

结合 LSTM 和 Self-Attention 的滚动轴承 剩余使用寿命预测方法

黄宇¹, 冯坤², 高俊峰³, 李周正², 江志农¹, 高金吉¹

(1. 北京化工大学信息科学与技术学院, 北京 100029;

2. 北京化工大学高端机械装备健康监控与自愈化北京市重点实验室, 北京 100029;

3. 中国石油天然气股份有限公司炼油与化工分公司, 北京 100007)

摘要: 为了构建准确表征滚动轴承退化过程的趋势性健康度指标, 提高滚动轴承剩余使用寿命(Remaining Useful Life, RUL)的预测精度, 提出了一种结合长短期记忆(Long-Short Term Memory, LSTM)和自注意力(Self-Attention)机制的神经网络模型(LSTM-SA)用于滚动轴承 RUL 预测。利用包络解调获得原始信号的包络谱, 再将包络谱分段并计算对应频段的皮尔逊相关系数, 得到具有单调性和趋势性的退化特征; 将退化特征归一化处理作为 LSTM-SA 模型的输入, 并利用 LSTM 自适应提取退化特征时间上的内部相关性以及 Self-Attention 对关键信息的筛选, 消除无用信息的干扰, 挖掘深层次特征, 构建健康度指标并得到退化曲线; 确定失效阈值, 利用最小二乘法拟合退化曲线, 预测寿命失效点, 实现滚动轴承的 RUL 预测。在 PHM2012 数据集上的实验结果表明, 所提出的方法相比于其他文献, 平均绝对误差分别降低了 43.18%, 62.57% 和 59.44%, 平均得分分别提高了 10.87%, 45.71% 和 34.21%; 在工程实际数据中的实验结果表明, 所提出方法的平均预测误差分别比 Standard-RNN 和 CNN 方法降低了 39.58% 和 74.86%。

关键词: 剩余使用寿命预测; 滚动轴承; 长短期记忆网络; 自注意力机制; 包络谱特征

中图分类号: TB114.3; TH133.33 **文献标志码:** A **文章编号:** 1004-4523(2023)06-1744-10

DOI: 10.16385/j.cnki.issn.1004-4523.2023.06.029

引言

滚动轴承是机械设备中最常见且极其重要的关键零部件。在设备运行的过程中由于环境、工况等因素的影响, 滚动轴承因易出现点蚀、磨损等现象, 而成为工业上最易损坏、可靠性最差的零部件之一。在旋转设备中, 约有 30% 的故障是由滚动轴承引起的^[1]。因此, 对滚动轴承的运行状态进行监测, 预测其剩余使用寿命并确定最佳维修点, 可以为决策者建立维修方案提供支撑, 具有重要意义。

对滚动轴承的剩余使用寿命预测主要有三种主流的方法: 基于模型的方法、数据驱动方法和二者混合的方法^[2]。基于模型的方法需要大量的专家知识和先验知识, 不仅浪费时间和劳动力, 而且不具有通用性。因此, 研究数据驱动的剩余使用寿命预测方法, 实现轴承精准的 RUL 预测, 是目前的研究热点。基于数据驱动的轴承 RUL 预测主要有三个步骤: ①从加速度振动信号中提取具有单调性和趋势性的退

化特征, 表征滚动轴承的衰退过程; ②构建健康度指标模型, 利用传统人工智能或深度学习的方法学习退化特征与轴承健康度指标之间的复杂映射关系, 得到轴承的退化曲线; ③对轴承退化曲线进行拟合预测, 计算轴承寿命失效点, 最终得到轴承剩余使用寿命。

退化特征提取是轴承 RUL 预测的前提条件, 目的是从轴承原始加速度振动信号中提取出能反映其退化趋势的特征。目前常见的方法是利用深度学习的方法从时域、频域以及时频域中提取原始振动信号的退化特征。杨宇等^[3]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, 挖掘出原始振动信号深层本质特征。李少鹏^[4]将原始信号通过快速傅里叶变换转换为频域幅值信号, 然后利用一维卷积神经网络从频域信号中挖掘深层退化特征。WANG 等^[5]提出了通过将原始信号的时频表示作为输入, 并利用三维深度卷积神经网络提取退化特征的方法。CAO 等^[6]将原始信号边缘谱作为输入, 利用时间卷积网络提取高级退化特征

表示。

上述基于深度学习的特征提取方法往往需要大量标签对深度学习网络进行监督学习,标签的缺乏严重制约了利用深度学习进行退化特征提取,而传统的统计特征难以反映轴承的整个退化过程,单调性和趋势性不明显^[7-9]。

此外,利用退化特征构建健康度指标是轴承RUL预测至关重要的一步,目的是将退化特征映射为趋势性健康度指标。如果模型输出的健康度指标能够很好地反映轴承的退化趋势,便能准确地预测轴承的剩余使用寿命。轴承的振动数据是一种时间序列数据,而循环神经网络(Recurrent Neural Network,简称RNN)具有强大的时间序列处理能力,因此,RNN被广泛应用于滚动轴承剩余使用寿命预测。GUO等^[10]利用RNN将提取的8个时域特征和6个频域特征映射为健康度指标,并用双指数函数模型拟合预测,得到了很高的预测精度。韩林洁^[11]构建了频域幅值累计特征,并用门控循环单元网络进行轴承剩余使用寿命预测。LI等^[12]利用核主成分分析和指数加权移动平均获得了退化曲线,并利用分层门控循环单元网络进行剩余使用寿命预测。WU等^[13]提出了一种动态差异技术,从原始监测数据中提取特征,利用LSTM学习监测数据背后真正的物理退化机制,并通过实验说明了LSTM的性能比标准RNN和门控循环单元(GRU)更优越。

LSTM是RNN的变体,其性能比RNN更为优越。然而,LSTM仍然有一些不可忽视的缺点。首先,LSTM不能进行并行处理,LSTM必须等待前一个数据处理完成才能处理下一个数据,这不仅降低了模型的灵活度,还导致误差逐级累积^[14]。其次,LSTM在训练过程中存在梯度爆炸、梯度消失以及占用大量内存等问题,而这一系列问题目前没有很好的解决方法。此外,当输入序列超过一定长度后,LSTM会出现记忆衰退现象,即不能有效记忆很久之前的信息^[15]。而自注意力机制(Self-Attention)可以很好地解决上述问题。Self-Attention能让模型对关键信息重点关注并充分学习吸收,能学习任意长度的时间序列信息;且Self-Attention是并行计算方式,处理速度较LSTM大大加快^[16]。

