



北京航空航天大学学报

*Journal of Beijing University of Aeronautics and Astronautics*

ISSN 1001-5965, CN 11-2625/V

## 《北京航空航天大学学报》网络首发论文

题目: 基于 Transformer 模型的滚动轴承剩余使用寿命预测方法  
作者: 周哲韬, 刘路, 宋晓, 陈凯  
DOI: 10.13700/j.bh.1001-5965.2021.0247  
收稿日期: 2021-05-10  
网络首发日期: 2021-07-12  
引用格式: 周哲韬, 刘路, 宋晓, 陈凯. 基于 Transformer 模型的滚动轴承剩余使用寿命预测方法. 北京航空航天大学学报.  
<https://doi.org/10.13700/j.bh.1001-5965.2021.0247>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于 Transformer 模型的滚动轴承剩余使用寿命预测方法

周哲韬<sup>1</sup>, 刘路<sup>1</sup>, 宋晓<sup>2,✉</sup>, 陈凯<sup>3</sup>

(1. 北京航空航天大学 自动化科学与电气工程学院, 北京 100191; 2. 北京航空航天大学 网络空间安全学院, 北京 100191;

3. 南京航空航天大学 机电学院, 南京 210016)

\*通信作者 E-mail: songxiao@buaa.edu.cn

**摘要** 准确的滚动轴承剩余使用寿命(remaining useful life, RUL)预测对保证机械安全运行和减小维修损失起着至关重要的作用。为提高滚动轴承 RUL 预测准确率,提出一种基于 Transformer 模型的轴承 RUL 预测方法,充分利用其自注意力机制与编码器-解码器结构的优势,解决轴承 RUL 预测中序列过长而导致的记忆力退化问题,挖掘出输入特征与轴承 RUL 之间复杂映射关系。同时,采用三角函数变换与累积变换来修正输入特征的单调性与趋势性,使其能更好地表征滚动轴承的退化过程。在 PHM2012 数据集上的实验结果表明,所提出的基于 Transformer 模型的滚动轴承 RUL 预测方法相比于其他文献平均绝对误差分别降低了 9.25%、28.63%、34.14%,平均得分分别提高了 2.78%、19.79%、29.38%;在 XJTU-SY 数据集上的实验结果表明,所提出的预测方法相比于其他文献均方根误差降低了 17.4%,平均得分提高了 18.6%,进一步证明了其可行性与优越性。

**关键词** 滚动轴承; 剩余使用寿命预测; Transformer 模型; 自注意力机制; 累积变换

中图分类号 V229+2; TH133.33; TP183

文献标识码: A

DOI: 10.13700/j.bh.1001-5965.2021.0247

## Remaining useful life prediction method of rolling bearing based on Transformer model

ZHOU Zhetao<sup>1</sup>, LIU Lu<sup>1</sup>, SONG Xiao<sup>2,✉</sup>, CHEN Kai<sup>3</sup>

(1. School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;

2. School of Cyber Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;

3. College of Mechanical&Electrical Engineering, Nanjing University of Aeronautics And Astronautics, Nanjing 210016, China)

\*Tel.:13671289267 E-mail: songxiao@buaa.edu.cn

**Abstract** Accurate rolling bearing RUL(remaining useful life) prediction is extremely important to assure machine's safety working and decrease damage repair.To improve the accuracy for rolling bearing RUL prediction,proposed a bearing RUL prediction method based on Transformer model,made full use of its self-attention mechanism and the advantages of encoder-decoder structure,solved the memory degradation problem caused by too long sequence in bearing RUL prediction,found out the dependent relationship between the input feature and the bearing degradation degree.Meanwhile,trigonometric function transform and cumulative transform are used to correct the feature's monotony and tendency,represented the rolling bearing degradation process better.Experimental results on PHM2012 dataset showed that compared with other prediction methods, the average absolute error of RUL prediction based on Transformer model is reduced by 9.25%, 28.63% and 34.14%,and the average ss were increased by 2.78%,19.79% and 29.38%. Experimental results on XJTU-SY dataset showed that compared with other prediction methods, the Root-mean-square error of RUL prediction based on Transformer model is reduced by 17.4%,and the average ss were increased by 18.6%,which indicates higher feasibility and superiority.

**Key words** rolling bearing; remaining useful life; Transformer model; self-attention;cumulative operation

收稿日期: 2021-05-10

基金项目: 国家重点研发计划(2018YFB1702703)

Fund: The National Key Research and Development Program of China (2018YFB1702703);

网络首发时间: 2021-07-12 15:55:57 网络首发地址: <https://kns.cnki.net/kcms/detail/11.2625.V.20210712.1407.004.html>

滚动轴承被称为“工业的关节”，是旋转机械设备的关键零件，广泛应用于各个工业领域<sup>[1]</sup>。比如在航空航天领域，轴承是航空发动机承力传动系统<sup>[2]</sup>以及航天器控制力矩陀螺<sup>[3]</sup>等设备的重要组成部分，其性能与质量直接影响了航空航天设备的安全性及可靠性。轴承若发生故障，轻则将导致设备故障、经济损失，重则将引发安全事故、人员伤亡。因此，准确、及时地预测轴承的剩余使用寿命（remaining useful life, RUL）对工业生产安全而言具有重大研究意义。

随着传感器、存储、网络传输等新技术的快速发展，轴承运行过程产生了大量的监测数据，利用这些数据来挖掘轴承退化信息，实现精确的 RUL 预测，是当前的研究热点<sup>[4]</sup>。基于数据驱动的轴承剩余使用寿命预测主要有两个步骤<sup>[5]</sup>：①从原始信号中提取具有趋势性的特征，来表征轴承性能退化的过程；②构建预测模型，使用传统机器学习或是深度学习的方法，利用其拟合能力建立特征与剩余寿命间的关系。

特征提取是进行轴承 RUL 预测的前提，目的是提取出符合轴承退化趋势的特征用于表征退化过程。传统的特征提取方法是计算原始振动信号的统计特征如均方根、峭度、峰峰值等，以及构造一些新的统计特征，如王冰<sup>[6]</sup>采用的基于多尺度形态分解谱熵，李洪儒<sup>[7]</sup>采用的二元多尺度熵。近年来，深度学习凭借其卓越的非线性函数自动映射能力在滚动轴承特征提取领域得到广泛应用。如 Ren<sup>[8]</sup>提出使用时间卷积网络来从更大的振动信号感受野中提取特征；杨宇<sup>[9]</sup>提出一种改进的深度信念网络，直接以滚动轴承原始振动信号作为网络输入，经过逐层抽象表示，挖掘出原始振动信号深层本质特征。

但是上述时间卷积网络、深度信念网络等深度学习方法往往需要大量标签数据进行有监督微调，标签数据的缺乏与难以获取严重制约了深度学习特征提取，而传统统计特征在单调性与趋势性上相比于轴承退化趋势往往不尽人意。针对这些问题，鉴于轴承退化是一个累积退化过程，本文对传统特征采用累积变换的方式优化其单调性与趋势性，使其在表征轴承退化过程中有更好的表现。

在对原始振动信号进行特征提取的基础上，需要构建预测模型来对轴承 RUL 进行预测。轴承 RUL 预测本质上是时间序列预测问题，而循环神经网络（Recurrent Neural Network, RNN）在时间序列处理方面具有明显的优越性，从而被广泛应用于轴承 RUL 预测领域。Guo<sup>[10]</sup>等人采用 RNN 提取表征轴承退化的特征量进行轴承剩余寿命预测，并用指数模型验证预测效果，取得了很高的预测精度。Chen<sup>[11]</sup>提取频谱的五个带通能量值作为特征，提出了一种具有注意力机制的编解码框架的 RNN 用于轴承寿命预测。康守强<sup>[12]</sup>等利用稀疏自动编码器进行特征提取，采用双向长短期记忆神经网络进行轴承剩余寿命预测。

