

文章编号: 1671-7449(2026)04-0522-09

J. Test Measur. Technol., 2026, 40(4): 522-530.

基于多域特征融合与Transformer-BiLSTM的 轴承寿命预测方法

周彦龙, 黄晋英*, 牛亚旭, 范振芳, 白 雄

(中北大学 机械工程学院, 山西 太原 030051)

摘要: 针对轴承剩余使用寿命(Remaining Useful Life, RUL)预测中单一特征难以全面表征其退化状态的问题, 提出一种结合多域特征融合与Transformer-双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)并联网络的预测方法。首先, 从振动信号中提取时域、频域、时频域和熵值特征, 采用基于移动平均分解的综合评价指标筛选敏感特征, 再利用主成分分析降维构建综合健康指标。随后, 设计并联网络结构, 利用Transformer编码器捕捉全局长期依赖, 通过BiLSTM学习局部双向时序特征, 并引入注意力机制自适应融合双分支输出。在PHM2012轴承数据集上的实验结果表明, 该方法在全寿命周期预测中的平均绝对误差和均方根误差与其它模型(BiLSTM、CNN等)相比, 平均降低36.1%和29.8%, 决定系数(R^2)达到0.95, 全寿命周期预测误差低于10%。跨工况泛化实验进一步验证了该预测方法的有效性和稳定性。

关键词: 剩余寿命预测; 多域特征融合; 主成分分析; Transformer; BiLSTM; 注意力机制

中图分类号: TH133.3

文献标识码: A

DOI: 10.62756/csjs.1671-7449.2026050

引用格式: 周彦龙, 黄晋英, 牛亚旭, 等. 基于多域特征融合与Transformer-BiLSTM的轴承寿命预测方法[J]. 测试技术学报, 2026, 40(4): 522-530.

Zhou Yanlong, Huang Jinying, Niu Yaxu, et al. A method for predicting bearing life based on multi-domain feature fusion and Transformer-BiLSTM[J]. Journal of Test and Measurement Technology, 2026, 40(4): 522-530.

A Method for Predicting Bearing Life Based on Multi-Domain Feature Fusion and Transformer-BiLSTM

Zhou Yanlong, Huang Jinying*, Niu Yaxu, Fan Zhenfang, Bai Xiong

(School of Mechanical Engineering, North University of China, Taiyuan 030051, China)

Abstract: To address the limitation of single-domain features in comprehensively characterizing the degradation state for bearing remaining useful life (RUL) prediction, a method integrated multi-domain feature fusion and a parallel Transformer-bidirectional long short-term memory (BiLSTM) network are proposed. Firstly, time-domain, frequency-domain, time-frequency and entropy features are extracted from vibration signals. A comprehensive evaluation index based on moving average decomposition is used to

【收稿日期】2025-09-06

【修回日期】2025-12-15

【录用日期】2025-12-30

【基金项目】山西省研究生科研创新资助项目(2024KY574); 山西省回国留学人员科研资助项目(2023-133); 山西省基础研究计划资助项目(202303021212194)

【作者简介】周彦龙(2000—), 男, 硕士生, 主要从事智能故障诊断与剩余寿命预测研究。E-mail: zoyanl@163.com。

*【通信作者】黄晋英(1971—), 女, 教授, 博士, 主要从事结构动力学分析、检测与控制、故障诊断与健康管理研究。E-mail: jyhuang@nuc.edu.cn。

select sensitive features, which are then reduced in dimensionality via principal component analysis to construct a composite health indicator. Subsequently, a parallel network architecture is designed. Global long-range dependencies are captured by a transformer encoder, while local bidirectional temporal patterns are learned via a BiLSTM network. An attention mechanism is introduced to adaptively fuse the outputs of the two branches. Experimental results on the PHM2012 bearing dataset show that the proposed method reduces the mean absolute error and root mean square error by 36.1% and 29.8% on average, respectively, compared to benchmark models (e.g., BiLSTM, CNN), and achieves a coefficient of determination (R^2) of 0.95 in full-lifecycle prediction. The prediction error for the entire life cycle is below 10%. The effectiveness and stability of the method are further validated by cross-condition generalization experiments.

Key words: remaining useful life prediction; multi-domain feature fusion; principal component analysis; Transformer; BiLSTM; attention mechanism

0 引言

轴承是旋转机械中的关键部件,其运行状态直接影响到整机的可靠性与安全性。在过载、冲击、疲劳和磨损等多种因素作用下,轴承会发生性能退化,甚至引发突发故障^[1]。因此,准确预测其剩余使用寿命(Remaining Useful Life, RUL),并据此制定视情维修策略,对避免重大经济损失与安全事故具有重要的工程意义。

目前,RUL预测方法主要可分为物理模型驱动与数据驱动两大类^[2-3]。物理模型方法是基于失效机理建立数学方程,物理意义明确,但在复杂系统中面临参数众多、建模困难等问题。随着传感技术的发展,数据驱动方法能够从历史监测数据中自动挖掘退化规律,已成为当前研究热点。其中,深度学习模型,凭借其强大的特征提取与序列建模能力,在RUL预测领域展现出显著优势。

