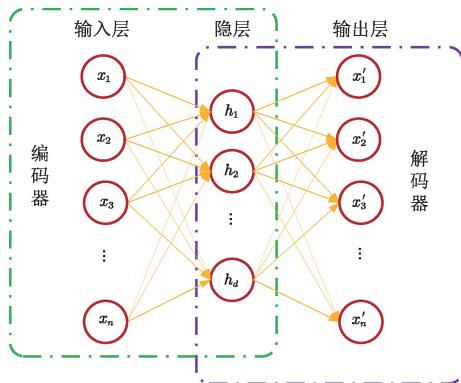


图 1 AE 模型结构

Fig. 1 AE model structure

$$g_2(x) = f(\mathbf{W}_2 g_1(x) + \mathbf{b}_2) \quad (2)$$

其中, $\mathbf{W}_1, \mathbf{W}_2$ 为编码矩阵; $\mathbf{b}_1, \mathbf{b}_2$ 为编码偏置向量; f 表示激活函数, 常用的有 sigmoid、tanh、ReLU 等。模型的损失函数可以表示为:

$$J(\mathbf{W}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \| g_2(x^{(i)}) - x^{(i)} \| ^2 \quad (3)$$

其中, N 表示样本数; i 表示第 i 个样本。

AE 模型由编码器网络和解码器网络两部分组成, 它们的主要功能在于高效地处理原始数据。编码器负责将高维原始数据进行压缩, 生成低维编码, 以最大程度地捕捉原始数据的关键特征。随后, 解码器将这些编码解析为新的数据, 最大限度地还原原始数据的结构。这一过程通常包括 3 个关键步骤: 首先, 编码器对未标记数据样本进行编码, 生成编码表示。其次, 解码器将编码还原为新的数据。最后, 通过计算新数据与原始数据之间的信息误差, 动态地调整编码器和解码器的权重参数, 以实现最小化重构误差的目标。最终, 生成的编码代表了原始数据样本的有效特征表达。

1.2 BiLSTM 模型

LSTM 是一种用于时间序列预测的神经网络算法, 它成功地解决了 RNN 的梯度爆炸或消失问题, 并在时间序列预测中表现得更好。该算法通过遗忘门、输入门和输出门的设计, 有效地调节了信息的流动和控制, 从而提高了模型的性能。

遗忘门用于确定可以从单元中丢弃的信息。其输出 f_t 与前一时刻隐层的输出 h_{t-1} 和当前时刻的输入 x_t 有关, 其表达式如下:

$$f_t = \sigma(\mathbf{W}_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

其中, \mathbf{W}_f 是输出层的权重矩阵; b_f 是遗忘门的偏置; σ 是 sigmoid 激活函数。

输入门决定单元状态存储的信息, 包括两个参数 i_t 和 \tilde{C}_t 。由 sigmoid 激活函数得到的 i_t 用于确定更新值; 由 tanh 激活函数得到的 \tilde{C}_t 表示新候选向量。其表达式如下:

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(\mathbf{W}_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

其中, $\mathbf{W}_i, \mathbf{W}_c$ 分别是输入层和存储单元的权重矩阵, b_i 和 b_c 是相应的偏差。然后可通过下式将先前单元状态 C_{t-1} 更新为 C_t :

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (7)$$

最终的输出值 h_t 取决于输出门, 可通过下式获得:

$$o_t = \sigma(\mathbf{W}_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (9)$$

其中, \mathbf{W}_o 和 b_o 分别是输出层的权重矩阵和偏置。

双向长短期记忆 (bidirectional long short-term memory, BiLSTM) 模型同时在相反的方向上训练两个并行的 LSTM 层, 并将它们的输出组合成单个输出。在训练过程中, 从输入到输出和从输出到输入的数据都被训练。与简单的 LSTM 模型相比, BiLSTM 模型可以同时从前向两个方向读取和整合序列信息, 在解决大规模时间序列问题方面具有很好的优势, 其结构如图 2 所示。当前时间步的前向 LSTM 的隐藏状态和单元状态可以表示如下:

$$h_t^f, c_t^f = \text{LSTM}^f(x_t, h_{t-1}^f, c_{t-1}^f) \quad (10)$$

其中, LSTM^f 是前向 LSTM, h_{t-1}^f 和 c_{t-1}^f 是前向 LSTM 在最后一个时间步的隐藏状态和单元状态。类似的, 当前时间步的后向 LSTM 的隐藏状态和单元状态表示如下:

$$h_t^b, c_t^b = \text{LSTM}^b(x_t, h_{t+1}^b, c_{t+1}^b) \quad (11)$$

其中, LSTM^b 是后向 LSTM, h_{t+1}^b 和 c_{t+1}^b 是后向 LSTM 在最后一个时间步的隐藏状态和单元状态。最后, BiLSTM 模型在当前时间步的输出可以通过拼接当前时间步的前向 LSTM 和后向 LSTM 的隐藏状态来获得:

$$y_t = \text{concat}(h_t^f, h_t^b) \quad (12)$$

其中, y_t 是 BiLSTM 模型在时间步长 t 的输出。

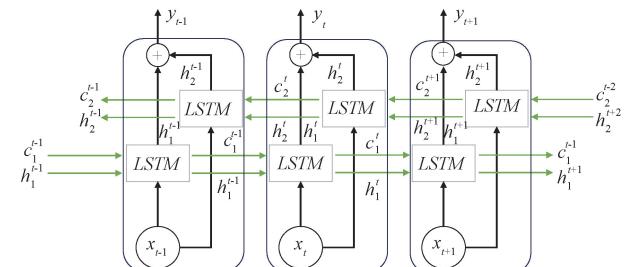


图 2 BiLSTM 模型结构

Fig. 2 BiLSTM model structure

1.3 STA 模型

STA 模型是一种用于处理时空数据的深度学习模型, 主要用于分析和建模涉及时空关系的数据, 结合了空间和时间信息, 以便更好地理解和预测数据中的时空关联。

1) 卷积层

在深度学习领域, 卷积运算被广泛用于特征提取。与原始数据相比, 卷积运算提取的特征能更好地表征机械设备的

运行状态。因此,使用 1 维卷积核对数据矩阵 $\mathbf{X}_t \in \mathbb{R}^{w \times N}$ 进行预处理,然后可以获得卷积特征矩阵,表示为 $\mathbf{H}_t \in \mathbb{R}^{w \times N}$ 。 \mathbf{H}_t 的行方向表示特征维度,列方向表示时间维度。

2) 空间注意力机制

空间注意力机制用于在同一时间步长中权衡不同特征维度的重要性,原理图如图 3 所示。对于不同的时间步长,相同特征维度的重要性可能不相同。因此,对卷积特征矩阵 $\mathbf{H}_t \in \mathbb{R}^{w \times N}$ (其中每行表示一个时间步)执行逐行注意机制,其过程可以通过以下公式描述:

$$\mathbf{H}'_t = \tanh(\mathbf{H}_t \mathbf{W}_1^{feature}) \quad (13)$$

$$\mathbf{A}_t^{feature} = \text{softmax}_{row}(\mathbf{H}'_t \mathbf{W}_2^{feature}) \quad (14)$$

$$\mathbf{H}_t^{feature} = \mathbf{A}_t^{feature} \odot \mathbf{H}_t \quad (15)$$

其中, $\mathbf{W}_1^{feature} \in \mathbb{R}^{N \times N}$, $\mathbf{W}_2^{feature} \in \mathbb{R}^{N \times N}$ 都是执行关于行的特征变换可学习矩阵, \tanh 是非线性激活函数, softmax_{row} 表示逐行 softmax 函数, \odot 表示逐元素乘法运算符。 $\mathbf{A}_t^{feature} \in \mathbb{R}^{w \times N}$ 是行归一化的特征权重矩阵,其中每行求和为 1。每行中的权重反映了该行所表示的时间步长处的不同特征维度的重要性。通过将 $\mathbf{A}_t^{feature}$ 和 \mathbf{H}_t 进行逐个元素相乘,可以将权重分配给 \mathbf{H}_t 的每一行,以获得空间加权卷积特征矩阵 $\mathbf{H}_t^{feature} \in \mathbb{R}^{w \times N}$ 。

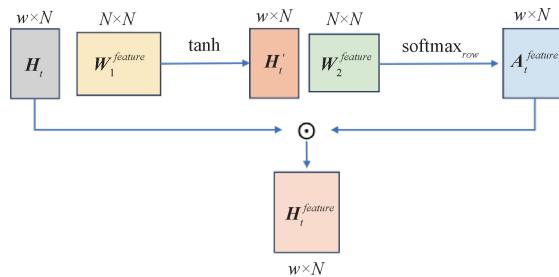


图 3 空间注意力机制原理

Fig. 3 Principle of spatial attention mechanism

3) 时间注意力机制

时间注意机制用于权衡不同时间步长对同一特征维度的重要性,其原理图如图 4 所示。由于相同的时间步长对不同的特征维度的重要性可能不同,对卷积特征矩阵 $\mathbf{H}_t \in \mathbb{R}^{w \times N}$ (其中每列表示一个特征维度)实行列式注意机制,其过程如下:

$$\mathbf{H}'_t = \tanh(\mathbf{W}_1^{time} \mathbf{H}_t) \quad (16)$$

$$\mathbf{A}_t^{time} = \text{softmax}_{col}(\mathbf{W}_2^{time} \mathbf{H}'_t) \quad (17)$$

$$\mathbf{H}_t^{time} = \mathbf{A}_t^{time} \odot \mathbf{H}_t \quad (18)$$

其中, $\mathbf{W}_1^{time} \in \mathbb{R}^{w \times w}$, $\mathbf{W}_2^{time} \in \mathbb{R}^{w \times w}$ 是可学习的矩阵,执行关于列的特征变换。 softmax_{col} 表示逐列 softmax 函数。 $\mathbf{A}_t^{time} \in \mathbb{R}^{w \times N}$ 是列归一化时间步长权重矩阵,其中每列的总和为 1。权重矩阵的每一列反映时间窗口中的不同时间步长对由该列表示的特征维度的重要性。以与空间注意力机制类似的方式,可以通过逐个元素地乘以 \mathbf{A}_t^{time} 和 \mathbf{H}_t 来获得时间加权卷积特征矩阵 $\mathbf{H}_t^{time} \in \mathbb{R}^{w \times N}$ 。

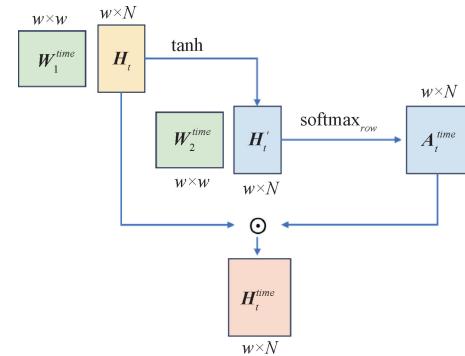


图 4 时间注意力机制原理

Fig. 4 The principle of time attention mechanism

1.4 数据处理

1) 数据归一化

数据归一化的主要目的是消除不同特征之间的尺度和范围差异,从而使数据具有可比性,进而提高数据分析的效果。由于数据集呈现非正态分布且具有一定的离散性,因此采用最小-最大归一化(min-max normalization)进行数据预处理。归一化公式如下:

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (19)$$

其中, X 表示原始数据, X_{new} 表示归一化后的数据, X_{min} 是数据集的最小值, X_{max} 是数据集的最大值。

2) 数据切片

数据切片是指根据特定标准或条件将数据集分割成多个子集,每个子集包含数据的特定部分。这种技术对于有效管理和分析大型数据集至关重要。例如,时间序列数据可以按特定时间段划分为多个区间,从而能够详细研究随时间变化的趋势。通过从单个数据点过渡到预定的片段,可以更准确地预测数据趋势。然而,在数据切片过程中,切片窗口大小的确定需要仔细考虑多种因素。这些因素的选择和调整通常需要根据数据集和分析目标进行优化和微调,以实现最佳的数据处理效率和结果。如果切片窗口过小,可能无法充分捕捉数据趋势;反之,如果切片窗口过大,可能会导致切片之间的重叠过多,从而降低数据处理效率和预测准确性。

1.5 RUL 定义

RUL 是指一个物理系统或设备在某个时间点上的剩余可靠使用时间。在预测维护和故障诊断领域,RUL 是指一个系统或设备在当前时间下仍能正常运行的剩余时间。通常通过监测系统或设备的性能和状态,结合历史数据、统计学方法和机器学习算法来预测其剩余寿命。机械 RUL 预测中有两个重要的数据,分别是当前时刻和最大寿命值。第 i 时刻的 RUL 可以表示为:

$$R_i = t_{max} - t_i \quad (20)$$

其中, t_{max} 表示设备失效时刻, t_i 表示为当前 i 时刻, R_i 表示为设备在当前 i 时刻下的 RUL。图 5 所示为当前时

刻和 RUL 的关系图。由于分阶段退化的健康指标构建方法更加符合轴承在实际应用中的退化过程,因此,当健康指标随时间的增加而降低时,表示轴承正在逐步退化,RUL 不断减小。当健康指标为 0 时,表示轴承已完全失效。

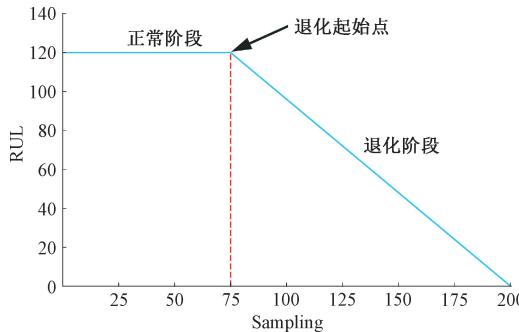


图 5 RUL 关系图

Fig. 5 Relationship diagram for RUL

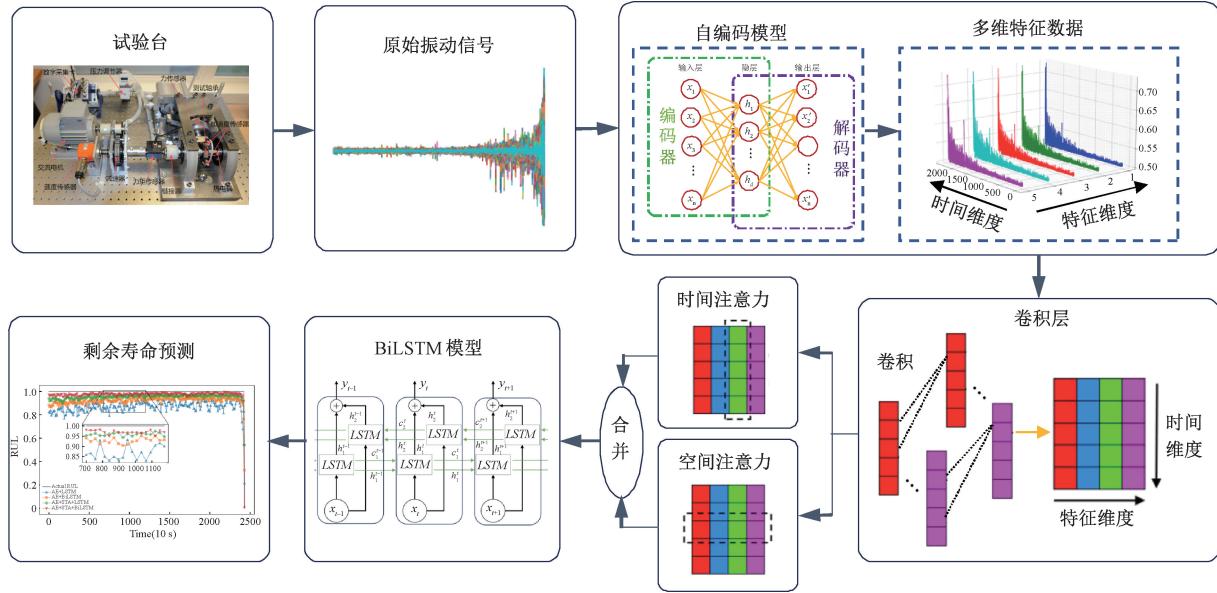


图 6 所提出方法的流程图

Fig. 6 Flow diagram of the proposed method

2.2 评价标准

为了评估预测模型的性能,本文选取 MAE (mean absolute error)、RMSE (root mean square error) 和 R^2 分数 (R-squared score) 三项评价指标。