在处理时间序列时，RNN 下一时刻的输出取决于先前时刻的输出以及当前时刻的状态，这在本质上是一种串行的运行方式，会严重制约模型运行速度<sup>[13]</sup>。同时，RNN 的每一次递归都伴随着信息的损耗，导致其在输入长序列的条件下捕捉依赖关系的能力迅速衰减，即出现记忆力退化现象。针对以上问题，本文将采用基于 Transformer 的预测模型来进行轴承 RUL 预测。Transformer 模型的核心原理是自注意力机制(self-attention)，其在具体实现上主要以矩阵乘法为基础，从而可以捕捉输入序列中任意向量之间的依赖关系，不受向量之间距离的影响。同时 self-attention 的这种运行原理实现了并行计算，相比于只能串行计算的 RNN 极大地提升了运行速度<sup>[14]</sup>。基于以上特性，Transformer 一经提出就引起了研究者的广泛关注，在机器翻译、阅读理解、文本摘要等多个领域展现出了优异的应用效果。本文将 Transformer 模型引入到滚动轴承 RUL 预测领域，提出一种基于 Transformer 模型的滚动轴承剩余使用寿命预测方法，属于 Transformer 模型的全新应用，为其在工业领域增加了新的应用场景。

综上所述，本文提出了一种基于 Transformer 模型的滚动轴承剩余使用寿命预测方法。首先对滚动轴承原始振动信号的统计特征进行改进，采用累积变换来优化特征的单调性与趋势性，使其能够更好地表征滚动轴承的退化过程。然后提出一种基于 Transformer 模型的轴承 RUL 预测模型，利用自注意力机制来解决轴承 RUL 预测中序列向量之间距离过大而导致的记忆力退化问题。最后利用一次函数对 RUL 预测值进行拟合，实现对滚动轴承的 RUL 预测。

## 1 累积特征提取

滚动轴承常用的状态监测数据包括振动、声波和温度信号，其中振动信号得到了广泛的分析，因为它提供了大量关于轴承内部异常的信息。但是滚动轴承的振动信号在采集时会不可避免地混入大量无意义的噪声信号，这对轴承 RUL 的准确预测是不利的。因此本文采用离散小波变换对振动信号进行去噪，通过小波分解、细节分量阈值处理和小波重构来剔除无意义的噪声，将去噪后的振动信号进行特征提取。

本文对去噪后的振动信号进行三角函数变换与统计特征提取。三角函数可以是单调递增或者单调递减的，如反双曲正弦(asinh)和反正切(atan)，它们能将原始数据转换为更低的尺度，所以将其作为振动信号的退化特征进行补充。三角函数变换按以下方式执行：用三角函数逐因素操作振动信号  $X(x_j, j=1, 2, \dots, n)$ ，并进行标准差计算以提取特征。同时，选取一些滚动轴承 RUL 预测常用的统计特征作为退化特征，具体信息如表 1 所示。

在现实生产环境中，滚动轴承会不可避免地经历退化过程直至故障，而能够反映轴承退化趋势的特征将有助于轴承 RUL 的准确预测。尽管表 1 所示的特征是基于专家知识提取的，并且在一些应用中被证明是有效的，但是传统的统计特征的单调性与趋势性相比于轴承退化趋势往往不尽如人意，不能很好地表征退化趋势。针对这个问题，鉴于轴承退化是一个累积退化过程，本文采用了一种有效的特征变换方法，即将提取的特征变换为相应的累积形式。具体来说，累积变换是通过在时间序列上应用一个累积函数来进行的，在这个时间序列中，同时逐点执行累加操作和缩放操作，然后使用累积特征来表征退化的趋势。累积变换执行如式(1)所示：

$$c_{jn} = \frac{\sum_{i=1}^n f_j(i)}{\sqrt{\sum_{i=1}^n f_j(i)}}, j = 1, 2, \dots, N \quad (1)$$

式中： $c_{jn}$  表示第  $j$  个特征  $f_j(i)$  在  $n$  个样本中的累积变换结果。从式中可以看出，如果存在噪声则其会随着累加操作而逐步被放大，所以累积变换算法容易受到不必要的噪声的影响，因此在变换前对特征进行平滑处理是非常必要的。本文采用 Savitzky-Golay 滤波器来提高数据的精度同时不失真信号的趋势。平滑过程通过线性最小二乘法采用低次多项式拟合相邻数据点的连续子集来实现。

在累积变换之后，设计度量来评估转换后特征的质量是至关重要的。一般来说，有两个常用的指标来评估构造的特征对 RUL 预测的适用性，即单调性和趋势性。

单调性反映了特征的增减趋势，是退化特征的重要组成部分。其计算方法如式(2)所示：

$$M = \frac{\left| \text{No.of } \frac{d}{dx} > 0 - \text{No.of } \frac{d}{dx} < 0 \right|}{n-1} \quad (2)$$

式中： $n$  表示总观测值。 $M=1$ ， $M=0$  分别表示特征的高度单调趋势和不单调趋势。

趋势性是用来描述退化状态如何随运行时间变化的。换句话说，这个度量表征了一个与特征及其与时间的相关性有关的函数。其计算方法如式(3)所示：

$$T = \frac{\left| \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \right|}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (3)$$

式中： $x_i$  和  $y_i$  分别代表时间和特征的值， $N$  为时间序列的长度。 $T$  的范围为-1 到 1，-1 代表特征值严格递减，1 表示特征值严格递增。

## 2 基于 Transformer 的滚动轴承 RUL 预测模型

在基于数据驱动的滚动轴承 RUL 预测的最新研究中，循环神经网络获得了广泛的应用（包括 RNN，长短期记忆神经网络，门控循环单元及各类变种模型），但其串行计算方式和长序列依赖捕捉



不敏感仍严重制约了模型运行速度与预测精度。因此,本文提出一种基于 Transformer 结构的滚动轴承 RUL 预测模型,旨在提高轴承 RUL 预测的精度与速度。

Transformer 是一种基于注意力机制的神经网络,在自然语言处理等领域取得了良好的效果,但是其独特的结构限制了直接应用于 RUL 预测领域,如轴承 RUL 预测的数据与自然语言的词向量间的有效映射等问题。所以本文针对滚动轴承 RUL 预测场景进行了改进,结构如图 1 所示,模型由四个部分组成,分别为位置编码、编码器、解码器以及全连接神经网络。

表 1 特征和相应的计算公式  
Table 1 Feature and corresponding formulas

特征	公式
反正切值标准差	$X_{\text{atan}} = \sigma \left( \log [x_i + \sqrt{x_i^2 + 1}] \right)$
反双曲正弦标准差	$X_{\text{asinh}} = \sigma \left( \frac{i}{2} \log \left( \frac{i + x_i}{i - x_i} \right) \right)$
标准差	$X_{\text{sd}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$
峰值	$X_{\text{p}} = \max  X $
峰峰值	$X_{\text{p-p}} = \max(x_i) - \min(x_i)$
均方根	$X_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
上边界	$X_{\text{upper}} = \max(x_i) + \frac{1}{2} \frac{\max(x_i) - \min(x_i)}{n-1}$
脉冲因子	$X_{\text{if}} = \frac{X_{\text{p}}}{\frac{1}{n} \sum_{i=1}^n  x_i }$
峰值因子	$X_{\text{cf}} = \frac{X_{\text{p}}}{X_{\text{rms}}}$
裕度系数	$X_{\text{mf}} = \frac{X_{\text{p}}}{\left( \frac{1}{n} \sum_{i=1}^n \sqrt{ x_i } \right)^2}$
能量	$X_{\text{e}} = \sum_{i=1}^n x_i^2$
峭度	$X_{\text{k}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^4}{X_{\text{sd}}^4}$
平均绝对值	$X_{\text{mav}} = \frac{1}{n} \sum_{i=1}^n  x_i $
偏度	$X_{\text{sk}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3}{X_{\text{sd}}^3}$