近年来,众多学者在此研究方向进行了深入探索^[4]。例如,张氢等^[5]采用时域特征结合长短期记忆网络(Long Short-Term Memory, LSTM)进行RUL预测;万忠义等^[6]构建了融合时域和特征域的多维时序特征融合模型;卢瑾等^[7]通过频域特征与注意力机制增强关键信息提取;蒋全胜等^[8]利用小波包分解构建敏感特征集;苏浩田^[9]和杨泽锐等^[10]分别采用主成分分析(Principal Component Analysis, PCA)和相关算法进行特征融合与降维。此外,Chen等^[11]采用编码器-解码器结构建立健康指标并预测RUL;Ding等^[12]将卷积模块引入Transformer以结合局部与全局特征;惠憬明等^[13]则将自注意力机制与BiLSTM结合。尽管上述方法取得了良好预测效

果,但仍存在两点局限:特征表征多局限于单一域(时域、频域或变换域),缺乏对多尺度退化信息的全面捕捉;预测模型多为单一结构(如LSTM、CNN或Transformer),难以同时建模长序列中的全局退化趋势与局部动态波动。

针对上述问题,本文提出一种双分支的Transformer-BiLSTM轴承RUL预测方法,创新性地融合多域特征(时域、频域、时频域和熵域)并设计并联网络结构。本文还提出一种基于移动平均分解的综合评价指标,用于筛选敏感性强、鲁棒性高的多域特征子集,并利用PCA构建能全面表征退化状态的健康指标(Health Indicator, HI)。该预测方法结合Transformer与BiLSTM并联的双分支网络结构,分别捕捉退化过程的全局长期依赖与局部双向时序模式,并通过注意力机制实现双分支输出的自适应融合,以全面提升预测精度与泛化能力。

1 理论基础

本节将概述本文方法所涉及的核心理论,包括用于特征筛选的移动平均分解与评价指标、用于特征降维的PCA算法,以及用于构建预测模型的双向长短期记忆网络(BiLSTM)、Transformer编码器和自适应注意力机制。

1.1 基于移动平均分解的特征筛选

移动平均法将特征序列分解为趋势项与残差项两项,趋势项反映退化演变规律,残差项则表示随机波动与噪声,通过分解结果可用于评估特征的关键性能,以剔除低质量特征^[14]。分解方法

如下：设某特征的时间序列为 $X = \{x_t | t = 1, 2, \dots, N\}$ ，将 X 分解为平稳趋势序列 $T = \{T_t | t = 1, 2, \dots, N\}$ 与随机余量序列 $R = \{R_t | t = 1, 2, \dots, N\}$ ，其满足 $x_t = T_t + R_t$ 。其中，平稳趋势序列 T 为

$$T = t_t = \frac{1}{w} \sum_{i=t-\lfloor w/2 \rfloor}^{t+\lfloor w/2 \rfloor} x_i, \quad (1)$$

式中： x_t 为该特征在第 t 时刻的取值； N 为时间序列的总长度； w 为移动窗口大小。

定义时间相关性 (C_{corr})、预测性 (C_{pred}) 和鲁棒性 (C_{rob}) 三项指标用于优选特征的筛选。其中， C_{corr} 量化特征序列与时间的关联性，反映特征趋势的单调性； C_{pred} 评价特征对未来状态的指示能力，基于趋势序列的首尾差异与波动程度； C_{rob} 反映特征序列的抗干扰 (抗噪声) 能力，波动越小，说明特征受噪声影响越小。各指标的计算公式为

$$C_{\text{corr}} = \frac{|\rho(X, t)| - \min(|\rho(X, t)|)}{\max(|\rho(X, t)|) - \min(|\rho(X, t)|)}, \quad (2)$$

$$C_{\text{pred}} = \frac{|T_N - T_1|}{\sigma_T}, \quad (3)$$

$$C_{\text{rob}} = \frac{1}{1 + \frac{\sigma_R}{\sigma_T}}, \quad (4)$$

式中： $\rho(X, t)$ 为 X 与 t 的秩相关系数； σ_T 为平稳趋势序列 T 的标准差； T_N 、 T_1 分别为 T 的末尾与首端取值； σ_R 为随机余量序列 R 的标准差。

1.2 PCA 算法

通过筛选得到的特征子集维度仍然较高且特征间可能存在多重共线性。采用 PCA 进行特征压缩消除特征冗余并降低维度，通过正交变换将高维特征映射至低维空间，在保留绝大部分方差的同时去除相关性。对于一个 $m \times n$ 维矩阵 X_{raw} ，进行 PCA 的具体步骤如下：

步骤一：数据标准化对特征子集按特征进行标准化，得到标准化矩阵 $X_{\text{std}} \in \mathbb{R}^{N \times m}$ 为

$$x_{ij}^{\text{std}} = \frac{x_{ij} - u_j}{\sigma_j}, \quad (5)$$

式中： u_j 和 σ_j 分别为第 j 维特征的均值与标准差。

步骤二：协方差矩阵与特征分解，计算协方差矩阵 $\Sigma \in \mathbb{R}^{m \times m}$ 为

$$\Sigma = \frac{1}{N-1} X_{\text{std}}^T X_{\text{std}} \quad (6)$$

步骤三：对 Σ 进行特征分解，得到特征值

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ 及对应的单位特征向量 $\Sigma u_j = \lambda_j u_j$ ($j = 1, 2, \dots, m$)。

步骤四：根据累计方差贡献率确定主成分数量

$$\eta(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^m \lambda_j} \times 100\%, \quad (7)$$

