DOI: 10.13382/j.jemi.B2508359

融合 SSM 分词器的轻量高斯注意力轴承 RUL 预测*

李嘉恒 常致远 樊 薇 陈 超 许桢英
(江苏大学机械工程学院 镇江 212013)

摘 要:轴承作为旋转机械系统的关键传动部件,其运行状态直接关系整体设备的安全性与运行效率。因此,实现轴承状态实时监测与剩余使用寿命(remaining useful life,RUL)的准确预测,对于预防潜在故障具有重要意义。当前基于深度学习的自注意力机制模型虽在寿命预测中得到广泛应用,但由于其主要依赖特征嵌入与位置编码机制,难以有效捕捉退化过程中的关键微观特征变化。尽管嵌入式高斯掩码能显著提升模型对局部细微退化特征的捕捉能力,但在处理数据时计算复杂度 3 次方增长,限制了使用效率。针对上述问题,提出融合状态空间模型(state-space model,SSM)与注意力机制的协同预测框架:通过将小波变换和倒谱滤波融入状态空间,构建新型特征分词器替代传统嵌入模块,提升退化特征表征能力;基于门控机制的动态筛选算法实时分析特征参数的单调性演变、趋势波动及抗干扰特性,实现关键退化指标的智能提取;结合振动信号的局部特性和全局退化规律,设计轻量化多尺度注意力模块,有效降低计算负荷并实现寿命映射解码。对比实验基于 PRONOSTIA 轴承寿命数据集工况 1、2 及江苏联益友仪器测控技术有限公司 000A1-3 轴承全寿命试验数据,结果表明所提方法在寿命预测平均绝对误差(MAE)和均方根误差(RMSE)指标上分别实现 11.4%、20%、15.4%和 15.2%、18.5%、27.4%的提升,消融实验表明关键模块可提升计算效率达 55.6%。

关键词: 滚动轴承;剩余寿命预测;Transformer 网络;高斯分布;自注意力机制;状态空间模型
中图分类号: TH133;TN06 **文献标识码**: A **国家标准学科分类代码**: 460.40

Rolling bearing remaining useful life prediction based on a SSM tokenizer and linear multi-scale gaussian attention

Li Jiaheng Chang Zhiyuan Fan Wei Chen Chao Xu Zhenying
(School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: Bearings are critical transmission components in rotating machinery whose operating conditions directly affect equipment safety and efficiency, making real-time monitoring and accurate prediction of remaining useful life (RUL) essential for preventing failures. Although deep learning-based self-attention models are widely used for life prediction, their reliance on feature embeddings and positional encoding hinders the capture of subtle degradation changes. Embedded Gaussian masks improve detection of delicate local degradation features, but their cubic computational complexity with data length limits practical efficiency. To overcome these issues, this study proposes a collaborative framework that integrates state-space model (SSM) with attention mechanisms. By incorporating wavelet transforms and cepstral filtering into the state-space process, the new feature tokenization module replaces traditional embeddings to enhance degradation representation. A gating-based dynamic selection algorithm then analyzes feature evolution, trend fluctuations, and noise resistance in real time to intelligently extract key degradation indicators, while a lightweight multi-scale attention module decodes life mapping by merging local vibration characteristics with global degradation patterns and reducing computational load. Comparative experiments on the PRONOSTIA dataset (conditions 1 and 2) and full-life test data from Jiangsu Lianyy Measurement and Control Technology Co., Ltd. show MAE improvements of 11.4%, 20%, and 15.4% and RMSE enhancements of 15.2%, 18.5%, and 27.4%, with ablation studies confirming up to a 55.6% boost in computational efficiency.

Keywords: rolling bearings; RUL prediction; transformer; Gaussian distribution; self-attention; SSM

收稿日期: 2025-05-05 Received Date: 2025-05-05
*基金项目:煤炭精细勘探与智能开发全国重点实验室开放研究课题(SKLCRSM24KF009)、国家自然科学基金(51905218)、第八届中国科协青年人才托举工程(2022QNRC001)、江苏大学青年英才培育计划、江苏省自然科学基金青年项目(BK20210772)资助

0 引 言

滚动轴承作为旋转机械的核心运转部件,其健康状况直接影响设备的可靠性[1]。然而,在高速、重载等工况下,轴承易出现渐进磨损或突发故障,可能引发意外停机,造成生产中断[2]。相比传统“故障后维修”的被动模式,基于数据驱动的剩余寿命预测技术能提前预警故障,已成为提升设备管理水平、保障连续化生产的关键手段,也是当前智能制造研究的热点方向。

近年来,注意力机制的提出突破了传统神经网络在并行计算效率与长程依赖建模方面的架构约束,其成功实践已从图像识别、机器翻译延伸至工业设备时序分析领域。其中基于自注意力机制的 Transformer 模型表现尤为突出,为轴承剩余寿命预测提供了端到端特征学习的创新解决方案。Mo 等[3]针对传统剩余寿命预测方法难以捕捉长时依赖性和局部退化特征不足的问题,提出了基于 Transformer 的预测方法,并引入门控卷积单元以增强局部特征提取,显著提高了预测精度和鲁棒性。Su 等[4]针对复杂退化过程中的特征提取困难,设计了一种自适应 Transformer 模型,该模型结合注意力机制与循环架构,有效捕捉非线性时序退化特征,降低了预测误差。Lim 等[5]为了解决多步时序动态预测中的信息融合与解释性不足问题,提出了 TFT(temporal fusion transformer),不仅实现了高预测性能,还通过内置解释模块具备可解释性。Zhu 等[6]利用自注意力机制构建了残差混合网络,专注于解决旋转机械退化过程中多尺度信息整合难题,提升预测准确率和稳定性。尽管 Transformer 凭借全局上下文建模优势在复杂工况数据表征中表现卓越,其位置信息不敏感难以区分振动信号局部冲击特征与长程共振衰减的差异性,且自注意力机制固有的指数级计算复杂度,严重限制其在工业关键系统中的实际工程部署。

改进状态空间模型(state space model, SSM)具备非线性动态表征能力及多尺度耦合效应解析,正成为寿命预测领域的研究前沿[7]。传统 SSM 虽通过线性状态-观测关系平衡建模可解释性与计算效率,但其线性假设难以表征复杂非线性动力学特征。为突破此局限,Krishnan 等[8]提出深度 SSM 混合架构,提升复杂退化过程的预测精度。Gu 等[9]提出了 Mamba 架构在深度学习框架下引入动态选择 SSM,使模型能根据输入特征自主调整内部状态演化路径。基于此,Dao 等[10]进一步揭示了 SSM 与各种注意力机制间的紧密关系,提出状态空间对偶框架,建立 SSM 与线性注意力的理论映射,证明 SSM 本质上具备隐式位置编码能力。当前基于 SSM 的轴承寿命预测研究仍处于探索阶段,Ran 等[11]提出一种结合深度潜变量状态空间模型与微分预变换的 SSM,有效利用深度学习的非线性建模能力显著提升轴承退化预测的长期准确性。Weikun 等[12]融合物理知识驱动与自监督学习机制提出液态网络 SSM,有效克服数据稀疏和物理先验不足。然而如何有效分解轴承非线性退化过程中的时变动态特征仍存在挑战。