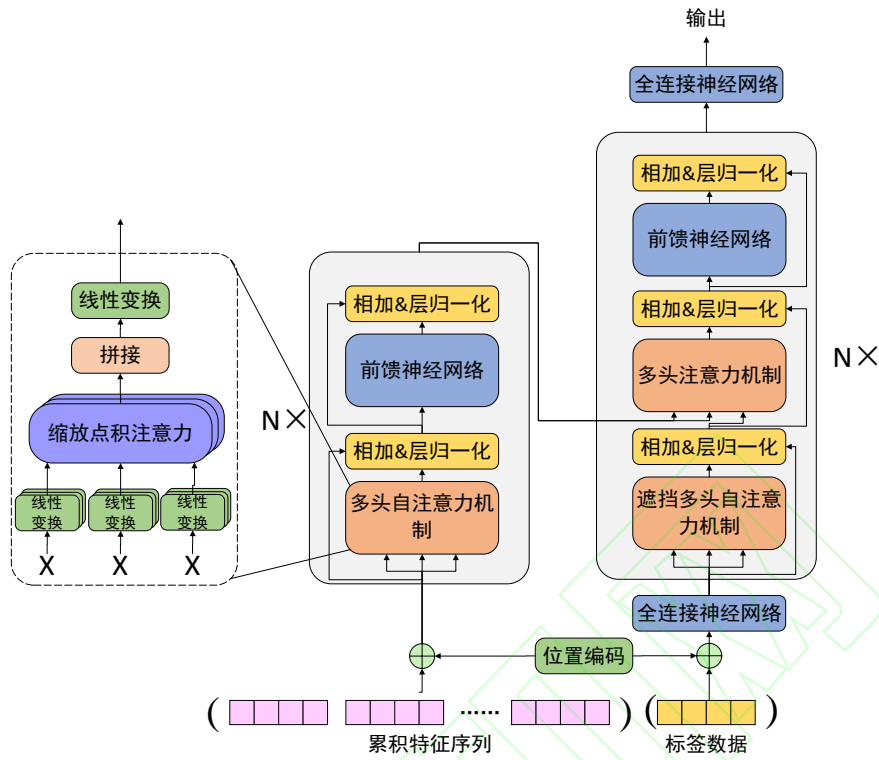


图 1 基于 Transformer 的轴承 RUL 预测模型  
Fig.1 Bearing RUL prediction model based on Transformer

基于 Transformer 的滚动轴承 RUL 预测模型的核心在于其编码器与解码器结构，编码器与解码器均由 6 层相同层堆叠而成，其中编码器每一层包含多头自注意力机制和前馈神经网络两个子层，解码器每一层包含遮挡多头自注意力机制、编码-解码器多头注意力机制和前馈神经网络三个子层。

编码器负责将输入的滚动轴承特征序列进行编码，映射为包含输入特征信息的中间向量，其核心原理为自注意力机制。自注意力机制是注意力机制的变体，减少了对外部信息的依赖，更擅长捕捉数据或特征的内部相关性，主要是通过计算特征向量间的相似度来表征相关性，以此解决长距离依赖捕捉问题。自注意力机制的目的是从输入的滚动轴承特征序列中筛选出少量重要的信息，使用权重代表信息的重要性，使得模型聚焦于更为重要的信息上。

自注意力机制采用缩放点积注意力来计算特征矩阵的注意力值，先对查询矩阵和键矩阵进行点积与 SoftMax 归一化来计算权重系数，再根据权重系数对值矩阵进行加权求和，如公式(4-7)所示。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (4)$$

$$\mathbf{Q} = \mathbf{X}_f \mathbf{W}^Q \quad (5)$$

$$\mathbf{K} = \mathbf{X}_f \mathbf{W}^K \quad (6)$$

$$\mathbf{V} = \mathbf{X}_f \mathbf{W}^V \quad (7)$$

式中： $\mathbf{Q}$  为查询矩阵， $\mathbf{K}$  为键矩阵， $\mathbf{V}$  为值矩阵，这三个矩阵由输入的特征矩阵  $\mathbf{X}_f$  分别与对应的权重矩阵  $\mathbf{W}^Q$ 、 $\mathbf{W}^K$ 、 $\mathbf{W}^V$  相乘得到， $d$  为  $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$  的维数。

自注意力机制可以让模型聚焦于输入特征的某些重要信息，但是单一的注意力机制只能在一个表现空间中学习到相关信息。为了综合输入序列包含的信息的重要性，采用多头自注意力机制在不同位置共同关注来自不同表现子空间的信息。多头自注意力机制其实是多个自注意力机制的拼接，利用多个自注意力头分别学习不同表现子空间的信息，再将多个注意力值进行拼接与线性变换，得到最终的注意力值，实现对不同约束条件的建模表达，如公式(8-9)所示。

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(h_1, \dots, h_m) \cdot \mathbf{W} \quad (8)$$

$$h_i = \text{Attention}(\mathbf{XW}_i^Q, \mathbf{XW}_i^K, \mathbf{XW}_i^V) \quad (9)$$

式中： $\mathbf{W}_i^Q$ 、 $\mathbf{W}_i^K$ 、 $\mathbf{W}_i^V$ 为第  $i$  个注意力头  $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$  的权重矩阵， $\mathbf{W}$  是多头注意力权重矩阵， $m$  是注意力头的数量，Concat 函数用于拼接各个注意力头计算得到的输出值。

解码器负责将编码器输出的中间向量解码为输出序列，其核心原理为编码-解码多头注意力机制以及遮挡多头自注意力机制。编码器中的多头自注意力机制注重学习输入特征序列自身向量之间的依赖关系，但是为了提高滚动轴承 RUL 预测的精准度，还应考虑输入特征序列与标签数据之间的依赖关系，所以在解码器中采用了编码-解码多头注意力机制。编码-解码多头注意力机制与编码器的多头自注意力机制不同，其键矩阵  $\mathbf{K}$  与值矩阵  $\mathbf{V}$  来自编码器的输出，包含输入序列的信息，查询矩阵  $\mathbf{Q}$  来自遮挡多头自注意力机制模块的输出，包含标签数据的信息。

遮挡多头自注意力机制模块的输入为标签数据，其目的是利用多头自注意力机制学习标签数据之间的依赖关系，并将依赖关系输入编码-解码多头注意力机制模块，使整个基于 Transformer 的滚动轴承 RUL 预测模型可以综合学习输入特征向量之间的依赖关系、标签数据之间的依赖关系以及二者相互的依赖关系。需要注意的是，在模型训练过程中，滚动轴承 RUL 的标记数据是已知的，解码器根据并行计算特性同时完成序列每个时间步解码。这会导致在缩放点积注意力计算时，每个时间步的计算会学习到未来的标签数据信息，这是不符合现实规律的，所以需要在解码器的多头自注意力机制中加入遮挡操作，屏蔽未来的标签数据信息。具体的操作是在计算缩放点积注意力时，引入一个下三角及对角线均为 1 上三角为 0 的矩阵，与  $\mathbf{QK}^T$  相乘，使得未来的序列信息置零。

为了解决网络深度增加反而影响轴承 RUL 预测准确率的问题，编码器与解码器的各个子层之间均添加了残差连接操作，关注训练前后差异部分变化，提升训练效果。同时，为了加快网络收敛，提升网络泛化能力，每个子层还同时采用了层归一化操作，如公式(10)所示：

$$o = \text{LayerNorm}(x + \text{Sublayer}(x)) \quad (10)$$

式中：Sublayer 是每个子层内部的函数，本文中为各个注意力机制层处理函数和全连接前馈神经网络处理函数。LayerNorm 为层归一化处理函数。

Transformer 的一系列优点得益于其纯粹的注意力机制构造，但这使其失去了学习序列位置信息的能力。而在滚动轴承 RUL 预测场景中，序列中向量的位置信息代表着时刻信息，对 RUL 预测起到至关重要的作用。针对这个问题，在编码器与解码器的输入序列中加入位置编码操作，将位置信息整合到输入序列当中，如公式(11-12)所示：

$$\text{PE}_{(p, 2i)} = \sin\left(\frac{p}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (11)$$

$$\text{PE}_{(p, 2i+1)} = \cos\left(\frac{p}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (12)$$

式中： $p$  表示特征向量的位置， $d_{\text{model}}$  表示特征向量的维度，每一个特征向量的位置编码都是由不同频率的余弦正弦函数组成的，波长逐渐由  $2\pi$  增长到  $10000 \cdot 2\pi$ 。