式中： $\eta(k)$ 为设定阈值 (本文取 95%)，取满足条件的最小 k 。

步骤五：将标准化后的数据投影前 k 个主成分方向，得到降维后的特征矩阵 $U = [u_1, \dots, u_k] \in \mathbb{R}^{m \times k}$ ，将数据映射到低维空间 Y ，

$$Y = X_{\text{std}} \times U. \quad (8)$$

1.3 BiLSTM 模型

LSTM 是循环神经网络 (Recurrent Neural Network, RNN) 的一种变体，通过引入门控机制 (输入门、遗忘门和输出门)，有效解决了长序列训练中的梯度消失和爆炸问题，结构图如图 1 所示，其单个单元在时刻 t 的计算过程如下：

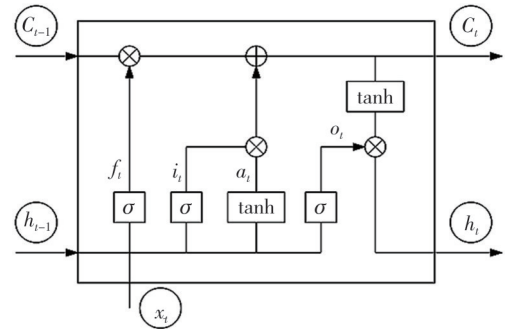


图 1 LSTM 结构图

Fig. 1 LSTM structure diagram

输入门

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (9)$$

遗忘门

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (10)$$

细胞状态

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (11)$$

输出门

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (12)$$

隐状态输出

$$h_t = o_t \odot \tanh(c_t) \text{ 输出门}, \quad (13)$$

式中： x_t 为当前时刻 LSTM 的输入向量； h_{t-1} 为前一时刻的隐状态； C_{t-1} 为前一时刻的记忆状态； W_i 、 W_f 、 W_c 、 W_o 分别为输入门、遗忘门、细胞状

态、输出门的权重矩阵; b_i, b_f, b_c, b_o 为对应的偏置向量; σ 为 Sigmoid 函数; $\odot$ 表示逐元素乘法。

BiLSTM 则进一步引入前向与后向两个 LSTM 层, 可全面捕捉时间序列的双向依赖模式。BiLSTM 的结构如图 2 所示。

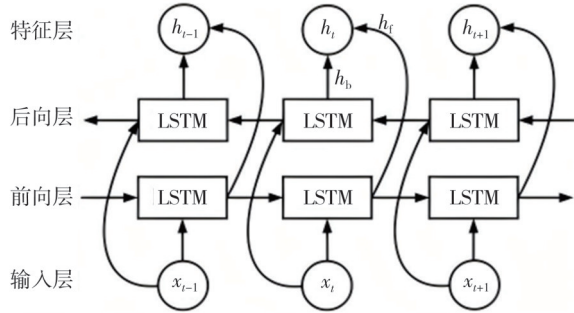


图 2 BiLSTM 网络示意图

Fig. 2 BiLSTM structure diagram

前向 LSTM 沿时间方向 ($t=1 \rightarrow t=T$) 处理序列, 输出隐状态 ($\vec{h}_t$); 后向 LSTM 沿逆时方向 ($t=T \rightarrow t=1$) 处理, 输出 ($\overleftarrow{h}_t$)。最终时刻 t 的输出由前后向隐状态拼接而成: ($\text{bilstm}_t = [\vec{h}_t; \overleftarrow{h}_t]$)。

1.4 Transformer 模型

Transformer 采用编码器-解码器 (Encoder-Decoder) 作为主要架构, 依赖注意力机制建立输入序列中元素之间的全局依赖关系, 从而实现对长期上下文的高效捕捉, 结构如图 3 所示。

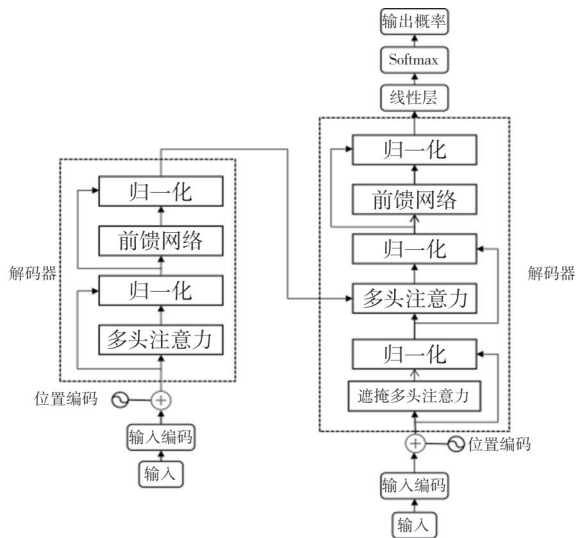


图 3 Transformer 网络结构图

Fig. 3 Transformer network structure diagram

在轴承 RUL 预测任务中, 主要使用其编码器部分, 提取具有全局信息的特征。编码器由多个结构相同的层堆叠而成, 每层包含一个多头自注意力机制和一个前馈神经网络, 并辅以残差连接和层归一

化促进训练稳定性和梯度流动。模型通过位置编码为输入序列注入位置信息, 弥补自注意力机制本身对序列顺序的不敏感性。此后, 将位置编码与输入特征向量按元素相加, 使模型能够感知序列顺序。

自注意力机制是 Transformer 的核心, 通过对序列内部元素间相关性进行加权聚合, 实现对关键信息的强化提取, 使得模型更加注重自身数据所含重要信息, 减少对外部信息的依赖。其计算过程表示为

$$\text{Attention}(Z) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{\text{key}}}}\right)V, \quad (14)$$

式中: $Q = ZW_q$, $K = ZW_k$, $V = ZW_v$, d_k 为键向量的维度, $\sqrt{d_{\text{key}}}$ 起到缩放作用以稳定梯度。