针对上述挑战,本文提出一种融合 Mamba 架构与注意力机制的新型深度学习网络,旨在提升滚动轴承退化预测精度并降低训练成本。针对 Transformer 在轴承剩余寿命预测中存在高计算复杂度与局部-全局特征辨识不足问题,创新设计轻量化多尺度高斯掩码自注意力机制,通过多尺度高斯掩码构建特征空间集成局部与全局特征,采用线性投影与滑动窗口联合计算策略降低复杂度。同时,为应对自注意力机制在特征嵌入和跨时间尺度建模中解释性不足的缺陷,提出基于 SSM 的轴承特征分词器,通过改进型 Mamba 架构结合小波包变换与倒谱滤波构建时频联合嵌入空间,并引入动态特征选择门控网络实时滤除噪声、增强有效特征表征,精准捕捉轴承退化过程的非线性演化规律。

1 基于 SSM 的轴承退化特征分词器结构

1.1 选择性 SSM 特征建模

1)选择性 SSM 模型

Mamba 架构在 SSM 建模的基础上,通过对每组特征做出有选择性的信息传递与动态调控,从而突破传统 SSM 中参数时间不变的局限[13]。图 1 所示为基于 S4 模型的选择性结构化 SSM。

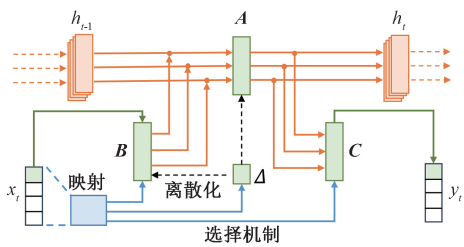


图 1 基于 S4 模型的选择性结构化 SSM

Fig. 1 Selective structured SSM based on the S4 framework

模型的具体过程如下:

$$B(x) = Linear_N(x) \tag{1}$$

$$C(x) = Linear_N(x) \tag{2}$$

$$\Delta(x) = softplus(Broadcast_D(Linear_1(x))) \tag{3}$$

$$softplus(x) = \ln(1 + e^x) \tag{4}$$

式中:$Linear_N(x)$ 表示把输入 x 通过线性变换映射到 N 维向量;$softplus$ 为激活函数;$Broadcast_D$ 表示将较低维度的输出扩展或复制到目标维度 D。这样,最终的 Δ 参数能够随输入 x 动态更新,从而实现对信息传递和控制的

精细调控。

完整的 Mamba 结构如图 2 所示，以 SiLU 与 Swish 激活函数替代传统乘法门，从而获得更平滑的梯度信息[14]。激活函数如下：

$$SiLU(x)=\frac{x}{1+e^{-x}} \tag{5}$$

$$Swish(x)=x\cdot\left(\frac{1}{1+e^{-x}}\right) \tag{6}$$

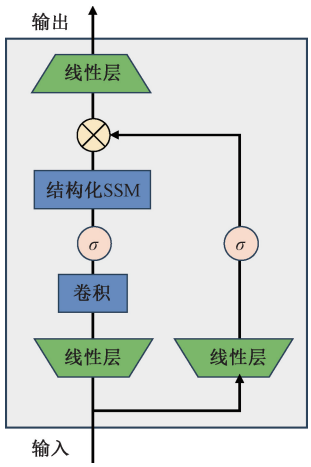


图 2　Mamba 模型框架

Fig. 2　The structure of Mamba

2）小波倒谱 SSM 过程

使用最广泛的倒谱形式：

$$\hat{y}_c(t)=j^{-1}\{\log(|Y(f)|)+j\angle(Y(f))\} \tag{7}$$

式中：$Y(f)$ 为原始时域信号的傅里叶变换；j 和 j^{-1} 分别代表傅里叶变换和逆变换；$|Y(f)|$ 和 $\angle(Y(f))$ 为信号的幅值和相位信息。考虑退化为非线性，采用 db4 小波变换进行分析：

$$\varphi(t)=\sqrt{2}\sum_{k=0}^{3}h_k\cdot\varphi(2t-k),\quad k=0,1,2,3 \tag{8}$$

$$h_i=\begin{cases}(1+\sqrt{3})/(4\sqrt{2}), & i=0\\(3+\sqrt{3})/(4\sqrt{2}), & i=1\\(3-\sqrt{3})/(4\sqrt{2}), & i=2\\(1-\sqrt{3})/(4\sqrt{2}), & i=3\end{cases} \tag{9}$$

式中：h_k 为缩放系数。小波函数 $\psi(t)$ 为：

$$\psi(t)=\sqrt{2}\sum_{k=0}^{3}g_k\cdot\varphi(2t-k) \tag{10}$$

$$g_k=(-1)^k\cdot h_{3-k},\quad k=0,1,2,3 \tag{11}$$

为得到能突出轴承退化的特征主导生成层，从小波倒谱角度重构 SSM 观测状态：

$$h_c(t)=W_{db4}^{-1}\log(|W_{db4}h_t|) \tag{12}$$

$$h_c(t)=A_t\cdot h_c(t-1)+B_t\cdot h_c(t-1) \tag{13}$$

$$y(t)=h_c(t)\cdot e^{-\sigma_0 t} \tag{14}$$

式中：W_{db4} 和 W_{db4}^{-1} 代表 db4 小波变换和逆变换；$h_c(t)$ 为当前时间步 t 的状态；$y(t)$ 为 $h_c(t)$ 指数滤波后的结果。

本文所提小波倒谱 SSM 过程如图 3 所示。该生成层利用倒谱指数滤波技术，首先通过倒谱方法将激励函数与传递函数分离，再借助指数滤波进一步强化系统固有模态特性，最终实现关键故障信息的提取和增强。

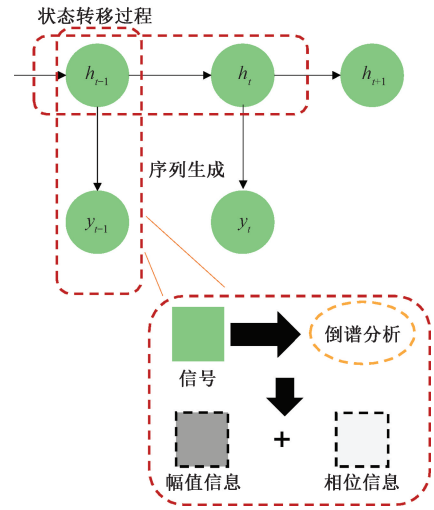


图 3　特征增强层架构

Fig. 3　Architecture of the feature enhancement module

1.2　门控特征选择层

尽管信号重构能有效揭示轴承退化过程的潜在故障特征，但重构过程中由算法过度拟合及信号分量相似性引发的冗余信息会稀释关键退化指标的信噪比，并增加计算复杂度。为此，在小波倒谱-SSM 重构框架后端嵌入门控特征选择机制，通过构建复合筛选指标，建立动态特征蒸馏路径。