Transformer 模型最初是针对自然语言处理领域提出的，其输入特征与标签数据均为维度一致的词向量，相同的维度便于编码-解码多头注意力机制的直接计算。但在轴承 RUL 预测场景中，输入特征为多维数据，而标签则是一维的剩余寿命，两者的维度冲突导致编码-解码多头注意力机制无法运算。针对这个问题，本文在编码器输入端直接采用累积特征序列作为输入，在解码器输入端加入一层全连接网络，将输入标签序列进行升维，在输出端同样加入一层全连接网络进行降维，解决了基于 Transformer 的轴承 RUL 预测模型中输入特征与标签数据维度不一致的冲突。

### 3 滚动轴承 RUL 预测流程

基于累积特征与 Transformer 的滚动轴承 RUL 预测方法流程如图 2 所示，具体步骤为：

1)信号去噪：首先对滚动轴承原始振动信号进行离散小波变换，包括小波分解、细节系数阈值处理和小波重构，在消除异常的噪声信号的同时保留有用的退化信息。

2)特征提取：先根据已有的专家知识，在经过去噪的重构信号上计算出常用的轴承 RUL 预测统计特征，包括峰峰值、标准差、均方根、三角特征等；再利用 Savitzky-Golay 滤波器对所有统计特征进行平滑处理，以减少提取特征的波动和进一步滤除不需要的噪声；然后对结果平滑处理的统计特征进行累积变换，优化特征的单调性与趋势性；最后计算所有累积特征的单调性与

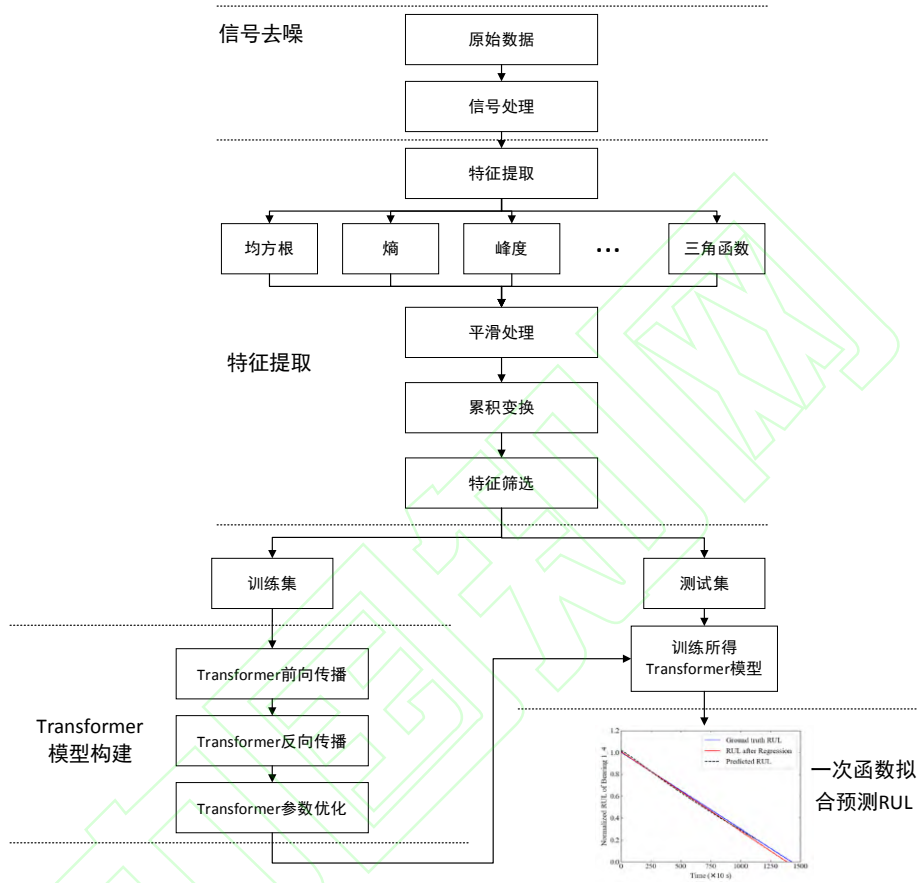


图2 滚动轴承 RUL 预测流程  
Fig.2 Flow chart of RUL prediction for rolling bearings

趋势性值，以此来筛选用于训练与测试的特征，并划分训练集与测试集。

3)Transformer 模型构建：将训练集的特征进行最大最小值归一化作为 Transformer 模型的输入，将寿命比值  $P$ （剩余使用寿命与全寿命的比值）作为模型的标签数据，与轴承运行时间满足一次函数模型<sup>[15-18]</sup>。计算训练模型的均方误差(Mean Squared Error, MSE)作为损失函数，采用 Adam 优化算法进行模型训练与优化，并采用 Dropout 技术防止模型过拟合。

4)测试集验证：将测试集的累积特征输入到已经训练好的 Transformer 模型中，预测出特征对应的寿命比值  $P$ 。由于案例一中 PHM2012 测试集提供的是非全寿命数据（即记录滚动轴承从运行开始到某个时间点的数据），缺少退化过程末期的数据，所以需要前半部分特征预测的寿命比值  $P$  进行拟合，以此预测滚动轴承的全寿命运行时间。由于  $P$  为寿命比值，与轴承运行时间满足一次函数模型，所以本文采用一次线性函数拟合来预测滚动轴承的 RUL。



## 4 实验验证

### 4.1 案例一：PHM2012 数据集验证

#### 4.1.1 数据集介绍

为了验证所提出的 RUL 预测方法的有效性与先进性, 本文首先采用 IEEE 协会在 PHM2012 数据挖掘挑战赛中提供的数据集, 该数据集由 PRONOSTIA 实验平台采集获得, 采集装置如图 3 所示。采集装置通过速度传感器和力矩传感器实现了 3 种不同工况下的滚动轴承加速退化, 进行了 17 次轴承全寿命周期试验, 利用加速度传感器采集了轴承水平方向与垂直方向共计 17 组原始振动信号数据集, 并将其划分为 6 组训练集与 11 组测试集。采集装置每隔 10 s 采集 0.1 s 内的原始振动信号, 采集频率为 25.60 kHz, 即每 10 s 采集 2560 组样本数据。数据集的详细信息如表 2 所示, 具体实验内容参考文献[19]。根据文献[20,21]中的相关研究, 水平振动信号通常比垂直振动信号给出更多的有用信息来跟踪轴承退化。因此, 本研究仅使用水平振动信号进行实验。

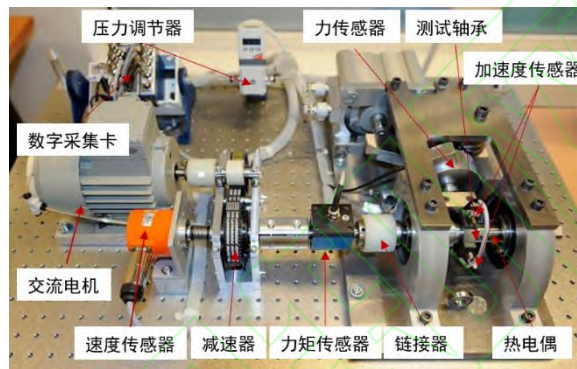


图 3 PRONOSTIA 采集平台  
Fig.3 The PRONOSTIA platform

表 2 PHM2012数据集工况信息

Table 2 Operating condition of PHM2012 dataset

工况	径向力(N)	转速 (r/min)	训练集	测试集
工况 1	4000	1800	Bearing1-1 Bearing1-2	Bearing1-3 Bearing1-4 Bearing1-5 Bearing1-6 Bearing1-7
工况 2	4200	1650	Bearing2-1 Bearing2-2	Bearing2-3 Bearing2-4 Bearing2-5 Bearing2-6 Bearing2-7
工况 3	5000	1500	Bearing3-1 Bearing3-2	Bearing3-3

#### 4.1.2 特征提取与筛选

在对滚动轴承的原始振动信号进行特征提取之前, 本文采用离散小波变换对其进行去噪。其中小波分解采用多贝西四阶小波(db4)为母小波, 分解层数为 4 层, 细节分量阈值处理中采用软阈值函数和最大最小阈值, 最后通过小波重构得到去噪后的信号。

根据第 3 章的预测流程, 原始振动信号去噪之后需要应用三角函数、提取统计特征和进行累积变化。从图 4 可以看出, 传统的统计特征表现出较低的单调性与趋势性, 这不利于 RUL 预测模型学习轴承的退化趋势, 同时较高的尺度也不利于预测模型的训练过程。而结合了标准差与三角函数(提取方案见表 1)的三角特征具有更高的单调性与趋势性以及更低的尺度, 有利于预测模型的训练过程。