综上,提出一种利用包络谱特征并结合LSTM和Self-Attention构建趋势性健康度指标来实现轴承RUL预测的方法。LSTM将退化特征编码为高阶特征向量,Self-Attention对LSTM隐藏层的输出计算权重系数,并且选择性地保留中间结果,使隐藏层输出向量之间的联系更加紧密,强调关键信息和减少对无效信息的关注,实现了滚动轴承RUL的准确预测。

1 理论背景

1.1 包络谱特征和皮尔逊相关系数

1.1.1 包络谱特征

包络解调可以有效地将轴承故障信号从高频信号中分离并提取出来^[17]。轴承在出现故障时,由轴承故障引起的特征信号被调制到高频带段,此时时域波形和频谱均难以明显体现其故障特征。包络解调方法对信号进行希尔伯特变换和快速傅里叶变换后得到原始信号的包络谱,从包络谱上可以获得清晰的故障特征。同时,包络解调还可以发现轴承更早期的缺陷,从而提取轴承初期故障特征^[18]。

轴承在退化过程中,多数情况下不止存在一种故障即单一故障模式,而是存在多种故障即复合故障模式。因此,各种故障频率会在高频段与多部件的固有频率发生共振。对信号包络解调后,可以在低频段得到故障特征频率。传统的包络解调在得到包络谱后利用低通滤波器滤掉高频段信号,消除高频段信号对故障分析的干扰,但鉴于复合故障模式的复杂性,包络谱的高频段仍可能包含故障特征^[19-20]。因此,本文尝试将包络谱按频率均分成 n 段。这使得在分析低频段故障特征时不会引入高频段信号的干扰,同时还保留了高频段中的故障信息;由于高频段故障特征相对较弱,按频率分段后再分析高频段信号,减弱了低频段故障信号对机器学习模型特征提取时的绝对主导性,一定程度上放大了高频段信号的故障特征。

1.1.2 皮尔逊相关系数

皮尔逊相关系数用于度量两个向量 X 和 Y 之间的相关性,定义如下式所示:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \quad (1)$$

式中 Cov 表示协方差; D 表示方差; ρ_{XY} 表示向量 X 和向量 Y 之间的皮尔逊相关系数,取值范围为 $[-1, 1]$, $|\rho_{XY}|$ 越大,说明 X 和 Y 的相关性越大。

皮尔逊相关系数消除了不同变量量纲上的差别,即两个变量的位置和尺度变化并不会改变其皮尔逊相关系数,所以通过皮尔逊相关系数理论计算得到的相关性所衡量的是趋势^[21]。因此,将包络谱分段后各子频段与标准样本的皮尔逊相关系数作为退化特征,更有利于模型准确表征轴承退化过程的健康度指标。

1.2 长短期记忆网络

LSTM网络是从标准RNN改进得来的。

LSTM通过其内部复杂的门运算和引入细胞态,有效缓解标准RNN的长期依赖问题^[22]。遗忘门 f_t 决定上一时刻的细胞态 c_{t-1} 有多少信息保留到当前时刻的细胞态 c_t ;输入门 i_t 决定当前时刻的输入有多少信息存储到当前时刻的细胞态 c_t 中;输出门 o_t 控制当前细胞态 c_t 有多少信息保留到当前时刻输出 h_t 中。图1为LSTM单元的内部结构, x_{t-1} 、 x_t 和 x_{t+1} 分别指 $t-1$ 、 t 和 $t+1$ 时刻的输入信息; h_{t-1} 、 h_t 和 h_{t+1} 分别指 $t-1$ 、 t 和 $t+1$ 时刻的输出信息。

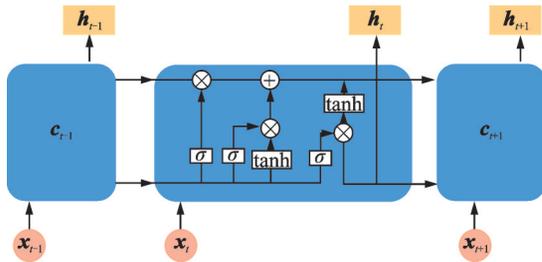


图1 LSTM结构

Fig. 1 The structure of LSTM

遗忘门 f_t 、输入门 i_t 、输出门 o_t 、细胞态 c_t 和输出 h_t 的计算公式分别如下式所示:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(i_f \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$o_t = \sigma(o_f \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

$$h_t = o_t \otimes \tanh c_t \quad (6)$$

式中 x_t 为当前时刻的输入; h_{t-1} 为上一时刻的输出; W 为权重矩阵; b 为偏置; $\sigma(x) = 1/(1 + e^x)$ 为 Sigmoid 激活函数; 下标“ f ”, “ i ”, “ c ”分别表示矩阵 W 和偏置 b 为分别代表遗忘门、输入门、细胞态的参数; \otimes 表示逐元素相乘法。

1.3 自注意力机制

Self-Attention 是注意力机制的改进,不仅可以快速筛选出关键信息,减少对其他无关信息的关注,还可以降低对外部信息的依赖,更易捕捉输入数据的内部相关性^[23]。神经网络通过引入自注意力机制,在解决模型信息过载问题的同时,还提高了网络的准确率和鲁棒性^[24]。

Self-Attention 的计算分为两步。步骤 1: 计算输入序列任意向量之间的注意力权重; 步骤 2: 根据注意力权重计算输入序列的加权平均值。自注意力机制如图 2 所示, $a^i (i=1, 2, 3, \dots, t)$ 表示输入序列; $v^i (i=1, 2, 3, \dots, t)$ 表示由输入序列生成的值向量; $\alpha_i (i=1, 2, 3, \dots, t)$ 表示输入序列与各自的向量 q 和 k 做运算并经过 Softmax 函数后的结果; $b^i (i=1, 2,$