为进一步增强模型表征能力, 将自注意力扩展为多头机制, 计算公式为

$$\text{MultiHead}(Q, K, V) =$$

$$\text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_H) \cdot W^o, \quad (15)$$

式中: 每一个注意力头 head_i 负责在不同表示子空间中捕捉序列模式; H 为头数; W^o 为输出投影矩阵。这种设计使模型能够并行关注序列的不同方面, 显著提升对轴承退化多尺度动态特征的提取能力。最终, 编码器输出融合全局依赖关系的高层特征表示, 为后续 RUL 预测提供信息基础。

1.5 自适应注意力机制

自适应注意力在两个序列分支上计算自适应注意力权重, 通过一个轻量的参数化网络自动计算最优融合权重, 实现对局部特征和全局特征的加权整合。

$$e_i = v^T \cdot \sigma(W \cdot [Z_{\text{trans}}^i; Z_{\text{bilstm}}^i] + b), \quad (16)$$

$$H_{\text{final}}^i =$$

$$\sigma(e_i) \cdot Z_{\text{trans}}^i + (1 - \sigma(e_i)) \cdot Z_{\text{bilstm}}^i, \quad (17)$$

式中: W, b, v 为可学习的参数矩阵与向量。 σ 函数将得分 e_i 压缩至 $(0, 1)$ 区间, 产生最终权重 $\sigma(e_i)$ 。

2 预测方法与流程

2.1 多域特征提取

为全面评估轴承性能退化过程, 本文从时域、频域、时频域和熵域 4 个维度构建了 27 维特征集, 以捕捉不同方面的退化特性。特征集如表 1 所示。

时域特征直接从振动信号波形中提取, 该类特征计算高效且具有明确的物理意义。频域特征

通过快速傅里叶变换(FFT)获得,用于描述振动信号的频率分布特性。时频域特征采用小波包分解得到8个子频带能量比(F17-F24),反映信号能

量在时频联合域中的分布变化。熵域特征包括模糊熵、递归熵和样本熵(F25-F27),用于量化信号复杂度,对早期微弱退化特征敏感。

表 1 多域特征集

Tab. 1 Multi-domain feature set

特征类型	特征名称及编号
时域	均值F1、方根幅值F2、标准差F3、峰差F4、峰峰值F5、均方根值F6 偏度F7、峭度F8、峰值因子F9、振幅因子F10、波形因子F11、裕度因子F12
频域	平均频率F13、中心频率F14、频率均方根F15、频率方差F16
时频域	8个子频带能量比F17~F24
熵域	模糊熵F25、递归熵F26、样本熵F27

特征提取采用滑动窗口法进行:窗口长度为2 048,步长为1 024,并确保时序连续性。在进入模型前需要进行归一化处理,消除不同特征量纲差异,使特征具有相同的度量尺度,归一化计算公式为

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (18)$$

式中: x_{norm} 为归一化后的数据; x 为原始数据; $x_{\max}$ 和 $x_{\min}$ 分别为该组数据中的最大值和最小值。为便于模型训练,将轴承RUL归一化至[0,1]区间,其中1代表轴承处于全新状态,0代表完全失效。

2.2 特征筛选与健康指标构建

高维特征集中存在冗余且不敏感特征,若直接将其输入模型中,会降低模型效率和性能,因此需进行特征筛选与融合。理想的特征参数应具备良好的 C_{corr} 、 C_{pred} 和 C_{rod} 。本文采用1.1节中所述的3项指标综合评价各特征的性能,并通过交叉验证的方法筛选出加权综合评分(权重: C_{corr} 为0.35、 C_{pred} 为0.4、 C_{rod} 为0.25)排名前40%的特征。筛选出的特征构成了优化后的敏感特征子集,但该特征子集仍可能存在冗余或相关性问题的。为此,本文采用1.2节中所述的PCA方法对特征子集进行降维处理,保留累计方差贡献率 $\geq 95\%$ 的前 k 个主成分。将其作为综合HI,将高维特征空间映射为低维、线性无关,且包含绝大部分退化信息的健康指标序列,为后续预测模型提供高质量的输入数据。

2.3 预测流程

基于多域特征融合与Transformer-BiLSTM的轴承剩余寿命预测流程如图4所示,具体步骤如下:

步骤1:信号预处理。原始振动信号通常包含噪声与异常点,采用异常值修复与线性插值方法对其进行去噪处理,为后续特征提取提供高质量的数据基础。

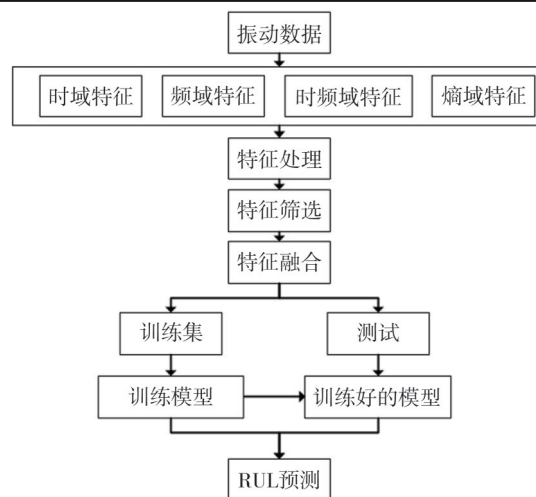


图 4 预测流程图

Fig. 4 Flowchart of prediction

步骤2:多域特征提取与平滑。对去噪后的信号按2.1节中所述方法提取27维多域特征,并使用Savitzky-Golay滤波器对特征序列进行平滑处理,进一步抑制噪声波动。