目前仍有大量数据驱动的滚动轴承 RUL 预测方法采用传统的时域、频域以及时频域统计特征, 但是这些传统的统计特征往往无法很好地表征轴承的退化趋势, 如图 5 的(a)和(c)所示, 仅经过平滑处理的特征单调性与趋势性各不相同, 单调性较差。而从图 5 的(b)和(d)不难看出, 累积变换后的特征表现出了更好的单调性与趋势性, 说明本文采用的累积变换对于传统的统计特征的趋势性与单调性具有很好的优化作用。

为了更好地说明累积变换的优势与必要性, 本文采用公式(2)(3)量化计算每种统计特征累积变化前后的单调性与趋势性, 表 3 显示了 6 个训练集的统计特征累积变化前后的单调性与趋势性平均值,

从表中可以看出, 累积变化对本文选取的统计特征的单调性与趋势性具有非常明显的改进作用。由于在滚动轴承 RUL 预测中, 特征的单调性与趋势性越好, 越能表征轴承的退化趋势, RUL 预测的精度就越高。基于以上要求, 本文舍弃了经过累积变换后单调性和趋势性仍然较差的平均绝对值, 将剩余 12 个统计特征的累积值输入 Transformer 预测模型进行模型训练。

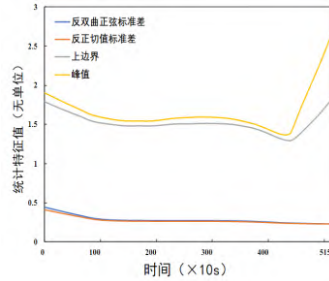
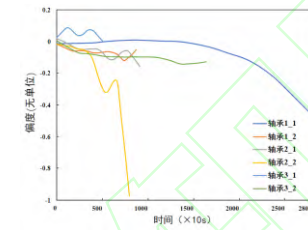
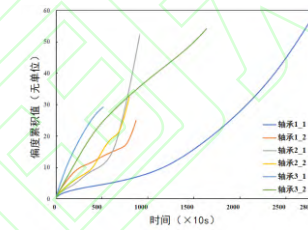


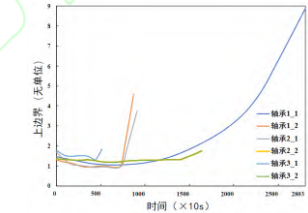
图 4 三角特征与传统特征对比图  
Fig.4 Contrast diagram of trigonometric features and classical features



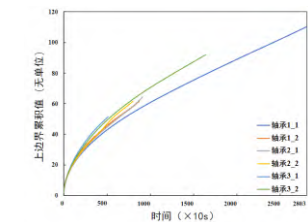
(a) 训练集的偏度对比图  
(a)The skewness comparison diagram of the training set



(b) 训练集的偏度累积值对比图  
(b)The C-skewness comparison diagram of the training set



(c) 训练集的上边界对比图  
(c)The upperbound comparison diagram of the training set



(d) 训练集的上边界累积值对比图  
(d)The C-upperbound comparison diagram of the training set

图 5 累积特征与传统特征对比图  
Fig.5 Contrast diagram of cumulative features and classical features

4.1.3 模型训练与预测结果分析

本文遵照 PHM2012 数据集的划分, 将轴承 1\_1, 1\_2, 2\_1, 2\_2, 3\_1, 3\_2 的全寿命数据作为训练集, 将剩余轴承的非全寿命数据作为测试集, 具体划分信息如表 4 所示。训练集的经过累积变换与筛选后的 12 个累积特征需要输入 Transformer 预测模型进行模型训练。Transformer 预测模型是序列到序列(Sequence-to-Sequence)类型的模型, 需要确定输入数据与标签数据的序列长度, 过短的序列长度会导致 Transformer 的注意力机制不容易学习到各个时间步中的依赖信息, 过长的序列长度则需要巨大的计算量, 降低运算速度。因此, 本文选择 10 个时间步的累积特征作为输入序列, 10 个时间步的寿命比值  $P$  作为标签数据, 寿命比值  $P$  的计算公式如式(13)所示:

$$P = \frac{\text{actRUL}_t}{\text{actRUL}_0}$$

(13)

式中:  $\text{actRUL}_0$  表示轴承实际的全寿命数值, 例如轴承 1\_1 从投入使用到报废共采集了 2803 组数据, 其寿命数值为 0~2802, 则其  $\text{actRUL}_0$  为 2802, 由于数据采样间隔为 10 s 一组, 所以轴承 1\_1 真实的全寿命时间为 28030 s;  $\text{actRUL}_t$  则表示  $t$  时刻时轴承的剩余使用寿命数值。从寿命比值  $P$  的计算公式不难看出, 其符合一次函数关系式, 所以在模型验证阶段本文也采用了拟合一次线性函数的方式来预测 RUL。

表 3 累积变换前后比较  
Table 3 Cumulative transformation comparison

传统特征	单调性	趋势性	累积特征	单调性	趋势性
反双曲正弦标准差	0.57	0.71	反双曲正弦标准差	1	1
反正切值标准差	0.58	0.76	反正切值标准差	1	1
能量	0.50	0.64	能量	1	1
偏度	0.60	0.83	偏度	1	1
峭度	0.39	0.80	峭度	1	1
上边界	0.26	0.36	上边界	1	1
均方根	0.50	0.61	均方根	1	1
脉冲因子	0.54	0.78	脉冲因子	1	1
峰值因子	0.52	0.75	峰值因子	1	1
峰峰值	0.27	0.37	峰峰值	1	1
裕度系数	0.51	0.77	裕度系数	1	1
标准差	0.50	0.60	标准差	1	1
平均绝对值	0.59	0.70	平均绝对值	0.87	0.96

为了更好地让预测模型学习到累积特征与轴承 RUL 之间的映射关系, 本文将输入序列的步长设置为 1 个时间步, 如第一组输入为 1~10 时间步的累积特征序列, 第二组输入则为 2~11 时间步的累积特征序列, 每一组输入序列只留取输出序列中最后一个时间步的寿命比值  $P$  作为预测结果, 以此达到数据增强的效果。

表 4 实验数据 (PHM2012轴承数据集)  
Table 4 Experimental data(PHM2012 Datasets)

数据集划分	轴承编号	非全寿命时间 (×10s)	全寿命时间 (×10s)
训练集	1_1	—	2803
	1_2	—	871
	2_1	—	911
	2_2	—	797
	3_1	—	515
	3_2	—	1637
测试集	1_3	1802	2375
	1_4	1139	1428
	1_5	2302	2463
	1_6	2302	2448
	1_7	1502	2259
	2_3	1202	1955
	2_4	612	751
	2_5	2002	2311
	2_6	572	701
	2_7	172	230
	3_3	352	434

本文 Transformer 预测模型的编码器与解码器均由 6 层相同的模块层堆叠而成，所有的多头注意力层的头数均为 2，所有前馈神经网络隐藏层神经元数量为 256，初始学习率设置为 0.0001 并随机初始化权重矩阵，使用均方误差（MSE）作为损失函数并通过 Adam 优化算法进行训练，训练与测试环境中 CPU 为 Intel Core i7-10700F，GPU 为 NVIDIA GeForce RTX 2060 SUPER，内存为 16GB，深度学习框架为 Pytorch 1.6.0。

模型训练完毕后需要对其进行验证，与模型训练过程相同，将测试集的 12 种累积特征以 10 个时间步的长度输入预测模型中，预测出寿命比值  $P$ 。

前文中提到 Transformer 模型相比于循环神经网络具有速度上的优势，为此本文在使用累计特征的基础上，采用了 LSTM<sup>[22]</sup>、GRU<sup>[11]</sup>、BiLSTM<sup>[23]</sup> 三种循环神经网络作为预测模型进行对比实验，实验结果如表 5 所示，其中的运行时间是根据 11 个测试集中每个样本的平均时间计算的。从表 5 的对比结果看出，Transformer 模型在运行速度上具有明显的优势。