$3, \dots, t)$ 则表示输入序列中第 i 个位置信息与所有位置信息进行注意力机制运算后的结果。

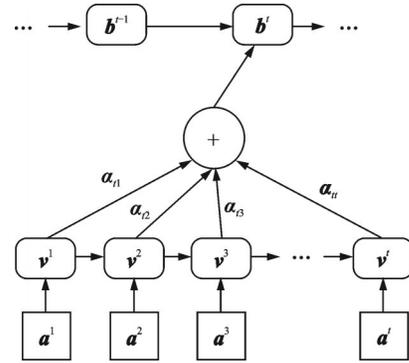


图2 自注意力机制结构

Fig. 2 The structure of Self-Attention mechanism

Self-Attention 具体运算如下式所示:

$$Q = XW^q \quad (7)$$

$$K = XW^k \quad (8)$$

$$V = XW^v \quad (9)$$

$$\text{Attention}(Q, K, V) = \text{Soft max} \left(\frac{QK^T}{\sqrt{\dim}} \right) V \quad (10)$$

式中 Q 、 K 和 V 分别为查询矩阵、键矩阵和值矩阵,由输入 X 分别与相应的权重矩阵 W^q 、 W^k 和 W^v 相乘得到; \dim 表示 Q 、 K 和 V 的维数。

2 RUL 预测方法及流程

针对基于深度学习的退化特征提取方法的不足,提出了一种基于包络谱特征和皮尔逊相关系数的退化特征提取方法;并结合 LSTM 和 Self-Attention 各自的优点,利用退化特征构建健康度指标,准确表征轴承的退化过程,实现滚动轴承 RUL 的预测,流程框图如图 3 所示。

具体流程如下:

步骤 1: 计算信号的包络谱,将包络谱按频率平均划分为 n 段。以轴承正常运行的振动信号作为标准样本,分别计算各样本每个子频段及标准样本对应频段的皮尔逊相关系数,将计算得到的相关系数作为轴承退化特征。

步骤 2: 模型的训练数据集是轴承全寿命周期振动信号所提取的退化特征。同时,由于轴承的退化是一个渐变的过程,因此设时刻为 t 的训练样本的健康度指标 HI 为:

$$HI = \frac{T-t}{T} \quad (11)$$

式中 T 为轴承的全寿命时间; t 为轴承当前运行的时间。将式(11)的计算结果作为训练集的标签,标

签的取值范围为 0~1。

赋予训练集标签后,利用其训练 LSTM-SA 模型。网络随机初始化参数,根据预设的误差函数,计算网络输出与标签之间的相对误差,并朝着误差减小的梯度方向训练参数,直至误差值降到预设的阈值以下或训练次数达到预设的迭代次数为止。训练过程中,将学习率初始化为较小的值,采用 Adam 优化器训练网络并自适应调整学习率。

步骤 3:利用最小二乘法多项式拟合健康度指标曲线^[25],得到轴承的 RUL。LSTM-SA 模型的输出是轴承每个时刻的健康度指标,各时刻的健康度指标形成退化曲线。将非全寿命测试轴承信号输入模型,得到非全寿命的健康度退化曲线,拟合该曲线并计算其达到失效阈值的时间点,该时间点就是测试轴承的寿命结束点。轴承的健康度指标取值在 0~1 之间,0 表示轴承失效,1 表示轴承完全正常,因此将失效阈值设为 0。

3 实验数据验证

3.1 数据描述

为了验证所提方法的有效性,采用 IEEE 协会提供的 PHM 2012 数据集进行验证^[26]。该数据集包含利用加速度传感器采集的 17 组轴承的全寿命周期振动信号,其中 6 组训练集和 11 组测试集,如表 1 所示。信号的采样频率为 25.6 kHz,采样间隔为 10 s,单次采样时间为 0.1 s。采集设备共采集水平和垂直两个方向的振动信号,根据文献[27-28]的研究,水平方向的振动信号比垂直方向的振动信号提供的有效信息更多。因此,本文采用水平方向的振动信号进行验证。

3.2 滚动轴承 RUL 预测

实验首先对数据集进行退化特征提取,第一步需要确定包络谱的分段数 n 。如前所述,将包络谱分段的退化特征提取方法,可以有效提取轴承早期损伤的故障特征以及优化退化特征的单调性和趋势性,这种性能改进与分段数 n 有关。因此,接下来将研究 n 对退化特征的单调性和趋势性的影响。 n 的初始值分别设置为 1, 2, 3, 4, 5, 6, 7 和 8, 利用主成分

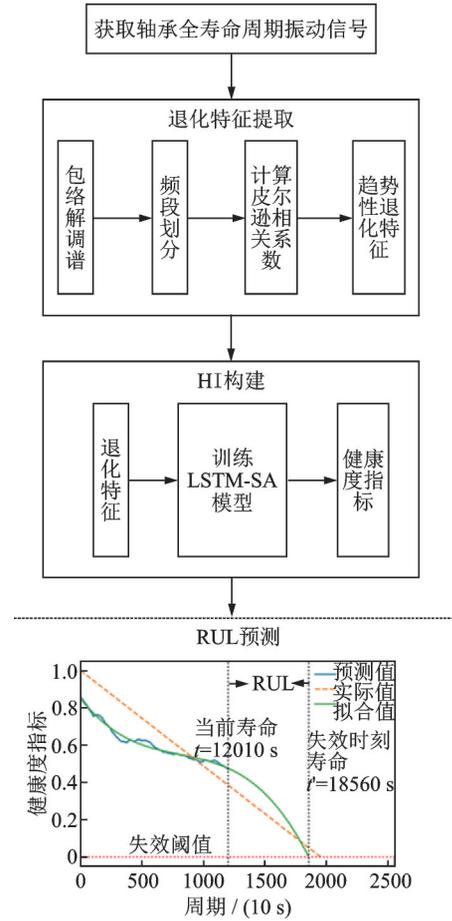


图 3 剩余使用寿命预测方法流程图

Fig. 3 The flowchart of RUL prediction method

分析^[29]对不同分段数所提取到的特征降维至一维,计算并比较主成分的单调性和趋势性,选出最优的 n 。

趋势性和单调性的计算分别如下式所示:

$$Trend = \frac{\left| \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \right|}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (12)$$

式中 x_i 和 y_i 分别表示时间和特征的值; \bar{x} 和 \bar{y} 分别表示 x 和 y 的平均值。Trend 的取值在 0~1 之间,值越大,趋势性越强。

$$Mono = \frac{\left| \text{No. of } \frac{d}{dx} > 0 - \text{No. of } \frac{d}{dx} < 0 \right|}{N - 1} \quad (13)$$