步骤3:特征筛选与融合。基于1.1节所述的 C_{corr} 、 C_{pred} 和 C_{rod} 指标对初始特征集进行综合评价与筛选,剔除低价值特征;再利用PCA算法对筛选后的特征进行融合与降维,构建低维主成分特征作为综合HI,并以此特征集作为模型的输入。

步骤4:模型训练。将所得的健康指标序列作为输入,寿命比值 P (剩余使用寿命与全寿命周期的比值)作为标签数据,采用均方误差(MSE)作为损失函数,并利用Adam优化算法对如图5所示的Transformer-BiLSTM并联网络进行训练与优化。该网络主要由Transformer编码器模块、BiLSTM模块及全连接模块构成。

步骤5:测试与预测。将测试集数据输入训练好的Transformer-BiLSTM模型,预测其对应的寿命比值 P ,从而实现轴承剩余寿命的准确预测。

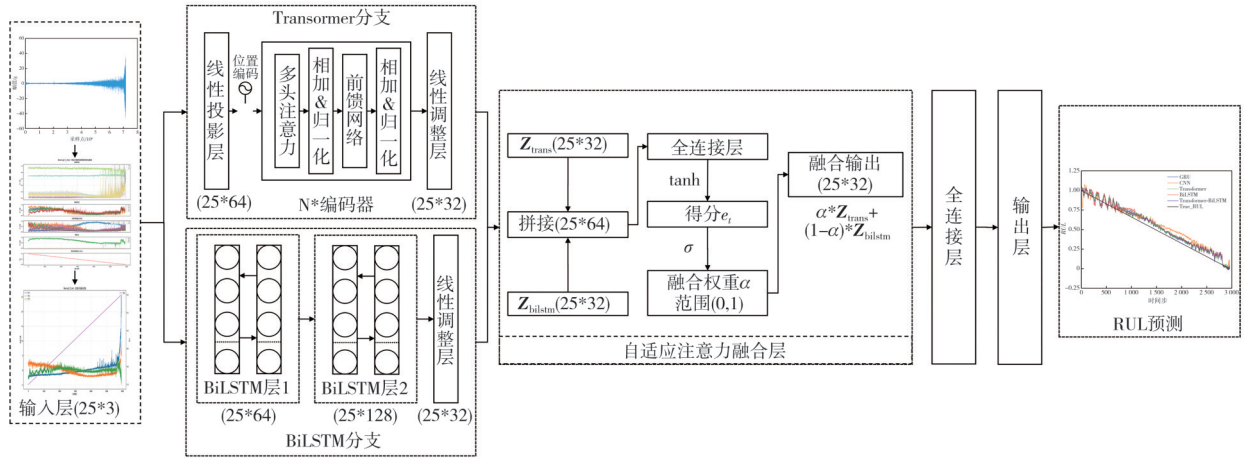


图 5 网络结构图

Fig. 5 Network structure diagram

3 实验验证与结果分析

3.1 数据集

实验采用PHM2012数据挑战赛提供的FEMTO-ST轴承全寿命振动数据集,试验台如图6所示^[15]。该数据集包含3种不同工况(转速与载荷)下共17个轴承的振动信号,记录了FEMTO-ST轴承从全新至完全失效的完整过程。具体工况参数见表2。采样频率为25.6 kHz,每10 s采集一次数据,每次采集0.1 s(即2 560个数据点)。鉴于水平方向振动信号通常对轴承径向退化更敏感^[16],本文后续分析均基于水平方向加速度数据。

表 2 实验数据工况详情

Tab. 2 Details of the experimental data conditions

工况	转速/(r·min ⁻¹)	载荷/N
1	1 800	4 000
2	1 650	4 200
3	1 500	5 000

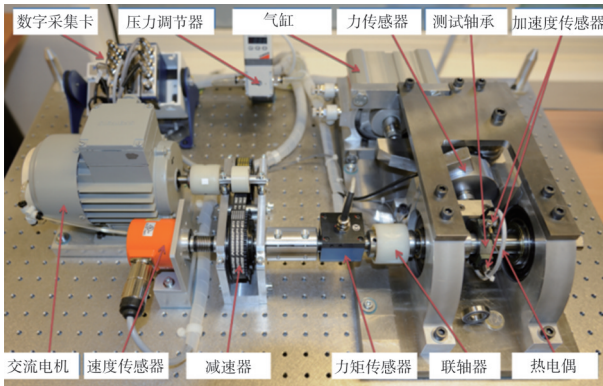


图 6 轴承寿命试验台

Fig. 6 Bearings life test bench

3.2 参数设置与评价指标

实验软件环境为TensorFlow2. 10, 运行该实验

环境的计算机配置为Windows 10 操作系统, Intel Core i7-10700KL CPU, 运行内存为16 GB。实验中设置批次大小为32, 训练轮数为100, 学习率为0.004, 损失函数为MSE, 使用Adam优化器进行优化。在实验划分方面, 选用工况1的轴承B1-1和B1-2作为训练集, 轴承B1-3作为测试集, 用于验证所提方法的性能。此外, 工况2的数据被用于后续的泛化能力检验。实验模型的超参数设置如表3所示。