1) MAE

MAE 代表平均绝对误差,可以用来评估模型的精度和偏差程度。其计算公式如下:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (21)$$

其中, n 表示样本数量, y_i 表示第 i 个样本的真实值, \hat{y}_i 表示模型预测的值。

2) RMSE

RMSE 代表均方根误差,是用来衡量预测值与真实值

2 基于 STA 机制的 AE-BiLSTM 的滚动轴承剩余寿命预测流程及评价指标

2.1 所提出算法的流程

本文基于空间时间注意力机制的 AE-BiLSTM 的滚动轴承剩余寿命预测的流程如图 6 所示,具体步骤如下:

- 1) 采集滚动轴承全寿命试验台中轴承的振动信号,对 3 种传感器信号进行归一化预处理;
- 2) 利用自编码器提取振动信号特征,深度挖掘隐藏特征,减少手动特征提取中主观因素的影响;
- 3) 将多维特征输入到 STA 模型中,通过卷积层将空间特征维度信息和注意力加权的时间步长信息合并,对特征在时间和空间维度上进一步优化;
- 4) 对多维特征数据进行切片,再输入到 BiLSTM 模型进行剩余寿命预测。

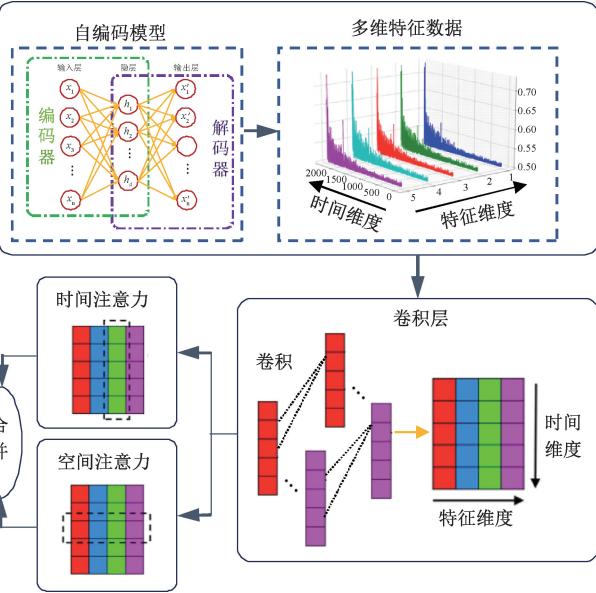


图 6 所提出方法的流程图

Fig. 6 Flow diagram of the proposed method

之间误差大小的一种指标。在机器学习和数据分析中, RMSE 通常用于评估模型的精度和预测准确性。

RMSE 的计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (22)$$

其中, n 表示样本数量, y_i 表示第 i 个观测值, \hat{y}_i 表示第 i 个预测值。

与 MAE 相比, RMSE 更加敏感于较大的误差,因为它对误差进行了平方操作。因此,当预测值与真实值之间存在较大差异时, RMSE 的值会比 MAE 更大,从而更加凸显出误差的大小。

3) R^2 分数

R^2 分数是一种用于评估回归模型性能的指标,用于衡

量模型对目标变量的拟合程度。其计算公式如下：

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (23)$$

其中, y 表示真实值, \hat{y} 表示预测值, \bar{y} 表示均值。 R^2 的值为 1, 表示模型完美拟合数据, 能够解释目标变量的所有方差; R^2 的值为 0, 表示模型的拟合效果较差, 无法解释目标变量的任何方差; R^2 的值为负数, 表示模型的拟合效果非常差。

通过比较 R^2 分数的取值, 可以评估模型的拟合效果, 了解模型对数据的解释方差程度。

3 试验研究与分析

3.1 IEEE 2012 PHM 预测挑战数据集

1) 数据集介绍

采用 PHM2012 挑战赛的滚动轴承数据集, 该数据集由 PRONOSTIA 加速寿命试验平台通过实验获得, 试验平台如图 7 所示。PHM2012 数据集是一个在健康管理与预测性维护领域中被广泛应用的时间序列数据集, 它可以用于进行轴承故障振动和剩余寿命预测等领域。

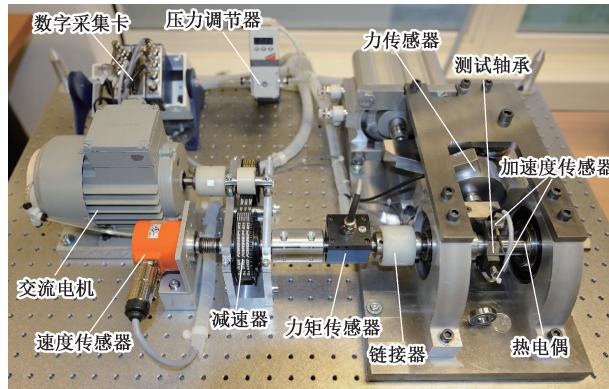


图 7 PRONOSTIA 平台

Fig. 7 PRONOSTIA platform

在实际工作过程中, 滚动轴承拥有很长的使用寿命, 为了能够更快速的了解滚动轴承的寿命情况, 该平台采用了加速轴承的使用寿命的方法。通过在实验过程中增加转速及载荷从而加速其全寿命过程。在被测试的轴承附近分别按照了横向和纵向两个方向的振动加速度传感器, 每隔 10 s 对两个方向上的振动信号进行采集, 每次采集 0.1 s 的数据, 即记录 2 560 个样本。数据集在工况 1 下分为训练集和测试集, 如表 1 所示, 训练数据集为 Bearing1_1 和 Bearing1_2, 测试数据集为 Bearing1_3~Bearing1_7。

