基于嵌套长短期记忆网络的机械装备剩余使用寿命预测方法

程一伟^{1,2}, 朱海平², 吴军^{1*}, 邵新宇²

1. 华中科技大学船舶与海洋工程学院, 武汉 430074;

2. 华中科技大学机械科学与工程学院, 武汉 430074

* E-mail: wuj@hust.edu.cn

收稿日期: 2021-07-30; 接受日期: 2021-09-06; 网络版发表日期: 2021-11-17

国家自然科学基金(批准号: 51875225, 52075202)资助项目

摘要 剩余使用寿命(remaining useful life, RUL)预测是保障机械装备可靠性、可用性和安全性的重要技术. 本文提出一种基于嵌套长短期记忆(nested long short-term memory, NLSTM)网络的机械装备RUL预测方法, 它通过融合多传感器监测信号, 实现对机械装备RUL的精确预测. 区别于普通LSTM网络, NLSTM将存储单元进一步加深, 将一个LSTM神经元结构嵌套在原有LSTM的存储空间中, 实现对多传感器时间序列信号中长期依赖性的深度捕捉. 本文使用涡扇发动机和加工刀具两个实验案例来验证NLSTM的预测性能; 从涡扇发动机案例验证可知, 相比于LSTM, NLSTM的预测性能在两个指标上分别整体提升了4.66%和15.18%, 且NLSTM的预测结果也优于文献中的其他先进方法; 从加工刀具案例验证可知, NLSTM的预测结果在六个刀具上的预测结果均优于LSTM.

关键词 循环神经网络, 嵌套长短期记忆网络, 剩余使用寿命预测, 多传感监测数据, 机械装备

1 引言

机械装备的可靠性、可用性和安全性一直是装备制造、航空航天和交通运输等行业的关键问题. 开展机械装备的剩余使用寿命(remaining useful life, RUL)预测可以发现其健康退化规律, 估计和预测其剩余有效工作时间, 以便在有限的时间内安排合适的运维计划, 避免机械装备出现突发故障; 同时, 也可以根据预测的RUL提前采取预防性维修策略, 以缩短停机时间, 提升运行效率^[1~4].

据文献[5~7], 现有的RUL预测方法大致可以分为两类, 即基于模型的方法和数据驱动的方法. 基于模型的方法通过分析装备运行原理和退化机理, 构建物理或数学模型来描述装备健康退化过程. 然而, 在实际场景中装备的健康退化过程受多种因素的影响, 导致退化建模中存在许多不确定的参数, 这限制了基于模型的方法的广泛应用. 相反, 数据驱动的方法对采集到的系统退化过程监测信号进行深度分析和挖掘, 从数据维度预测装备RUL, 无需构建装备的物理或数学退化模型. 因此, 数据驱动的RUL预测方法已经成为

引用格式: 程一伟, 朱海平, 吴军, 等. 基于嵌套长短期记忆网络的机械装备剩余使用寿命预测方法. 中国科学: 技术科学, 2022, 52: 76–87
Cheng Y W, Zhu H P, Wu J, et al. A remaining useful life prediction method based on nested long short-term memory network for mechanical equipment (in Chinese). Sci Sin Tech, 2022, 52: 76–87, doi: 10.1360/SST-2021-0332

研究热点之一。

时间序列信号是数据驱动的RUL预测中最常用的信号类型, 这是因为它们能够很好地反映出装备健康退化过程^[8,9]。许多使用时间序列数据的RUL预测方法已经成功地实现了较好的预测效果。Ompusunggu等人^[10]提出了一种基于卡尔曼滤波器的RUL预测方法, 并成功将其应用于自动变速器离合器。Wu等人^[11]利用极限学习机对锂离子电池进行了RUL预测, 并采用萤火虫群优化算法对模型参数进行优化。Khelif等人^[12]提出了一种基于支持向量回归的直接RUL估计方法, 该方法避免了退化状态估计或故障阈值设置。奚立峰等人^[13]综合利用自组织映射网络和BP神经网络从不同维度提取球轴承的退化指标, 进而实现剩余寿命预测。Wang等人^[14]利用深度自编码器提取滚动轴承的深度退化指标, 并使用一个线性映射方法实现RUL预测。Sun等人^[15]提出了一种改进的逆高斯过程模型用于液压活塞泵的RUL预测, 该模型考虑了随机效应和测量误差来描述磨损退化。Wang等人^[16]使用三层的神经网络来预测未来的健康指标, 并将预测的健康指标输入比例风险回归模型来估计风险概率和生存概率。尽管这些方法实现了RUL预测, 但它们均通过探索实时监测信号和RUL值之间的映射关系来实现RUL预测, 没有考虑监测信号不同时刻的健康状态间的相关性, 即数据之间的时间依赖性。