表 3 超参数设置

Tab. 3 Hyper parameters

编号	网络层	参数说明
1	Transformer 编码层	Transformer头数2, 前馈神经元数量128, 编码器层数2
2	自适应平均池化层	输出大小1
3	BiLSTM 层1	神经元数量32, 双向, 层数1
4	BiLSTM 层2	神经元数量64, 双向, 层数1
5	特征拼接层	输入维度160
6	融合全连接层	神经元数量64, 激活函数ReLU
7	输出层	神经元数量1, 激活函数Linear

为全面客观地评估预测模型的性能, 本研究选用以下3种广泛使用的评价指标: 均方根误差(RMSE)、平均绝对误差(MAE)、决定系数(R^2), 计算公式如下

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (19)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (20)$$

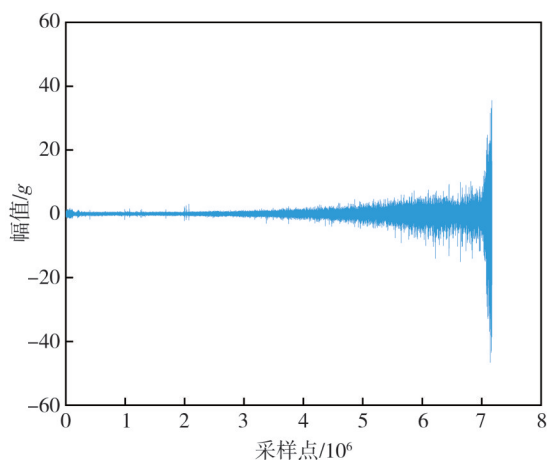
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (21)$$

式中: n 为样本数量; y_i 为样本的真实值; $\hat{y}_i$ 为样本的预测值; $\bar{y}$ 为所有样本真实值的均值。RMSE 和 MAE 越小或 R^2 接近 1 时, 模型性能越好。

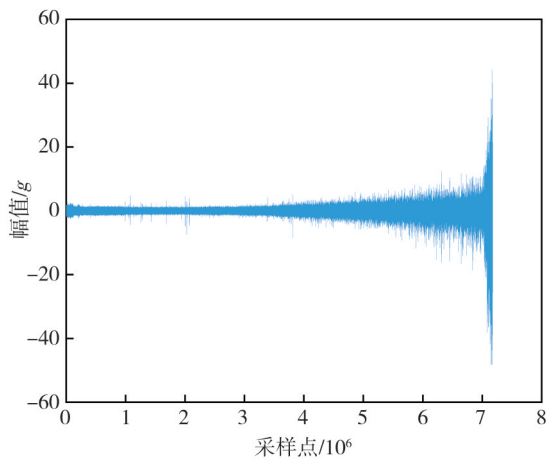
3.3 实验结果与分析

3.3.1 特征提取与 PCA 降维

原始振动信号不可避免地含有大量噪声(如图 7(a))。经过中位数绝对偏差异常值检测与线性插值处理后, 信号中的脉冲噪声被有效抑制, 趋势平滑性得到显著提升(见图 7(b)), 为后续特征提取奠定了可靠的数据基础。



(a) 原始振动信号



(b) 处理后的振动信号

图 7 原始数据和处理后的数据对比

Fig. 7 Comparison of original data and processed data

从处理后的信号中提取出的 27 维初始特征经过 Savitzky-Golay 滤波平滑后, 计算其综合评估得分(见图 8)。

依据综合评分排名, 选取排名前 40% 的特征(共 10 个)作为敏感特征子集, 其包括方根幅值 F2、标准差 F3 等时域特征, 也包括中心频率 F14

等频域特征。这些特征在单调性、预测性和抗噪性上均表现优异。该筛选比例(40%)所对应的评分阈值约为 0.654, 该阈值是通过在训练集上进行网格搜索, 以模型预测性能为导向优化确定, 可有效平衡特征质量与数量, 并避免信息冗余或丢失。随后, 对筛选出的特征进行 PCA 降维。降维效果通过方差贡献率来评估, 结果如图 9 所示。该图展示了各主成分的方差贡献率, 其中第一个主成分(PC1)的贡献率最高, 约为 65%, 这意味着该成分独自承载了原始特征集中约 65% 的退化信息。图 9 可以清晰地显示随着主成分数量的增加, 所保留信息总量的变化趋势。

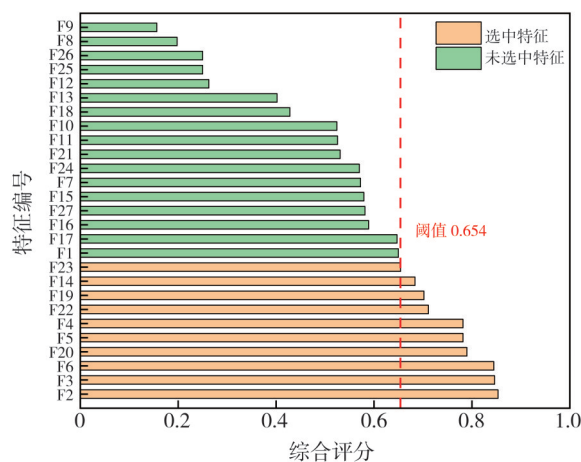


图 8 特征综合评分

Fig. 8 Composite feature score

如图 9 所示, 前 3 个主成分的累积方差贡献率已达 96.5%, 因此本文模型选择保留 3 个主成分, 将特征维度从 27 维降至 3 维, 构建为健康指标。该过程在极大降低数据复杂度和计算开销的同时, 最大程度地保留了表征性能退化的关键信息。