2) 实验环境

实验选用了 TensorFlow 和 Keras 作为深度学习框架来训练和测试网络模型, 使用 Python 作为开发语言, 具体实验环境如表 2 所示。

表 1 工况信息和数据集划分

Table 1 Work condition information and data set division

工况	转速/N	负载/(r·min⁻¹)	训练数据	测试数据
工况 1	4 000	1 800	Bearing1_1 Bearing1_2	Bearing1_3 Bearing1_4 Bearing1_5 Bearing1_6 Bearing1_7
				Bearing1_1 Bearing1_2
				Bearing1_5 Bearing1_6 Bearing1_7

表 2 具体实验环境

Table 2 Specific experimental environment

编号	项目	具体配置信息
1	CPU	Intel i7-13620H
2	GPU	RTX 4050
3	内存	16 GB DDR5 5200 MHz
4	操作系统	Windows 11
5	开发语言	Python
6	开发环境	Anaconda、PyCharm
7	深度学习框架	TensorFlow、Keras

3) 网络模型

本次实验所用网络模型的主要参数如表 3 所示。

表 3 AE-STA-BiLSTM 模型参数

Table 3 AE-STA-BiLSTM model parameters

网络	参数	数值	描述
AE	Train_epochs	50	训练最大迭代次数
	Learning_rate	0.001	学习率
	Size	1000	输入尺寸
	Batch_size	256	输入一批训练数据的大小
	Hidden layer neurons (two layers)	500/125	编码器隐藏层单元数
	Bottleneck layer neurons	5	编码器输出层单元数
STA	Train_epochs	200	训练最大迭代次数
	Batch_size	20	输入一批训练数据的大小
	Learning_rate	0.003	学习率
	Kernel_size	2	卷积核大小
	Kernel_num	64	卷积核个数
	Train_epochs	50	训练最大迭代次数
	Learning_rate	0.001	学习率
BiLSTM	Batch_size	60	输入一批训练数据的大小
	Hidden layer neurons	30	隐藏层单元数

4) 寿命预测结果分析

本次实验使用工况 1 下的 Bearing1_1 和 Bearing1_2 作为训练集来训练模型, 并且对工况 1 下的所有测试集 Bearing1_3~Bearing1_7 进行预测, 并对其结果进行评估, 评估结果如表 4 所示。

表 4 所提出模型的预测误差

Table 4 The prediction error of the proposed model

轴承	RMSE	MAE	R^2
Bearing1_3	0.089 1	0.064 9	0.927 9
Bearing1_4	0.061 5	0.036 9	0.950 5
Bearing1_5	0.067 4	0.036 6	0.412 5
Bearing1_6	0.301 2	0.251 2	-0.073 9
Bearing1_7	0.041 7	0.021 1	0.789 4
Average	0.112 1	0.082 1	0.601 3

5) 模型参数分析

在特征数据输入到 BiLSTM 预测模型之前, 会对数据进行切片处理, 即将特征数据序列分成不同长度的子序列或窗口, 以便对数据进行分析、处理, 通常不同的窗口长度会对预测的效果有不同的影响。一般较短的窗口能提供更高的时间或空间分辨率, 捕捉到更多的细节, 但可能在数据中引入更多的噪音; 较长的窗口则能提供更广泛的视图, 会平滑数据, 减少噪音, 可能有助于捕捉长期趋势, 对时间序列预测模型可能更有效。

为了选择合适的窗口长度, 在实验过程中依次设定窗口长度为 10、20、30、40、50 对模型进行训练, 并对测试集的 RMSE 值进行记录和比较, 得出不同窗口长度下测试集 RMSE 值的分布情况如图 8 所示。可以看出当窗口长度为 30 时, RMSE 值最大, 误差也就越大。当接着增大窗口长度时, RMSE 值开始出现减小的趋势, 在窗口长度为 50 时, RMSE 值达到最小, 预测的误差最小。所以, 本次实验选取窗口长度为 50 对数据进行切片, 所达到的预测效果也为最佳。

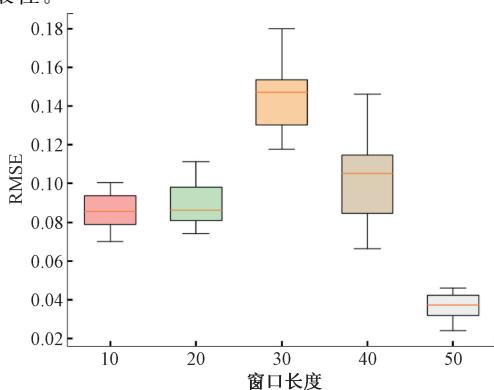


图 8 不同窗口长度下 RMSE 值的分布

Fig. 8 Distribution of RMSE values for different window lengths

6) PHM 数据集消融实验研究

消融实验用于确定特定因素对研究对象的影响, 这些因素可以是变量、条件或处理。本次实验为了探究 STA 模型是否对预测效果有明显的提升, 选取了 3 种对比模型, 分别为 AE+LSTM(M1)、AE+BiLSTM(M2)、AE+STA+LSTM(M3)。在实验中, 选取 Bearing1_1 和 Bearing1_2 作为训练集, 选取 Bearing1_3~Bearing1_7 作为测试集, 分别用上述 4 种模型进行 RUL 预测。Bearing1_5 的预测曲线对比如图 9(a)所示。从预测曲线可以看出, AE+LSTM 模型组合的预测效果最差, AE+BiLSTM 模型组合的预测效果有所提升, BiLSTM 模型相较于 LSTM 模型有更好的预测能力。当在 AE+LSTM 模型中添加时间空间注意力即 STA 模型后, 其预测的准确度比 AE+LSTM 模型、AE+BiLSTM 模型更高, 可见 STA 模型中对预测效果提升的影响因子作用明显。从最后的结果图可以看出, 所提出的 AE+STA+BiLSTM 模型组合的拟合程度最高, 该模型组合结合了 BiLSTM 模型对长序列数据的预测优势和 STA 模型中对时间和空间信息注意力提升的特点。从消融实验的预测曲线图可以看出 STA 模型对预测效果有明显的提升作用, 可广泛应用于预测任务中。