式中 N 表示样本的数量。Mono 取值在 0~1 之间,值越大,单调性越强。

表 1 轴承数据集描述

Tab. 1 Description of bearing datasets

数据集	径向力/N	转速/(r·min ⁻¹)	轴承集						
数据集 1	4000	1800	轴承 1_1	轴承 1_2	轴承 1_3	轴承 1_4	轴承 1_5	轴承 1_6	轴承 1_7
数据集 2	4200	1650	轴承 2_1	轴承 2_2	轴承 2_3	轴承 2_4	轴承 2_5	轴承 2_6	轴承 2_7
数据集 3	5000	1500	轴承 3_1	轴承 3_2	轴承 3_3	—	—	—	—

除此之外,为了从整体上对单调性和趋势性进行度量,定义 $Cori$ ^[30] 作为退化特征度量指标,其值越大,说明退化特征越能反映轴承的退化趋势。计算公式如下:

$$Cori = \frac{Trend + Mono}{2} \quad (14)$$

退化特征的 $Cori$ 对比结果如图 4 所示。从图 4 中可以看出, $Cori$ 随着 n 的增加而增大,并在 $n=4$ 时达到最高点。之后, $Cori$ 随着 n 的增加而减小。表 2 以轴承 1_1 和 2_2 为例,对比了 n 取不同值时,退化特征的 $Cori$ 值。当分段数 n 太小时,会在退化特征所在频段混入太多干扰信息;相反,当 n 太大时,会将退化特征所在频段切碎,难以提取出单调性和趋势性明显的退化特征。基于以上分析,在实验中将包络谱分段数 n 设置为 4。

分段数 n 确定后,计算原始振动信号包络谱,随后划分子频段并分别计算各频段与标准样本的皮尔逊相关系数,将相关系数作为退化特征。将包络谱划分为 n 个子频段的方法如下式所示:

$$s_1 = \{0 \sim \frac{F}{n}\} \text{ kHz} \quad (15)$$

$$s_2 = \{\frac{F}{n} \sim \frac{2F}{n}\} \text{ kHz} \quad (16)$$

⋮

$$s_n = \{\frac{(n-1)F}{n} \sim F\} \text{ kHz} \quad (17)$$

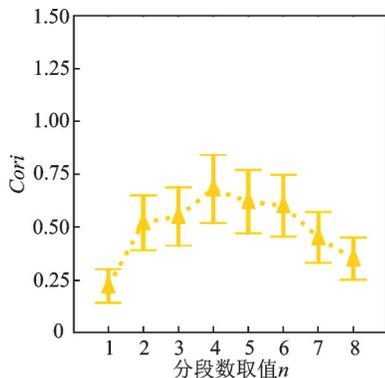


图 4 退化特征的 $Cori$ 对比

Fig. 4 Comparison of $Cori$ of degradation features

表 2 退化特征单调性和趋势性对比结果

Tab. 2 Comparison results of monotonicity and tendency of degradation characteristics

n	轴承 1_1			轴承 2_2		
	Mono	Trend	$Cori$	Mono	Trend	$Cori$
1	0.17	0.05	0.22	0.50	0.66	0.58
2	0.10	0.93	0.52	0.65	0.89	0.77
3	0.23	0.87	0.55	0.25	0.74	0.50
4	0.47	0.88	0.68	0.74	0.90	0.82
5	0.33	0.91	0.62	0.36	0.75	0.56
6	0.26	0.94	0.60	0.25	0.81	0.53
7	0.10	0.80	0.45	0.27	0.75	0.51
8	0.10	0.60	0.35	0.17	0.77	0.47

式中 s_1, s_2, \dots, s_n 分别表示 n 个子频段的频率范围; F 为包络谱的分析频率。将各子频段按式(1)计算皮尔逊相关系数,得到退化特征。

在本实验中,包络谱分析频率为 12.8 kHz,如图 5 所示。将包络谱按频率划分为 4 个子频段,每个子频段的频率范围分别为 0~3.2 kHz, 3.2~6.4 kHz, 6.4~9.6 kHz 和 9.6~12.8 kHz,如图 6 所示。分别计算每个样本各子频段、全频段及标准样本相应频段的皮尔逊相关系数,计算得到的相关系数作为各样本的退化特征。利用上述退化特征训练 LSTM-SA 模型,从而构建健康度指标模型。

为了进一步说明退化特征的单调性和趋势性,以训练数据轴承 1_1 为例,利用 t-SNE 对提取到的退化特征进行可视化分析,如图 7 所示。从图 7 中可以发现,所提取的特征随着时间有序变化,能够反映出轴承的退化过程。

本文按照 PHM 2012 数据集的划分,将轴承 1_1, 1_2, 2_1, 2_2, 3_1 和 3_2 全寿命周期数据作为训练集,其余轴承作为测试集。训练集提取退化特征后输入到 LSTM-SA 模型进行训练。所提模型由 2 层 LSTM 层、1 层 Self-Attention 运算层和 2 层全连接层组成,所有前馈神经网络的神经元个数均为 64。初始学习率为 0.005,随机初始化权重参数,以均方误差(MSE)作为损失函数,并利用 Adam 优化器进行训练。训练和测试的 CPU 环境为 Intel Core i5,内存为 16 GB,深度学习框架为 Pytorch 1.10.1。

轴承在运行过程中逐渐发生退化,但其早期运行过程的退化状态表现不明显。以轴承 1_7 和 2_3 为例,分别对两组轴承全寿命时域波形图进行分析,分别求各自的时域特征 RMS,如图 8 和 9 所示。从图 8 和 9 中可以看出,轴承 1_7 和 2_3 在运行的前期和中期 RMS 变化平稳,直到轴承损坏后期才出现 RMS 的跳变。由此可知,轴承早期退化不明显,难以捕捉其退化趋势。

轴承 1_7 和 2_3 的健康度指标如图 10 和 11 所示。从图 10 和 11 中可知,轴承的健康度指标缓慢变化,较好地反映了轴承在运行过程中的退化状态。由于存在局部振荡,采用 Savitzky-Golay 滤波器^[31]