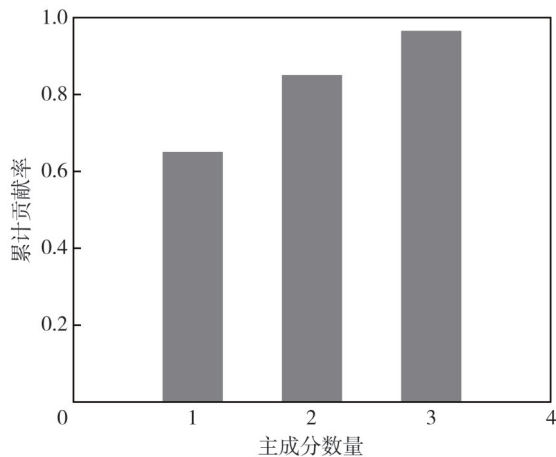


图 9 PCA 方差贡献率

Fig. 9 PCA variance contribution rate

3.3.2 对比实验

为验证本文所提模型的有效性,在工况 1 的测试轴承 B1-3 上,将 Transformer-BiLSTM 模型与 GRU^[17]、CNN^[18]和 BiLSTM^[19]基线模型进行了对比实验。同时,为验证双分支结构的有效性,对比实验中也包含对单一 Transformer 分支和单一 BiLSTM 分支的性能对比。所有模型均采用经上述 PCA 降维后的相同特征作为输入,并在完全相同的训练集和测试集,以及统一的超参数设置(窗口长度:25,训练轮数:200,批次大小:32,优化器:Adam,初始学习率:0.004)下进行训练与评估。各模型在轴承 B1-3 上的预测曲线如图 10 所示,定量评价指标对比结果见表 4。

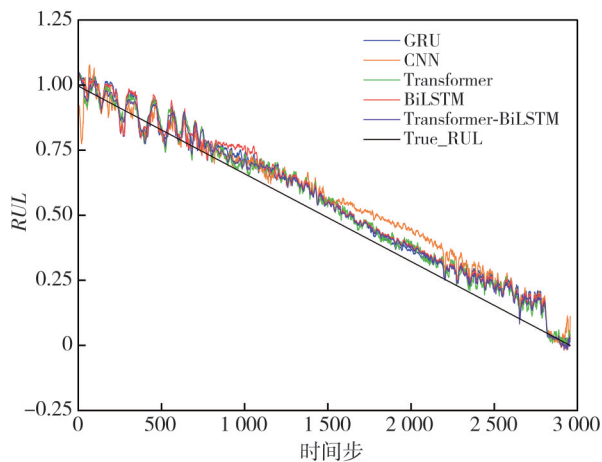


图 10 轴承 B1-3 预测结果

Fig. 10 Prediction results of bearing B1-3

从预测结果来看,本模型的预测曲线较其他模型波动更小,且更接近于真实寿命,说明该预测方法学习到了退化特征与寿命间的关系。从评价指标来看,本文方法 MAE 为 0.046,较 GRU、CNN、BiLSTM 分别降低 49.5%、47.7% 和 36.1%; RMSE 为 0.066,较这 3 类模型分别降低 44.5%、41.1% 和

29.8%; R^2 达 0.950,较这 3 类模型分别提升 15.0%、12.4% 和 6.4%。实验结果显示 Transformer 或 BiLSTM 分支性能均优于基线模型,但双分支结构进一步降低了误差,证明全局与局部特征互补的有效性。本文方法在退化全阶段($RUL=1.0 \rightarrow 0$)误差均低于 10%,避免了传统模型在严重退化阶段($RUL < 0.3$)误差骤升问题。在退化中期($RUL \approx 0.6$)和严重退化阶段($RUL < 0.3$),基线模型(如 CNN)的预测曲线出现明显偏差与抖动,而本文方法的预测轨迹始终与真实值高度吻合,证明了其优异的稳定性。

表 4 对比实验结果

Tab. 4 Results of comparative experiment

方法	MAE	RMSE	R^2
GRU	0.091	0.119	0.826
CNN	0.088	0.112	0.845
BiLSTM	0.072	0.094	0.893
Transformer	0.068	0.089	0.918
本文方法	0.046	0.066	0.950

预测性能差异源于模型与任务适配度:GRU 门控机制相对简单,难以有效为长周期退化趋势建模;BiLSTM 为双向建模,但难捕捉全局依赖;CNN 仅能提取局部特征,且无时序建模能力;而本文双分支架构中,Transformer 能捕捉全局退化趋势,BiLSTM 可挖掘局部动态模式,二者互补适配轴承 RUL 预测,可满足全局-局部时序依赖兼顾的需求,为工业场景轴承全寿命健康管理与维修决策提供可靠支撑。

3.3.3 泛化实验

为检验本文所提方法在不同运行条件下的泛化能力,将其在工况 2(转速为 1 650 r/min,载荷为 4 200 N)下的 4 个轴承全寿命数据上进行测试。所有模型设置与工况 1 实验保持完全一致。泛化实验结果如表 5 所示。

表 5 泛化实验对比结果

Tab. 5 Generalization experiment comparison results

方法	MAE				RMSE				R^2			
	B2-3	B2-4	B2-5	B2-6	B2-3	B2-4	B2-5	B2-6	B2-3	B2-4	B2-5	B2-6
GRU	0.102	0.109	0.096	0.105	0.128	0.135	0.121	0.131	0.821	0.805	0.812	0.793
CNN	0.095	0.101	0.089	0.098	0.117	0.123	0.110	0.120	0.852	0.838	0.845	0.818
BiLSTM	0.081	0.088	0.075	0.084	0.102	0.108	0.095	0.104	0.887	0.875	0.882	0.860
Transformer	0.075	0.082	0.070	0.078	0.095	0.101	0.098	0.088	0.902	0.889	0.898	0.863
本文方法	0.049	0.052	0.045	0.048	0.067	0.070	0.063	0.066	0.941	0.935	0.938	0.945