本文采用一次线性函数拟合预测的  $P$  值，得到  $P$  值的未来趋势，以此计算轴承 RUL 的预测值。部分测试轴承的测试及拟合结果如图 6 和图 7 所示。本文采用误差  $E_i$  来验证预测模型的好坏，如式 (14) 所示：

$$E_i = \frac{\text{actRUL}_i - \text{preRUL}_i}{\text{actRUL}_i} \times 100\% \quad (14)$$

式中： $\text{actRUL}_i$  表示第  $i$  个轴承实际的剩余使用寿命， $\text{preRUL}_i$  表示第  $i$  个轴承剩余使用寿命的预测结果。以轴承 1\_5 为例，由表 4 可知，轴承 1\_5 的全寿命数据为 2463 组，非全寿命数据为 2302 组，其中预测值的拟合直线与 X 轴的交点（即  $P=0$  轴承完全报废）为 2447 组。由于数据采样间隔为 10 s 一组，所以  $\text{actRUL}_i$  实际剩余使用寿命为  $(2463-2302) \times 10=1610$  s， $\text{preRUL}_i$  剩余使用寿命预测值为  $(2447-2302) \times 10=1450$  s，则误差为  $(1610-1450)/1610=9.9\%$ 。

表 5 运行速度对比结果  
Table 5 Running speed comparison results

预测模型	Transformer	LSTM	GRU	BiLSTM
运行时间 (ms)	2.15	3.47	3.33	4.00

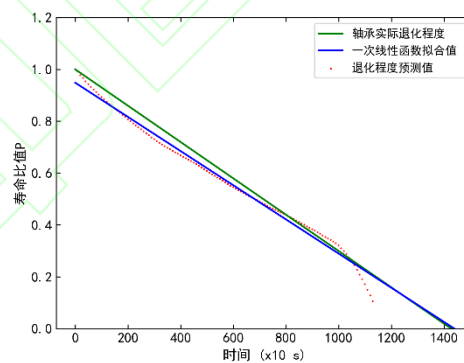


图 6 轴承 1-4 预测结果  
Fig.6 Bearing 1-4 prediction results



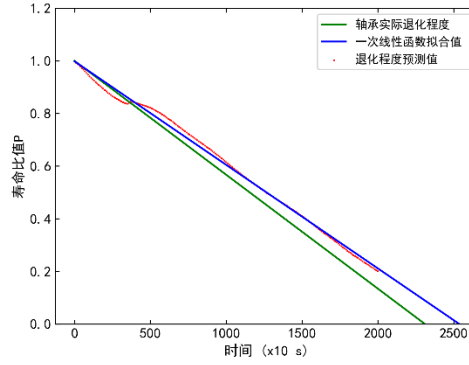


图7 轴承 2-5 预测结果  
Fig.7 Bearing 2-5 prediction results

为了证明本文提出的累积特征+Transformer 模型的滚动轴承 RUL 预测方法的有效性与必要性, 本文设置了另外两种实验方案来与本文的预测方法进行消融实验, 两种实验方案详情如表 6 所示。其中经典统计特征指的是将累积特征中选取的 12 种特征未经累积变换的值输入预测模型中。文献[24]提出将 Transformer 模型用于航天涡扇发动机的 RUL 预测中, 但其仅使用了 Transformer 模型的编码器结构, 本文采用的是完整的 Transformer 编码器-解码器结构, 相比于单编码器结构上进行了改进, 所以在消融实验方案中将单编码器结构作为对比实验条件之一。

表 6 所提的方法与其它两种方案的构成

Table 6 The composition of the proposed prediction method and other two schemes

预测方法	特征提取模型	预测模型
本文提出方法	累积特征	编码器-解码器结构 Transformer 模型
方案一	经典统计特征	编码器-解码器结构 Transformer 模型
方案二	累积特征	单编码器结构 Transformer 模型

为了更好地将本文提出的轴承 RUL 预测方法与消融实验方案以及其它先进的预测方法进行比较, 需要建立一个统一的评估指标来衡量预测结果的准确性。因此, 本文采用 IEEE PHM2012 挑战赛规定的平均得分  $s$  作为方法比较时的衡量标准, 平均得分  $s$  的计算方法如公式(15-16)所示:

$$A_i = \begin{cases} \exp^{-\ln(0.5) \cdot (E_i/5)} & E_i \leq 0 \\ \exp^{+\ln(0.5) \cdot (E_i/20)} & E_i > 0 \end{cases} \quad (15)$$

$$s = \frac{1}{n} \sum_{i=1}^n A_i \quad (16)$$

图 8 显示了误差  $E_i$  与得分  $A_i$  之间的函数关系。从图中可以看出, 正的误差比负的误差具有更高的得分。正的误差代表预测的 RUL 小于实际的 RUL, 即超前预测, 而负的误差代表预测的 RUL 大于实际的 RUL, 即滞后预测。在实际生产环境中, 超前预测带来的风险远远低于滞后预测, 所以这种得分计算方法对滞后预测进行了惩罚, 是公平且合理的。

同时, 在表示 11 个测试轴承的整体预测误差的时候, 如果直接使用误差  $E_i$  的算数平均值, 有时会导致正负误差相互抵消, 出现较差的预测结果计算出较低的平均误差的情况, 针对这个问题, 本文采用平均绝对误差  $\overline{|E|}$  表示测试轴承的整体预测误差, 如公式(17)所示:

$$\overline{|E|} = \frac{1}{n} \sum_{i=1}^n |E_i| \quad (17)$$

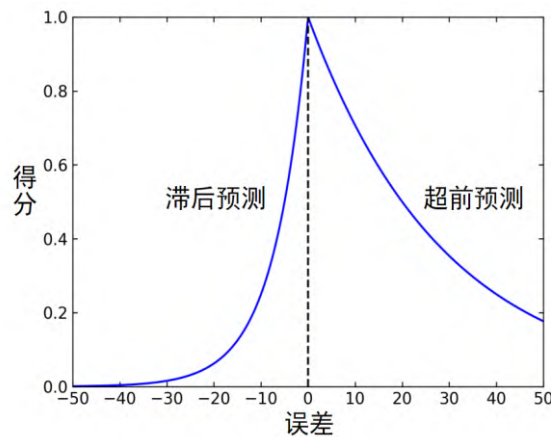


图 8 得分  $A_i$  与误差  $E_i$  的函数关系图  
Fig.8 Function diagram of  $s A_i$  and error  $E_i$

表 7 列出了测试集中 11 个轴承的数据在本文预测模型中的平均绝对误差与平均得分，同时也列出了与两种消融实验方案以及文献[11,25,26]的对比实验结果，从对比实验结果可以看出：

1)在同样使用编码器-解码器结构 Transformer 模型作为预测模型的情况下，由于累积变换之前的传统统计特征单调性和趋势性很低，无法表征滚动轴承的退化趋势，所以表现出了极低的预测性能。而使用累积特征则大幅改善了这种情况，使得平均绝对误差降低了 92.63%，平均得分提高了 558.39%，可见本文所采用的累积变换对于传统统计特征的单调性与趋势性起到了极强的修正作用，对于轴承 RUL 的准确预测起到了至关重要的作用。

2)在同样使用累积特征作为输入的情况下，完整编码器-解码器结构的 Transformer 模型相比于单编码器结构模型的平均绝对误差降低了 19.84%，平均得分提高了 32.6%，这是由于单编码器结构的 Transformer 模型只学习到了输入序列中各个特征向量之间的依赖关系，而解码器结构的引入可以使得 Transformer 模型学习到标签数据中各个向量之间以及输入特征与标签数据相互之间的依赖关系，使得 RUL 的预测更为准确，由此证明了本文所提的 Transformer 模型在轴承 RUL 预测领域的优越性。

3)综合前两点消融实验的对比分析可以看出，本文提出的轴承 RUL 预测方法的优越性是累积特征与 Transformer 模型综合作用的结果，累积特征确保 Transformer 模型能够挖掘出原始振动和轴承健康状况之间的潜在关系，使其能够学到绝大部分测试轴承的退化趋势，并将滞后预测 ( $E_i < 0$ ) 的情况控制在较低范围内。但是仍有个别测试轴承的预测误差较大，如轴承 1-7，这是由于轴承 1-7 的测试数据占全寿命周期数据的比重较小，少量退化前期的数据不足以让模型准确预测到退化后期的 RUL。