1）单调性：

$$M_i=\left|\frac{1}{(N-1)}\sum_{j=1}^{N-1}\operatorname{sgn}(T_{i,j+1}-T_{i,j})\right| \tag{15}$$

$$\operatorname{sgn}(x)=\begin{cases}0, & x=0\\1, & x<0\\-1, & x>0\end{cases} \tag{16}$$

式中：$T_{i,j}$ 为 T_i 在第 j 个位置的值；N 为样本总数。

2）趋势性：

$$\bar{T}_i=\left(\frac{1}{N}\right)\sum_{j=1}^{N}T_{i,j} \tag{17}$$

$$TR_i=\left|\frac{\left(\sum_{j=1}^{N}(j-\bar{j})(T_{i,j}-\bar{T}_i)\right)}{\left(\sqrt{\left(\sum_{j=1}^{N}(j-\bar{j})^2\right)}\cdot\sqrt{\left(\sum_{j=1}^{N}(T_{i,j}-\bar{T}_i)^2\right)}\right)}\right| \tag{18}$$

式中：$\bar{T}_i$ 表示第 i 组特征的均值；$\bar{j}=\frac{1}{N}\sum j$。

3)鲁棒性:

$$R_i = \frac{Q_3(T_i) - Q_1(T_i)}{\max(T_i) - \min(T_i)} \tag{19}$$

式中:$Q_3(T_i)$ 和 $Q_1(T_i)$ 分别为特征数据的第 3 和第 1 分位数。

采用加权平均构造复合指标。本文给单调性、趋势性和鲁棒性分别赋予 0.6、0.2、0.2 的权重。因此,复合指标 PZT_i 为:

$$PZT_i = 0.6M_i + 0.2Tr_i + 0.2R_i \tag{20}$$

加入复合指标门控单元调整信息流的门控过程:

$$y = (W_1x + b_1) \odot \text{Sigmoid}((W_2x + b_2) + (W_3PZT_i + b_3)) \tag{21}$$

$$\text{Sigmoid}(x) = \frac{1}{(1 + e^{-x})} \tag{22}$$

式中:W_1、W_2、W_3 与 b_1、b_2、b_3 分别为全连接层对应的权重和偏置;⊙表示逐元素相乘。

2　轻量化多尺度高斯掩码注意力

2.1　多尺度高斯分布距离掩码

高斯距离掩码算法利用高斯函数计算距离权重,通过在自注意力中赋予邻近数据更高权重来增强局部依赖、抑制远程干扰[15]。然而,轴承退化特征中存在的周期性关系要求模型能有效关联非邻近数据。该算法过度侧重于局部邻域,难以充分建模此类长程周期性依赖,可能导致预测精度下降。

为解决上述问题,本文将自注意力机制与高斯距离掩码有机结合,并引入多尺度策略[16],如图 4 所示,该模块能协同捕捉周期数据中的短期局部与长期全局依赖,使模型可动态调整对非邻近数据点的关注权重。高斯分布掩码矩阵使用的分段高斯密度函数为:

$$M_{dis}^{\sigma}(\boldsymbol{D}) = \begin{cases} G(r,\sigma), r \leqslant l_w/2 \\ G\left(\frac{l_w}{r},\sigma\right), l_w/2 < r \leqslant l_w \\ G\left(\frac{r}{\left|\frac{r}{l_w}\right|} \cdot l_{sw},\sigma\right), r > l_w,\ r \bmod l_w \leqslant l_w \\ G\left(\frac{l_w}{r} + \left|\frac{r}{l_w}\right| \cdot l_w,\sigma\right), r > l_w, r \bmod l_w > l_w \end{cases} \tag{23}$$

式中:$G(r,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$;$l_w$ 表示输入数据的滑动窗口长度;r 表示相对位置距离;超参数 σ 和 μ 被设置为通过网络学习的参数。

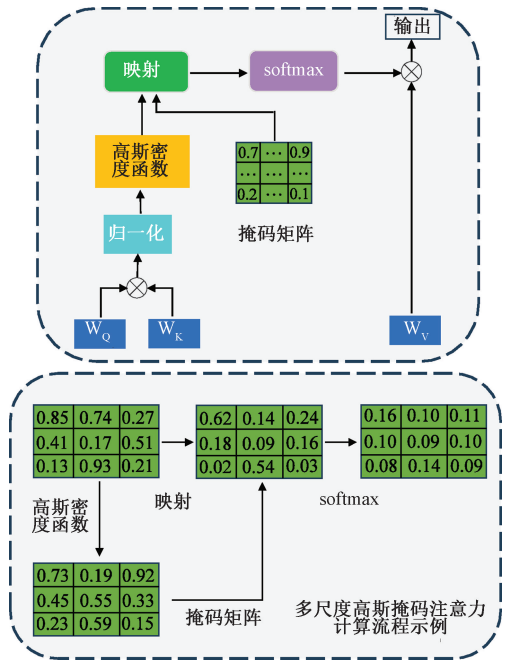


图 4　多尺度时序高斯分布注意力

Fig. 4　Multiscale temporal gaussian distribution attention

2.2　基于线性注意力机制与滑动窗口的轻量化注意力解码器

Based 模型是一种简化线性语言模型,其核心在于融合线性注意力与滑动窗口注意力,实现高效的上下文捕捉和序列建模。线性注意力通过递归状态捕捉长距离依赖,并将计算复杂度降至线性,显著降低了推理开销;而滑动窗口注意力则通过限定窗口 $W(i)$ 内的注意力得分,在增强局部细节捕捉的同时,进一步控制了计算成本[17]。

受此启发,本文注意力模块选择了类似的协同设计。其中,线性注意力通过泰勒二阶展开近似替代 softmax 激活函数,从而实现注意力计算线性化,加速了长序列的处理:

$$a_{ij} = \frac{(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}})_{ij}}{\sqrt{d}} + [M_{dis}^{\sigma}(\boldsymbol{D})]_{ij} \tag{24}$$

$$\phi(a_{ij}) = 1 + a_{ij} + \frac{a_{ij}^2}{2} \tag{25}$$

$$\alpha_{ij} = \frac{\phi(a_{ij})}{\sum_{k=1}^{N}\phi(a_{ik})} \tag{26}$$

$$L - Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \sum_{j=1}^{N}\alpha_{ij}V_j \tag{27}$$

式中:a_{ij} 代表了序列中任意位置 i 和 j 的注意力得分;$[M_{dis}^{\sigma}(\boldsymbol{D})]_{ij}$ 为位置 i 和 j 的掩码计算值;$\phi(a_{ij})$ 是泰勒二阶展开线性化映射;α_{ij} 为归一化后的注意力权重;N 为序列总个数。同时,模型引入滑动窗口机制,精确定义局