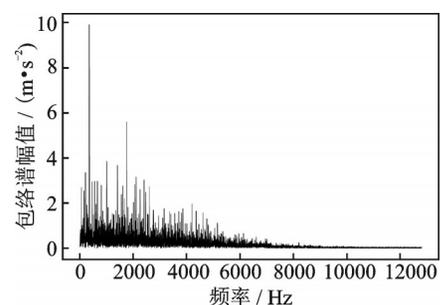


图 5 全频段包络谱

Fig. 5 Envelope spectrum for full band

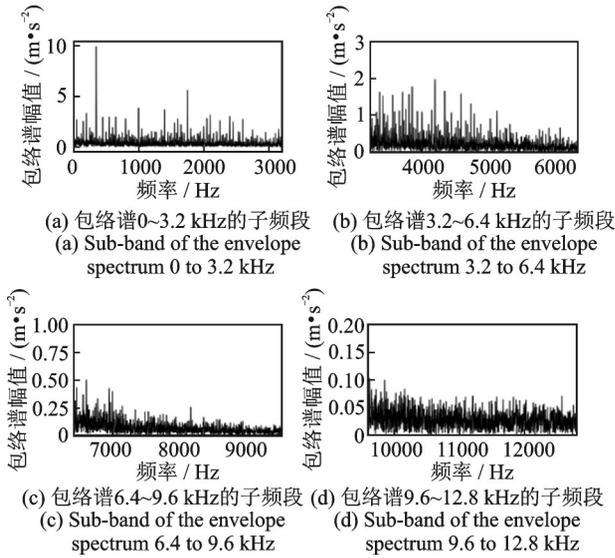


图 6 包络谱各子频段

Fig. 6 Each sub-band of the envelope spectrum

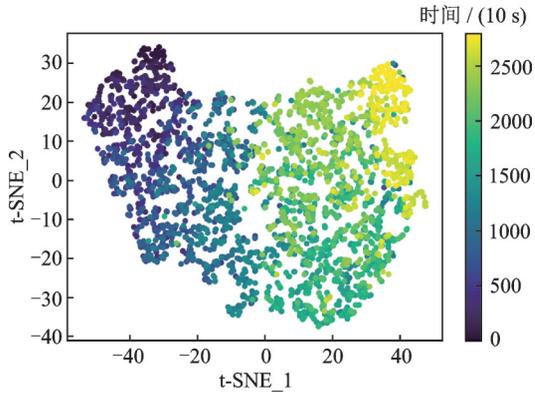


图 7 t-SNE降维可视化退化特征

Fig. 7 t-SNE reduced dimensional visualization of degradation features

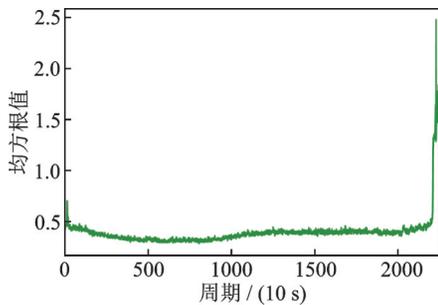


图 8 轴承 1_7 原始振动信号均方根值

Fig. 8 RMS value of original vibration signal of bearing 1_7

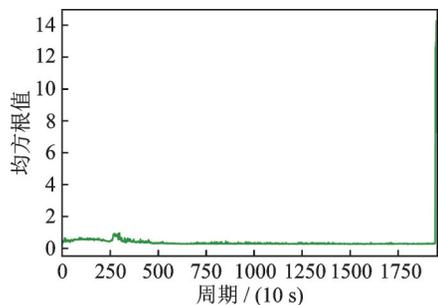


图 9 轴承 2_3 原始振动信号均方根值

Fig. 9 RMS value of original vibration signal of bearing 2_3

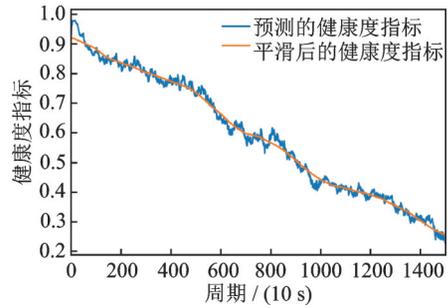


图 10 轴承 1_7 的健康度指标

Fig. 10 Health indicator of bearing 1_7

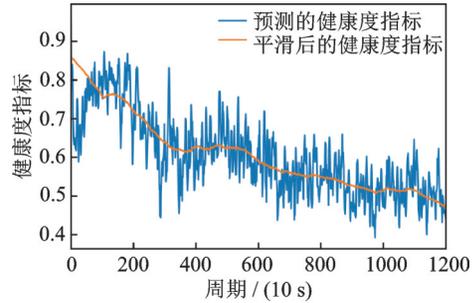


图 11 轴承 2_3 的健康度指标

Fig. 11 Health indicator of bearing 2_3

对健康度指标进行平滑处理,消除局部振荡影响并使健康度指标的退化趋势更为平缓。由图中健康度指标的变化趋势可知,所提方法构建的健康度指标具有明显单调性,能够反映轴承的退化过程,且对轴承早期退化特征敏感。图 12 展示了训练集中 6 个轴承的健康度指标,针对轴承每个时刻的运行状态,均给出 0~1 之间的量化指标。

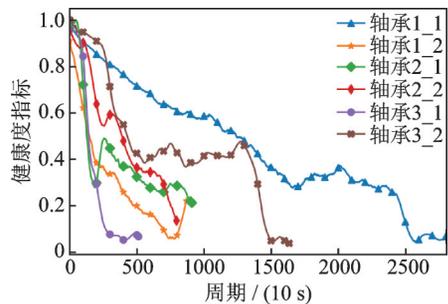


图 12 训练集轴承的健康度指标

Fig. 12 Health indicators of bearings in training set

为验证所提方法构造的健康度指标对提高滚动轴承 RUL 预测精度的作用,利用最小二乘法多项式拟合由健康度指标得到的轴承退化曲线,预测轴承的失效点,从而得到轴承的 RUL。选取三次多项式进行退化曲线拟合,其公式为:

$$y = ak^3 + bk^2 + ck + d \quad (18)$$

式中 y 表示健康度指标; k 表示第 k 个样本, $k=1, 2, \dots, N$; a, b, c 和 d 表示待拟合的系数。

设 t 为轴承当前运行时间,当所预测的健康度指标达到失效阈值 0 时,求对应失效时间 t' ,二者之差即为预测的剩余使用寿命:

$$\text{predRUL}_i = \{t' - t | t' > t\} \quad (19)$$

以轴承 1_7 和 2_3 为例,图 13 和 14 分别为轴承 1_7 和 2_3 的剩余使用寿命预测结果。已知数据集中轴承 1_7 当前寿命为 15020 s,从图 13 可知,预测的失效时刻为 20160 s,由式(19)计算得到 RUL 的预测值为 5140 s,而真实剩余使用寿命为 7570 s。同理,已知轴承 2_3 的当前寿命为 12010 s,从图 14 可知,失效时刻为 18560 s,则预测 RUL 为 6550 s,而真实剩余使用寿命为 7530 s。