4)同时，本文所提方法相较于文献[11,25, 26]平均绝对误差分别降低了 9.25%、28.63%、34.14%，平均得分分别提高了 2.78%、19.79%、29.38%，由此进一步证明了本文所提方法在滚动轴承 RUL 预测方面的有效性与优越性。

表 7 PHM 2012数据集RUL预测结果与比较  
Table 7 Bearing RUL prognostics results and comparisons on PHM 2012 dataset

轴承型号	误差					
	本文采用模型	方案一	方案二	文献[11]	文献[25]	文献[26]
1_3	74.17	-4131	67.66	7.62	54.73	-1.04
1_4	-0.69	23.31	-98.81	69.77	38.48	85.81
1_5	9.9	336.75	10.45	-72.57	-99.4	-278.2
1_6	-12.33	102.39	-7.18	0.93	-120.0	19.18
1_7	83.62	132.32	82.48	85.99	70.65	-7.13
2_1	61.35	90.14	65.08	81.24	75.53	10.49
2_1	5.06	-196.3	14.49	9.04	19.81	51.8
2_1	-70.22	60.04	-53.99	28.19	8.2	28.8
2_1	0.78	102.82	5.13	24.92	17.87	-20.93
2_1	44.83	77.44	47.03	19.06	1.69	44.83
3_1	1.22	118.27	-1.23	2.09	2.93	-3.66
$ E_r $	33.05	448.30	41.23	36.42	46.31	50.18
s	0.4589	0.0697	0.3460	0.4465	0.3831	0.3547

4.2 案例二：XJTU-SY 数据集验证

4.2.1 数据集介绍

为了进一步验证本文提出的轴承 RUL 预测方法的有效性与先进性，本文还采用了 XJTU-SY 数据集进行验证。XJTU-SY 数据集由西安交通大学机械工程学院雷亚国教授团队设计并采集获得，其试验平台如图 9 所示，该平台由交流电动机、电动机转速控制器、转轴、支撑轴承、液压加载系统和测试轴承等组成，能够在不同运行条件下对轴承进行加速退化试验，并获得完整的运行至失效数据。径向力由液压加载系统产生并施加到被测轴承的壳体上，转速由交流感应电机的速度控制器设定并保持。图 10 显示了正常和退化轴承的照片，可以看出，被测轴承的失效是由不同类型的故障引起的，包括内圈磨损、保持架断裂、外圈磨损、外圈断裂等。为了采集被测轴承的振动信号，两个 PCB 352C33 单向加速度传感器分别通过磁座固定于测试轴承的水平和竖直方向上。采样频率设置为 25.6 kz，每 1 分钟记录 32768 个数据点(即 1.28 s)。

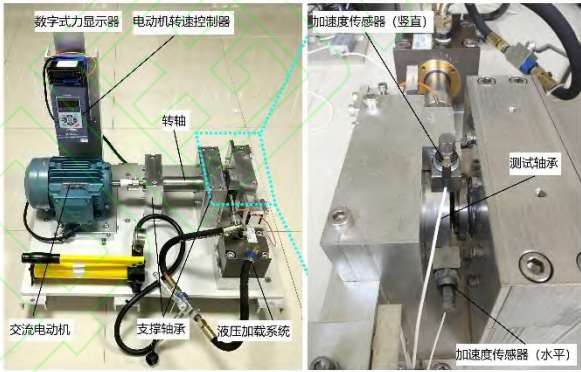


图 9 滚动轴承试验台  
Fig.9 Testbed of rolling element bearings.



图 10 退化轴承照片: (a)内圈磨损 (b)保持架断裂 (c)外圈磨损 (d)外圈断裂  
Fig.10 Photographs of normal and degraded bearings:(a) Inner race wear. (b) Cage fracture. (c) Outer race wear. (d) Outer race fracture.

#### 4.2.2 实验结果对比与分析

在 XJTU-SY 数据集的验证过程中, 本文使用了与案例一相同的实验环境与操作流程进行训练与测试。为了验证提出的轴承 RUL 预测方法的有效性与先进性, 本文将实验结果与文献[28]的实验结果进行了比较, 为了便于比较, 本文采用了与该文献相同的实验方案, 如表 8 所示。文献[28]得到了 4 种预测模型在 XJTU-SY 数据集上的实验结果, 分别为基于剩余自注意力机制的时间卷积网络(TCN-RSA)、深度可分卷积网络(DSCN)、递归卷积神经网络(RCNN)以及相关向量机(RVM)。采用均方根误差(RMSE)与平均得分  $s$  作为实验结果对比的衡量标准, RMSE 的计算公式如公式(18)所示

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{act}P_i - \text{pre}P_i)^2} \quad (18)$$

式中:  $\text{act}P_i$  为第  $i$  个测试样例的寿命比值  $P$  的实际值,  $\text{pre}P_i$  为第  $i$  个测试样例的寿命比值  $P$  的预测值,  $n$  为测试样例数量。

平均得分  $s$  的计算公式如式(14)(15)(16)所示,但是由于 XJTU-SY 数据集提供的均为全寿命周期数据,并未像案例一中 PHM2012 数据集一样截断了测试集的一部分数据用于规定寿命预测时间点,所以本文在案例二的  $s$  计算中参照文献[28]的做法,对所有测试样例均进行了得分  $s$  计算并取平均值。

表 8 实验方案  
Table 8 The experimental scheme

径向力(N)	转速 (r/min)	训练集	测试集
11000	2250	Bearing2-1 Bearing2-2 Bearing2-3	Bearing2-4 Bearing2-5
10000	2400	Bearing3-1 Bearing3-2 Bearing3-3	Bearing3-4 Bearing3-5

实验的对比结果如表 9 所示, 从表中可以看出, 本文提出的预测模型相比于其它四种模型具有更低的均方根误差以及更高的得分, 相比于 TCN-RSA 模型 RMSE 降低了 17.4%,  $s$  提高了 18.6%, 这进一步验证了本文提出的基于 Transformer 模型的滚动轴承剩余使用寿命预测方法的有效性与先进性。进一步对比可以发现, 深度学习方法(Transformer、TCN-RSA、DSCN、RCNN)的表现均优于传统的机器学习方法(RVM), 这是由于传统机器学习方法局限于浅层学习, 对滚动轴承的退化过程学习能力有限, 而深度学习模型由于其突出的深层次学习能力, 可以更好地挖掘轴承退化信息, 建立更准确的滚动轴承 RUL 预测模型。而在这其中, Transformer 由于其强大的自注意力机制, 能够挖掘出输入特征与轴承退化程度自身及相互之间的依赖关系, 因此其表现优于其他三种深度学习模型。



表 9 XJTU-SY数据集RUL预测结果与比较

Table 9 RUL prognostics results and comparisons on XJTU-SY dataset

模型	RMSE	平均得分
DSCN	0.0749	0.4165
RCNN	0.0803	0.3586
RVM	0.1082	0.2911
TCN-RSA	0.0659	0.4803
Transformer	0.0544	0.5697

## 5 结论

本文提出了一种基于累积特征与 Transformer 模型的滚动轴承剩余使用寿命预测方法,通过实验验证得到以下结论:

1)本文采用的累积变换对于传统统计特征的单调性与趋势性具有很强的修正作用,使得累积特征能够很好地表征轴承的退化趋势。实验证明,在同样使用 Transformer 预测模型的情况下,累积特征相比于传统统计特征平均绝对误差降低了 92.63%,平均得分提高了 558.39%。

2)本文提出的基于编码器-解码器结构 Transformer 模型的轴承 RUL 预测模型,能够更好地挖掘出输入特征与轴承 RUL 之间复杂映射关系,使得 RUL 的预测更为准确。实验证明,在同样使用累积特征的情况下,本文的预测模型相比于单解码器结构的 Transformer 预测模型平均绝对误差降低了 19.84%,平均得分提高了 32.6%。