部关注区域,并与之前采用的线性计算部分相结合。图 5 所示为模型中注意力机制的整体流程。对滑动窗口 $W(i)$ 范围内的注意力计算如下:

$$\text{SW-Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \sum \text{softmax}\left(\frac{(\boldsymbol{QK}^{\mathrm{T}})_{ij}}{\sqrt{d}} + [M_{dis}^{\sigma}(\boldsymbol{D})]_{ij}\right) V_j, \quad j \in W(i) \tag{28}$$

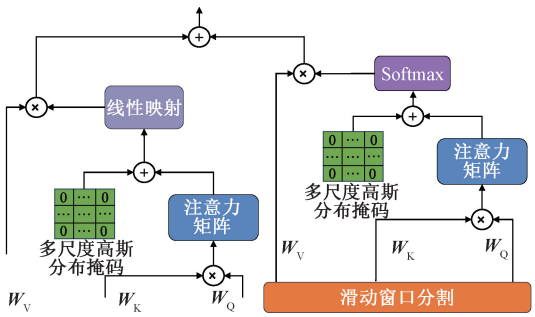


图 5　轻量化后的高斯多尺度注意力模块

Fig. 5　Lightweight Gaussian multi-scale attention module

3　算法构建

3.1　特征提取与健康指标构建

为降低特征工程复杂度并利用分词器模块的深度特征挖掘与特征增强能力,在构建数据表示时直接采用基础时域特征统计量作为输入,避免了频域信息提取所需的复杂变换。表 1 为所选主要时域特征。尽管传统时域特征能反映整体趋势并辅助深度特征挖掘,其在捕捉局部、瞬变及多尺度细节方面仍存在局限。为此,本文基于 SSM 分词器中的小波倒谱观测方程,从小波域提取多尺度特征,并借助函数型主成分分析降维整合,构建健康指标,如表 2 所示。其中, D_i 表示第 i 个小波分量, N 表示样本数量。

3.2　模型训练与优化

图 6 所示本文模型的训练流程。首先,基于监督数据生成特征集合 $\{f_{11},\cdots,f_{mn}\}$ 与健康指标序列 $\{H_1,\cdots,H_m\}$。特征数据输入小波倒谱 Mamba 分词器,并利用小波倒谱 SSM 及时序堆叠结构增强退化特征的表达,再通过门控机制筛除冗余信息,而健康指标则作为设备状态变化的关键参考,辅助模型捕捉退化趋势。训练中使用 K 折交叉验证策略,同时采用 32 大小的分块技术优化内存管理。

表 1　时域特征

Table 1　Selected time-domain features

名称	公式	名称	公式	名称	公式
峰值	$T_1 = \max\lvert x(i)\rvert$	峰峰值	$T_2 = \max\lvert x(i)\rvert - \min\lvert x(i)\rvert$	平均绝对值	$T_3 = \frac{1}{N}\sum_i^N \lvert x(i)\rvert$
方差	$T_4 = \frac{1}{N}\sum_i^N (x(i)-\bar{x})^2$	标准差	$T_5 = \sqrt{\frac{1}{N}\sum_i^N (x(i)-\bar{x})^2}$	均方根值	$T_6 = \sqrt{\frac{1}{N}\sum_i^N x(i)^2}$
峰值因子	$T_7 = \frac{\max\{\lvert x(i)\rvert\}}{\sqrt{\frac{1}{N}\sum_{i=1}^N x(i)^2}}$	冲击因子	$T_8 = \frac{\max\{\lvert x(i)\rvert\}}{\frac{1}{N}\sum_{i=1}^N \lvert x(i)\rvert}$	偏度	$T_9 = \frac{1}{N}\sum_{i=1}^N \left(\frac{x(i)-\bar{x}}{T_5}\right)^3$
峭度	$T_{10} = \frac{1}{N}\sum_{i=1}^N \left(\frac{x(i)-\bar{x}}{T_5}\right)^4$				

表 2　小波域多尺度平均绝对值特征指标

Table 2　Multi-scale wavelet mean absolute value feature metrics

名称	公式	名称	公式	名称	公式
一阶平均绝对值	$F_1 = \frac{1}{N_{D1}}\sum_{k=1}^{N_{D1}} \lvert D_1(k)\rvert$	二阶平均绝对值	$F_2 = \frac{1}{N_{D2}}\sum_{k=1}^{N_{D2}} \lvert D_2(k)\rvert$	三阶平均绝对值	$F_3 = \frac{1}{N_{D3}}\sum_{k=1}^{N_{D3}} \lvert D_3(k)\rvert$
四阶平均绝对值	$F_4 = \frac{1}{N_{D4}}\sum_{k=1}^{N_{D4}} \lvert D_4(k)\rvert$	五阶平均绝对值	$F_5 = \frac{1}{N_{D5}}\sum_{k=1}^{N_{D5}} \lvert D_5(k)\rvert$		

3.3　剩余使用寿命预测结果评估

模型的评价指标包括均方根误差(RMSE)、平均绝对误差(MAE)以及 IEEE PHM 2012 滚动轴承剩余使用寿命预测挑战赛[18]所定义的官方评分函数 Socre:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{28}$$

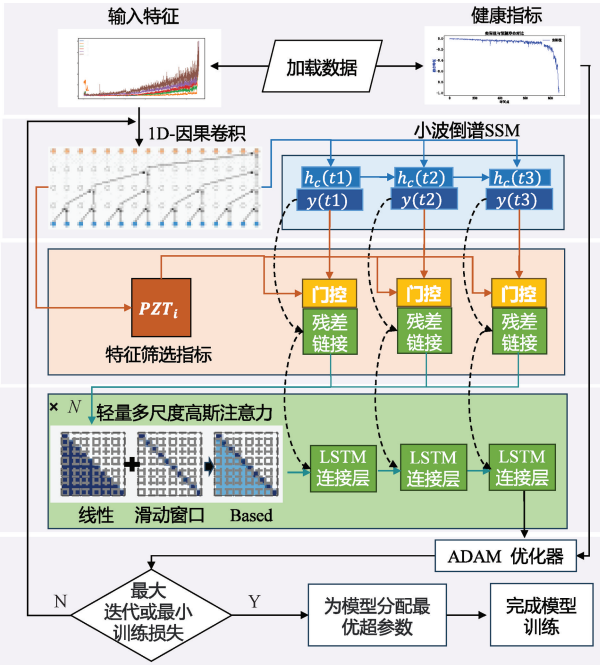


图 6　模型训练流程

Fig. 6　Flowchart of the model training

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{29}$$

$$Socre = \frac{1}{20}\sum_{i=1}^{20} A_i \tag{30}$$

$$A_i = \begin{cases} \exp\left(-\ln(0.5) \times \left(\frac{Er_i}{5}\right)\right), Er_i \leqslant 0 \\ \exp\left(+\ln(0.5) \times \left(\frac{Er_i}{20}\right)\right), Er_i > 0 \end{cases} \tag{31}$$

$$Er_i = \frac{Rel_i - \overline{Rel_i}}{Rel_i} \tag{32}$$

式中：m 为样本数量；y_i 表示样本真实值；$\hat{y}_i$ 表示预测值；Rel_i 和 $\overline{Rel_i}$ 分别代表实际与预测的 RUL。