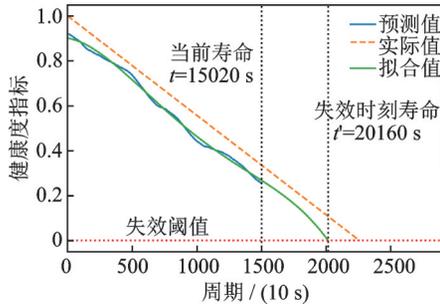


图 13 轴承 1_7 的 RUL 预测结果

Fig. 13 RUL prediction results of bearing 1_7

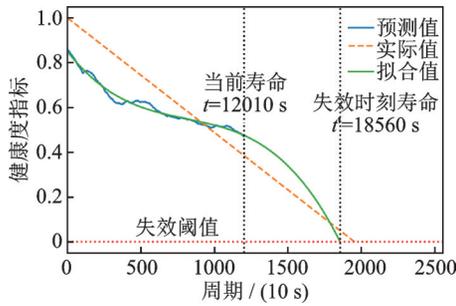


图 14 轴承 2_3 的 RUL 预测结果

Fig. 14 RUL prediction results of bearing 2_3

3.3 模型评估与误差分析

本文采用误差 E_i 来评估模型的预测精度,如下式所示:

$$E_i = \frac{\text{actRUL}_i - \text{predRUL}_i}{\text{actRUL}_i} \times 100\% \quad (20)$$

式中 actRUL_i 表示第 i 个测试轴承的真实剩余使用寿命; predRUL_i 表示第 i 个测试轴承的预测剩余使用寿命。

除评估模型的预测精度之外,还需建立一个总体指标来评估模型的有效性,即是否超前预测或滞后预测。因此,本文采用 PHM 2012 数据集设立的模型平均得分 $Score$ 来评估模型的有效性。计算方法如下式所示:

$$A_i = \begin{cases} e^{-\ln \frac{1}{2} \cdot \left(\frac{E_i}{5}\right)}, & E_i \leq 0 \\ e^{+\ln \frac{1}{2} \cdot \left(\frac{E_i}{20}\right)}, & E_i > 0 \end{cases} \quad (21)$$

$$Score = \frac{1}{n} \sum_{i=1}^{11} A_i \quad (22)$$

式中 A_i 表示第 i 个轴承的得分。

图 15 展示了误差 E_i 与得分 A_i 之间的关系。从图 15 中可以看出,当超前预测,即 $E_i > 0$ 时,模型的得分更高;反之,当滞后预测,即 $E_i \leq 0$ 时,模型的得分较超前预测低。这是因为在实际生产过程中,超前预测比滞后预测更有意义,因此该得分计算方法对滞后预测进行了惩罚,具有公平性与合理性。

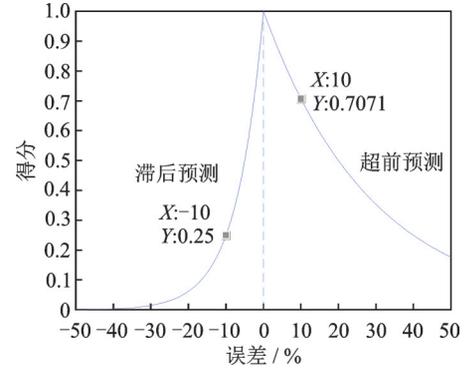


图 15 误差 E_i 与得分 A_i 的函数关系

Fig. 15 Function relationship of error E_i and score A_i

表 3 列出了测试集中 11 个轴承的平均绝对预测误差以及模型平均得分,采用 Transformer Encoder 模型^[7]、SOM 模型^[32]以及 CNN-LSTM 模型^[33]进行对比实验。其中文献[7]和[32]先提取退化特征建立健康度指标,再进行 RUL 预测,文献[33]属于“端到端”的 RUL 预测方法。从对比结果可以看出,本文提出的方法相较于文献[7],[32]和[33],平均绝对误差分别降低了 43.18%,62.57%和 59.44%,平均得分分别提高了 10.87%,45.71%和 34.21%。此外,为了进一步说明利用包络谱分段并计算皮尔逊相关系数提取退化特征的有效性,需要增加对比实验,即不对包络谱分段,对整个包络谱计算皮尔逊相关系数提取退化特征,并进行 RUL 预测,对比结果如表 4 所示。由表 4 可知,包络谱分段比不分段的平均绝对预测误差降低了 35.56%, $Score$ 提高了 49.02%。

表 3 RUL 预测结果与比较

Tab. 3 RUL prediction results and comparisons

轴承	平均绝对预测误差 / %			
	本文模型	文献[7]	文献[32]	文献[33]
1_3	33.68	74.17	-1.04	54.73
1_4	22.76	-0.69	85.81	38.48
1_5	-8.07	9.90	-278.20	-99.40
1_6	28.08	-12.33	19.18	-120.00
1_7	29.19	83.62	-7.13	70.65
2_3	13.01	61.35	10.49	75.53
2_4	5.04	5.06	51.80	19.81
2_5	22.33	-70.22	28.80	8.20
2_6	-1.55	0.78	-20.93	17.87
2_7	15.52	44.83	44.83	1.69
3_3	24.39	1.22	-3.66	2.93
$ E_i $	18.78	33.05	50.18	46.31
$Score$	0.51	0.46	0.35	0.38

表 4 对比实验结果

Tab. 4 Comparison of experimental results

对比实验条件	平均绝对预测误差/%	Score
包络谱分段($n=4$)	18.78	0.51
包络谱不分段($n=1$)	54.34	0.26

4 工程实际数据验证

本节将所提方法应用于实际现场设备数据,进一步验证方法的有效性。该数据来源于某石化企业离心泵,该泵为交流异步电机驱动、滚动轴承双支撑结构,额定转速 2980 r/min。如图 16 所示,泵两端的轴承座(3[#]轴承和 4[#]轴承)上均安装了 IEPE 振动加速度传感器,型号为 PCB 608A11,线性频响范围为 0.5~10 kHz,量程为±50g。图 16 中,Ha 表示水平方向的加速度;Va 表示垂直方向的加速度。每个轴承均采集了垂直和水平方向的振动信号。每个加速度传感器的采样频率均被设定为 25.6 kHz,采样间隔为 10 s,单次采样时间为 0.64 s。在监测的过程,泵两端的轴承发生了退化,最终轴承彻底损坏。图 17 为 3[#]轴承振动信号波形图,从图 17 中可以看出,信号的幅值在轴承损伤的最后阶段随着时间而增加。