在 PHM2012 工况 2 的 4 条轴承上,本文方法平均 MAE、RMSE 分别为 0.050 和 0.068,较 BiLSTM 再降 40.5% 和 34.6%, R^2 均值为

0.943,充分验证了所提多域特征融合与 Transformer-BiLSTM 架构在不同工况下的强泛化能力与工程推广价值。

4 结 论

本文提出了一种基于多域特征融合与Transformer-BiLSTM并联网路的轴承剩余寿命预测方法。该方法首先从振动信号中提取时域、频域、时频域和熵域等多域特征,通过基于移动平均法的特征筛选策略保留敏感性高的特征子集,并借助PCA构建出累积贡献率达95%的健康指标,有效增强了退化信息的表达与鲁棒性。在此基础上,本文设计了一种双分支并联结构:利用Transformer编码器捕捉全局长期退化趋势,通过BiLSTM挖掘局部双向时间模式。在PHM2012数据集上的实验表明,所提方法在全寿命周期预测误差低于10%,与BiLSTM、CNN等模型相比,本文模型的MAE和RMSE分别降低36.2%和30.4%, R^2 达到0.950。两个分支的对比进一步验证了双分支结构的有效性,在不同工况下也表现出良好的稳定性与泛化能力。未来的研究工作将集中于跨设备迁移预测、噪声环境下的模型鲁棒性优化,并结合轻量化技术推动其在工程实际中的部署应用。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献

- [1] Li J, Mba D, Li X, et al. An adaptive synchroextracting transform for the analysis of noise contaminated multi-component nonstationary signals[J]. Applied Acoustics, 2023, 202: 109169.
- [2] 王东峰, 赵翀, 李超强, 等. 滚动轴承设计技术研究现状及发展趋势[J]. 轴承, 2023(5): 1-5.
Wang Dongfeng, Zhao Chong, Li Chaoqiang, et al. Research status and development trends of design technology for rolling bearings[J]. Bearing, 2023(5): 1-5. (in Chinese)
- [3] Zhou J, Ma Y, Jiang B, et al. A bearing remaining useful life prediction method based on residual convolutional network and LSTM[C]//2023 6th International Symposium on Autonomous Systems (ISAS), 2023: 1-6.
- [4] Jena B, Saxena S, Nayak G K, et al. Artificial intelligence-based hybrid deep learning models for image classification: the first narrative review[J]. Computers in Biology and Medicine, 2021, 137: 104803.
- [5] 张氢, 江伟哲, 肖慧灵, 等. 基于LSTM的滚动轴承寿命预测[J]. 港口装卸, 2021(5): 1-5.
Zhang Qing, Jiang Weizhe, Xiao Huiling, et al. Life prediction of rolling bearing based on LSTM[J]. Port Operation, 2021(5): 1-5. (in Chinese)
- [6] 万忠义, 吴开平, 徐彬, 等. 基于多维时序融合特征的滚动轴承剩余使用寿命预测[J]. 制造技术与机床, 2025(2): 9-16.
Wan Zhongyi, Wu Kaiping, Xu Bin, et al. Residual useful life prediction of rolling bearings based on multi-dimensional temporal mixing features[J]. Manufacturing Technology & Machine Tool, 2025(2): 9-16. (in Chinese)
- [7] 卢瑾, 张永平. 基于注意力机制的滚动轴承剩余使用寿命预测方法[J]. 机电工程, 2023, 40(4): 516-521.
Lu Jin, Zhang Yongping. Remaining useful life prediction of rolling bearings based on attention mechanism[J]. Journal of Mechanical & Electrical Engineering, 2023, 40(4): 516-521. (in Chinese)
- [8] 蒋全胜, 许伟洋, 朱俊俊, 等. 基于动态加权卷积长短时记忆网络的滚动轴承剩余寿命预测方法[J]. 振动与冲击, 2022, 41(17): 282-291.
Jiang Quansheng, Xu Weiyang, Zhu Junjun, et al. Residual life prediction method of rolling bearing based on DW-CNN-LSTM networks[J]. Journal of Vibration and Shock, 2022, 41(17): 282-291. (in Chinese)
- [9] 苏浩田. 基于PCA-GRU的轴承剩余使用寿命混合预测方法研究[D]. 大连: 大连交通大学, 2022: 11.
Su Haotian. Research on hybrid prediction method of bearing remaining useful life based on PCA-GRU[D]. Dalian: Dalian Jiaotong University, 2022: 11. (in Chinese)
- [10] 杨泽锐, 胡红萍. 基于改进的减法平均优化算法与BP神经网络的人脸识别[J]. 中北大学学报(自然科学版), 2024, 45(6): 725-736.
Yang Zerui, Hu Hongping. Face recognition based on improved subtraction average-based optimizer and BP neural network [J]. Journal of North University of China (Natural Science Edition), 2024, 45(6): 725-736. (in Chinese)
- [11] Chen Y, Peng G, Zhu Z, et al. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction[J]. Applied Soft Computing, 2020, 86: 105919.

(下转第539页)