4　实验与分析

4.1　对比试验

1) PRONOSTIA 数据集验证

为验证本文方法的有效性，采用 IEEE PHM 2012 挑战 PRONOSTIA 数据集。该数据集由法国 FEMTO-ST 研究所开发。该数据集包含 17 个滚动轴承从初始状态到失效的完整加速试验数据，采集了在整个退化过程中的振动信号和温度信息。PRONOSTIA 实验平台的结构和采集数据具体参数分别如图 7 和表 3 所示。

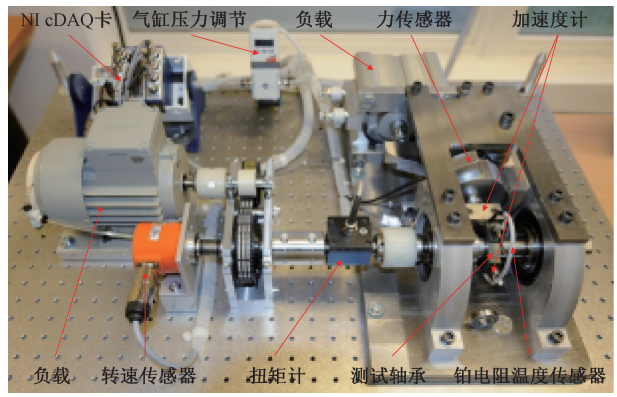


图 7　PRONOSTIA 轴承退化试验台

Fig. 7　PRONOSTIA bearing degradation test bench

表 3　PRONOSTIA 数据集参数

Table 3　PRONOSTIA dataset

工况	转速/(r·min^{-1})	负载/N	采样频率/kHz	训练集	测试集
工况 1	1 800	4 000	25.6	轴承 1-1, 1-2	轴承 1-3, 1-4, 1-5, 1-6, 1-7
工况 2	1 650	4 200	25.6	轴承 2-1, 2-2	轴承 2-3, 2-4, 2-5, 2-6, 2-7
工况 3	1 500	5 000	25.6	轴承 3-1, 3-2	轴承 3-3

表 4 为模型的完整隐藏层参数配置。为确保公平比较，对比模型均采用相同关键参数：ADAM 优化器初始学习率为 0.000 1，滑动窗口长度初始化为 1。在轴承工况 1(1-1、1-2 训练、1-3 测试) 和工况 2(2-1、2-2 训练、2-3 测试) 下测试模型性能。

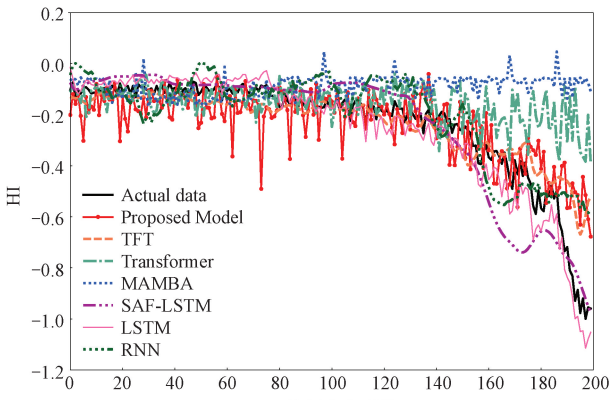
为了验证模型性能，本文选择了 TFT、MAMBA、Transformer、SAF-LSTM 等模型进行对比。图 8(a)、(b) 所示为模型及对比方法基于 50 次测试结果得到的平均退化趋势预测效果的对比。提出模型通过引入特征分词器和高斯过程提升了退化趋势预测精度，其预测曲线与实际设备退化过程高度吻合。相比之下，TFT 和 Transformer 模型虽通过注意力机制捕捉到峰值特征，但缺乏有效额外特征输入和局部注意力捕捉导致偏差；MAMBA 模型受限于线性状态空间建模，长序列预测稳定性不足；SAF-LSTM 虽在局部依赖关系建模中表现优秀，但对长程退化趋势捕捉能力较弱。LSTM 模型由于依赖门控机制处理短期序列依赖，在长程退化趋势建模中易出现误差累积；RNN 模型则因梯度消失问题，难以有

效学习轴承退化过程中的非线性动态演化。

表 4　本文模型超参数设置

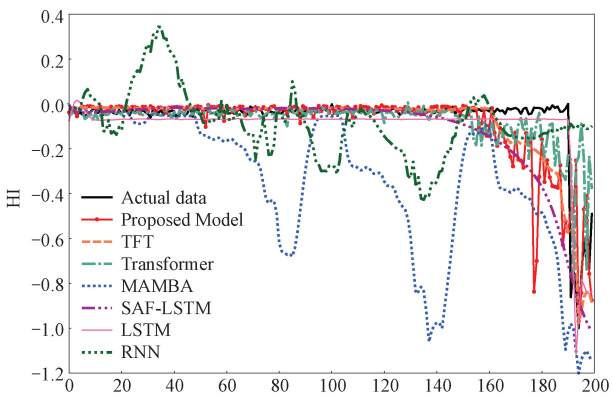
Table 4　Proposed model configuration

模块	编号	层类型	隐层维度	激活函数
分词器模块	1	Linear	512	GeLU
	2	1D-CNN	4×128	ReLU
	3	小波倒谱 SSM×4	128	Tanh, Sigmoid
	4	Linear	512	GeLU
门控模块	1	Gated Linear	512	Sigmoid
	2	线性多尺度高斯注意力×4	512	\
	3	滑动窗口高斯注意力×4	512	SoftMax
解码器	4	深度 LSTM×4	512	Tanh, Sigmoid
	5	Full-connection layer	256	GeLU
	6	Full-connection layer	M	Sigmoid



(a) 轴承1-3预测曲线

(a) Prediction curve for bearing 1-3



(b) 轴承2-3预测曲线

(b) Prediction curve for bearing 2-3

图 8　PRONOSTIA 数据集上模型及对比方法表现

Fig. 8　Performance of models and comparative methods on the PRONOSTIA dataset

PRONOSTIA 数据集测试结果如表 5 所示，在 PRONOSTIA 1-3 数据集中，改进模型的 MAE 较 TFT 降低约 11.4%、较 MAMBA 降低约 33.9%、较 Transformer 降低约 58.1%、较 SAF-LSTM 降低约 15.2%、较 LSTM 和 RNN 分别降低约 20.4%和 54.7%。RMSE 分别较 TFT 降低约 23.6%、较 MAMBA 降低约 53.0%、较 Transformer 降低约 55.6%、较 SAF-LSTM 降低约 8.3%、较 LSTM 和 RNN 分别降低约 5.1%和 53.8%。在 PRONOSTIA 2-3 数据集中，改进模型的 MAE 较 TFT、MAMBA、Transformer、SAF-LSTM、LSTM 和 RNN 分别降低约 20.0%、53.8%、48.9%、47.8%、14.8%和 84%，RMSE 降幅分别为约 31.2%、51.8%、18.5%、48.5%、24.3%和 70.4%。实验结果表明，模型在退化过程建模中展现出更高的预测精度与鲁棒性。