长短期记忆神经网络(long short-term memory, LSTM)是解决上述问题的一种方案。LSTM拥有独特的记忆机制, 允许模型在任意的时间尺度上记忆过去的信息, 使其可以跨时间尺度地从先前处理的数据中提取有用的重要信息, 并将其循环到当前神经元状态中, 从而对时间序列数据进行建模^[17]。目前, LSTM被广泛用于装备RUL预测中。Guo等人^[18]介绍了一种基于LSTM的预测方法, 通过融合多个特征数据, 构建了可用于滚动轴承RUL预测的健康指标。Cheng等人^[19]利用LSTM对滚动轴承进行了失效预测, 实现了较高的预测精度。Qin等人^[20]提出一种基于宏观-微观注意力的改进型LSTM网络, 并将其成功应用于齿轮箱的RUL预测。Xiang等人^[21]构建了一种注意力引导的有序神经元LSTM网络, 并提出了基于该模型的旋转机械RUL预测方法。尽管LSTM可以实现对时间抽象特征的学习, 但有效捕捉时间序列信号中的长期依赖性仍然是非常有挑战性的。

为了强化模型的长期依赖特征捕捉能力, 实现高精度的RUL预测, 本文提出一种基于嵌套长短期记忆网络(nested LSTM, NLSTM)的RUL预测方法。该方法通过融合多传感器监测信号, 实现对设备RUL的精确预测。NLSTM模型对LSTM神经元中的存储空间进行改造, 将LSTM神经元嵌套在存储空间中, 实现对时间序列信号中的长期依赖性的深度捕捉。同时, 采用涡扇发动机和加工刀具两个实验案例对所提出方法的有效性进行验证。

本文的创新点如下: (1) 首次将NLSTM引入RUL预测场景中, 提出一种基于NLSTM网络的RUL预测方法, 实现机械装备RUL的精准预测; (2) 构建一种新的NLSTM网络结构, 可以更好地挖掘时间序列信号数据中的长期依赖特性, 获取更好的预测效果; (3) 所提出的方法能适用于多传感器监测场景, 后文的两个案例验证可以证明该方法在多传感监测场景的适用性和有效性。

2 模型原理

2.1 LSTM

与普通神经网络不同, 循环神经网络(recurrent neural network, RNN)具有独特的循环结构, 即隐藏层的可循环返回本层作为下一时刻的输入, 这意味着隐藏层在时间维度与自身具有自连接。因此, RNN有很强的处理顺序和时间相关数据的能力^[22]。RNN隐藏层的输出可以表示为

$$\mathbf{h}_t = \varphi(\mathbf{w}_{hx}\mathbf{x}_t + \mathbf{w}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h), \quad (1)$$

其中, $\mathbf{w}_{hx}$ 和 $\mathbf{w}_{hh}$ 是隐藏层的输入数据 $\mathbf{x}_t$ 和隐藏层上一时刻的输出 $\mathbf{h}_{t-1}$ 的权重系数, $\mathbf{b}_h$ 表示隐藏层的偏置。然而, 由于模型训练过程中反向传播过程中存在的梯度消失问题, 使得RNN无法捕获数据中的长期相关性。

为了缓解这个问题, Hochreiter和Schmidhuber^[23]在RNN架构的基础上提出了LSTM。在LSTM网络结构中, LSTM神经元代替传统的RNN隐藏层神经元来构造LSTM层。每个LSTM神经元都有记忆单元以及特殊设计的门结构, 即遗忘门、输入门和输出门。这种结构保证了LSTM神经元具有发现和记忆长期依赖关系的能力。图1描绘了LSTM神经元的结构。

如图1所示, LSTM神经元中的三个门结构控制了

图 1 LSTM神经元结构图

Figure 1 The structure of LSTM neuron.