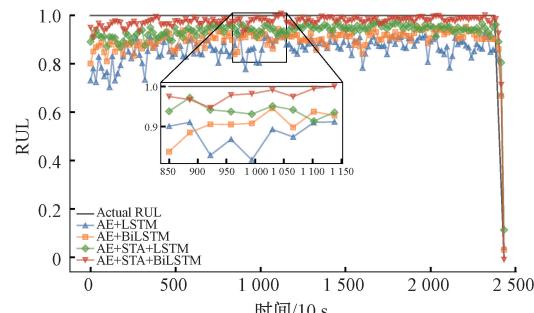


图 9 Bearing1_5 的剩余寿命预测结果

Fig. 9 RUL prediction results of Bearing1_5

在消融实验中, 通过利用 4 种模型组合对轴承数据工况 1 下的所有测试集进行预测, 并使用 RMSE、MAE、 R^2 三种评价指标来评估模型, 得出的结果如表 5 所示。从结果中可以看出, 所提模型 AE+STA+BiLSTM 组合经过所有测试集的预测, 平均误差达到最小, 可见加入 STA 模型后明显提升了 RUL 预测能力。

7) 与已有算法的比较分析

为了进一步说明所提的模型的优越性, 分别与已有的文献 [13] 提出的 CPSS-DCN 算法、文献 [14] 提出的 KPCA-GRU 算法、文献 [15] 提出的 TFR-MSCNN 算法、文献 [16] 提出的 RNN-AM 算法、文献 [17] 提出的 SE-DSCN 算法进行对比分析, 结果如表 6 所示。从表 6 中可以看出所提出的模型的 RMSE 和 MAE 更小, 预测效果更好。

8) STA 模块分析

图 10 给出了轴承 1_5 的特征维度和时间维度的权重

表 5 PHM 数据集消融实验结果

Table 5 Ablation study results of the PHM dataset

模型	评价方法	Bearing1_3	Bearing1_4	Bearing1_5	Bearing1_6	Bearing1_7	Average
M1	RMSE	0.175 4	0.078 5	0.128 2	0.428 6	0.075 3	0.177 2
	MAE	0.152 7	0.048 9	0.118 3	0.395 5	0.069 7	0.157 0
	R ²	0.720 7	0.919 4	-1.125 6	-1.174 9	0.315 5	-0.068 9
M2	RMSE	0.135 0	0.075 0	0.095 1	0.351 7	0.065 1	0.144 3
	MAE	0.111 5	0.053 2	0.083 0	0.322 3	0.049 4	0.123 8
	R ²	0.834 5	0.926 3	-0.168 8	-0.464 3	0.488 4	0.323 2
M3	RMSE	0.101 0	0.069 1	0.079 7	0.339 6	0.049 4	0.127 7
	MAE	0.065 6	0.053 7	0.065 3	0.304 6	0.045 6	0.106 9
	R ²	0.907 4	0.937 4	0.179 2	-0.365 1	0.705 6	0.472 9
所提出的方法	RMSE	0.089 1	0.061 5	0.067 4	0.301 2	0.041 7	0.112 1
	MAE	0.064 9	0.036 9	0.036 6	0.251 2	0.021 1	0.082 1
	R ²	0.927 9	0.950 5	0.412 5	-0.073 9	0.789 4	0.601 3

表 6 与已有算法对比实验结果

Table 6 Comparison results with existing algorithms

算法	RMSE	MAE
CPSS-DCN ^[13]	0.338 5	0.271 6
KPCA-GRU ^[14]	0.265 7	0.205 1
TFR-MSCNN ^[15]	0.198 7	0.154 2
RNN-AM ^[16]	0.129 1	0.093 8
SE-DSCN ^[17]	0.147 5	0.128 3
所提出的模型	0.112 1	0.082 1

分布。可以看到,经过 STA 处理后,特征根据其在特征维度和时间维度上的重要性获得相应的权重系数,这些权重系数在图中通过颜色强度的来表示。

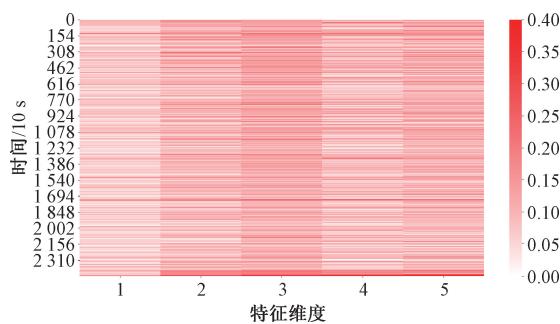


图 10 轴承 1_5 的 STA 权重分布

Fig. 10 Weight distribution of the STA for Bearing 1_5

3.2 ABLT-1A 轴承全寿命周期数据集

1) 实验台介绍

为了进一步验证所提出算法的有效性,在 ABLT-1A 轴承寿命强化试验台上进行了电机轴承的全寿命周期测试,实验平台如图 11 所示。轴承寿命强化测试台包括测试主轴、传动系统、加载系统、润滑系统、电气控制系统和计算机控制系统等各个组件。轴承在测试台上的旋转由