表 5　PRONOSTIA 数据集测试结果

Table 5　Evaluation metrics on PRONOSTIA datasets

轴承	模型	MAE	RMSE	Score
1-3	本文	**0.039±0.012**	**0.055±0.016**	**0.94**
	TFT	0.050±0.008	0.072±0.009	0.92
	MAMBA	0.059±0.013	0.117±0.023	0.85
	Transformer	0.093±0.021	0.124±0.027	0.84
	SAF-LSTM	0.044±0.009	0.060±0.012	0.93
	LSTM	0.049±0.008	0.058±0.022	0.90
	RNN	0.086±0.0022	0.119±0.016	0.87
2-3	本文	**0.023±0.002**	**0.053±0.012**	**0.85**
	TFT	0.030±0.004	0.077±0.010	0.83
	MAMBA	0.052±0.023	0.110±0.025	0.84
	Transformer	0.047±0.008	0.065±0.017	0.83
	SAF-LSTM	0.046±0.012	0.103±0.014	0.82
	LSTM	0.027±0.010	0.070±0.021	0.84
	RNN	0.145±0.008	0.201±0.015	0.69

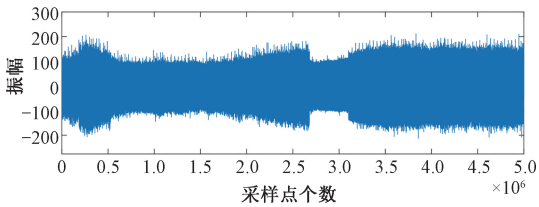
2) NSK-6007 轴承数据集验证

为验证所提方法的工程有效性，将所以方法应用于江苏联谊友仪器测控技术有限公司轴承全寿命试验台，基于 000A1-3 型轴承试验台采集工厂实际环境下(包括工厂环境噪声和设备定时启停)的 NSK-6007 轴承全周期退化数据，试验台如图 9 所示。数据集使用径向加速度传感器采集，转速为 6 000 r/min，加载 7 000 N。采样频率为 25.6 kHz，每分钟采集 1.5 s，共获取 3 个轴承的完整退化数据，其原始信号如图 10(a)~(c)所示。

图 9　000A1-3 轴承退化试验台

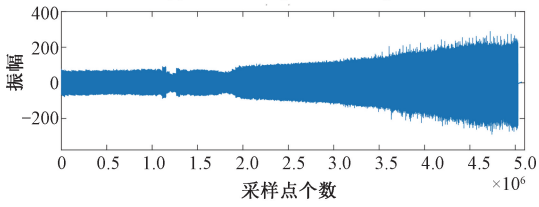
Fig. 9　000A1-3 bearing degradation test bench

对比实验保持参数与训练测试一直，选择轴承 1 和 2 训练，在轴承 3 上进行测试。在加速退化数据集上，各模



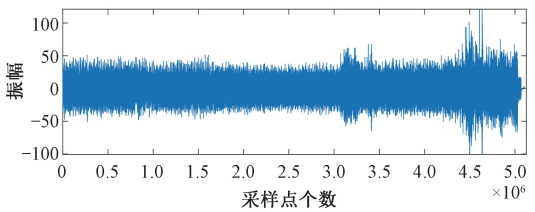
(a) 轴承1振动信号

(a) Vibration signal of bearing #1



(b) 轴承2振动信号

(b) Vibration signal of bearing #2



(c) 轴承3振动信号

(c) Vibration signal of bearing #3

图 10 轴承振动信号

Fig. 10 Vibration signal of bearing

型基于平均 50 次测试结果得到的平均退化趋势预测效果的对比如图 11 所示。图 11 表明,经过引入掩码机制和分词器特征协同优化后的模型,其预测结果最接近实际退化趋势。

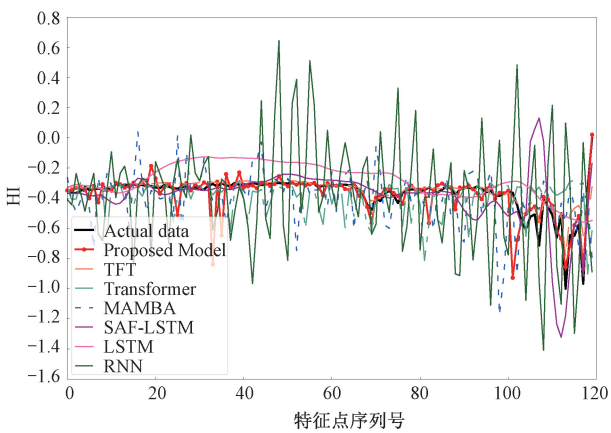


图 11 轴承 3 数据上模型及对比方法表现

Fig. 11 Performance of models and comparative methods on the bearing 3 dataset

表 6 为具体测试结果,所提方法在 MAE 和 RMSE 上分别比 TFT 模型降低约 26.7%和 51.6%,比 MAMBA 模型降低约 68.0%和 69.2%,比 Transformer 模型分别降低约 59.3%和 57.6%,而与 SAF-LSTM 模型相比,在 MAE 上降低约 15.4%、在 RMSE 上降低约 27.4%,与 LSTM 模型相比,在 MAE 上降低约 69.2%,在 RMSE 上降低约 68.1%,与 RNN 模型相比,在 MAE 上降低约 88.4%,在 RMSE 上降低约 88.1%。实验结果表明,模型在退化过程建模中展现出更高的预测精度与鲁棒性。

表 6 NSK-6007 轴承 3 数据测试结果

Table 6 Evaluation metrics on NSK-6007 bearing 3 datasets

模型	MAE	RMSE	Score
本文	**0.033±0.010**	**0.045±0.014**	**0.98**
TFT	0.064±0.009	0.103±0.012	0.96
MAMBA	0.103±0.008	0.146±0.013	0.91
Transformer	0.091±0.018	0.106±0.024	0.94
SAF-LSTM	0.043±0.012	0.072±0.023	0.95
LSTM	0.107±0.004	0.141±0.009	0.85
RNN	0.284±0.006	0.378±0.012	0.82

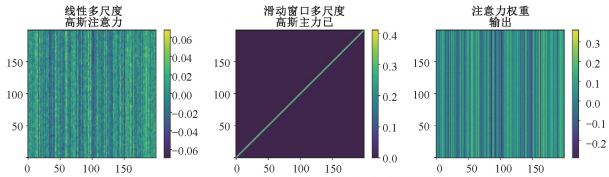
4.2 注意力可视化

对各数据集编码器中线性高斯注意力层的分析表明,其初始注意力分布较为稀疏,局部区域间权重联系较弱。随着网络层次加深,叠加滑动窗口计算后的输出层逐步为关键局部区域分配更高权重,增强了区域间的连接,并使注意力最终聚焦于固定深层特征。此外,各编码器在处理多尺度时频特征时,能够从各自专属区域捕获互补信息,确保模型在剩余寿命预测中始终聚焦于关键特征,呈现出稳定一致的注意力分布。不同数据集的注意力分布如图 12(a)~(c)所示,PRONOSTIA 1-3 集中于标记 110-140 与 150-157;PRONOSTIA 2-3 聚焦于 175-200;自采轴承 3 则主要分布在 105 号标记之后。这些结果表明,模型能自适应地捕捉不同数据集中的关键退化特征。