3)本文提出的累积特征+Transformer 模型的算法组合能够确保预测模型挖掘出输入特征与轴承 RUL 之间复杂映射关系,使其能够学到绝大部分测试轴承的退化趋势。案例一中的 PHM 2012 数据集验证实验证明,与文献[11,25,26]相比,本文提出的轴承 RUL 预测方法平均绝对误差分别降低了 9.25%、28.63%、34.14%,平均得分分别提高了 2.78%、19.79%、29.38%。

4)本文在案例二中采用了 XJTU-SY 数据集进行进一步的验证,实验表明,本文提出的方法相比于文献[28]均方根误差降低了 17.4%,平均得分提高了 18.6%,由此进一步证明了本文所提方法在滚动轴承 RUL 预测方面的有效性与优越性。

在后续的研究中,应该注意到滞后预测的危害,在实际生产环境中滞后预测相比于超前预测会带来更大的事故风险。本文提出的方法虽然提高了整体预测精准度,但在个别轴承的工况下还存在滞后预测(如案例一轴承 2-1 的实验结果),下一步的研究重点应该是着眼于如何降低滞后预测误差或者尽可能消除滞后预测。

## 参考文献 (References)

- [1] 刘建昌,权贺,于霞,等.基于参数优化 VMD 和样本熵的滚动轴承故障诊断[J/OL].自动化学报,(2019-08-12)[2021-05-09].<https://doi.org/10.16383/j.aas.190345>.  
Liu J C,Quan H,Yu X,et al.Rolling Bearing fault diagnosis bas-ed on parameter optimization VMD and sample entropy[J/OL].Acta Automatica Sinica,(2019-08-12)[2021-05-09].<https://doi.org/10.16383/j.aas.190345>(in chinese).
- [2] 张振良,刘君强,黄亮,等.基于半监督迁移学习的轴承故障诊断方法[J].北京航空航天大学学报,2019,45(11):2291-2300.  
Zhang Z L,Liu J Q,Huang L,et al.A bearing fault diagnosis method based on semi-supervised and transfer learning[J].Jour-nal of Beijing University of Aeronautics and Astronautics,2019,45(11):2291-2300(in chinese).
- [3] 田利梅,龚梦彤,唐获音,等.基于功耗残差的航天器 CMG 退化特征提取方法[J].北京航空航天大学学报,(2021-04-12)[2021-05-09].<https://doi.org/10.13700/j.bh.1001-5965.2021.0060>. Tian L M, Gong M T, Tang D Y, et, al. Degradation indicator extraction for aerospace CMG based on power consumption analysis[J]. Journal of Beijing University of Aeronautics and Astronautics, (2021-04-12)[2021-05-09]. <https://doi.org/10.13700/j.bh.1001-5965.2021.0060> (in chinese).
- [4] 袁焯,张永,丁汉.工业人工智能的关键技术及其在预测性维护中的应用现状[J].自动化学报,2020,46(10):2013-2030.  
Yuan Y,Zhang Y,Ding H.Research on Key Technology of Ind-ustrial Artificial Intelligence andIts Application in Predictive Maintenance[J].Acta Automatica Sinica,2020,46(10):2013-2030 (in chinese).
- [5] Lei Y,Li N,Guo L,et al.Machinery health prognostics:A systematic review from data acquisition to RUL prediction[J].Mechanical Systems and Signal Processing.2018,104:799-834.
- [6] 王冰,李洪儒,许葆华.基于多尺度形态分解谱熵的电机轴承预测特征提取及退化状态评估[J].振动与冲击,2013,32(22):124-128.  
Wang B,Li H R,Xu B H. Motor bearing forecast feature extracting and degradation status identification based on multi-scale- morphological decomposition spectral entropy[J].Journal of Vibration and Shock, 2013,32(22):124-128(in chinese).
- [7] 李洪儒,于贺,田再克,等.基于二元多尺度熵的滚动轴承退化趋势预测[J].中国机械工程,2017,28(20):2420-2425.  
Li H R,Yu H,Tian Z K,et al.Degradation Trend Prediction of Rolling Bearings Based on Two-element Multiscale Entropy.China Mechanical Engineering,2017,28(20):2420-2425(in chinese).
- [8] Ren L,Liu Y,Wang X,et al.Cloud-Edge based Lightweight Temporal Convolutional Networks for Remaining Useful Life Prediction in

- IIoT[J].IEEE Internet of Things Journal,2020:1.
- [9] 杨宇, 张娜, 程军圣. 全参数动态学习深度信念网络在滚动轴承寿命预测中的应用[J]. 振动与冲击, 2019, 38 (10): 199-205.  
Yang Y,Zhang N,Cheng J S.Global parameters dynamic learning deep belief networks and its application in rolling bearing life prediction[J].Journal of Vibration and Shock,2019,38(10):199-205(in chinese).
- [10] Guo L,Li N,Jia F,et al.A recurrent neural network based health indicator for remaining useful life prediction of bearings[J]. Neurocomputing,2017,240:98-109.
- [11] 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.
- [12] 康守强, 周月, 王玉静, 等. 基于改进 SAE 和双向 LSTM 的滚动轴承 RUL 预测方法[J]. 自动化学报, (2020-05-15) [2021-05-09]. <https://doi.org/10.16383/j.aas.c190796>. Kang S Q, Zhou Y, Wang Y J, et al. RUL Prediction Method of a Rolling Bearing based on Improved SAE and Bi-LSTM[J]. Acta Automatica Sinica, (2020-05-15) [2021-05-09]. <https://doi.org/10.16383/j.aas.c190796> (in chinese).
- [13] 邱大伟, 刘子辰, 周一青, 等. 基于 Transformer 神经网络的滚动轴承故障类型识别[J]. 高技术通讯, 2021, 31 (01): 1-11.  
Qiu D W,Liu Z C,Zhou Y Q,et al.A novel fault type detection method of rolling bearing using Transformer neural networks[J]. Chinese High Technology Letters,2021,31(01):1-11(in chinese).
- [14] Vaswani A,Shazeer N,Parmer N,et al.Attention is all you need[J]. arXiv preprint arXiv:1706.03762,2017.
- [15] Wang B,Lei Y,Yan T,et al.Recurrent convolutional neural net-work:A new framework for remaining useful life prediction of machinery[J].Neurocomputing,2020,379:117-129.
- [16] Ma M,Mao Z.Deep-Convolution-Based LSTM Network for Re-maining Useful Life Prediction[J].IEEE Transactions on Industrial Informatics,2021,17(3):1658-1667.
- [17] Wang B,Lei Y,Li N,et al.Deep separable convolutional network for remaining useful life prediction of machinery[J].Mechanical Systems and Signal Processing,2019,134:106330.
- [18] Zeng F,Li Y,Jiang Y,et al.An online transfer learning-based remaining useful life prediction method of ball bearings[J]. Measurement,2021,176:109201.
- [19] Nectoux P,Gouriveau R,Medjaher K,et al.PRONOSTIA:An experimental platform for bearings accelerated degradation tests[C]//IEEE International Conference on Prognostics and Health Management,PHM'12.IEEE Catalog Number:CPF12PHM-CDR,2012: 1-8.
- [20] Soualhi A,Medjaher K,Zerhouni N.Bearing Health Monitoring Based on Hilbert - Huang Transform,Support Vector Machine,and Regression[J].IEEE Transactions on Instrumentation and Measurement, 2015,64(1):52-62.
- [21] K Singleton R,G Strangas E,Aviyente S.Extended Kalman Filtering for Remaining-Useful-Life Estimation of Bearings[J].IEEE Transactions on Industrial Electronics,2015,62(3):1781-1790.
- [22] Wang F,Liu X,Deng G,et al.Remaining life prediction method for rolling bearing based on the long short-term memory network[J]. Neural Processing Letters,2019,50(3):2437-2454.
- [23] Luo J,Zhang X.Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction[J]. Applied Intelligence,2021:1-16.
- [24] 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,2021,1-10.
- [25] Hinch A Z,Tkhouat M.Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network[J].Procedia Computer Science,2018,127:123-132.
- [26] Hong S,Zhou Z,Zio E,et al. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method[J].Digital Signal Processing,2014,27:159-166.
- [27] Wang B,Lei Y,Li N,et al.A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J].IEEE Transactions on Reliability,2018,69(1):401-412.
- [28] Cao Y,Ding Y,Jia M,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,2021:107813.