传动系统驱动,实现对 4 个轴承的同时全寿命周期测试。

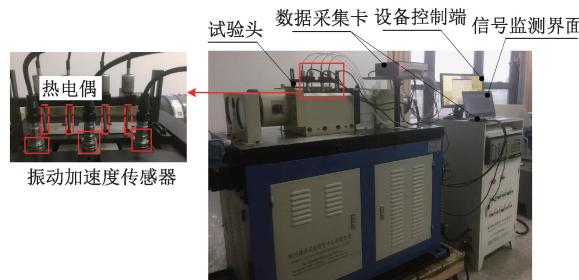


图 11 轴承强化疲劳实验台

Fig. 11 Bearing strengthening fatigue test bench

2) 数据集介绍

本实验选用了 6008 型单列深沟球轴承作为轴承模型进行实验。全寿命周期试验的具体实验条件如表 7 所示。轴承 A 和轴承 B 分别运行了 765 min 和 1 018 min。图 12(a)和(b)分别表示两组滚动轴承全寿命周期试验中收集到的原始振动信号。通过比较 Bearing_A 和 Bearing_B 的原始振动信号,两种轴承在正常运行阶段的振动较为平稳,振动幅值开始增加到最终失效的时间很短,这也表明这是一种快速退化模式。

表 7 试验使用轴承数据集的详细信息

Table 7 Details of the test using the bearing data set

轴承编号	轴承类型	负载/N	转速/rpm	样本时间/s	采样频率/Hz	保存间隔/s
Bearing_A	6008	5 000	7 000	2 282	25 600	10
Bearing_B	6008	5 000	7 000	4 594	25 600	10

3) ABLT-1A 数据集消融实验

本实验的实验环境和使用的网络模型与 PHM 数据集消融实验中的一致,如表 2 和 3 所示。Bearing_A 被用作

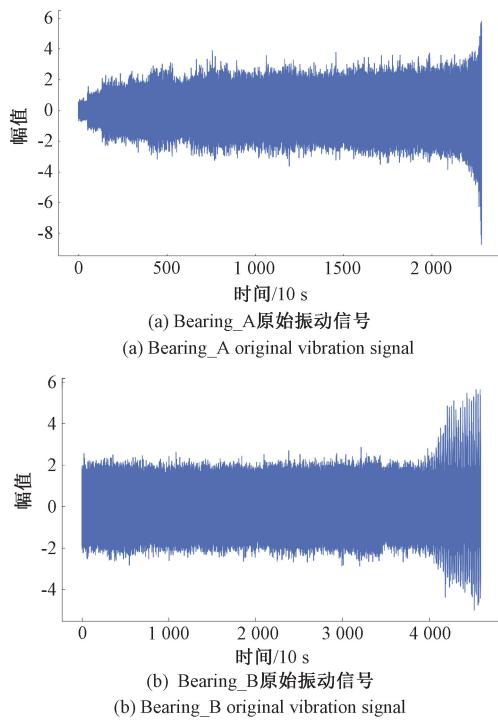


图 12 训练集和测试集的原始振动信号

Fig. 12 Raw vibration signals of training set and test set

训练集,而 Bearing_B 被用作测试集,以验证本文提出的算法的有效性。分别使用 AE+LSTM(M1)、AE+BiLSTM(M2)、AE+STA+LSTM(M3)、所提出的模型进行预测,图 13 是 Bearing_B 的 RUL 预测结果。采用 RMSE、MAE、 R^2 三种评价指标来评估模型的性能,实验结果如表 8 所示。从消融实验结果可以看出,本文所提出的算法 AE+STA+BiLSTM 具有良好的预测性能,能够为滚动轴承剩余寿命预测方面的工作提供有力的参考。

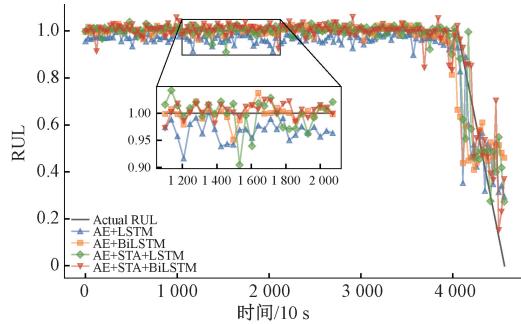


图 13 Bearing_B 预测曲线对比

Fig. 13 Comparison of prediction curves for Bearing_B

表 8 ABLT-1A 数据集消融实验结果

Table 8 ABLT-1A data set ablation results

模型	RMSE	MAE	R^2
M1	0.097 3	0.041 8	0.724 2
M2	0.090 3	0.051 5	0.762 5
M3	0.080 8	0.035 4	0.809 7
所提出的方法	0.069 1	0.031 5	0.860 9

4 结 论

本文提出了一种基于 STA 机制的 AE-BiLSTM 的滚动轴承 RUL 预测方法。首先,为了解决传统的基于振动信号的滚动轴承 RUL 预测方法中依赖专家经验的问题,利用 AE 模型对振动信号进行特征提取,进一步将提取的特征输入到 STA 模型中以实现特征数据空间和时间信息的加权融合,再结合 BiLSTM 预测模型实现对 RUL 的预测。最后,采用 PHM2012 挑战赛的滚动轴承全寿命数据集和 ABLT-1A 轴承全寿命周期数据集进行实验论证,结果表明:

通过 AE 模型提取的多维特征信息输入到空间时间机制中对特征信息的进一步优化,使得多维特征的时间步长信息和维度空间信息进行深度融合,再输入到 BiLSTM 模型中进行 RUL 的预测,能够精确捕捉到特征数据的核心信息,提高了 RUL 预测的准确度。

通过 PHM2012 Challenge 和 ABLT-1A 跨不同轴承类型和工况的全寿命周期试验,对所提出模型的有效性和泛化性进行了验证。从 ABLT-1A 实验结果中可以看出所提出的模型的 RMSE 平均降低了约 22.76%,MAE 平均降低了约 26.57%,而 R^2 平均提升了约 12.47%,表明所提出方法对 RUL 预测效果有明显的提升,为工业设备的状态监测和预测性维护提供了可靠的支持。

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.
- [2] PEI H, HU CH H, SI X SH, et al. Review of machine learning based remaining useful life prediction methods for equipment [J]. Journal of Mechanical Engineering, 2019, 55(8): 1-13.
- [3] 陈曦,王辉,陆思良,等. ALC-PFL: 基于个性化联邦学习的轴承寿命预测方法[J]. 仪器仪表学报, 2023, 44(11): 1-10.
- [4] CHEN X, WANG H, LU S L, et al. ALC-PFL: Bearing remaining useful life prediction method based on personalized federated learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 1-10.
- [5] 袁倩,孙冬梅,范文. 基于 D-S 信息融合技术的磁轴承转子故障诊断方法[J]. 仪表技术与传感器, 2016,(11): 118-122.
- [6] YUAN Q, SUN D M, FAN W. Fault diagnosis method of magnetic rotor based on D-S information fusion technology [J]. Instrument Technique and Sensor, 2016,(11): 118-122.
- [7] 冯兆熙,邱度金,孔令驹,等. 基于深度置信网络的轴承剩余使用寿命预测[J]. 电子测量与仪器学报,

- 2021, 35(10): 124-129.
- FENG ZH X, QIU D J, KONG L J, et al. Remaining useful life prediction of bearing based on deep belief network[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 124-129.
- [5] 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.
- YANG X D, JI G Y. Remaining useful lifetime prediction method of rolling bearing based on fusion-CNN [J]. Foreign Electronic Measurement Technology, 2021, 40(3): 62-67.
- [6] 兰杰, 李志宁, 李宁, 等. 基于注意力机制和残差深度分离卷积的 RUL 预测方法[J]. 电子测量技术, 2023, 46(15): 149-157.
- LAN J, LI ZH N, LI N, et al. Prediction method of remaining useful life based on attention mechanism and residual depthwise separation convolution [J]. Electronic Measurement Technology, 2023, 46(15): 149-157.
- [7] CAO Y D, DING Y F, JIA M P, et al. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings [J]. Reliability Engineering & System Safety, 2021, 215: 107813.
- [8] HU CH H, PEI H, SI X SH, et al. A prognostic model based on DBN and diffusion process for degrading bearing [J]. IEEE Transactions on Industrial Electronics, 2019, 67(10): 8767-8777.
- [9] LIU H, LIU ZH Y, JIA W Q, et al. Remaining useful life prediction using a novel feature-attention-based end-to-end approach[J]. IEEE Transactions on Industrial Informatics, 2020, 17(2): 1197-1207.
- [10] 付孟丹, 宣士斌, 王婷, 等. 基于时间和空间注意力机制的视频异常检测[J]. 计算机技术与发展, 2023, 33(8): 51-58.
- FU M D, XUAN SH B, WANG T, et al. Video anomaly detection based on temporal and spatial attention mechanism[J]. Computer Technology and Development, 2023, 33(8): 51-58.
- [11] 张志浩, 王坤侠. 基于 STA-CRNN 模型的语声情感识别[J]. 应用声学, 2022, 41(5): 843-850.
- ZHANG ZH H, WANG K X. Speech emotion recognition based on STA-CRNN model[J]. Applied Acoustics, 2022, 41(5): 843-850.
- [12] 李大湘, 陈梦思, 刘颖. 基于 STA-LSTM 的自发微表情识别算法[J]. 吉林大学学报(工学版), 2022, 52(4): 897-909.
- LI D X, CHEN M S, LIU Y. Spontaneous micro-expression recognition based on STA-LSTM [J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(4): 897-909.
- [13] REN L, SUN Y Q, WANG H, et al. Prediction of bearing remaining useful life with deep convolution neural network [J]. IEEE Access, 2018, 6: 13041-13049.
- [14] CHEN J L, JING H J, CHANG Y H, et al. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process[J]. Reliability Engineering & System Safety, 2019, 185: 372-382.
- [15] ZHU J, CHEN N, PENG W W. Estimation of bearing remaining useful life based on multiscale convolutional neural network[J]. IEEE Transactions on Industrial Electronics, 2018, 66(4): 3208-3216.
- [16] CHEN Y H, PENG G L, ZHU ZH Y, et al. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction [J]. Applied Soft Computing, 2020, 86: 105919.
- [17] WANG B, LEI Y G, LI N P, et al. Deep separable convolutional network for remaining useful life prediction of machinery[J]. Mechanical Systems and Signal Processing, 2019, 134: 106330.

作者简介

施彤,本科,主要研究方向为滚动轴承剩余寿命预测。

E-mail: 202213930045@nuist.edu.cn

张自豪,本科,主要研究方向为机电系统状态监测与寿命预测。

E-mail: 202213930045@nuist.edu.cn

邱晓慧(通信作者),硕士,主要研究方向为机器健康管理与智能运维。

张莞,博士,副教授,主要研究方向为机电系统状态监测、剩余寿命预测和智能运维。