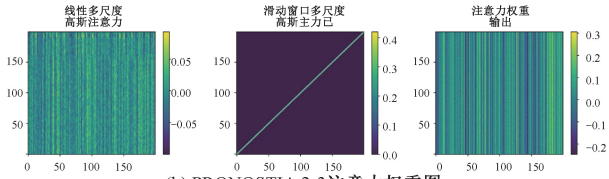
4.3 复杂度分析

为验证所提方法的优越性,本文将其与多种常用方法进行复杂度对比分析。所选对比模型包括传统 RNN、自注意力机制及状态空间模型。

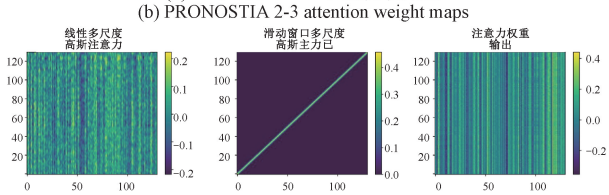
表 7 为各方法在训练阶段中关键网络层的时、空复杂度,包括所提小波倒谱 SSM 分词器层、轻量化注意力机制及其他对比模型中的典型网络层(其中 L 为序列长度,d 为输入维度)。得益于轻量化设计,掩码注意力层的复杂度已降至与传统无掩码注意力相当,未引入额外计算负担。滑动窗口注意力虽理论复杂度较高,但由于其计算范围限定于局部窗口,对整体资源与耗时影响有限。此外,在引入小波倒谱强化退化特征后,分词器仍维持了与基础 SSM 一致的低复杂度。该设计在增强特征表达能力的同时,有效控制了模型整体的计算开销。



(a) PRONOSTIA 1-3注意力权重图
(a) PRONOSTIA 1-3 attention weight maps



(b) PRONOSTIA 2-3注意力权重图
(b) PRONOSTIA 2-3 attention weight maps



(c) 自采轴承数据3注意力权重图
(c) Self-collected bearing data 3 attention weight maps

图 12　注意力权重图

Fig. 12　Attention weight maps

表 7　模型各层复杂度比较

Table 7　Complexity analysis for different models per layer

模型	各层时间复杂度	各层空间复杂度	串行操作时间复杂度	最大输入序列长度
线性高斯注意力层	$o(L^2 \times d)$	$o(L^2 \times d)$	$o(L)$	—
滑动窗口高斯注意力层	$o(L^3 \times d)$	$o(L^3 \times d)$	$o(L)$	—
小波倒谱 SSM 分词器	$o(L \times d)$	$o(L \times d)$	(L)	(L)
RNN	$(L \times d^2)$	$(L \times d^2)$	(L)	(L)
自注意力	$o(L^2 \times d)$	$o(L^2 \times d)$	$o(1)$	$o(1)$
状态空间	$o(L \times d)$	$o(L \times d)$	(L)	(L)

本文模型与 RNN、自注意力及状态空间模型在实际训练时间上的表现对比如表 8 所示。实验在配备 Intel Core i9-12900K 与 NVIDIA GeForce RTX 4090 的服务器上进行，PRONOSTIA 数据集选用轴承 1-1，NSK-6007 选用轴承 2 作为训练数据。结果表明，所提模型在训练效率上显著优于 RNN、LSTM、Transformer、SAF-LSTM 和 TFT 模型。尽管 MAMBA 模型因其结构简单训练耗时较短，本文方法在 PRONOSTIA 数据集上仍实现了约 20%～30%的训练时间降低，在 NSK-6007 数据集上则降低了约 15%～50%，验证了所提轻量化多尺度高斯注意力模块的有效性。

4.4　消融实验

为评估各模块的贡献，本文设计了消融实验，设置如下：1）移除小波倒谱 SSM 分词器，改用 FPCA 降维特征，并利用 512 维线性层完成词嵌入与位置嵌入；2）将特征筛选门控机制替换为普通线性层；3）将轻量化多尺度高斯注意力替换为普通线性注意力和滑动窗口注意力；4）将多尺度高斯掩码替换为均匀注意力掩码，以验证其尺度自适应能力的作用。

表 8　不同模型的训练时间

Table 8　Training time of different models　(s)

对比模型	PRONOSTIA	NSK-6007
RNN	13. 22	12. 53
TFT	15. 20	12. 49
SAF-LSTM	11. 03	10. 56
LSTM	12. 16	12. 78
Transformer	8. 41	6. 87
MAMBA	2. 42	2. 06
所提方法	9. 43	5. 88

消融实验结果如图 13 所示。各模块均对预测精度具有正向影响：移除小波倒谱 SSM 分词器显著削弱了模型对退化初期细微特征的捕捉能力，导致预测误差大幅上升；删除轻量化注意力层影响了全局特征提取，预测性能有所下降；移除特征筛选门控对序列依赖建模产生一定影响，但整体精度下降相对较小；而去除多尺度高斯掩码机制后，模型对周期性依赖关系的建模能力显著减弱，预测误差同样显著升高。

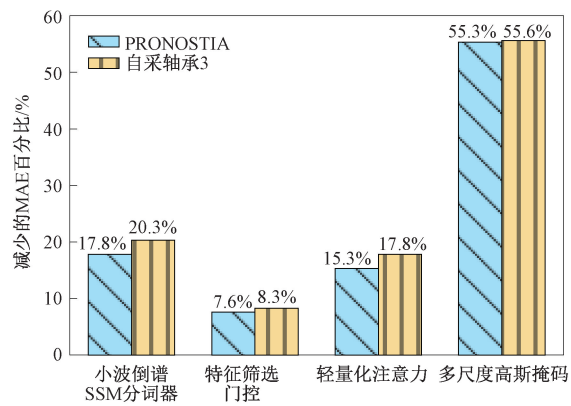


图 13　消融测试中 MAE 的变化

Fig. 13　Changes in MAE across ablation tests

5　结　论

本文融合了 SSM 分词器与轻量化多尺度高斯掩码注意力机制，实现了对滚动轴承退化特征的精准捕捉和剩余寿命的高效预测。针对传统方法在复杂非线性退化过程中难以捕捉多阶段动态变化与物理位置信息且计算复杂度高的问题，SSM 分词器通过引入小波倒谱和门控特征选择实现了信号中局部及多尺度信息的动态调控和降噪提取；同时，改进的注意力模块采用线性多尺度高斯掩码结合滑动窗口策略，不仅保证了全局长距离依赖的

捕捉,还兼顾了对局部细节的关注,从而大幅降低了长序列处理的成本。在对比实验中,该方法在误差控制和计算复杂度上均表现出明显优势,证明了其在轴承退化趋势学习和寿命预测上的优越性。然而,模型仍存在跨工况泛化能力不足及分词器内部决策不够直观的问题,未来可考虑引入迁移学习、元学习及因果推理和可视化技术以提升适应性和解释性。

参考文献

[1] 钟辉, 郭瑜, 高国泽. 参数自适应 SMHD 滚动轴承 IAS 信号特征提取方法[J]. 电子测量与仪器学报, 2023, 37(12):10-17.
ZHONG H, GUO Y, GAO G Z. Parameter adaptive SMHD rolling bearing IAS signal feature extraction method[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(12):10-17.