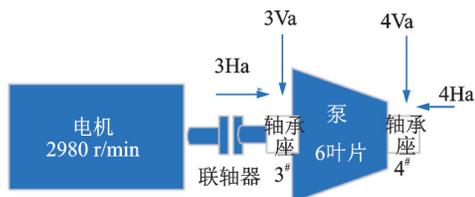


图 16 泵的测点布局图

Fig. 16 Survey point layout diagram of a pump

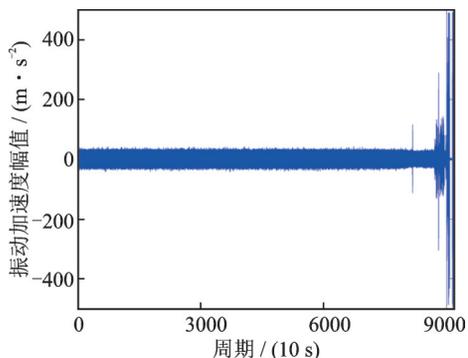


图 17 3[#]轴承时域振动信号

Fig. 17 Time domain vibration signal of 3[#] bearing

在现场设备数据的验证过程中,采用了与第 3 节相同的操作流程和相同的 n 值,在相同的实验环境下进行训练和验证。为了突出 LSTM-SA 模型的优势,文中用基于标准递归神经网络(Standard-RNN)和基于卷积神经网络(CNN)的 RUL 预测模型进行比较。图 18 为本文提出的基于 LSTM-SA 模型的 RUL 预测结果;图 19 和 20 分别为基于 Standard-RNN 和 CNN

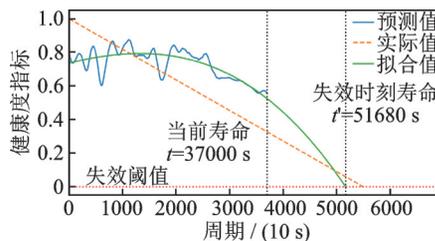


图 18 LSTM-SA 模型的 RUL 预测结果

Fig. 18 RUL prediction results of LSTM-SA model

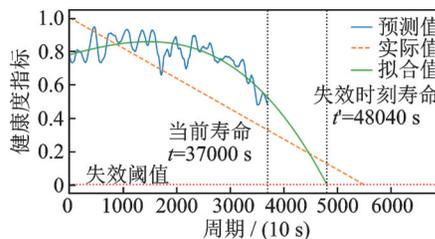


图 19 Standard-RNN 模型的 RUL 预测结果

Fig. 19 RUL prediction results of Standard-RNN model

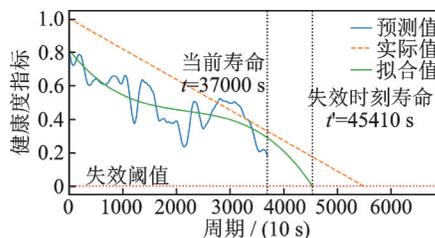


图 20 CNN 模型的 RUL 预测结果

Fig. 20 RUL prediction results of CNN model

的 RUL 的预测结果。结合表 5 三种模型比较结果可以看出,LSTM-SA 模型的平均预测误差分别比 Standard-RNN 和 CNN 模型降低了 39.58% 和 74.86%。该结果也进一步说明 LSTM-SA 模型可以有效地预测滚动轴承的剩余使用寿命。

表 5 RUL 预测结果与比较

Tab. 5 RUL prediction results and comparisons

测试轴承	当前时间/(10 s)	实际 RUL/(10 s)	LSTM-SA 预测 RUL/(10 s)	预测误差/%		
				LSTM-SA	Standard-RNN	CNN
3 [#]	3700	1798	1468	18.4	38.6	53.2
4 [#]	3900	1602	1341	16.3	18.9	-85.1
平均误差	—	—	—	17.4	28.8	69.2

5 结论

RUL 的预测精度很大程度上依赖于所构建的 HI。所提出的方法结合包络谱分段特征和 LSTM-SA 模型提高了滚动轴承 RUL 的预测准确率。在 HI 的构建过程中,提出了基于包络谱特征的退化特征提取方法。将退化特征输入到 LSTM-SA 中构建 HI。为了验证所提出方法的有效性,使用公

开的实验数据和真实现场数据进行验证。在实验数据集的验证中,所提出方法比文献[7],[32]和[33]中的方法表现更好。此外,真实现场数据的验证表明,所提出方法比基于 Standard-RNN 和基于 CNN 的方法更能有效地预测泵轴承的 RUL。

参考文献:

- [1] MOGHADDASS R, ZUO M J. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process[J]. *Reliability Engineering & System Safety*, 2014, 124: 92-104.
- [2] KHELIF R, CHEBEL-MORELLO B, MALINOWSKI S, et al. Direct remaining useful life estimation based on support vector regression[J]. *IEEE Transactions on Industrial Electronics*, 2017, 64(3): 2276-2285.
- [3] 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.
- [4] 李少鹏. 结合 CNN 和 LSTM 的滚动轴承剩余使用寿命预测方法研究[D]. 哈尔滨: 哈尔滨理工大学, 2019.
LI Shaopeng. Research on remaining useful life prediction method of a rolling bearing combined CNN and LSTM[D]. Harbin: Harbin University of Science and Technology, 2019.
- [5] WANG Xu, WANG Tianyang, MING Anbo, et al. Deep spatiotemporal convolutional-neural-network-based remaining useful life estimation of bearings[J]. *Chinese Journal of Mechanical Engineering*, 2021, 34(3): 115-129.
- [6] 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.
- [7] 周哲韬, 刘路, 宋晓, 等. 基于 Transformer 模型的滚动轴承剩余使用寿命预测方法[J]. *北京航空航天大学学报*, 2023, 49(2): 430-443.
ZHOU Zhetao, LIU Lu, SONG Xiao, et al. Remaining useful life prediction method of rolling bearing based on Transformer model[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2023, 49(2): 430-443.
- [8] LEI Yaguo, LI Naipeng, GUO Liang, et al. Machinery health prognostics: a systematic review from data acquisition to RUL prediction[J]. *Mechanical Systems & Signal Processing*, 2018, 104: 799-834.
- [9] TSUI K L, CHEN N, ZHOU Q, et al. Prognostics and health management: a review on data driven approaches[J]. *Mathematical Problems in Engineering*, 2015, 2015: 793161.
- [10] GUO L, LEI Y G, LI N P, et al. Deep convolution feature learning for health indicator construction of bearings[C]//2017 Prognostics and System Health Management Conference (PHM). 2017: 1-6.
- [11] 韩林洁. 基于 CNN 和 GRU 的滚动轴承退化状态评估及剩余寿命预测[D]. 北京: 机械科学研究总院, 2020.
HAN Linjie. Degradation state assessment and remaining life prediction of rolling bearing based on CNN and GRU[D]. Beijing: China Academy of Machinery Sciences and Technology, 2020.
- [12] LI X Q, JIANG H K, XIONG X, et al. Rolling bearing health prognosis using a modified health index based hierarchical gated recurrent unit network[J]. *Mechanism and Machine Theory*, 2019, 133: 229-249.
- [13] WU Yuting, YUAN Mei, DONG Shaopeng, et al. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks[J]. *Neurocomputing*, 2018, 275: 167-179.
- [14] Caterini A L, Chang D E. Deep Neural Networks in a Mathematical Framework[M]. Cham, Switzerland: Springer, 2018.
- [15] 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(1): 1-11.
QIU Dawei, LIU Zichen, Zhou Yiqing, et al. A novel fault type detection method of rolling bearing using transformer neural networks[J]. *Chinese High Technology Letters*, 2021, 31(1): 1-11.
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need[C]//Advances in neural information processing systems. Long Beach, CA, USA, 2017: 5998-6008.
- [17] RANDALL R B. Vibration-Based Condition Monitoring: Industrial, Automotive and Aerospace Applications[M]. New York: John Wiley & Sons, Inc., 2010.
- [18] RANDALL R B, ANTONI J. Rolling element bearing diagnostics—a tutorial[J]. *Mechanical Systems and Signal Processing*, 2011, 25(2): 485-520.
- [19] BORGHESEANI P, RICCI R, CHATTERTON S, et al. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions[J]. *Mechanical Systems and Signal Processing*, 2013, 38(1): 23-35.
- [20] RANDALL R B, ANTONI J, CHOBSAARD S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals[J]. *Mechanical Systems and Signal Processing*, 2001, 15(5): 945-962.
- [21] SCHOBER P, BOER C, SCHWARTE L A. Correlation coefficients: appropriate use and interpretation[J]. *Anesthesia & Analgesia*, 2018, 126(5): 1763-1768.
- [22] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. *Neural Computation*, 2000, 12(10): 2451-2471.
- [23] Zhao H S, Jia J Y, Koltun V. Exploring self-attention

- for image recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 19876591.
- [24] ZHANG H, ZHANG Q, SHAO S Y, et al. Attention-based LSTM network for rotatory machine remaining useful life prediction[J]. IEEE Access, 2020, 8(132): 188-199.
- [25] Mao W T, He J L, Tang J M, et al. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network[J]. Advances in Mechanical Engineering, 2018, 10(12): 1-18.
- [26] 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. Denver, USA, 2012: 1-8.
- [27] 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.
- [28] SINGLETON R K, STRANGAS E G, AVIYENTE S. Extended Kalman filtering for remaining-useful-life estimation of bearings[J]. IEEE Transactions on Industrial Electronics, 2015, 62(3): 1781-1790.
- [29] WOLD S, ESBENSEN K, GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1-3): 37-52.
- [30] GUO L, LI N P, JIA F, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings [J]. Neurocomputing, 2017, 240: 98-109.
- [31] PRESS W H, TEUKOLSKY S A. Savitzky-golay smoothing filters[J]. Computers in Physics, 1990, 4(6): 669-672.
- [32] 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.
- [33] HINCHI A Z, TKIOUAT 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.

Combination of LSTM and Self-Attention for remaining life prediction of rolling bearings

HUANG Yu¹, FENG Kun², GAO Jun-feng³, LI Zhou-zheng², JIANG Zhi-nong¹, GAO Jin-ji¹

(1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;

2. Beijing Key Laboratory of Health Monitoring and Self-Recovery of High-end Mechanical Equipment, Beijing University of Chemical Technology, Beijing 100029, China;

3. PetroChina Company Limited by Shares Refining & Chemical Branch, Beijing 100007, China)

Abstract: In order to construct a trend health index that accurately characterizes the degradation process of rolling bearings and improve the prediction accuracy of remaining useful life (RUL) of rolling bearings, a neural network model (LSTM-SA) combining long-short term memory (LSTM) and self-attention mechanism (Self-Attention) is proposed for RUL prediction of rolling bearings. The envelope spectrum of the original signal is obtained by using envelope demodulation, and then the envelope spectrum is segmented and the Pearson correlation coefficients of the corresponding frequency bands are calculated to extract the degradation features with monotonicity and trend. The degradation features are normalized and processed as the input of the LSTM-SA model, and the LSTM is used to adaptively extract the temporal internal correlation of the degradation features and the Self-Attention is used to screen key information. By eliminating the interference of useless information and mining deep-level features, healthiness indexes are constructed and degradation curves can be obtained. By determining the failure threshold, the degradation curves are fitted by the least squares method, and the life failure point is predicted, which realize the RUL prediction of rolling bearings. The experimental results on the PHM2012 dataset show that the proposed method reduces the average absolute error by 43.99%, 63.11% and 60.00%, respectively, and improves the average score by 10.87%, 45.71% and 34.21%, respectively, compared with other literature. The experimental results on the actual engineering data show that the average prediction error of the proposed method is higher than that of standard-RNN and CNN by 39.58% and 74.86%, respectively.

Key words: remaining useful life prediction; rolling bearing; long short-term memory network; Self-Attention mechanism; envelope spectrum characteristics

作者简介: 黄宇(1997—),男,硕士研究生。E-mail: buctyuhuang@163.com。

通讯作者: 冯坤(1982—),男,博士,教授。E-mail: kunfengphd@163.com。