[2] 黄星华, 吴天舒, 杨龙玉, 等. 一种面向旋转机械的基于 Transformer 特征提取的域自适应故障诊断[J]. 仪器仪表学报, 2022, 43(11):210-218.
HUANG X H, WU T SH, YANG L Y, et al. Domain adaptive fault diagnosis for rotating machinery based on Transformer feature extraction[J]. Chinese Journal of Scientific Instrument, 2022, 43(11):210-218.

[3] MO Y, WU Q, LI X, et al. Remaining useful life estimation via Transformer encoder enhanced by a gated convolutional unit [J]. Journal of Intelligent Manufacturing, 2021, 32: 1997-2006.

[4] SU X, LIU H, TAO L, et al. An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive Transformer model[J]. Computers & Industrial Engineering, 2021, 161: 107531.

[5] LIM B, ARIK SÖ, LOEFF N, et al. Temporal fusion Transformers for interpretable multi-horizon time series forecasting[J]. International Journal of Forecasting, 2021, 37(4): 1748-1764.

[6] ZHU J, JIANG Q, SHEN Y, et al. Res-HSA: Residual hybrid network with self-attention mechanism for rul prediction of rotating machinery [J]. Engineering Applications of Artificial Intelligence, 2023, 124: 106491.

[7] 刘晓磊, 刘连胜, 王璐璐,等. 基于状态空间模型的飞机 APU 在翼 RUL 预测方法[J]. 仪器仪表学报, 2021,42(2):45-54.
LIU X L, LIU L SH, WANG L L, et al. A state-space model-based method for predicting aircraft APU remaining useful life on wings [J]. Chinese Journal of Scientific Instrument, 2021, 42(2): 45-54.

[8] KRISHNAN R, SHALIT U, SONTAG D. Structured inference networks for nonlinear state space models[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017.

[9] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces [J]. ArXiv preprint arXiv: 2312.00752,2023.

[10] DAO T, GU A. Transformers are SSMS: Generalized models and efficient algorithms through structured state space duality[C]. Proceedings of the 41st International Conference on Machine Learning, 2024, 235: 10041-10071.

[11] RAN B, PENG Y, WANG Y. Bearing degradation prediction based on deep latent variable state space model with differential transformation[J]. Mechanical Systems and Signal Processing, 2024, 220: 111636.

[12] WEIKUN D, NGUYEN K T P, GOGU C, et al. Towards generalizable PHM: An interpretable liquid operators-informed selective state space model for cross-scenario prognostic applications [DB/OL]. Social Science Research Network, 2025-02-27. https://dx. doi. org/10. 2139/ssrn. 5149848.

[13] BAI R, NOMAN K, FENG K, et al. A two-phase-based deep neural network for simultaneous health monitoring and prediction of rolling bearings [J]. Reliability Engineering & System Safety, 2023, 238: 109428.

[14] NI Q, JI J C, HALKON B, et al. Physics-informed residual network for rolling element bearing fault diagnostics [J]. Mechanical Systems and Signal Processing, 2023, 200: 110544.

[15] HENDRYCKS D, GIMPEL K. Gaussian error linear units [J]. ArXiv preprint arXiv:1606.08415,2016.

[16] 周炜杰, 李智, 张绍荣,等. 融合多级注意力与多尺度信息的铁轨缺陷分割网络[J]. 电子测量与仪器学报,2025,39(7): 140-150.
ZHOU W J, LI ZH, ZHANG SH R, et al. Railway defect segmentation network integrating multi-level attention and multi-scale information [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(7): 140-150.

[17] ARORA S, EYUBOGLU S, ZHANG M, et al. Simple linear attention language models balance the recall-throughput tradeoff [J]. ArXiv preprint arXiv: 2402.18668, 2024.

[18] 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,

2012：1-8.

作者简介

李嘉恒，2025 年于江苏大学获得学士学位，现为江苏大学硕士研究生，主要研究方向为信号处理、机械系统智能故障诊断与寿命预测。

E-mail：2212503101@ stmail. ujs. edu. cn

Li Jiaheng received his B. Sc. degree from Jiangsu University in 2025. Now he is a M. Sc. candidate at Jiangsu University. His main research interests include signal processing and machinery fault diagnosis.

常致远，2022 年于南京理工大学紫金学院获得学士学位，2025 年于江苏大学获得硕士学位，主要研究方向为信号处理、智能故障诊断。

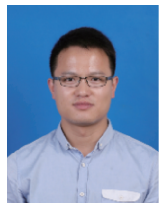
E-mail：2870643950@ qq. com

Chang Zhiyuan received his B. Sc. degree from Zijin College of Nanjing University of Science and Technology in 2022, and M. Sc. degree from Jiangsu University in 2025. His main research interests include signal processing and intelligent fault diagnosis.

樊薇（通信作者），分别在 2012 年和 2015 年于苏州大学获得学士学位和硕士学位，2018 年于香港城市大学获博士学位，现为江苏大学教授，主要研究方向为信号处理和机械故障诊断。

E-mail：weifan@ ujs. edu. cn

Fan Wei（Corresponding author）received her B. Sc. degree and the M. Sc. degree from Soochow University in 2012 and 2015, respectively, and Ph. D. degree from City University of Hong Kong in 2018. Now she is a professor at Jiangsu University. Her main research interests include signal processing and machinery fault diagnosis.



陈超，分别在 2011 年和 2014 年于江苏大学获得学士学位和硕士学位，2020 年于东南大学获博士学位，现为江苏大学讲师、硕士生导师，主要研究方向为信号处理、机械系统状态监测与智能故障诊断。

E-mail：chenchao@ ujs. edu. cn

Chen Chao received his B. Sc. degree and M. Sc. degree from Jiangsu University in 2011 and 2014, respectively, and Ph. D. degree from Southeast University in 2020. Now he is a lecturer and M. Sc. supervisor at Jiangsu University. His main research interests include signal processing, machine condition monitoring and intelligent fault diagnosis.

许桢英，1999 年于合肥工业大学获得学士学位，2004 年于合肥工业大学获得博士学位，现为江苏大学教授、博士生导师，主要研究方向为声、光无损检测理论与技术。

E-mail：xuzhenying@ ujs. edu. cn

Xu Zhenying received her B. Sc. degree from Hefei University of Technology in 1999, Ph. D. degree from Hefei University of Technology in 2004. Now she is a professor and Ph. D. supervisor at Jiangsu University. Her main research interests include the theory and technology of acoustic and optical nondestructive testing.