信息的输入和遗忘, 提供了良好的非线性控制机制. 输入门决定哪些信息将被输入到神经元的记忆单元, 遗忘门决定神经元记忆状态中需要丢弃的信息, 输出门决定从神经元状态输出什么信息. LSTM神经元的计算过程可以用数学方法表示为

$$\mathbf{i}_t = \sigma(\mathbf{w}_{ix}\mathbf{x}_t + \mathbf{w}_{ih}\mathbf{h}_{t-1} + \mathbf{b}_i), \quad (2)$$

$$\mathbf{f}_t = \sigma(\mathbf{w}_{fx}\mathbf{x}_t + \mathbf{w}_{fh}\mathbf{h}_{t-1} + \mathbf{b}_f), \quad (3)$$

$$\mathbf{o}_t = \sigma(\mathbf{w}_{ox}\mathbf{x}_t + \mathbf{w}_{oh}\mathbf{h}_{t-1} + \mathbf{b}_o), \quad (4)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tau(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{h}_{t-1} + \mathbf{b}_c), \quad (5)$$

$$\mathbf{h}_t = \tau(\mathbf{c}_t) \odot \mathbf{o}_t, \quad (6)$$

其中, $\mathbf{w}_{ix}$, $\mathbf{w}_{fx}$, $\mathbf{w}_{ox}$ 和 $\mathbf{w}_{cx}$ 是输入数据 $\mathbf{x}_t$ 的权重; $\mathbf{w}_{ih}$, $\mathbf{w}_{fh}$, $\mathbf{w}_{oh}$ 和 $\mathbf{w}_{ch}$ 是LSTM神经元上一时刻的输出 $\mathbf{h}_{t-1}$ 的权重系数; $\mathbf{b}_i$, $\mathbf{b}_f$, $\mathbf{b}_o$ 和 $\mathbf{b}_c$ 分别表示输入门、遗忘门、输出门和输入节点的偏差; $\mathbf{i}_t$, $\mathbf{f}_t$ 和 $\mathbf{o}_t$ 分别表示输入门、遗忘门、输出门的输出; σ 和 τ 表示sigmoid和tanh函数; $\mathbf{c}_t$ 和 $\mathbf{c}_{t-1}$ 是 t 和 $t-1$ 时的LSTM神经元状态; $\odot$ 表示矩阵点乘运算.

2.2 NLSTM

为了实现对时间序列信号中的长期依赖性的深度捕捉, 对LSTM神经元中的存储空间进行改造, 以构建如图2所示的NLSTM模型.

如图2所示, 相比于LSTM神经元, NLSTM神经元使用了LSTM神经元结构取代了传统LSTM神经元中的状态 $\mathbf{c}_t$, 实现了在LSTM神经元中嵌套了一个LSTM. 将原来简单的内部存储 $\mathbf{c}_t$ 升级成了更为复杂的LSTM

图 2 NLSTM神经元结构图

Figure 2 The structure of NLSTM neuron.

结构, 实现对长期依赖性的深度捕捉. 具体地, 将 $\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tau(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{h}_{t-1} + \mathbf{b}_c)$ 拆解成两部分, 分别为 $\mathbf{f}_t \odot \mathbf{c}_{t-1}$ 和 $\mathbf{i}_t \odot \tau(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{h}_{t-1} + \mathbf{b}_c)$, 分别作为内层LSTM的 $\tilde{\mathbf{h}}_t$ 和 $\tilde{\mathbf{x}}_t$. 因此, 内层LSTM可以表示为

$$\tilde{\mathbf{h}}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1}, \quad (7)$$

$$\tilde{\mathbf{x}}_t = \mathbf{i}_t \odot \tau(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{h}_{t-1} + \mathbf{b}_c), \quad (8)$$

$$\tilde{\mathbf{i}}_t = \sigma(\tilde{\mathbf{w}}_{ix}\tilde{\mathbf{x}}_t + \tilde{\mathbf{w}}_{ih}\tilde{\mathbf{h}}_{t-1} + \tilde{\mathbf{b}}_i), \quad (9)$$

$$\tilde{\mathbf{f}}_t = \sigma(\tilde{\mathbf{w}}_{fx}\tilde{\mathbf{x}}_t + \tilde{\mathbf{w}}_{fh}\tilde{\mathbf{h}}_{t-1} + \tilde{\mathbf{b}}_f), \quad (10)$$

$$\tilde{\mathbf{o}}_t = \sigma(\tilde{\mathbf{w}}_{ox}\tilde{\mathbf{x}}_t + \tilde{\mathbf{w}}_{oh}\tilde{\mathbf{h}}_{t-1} + \tilde{\mathbf{b}}_o), \quad (11)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tau(\tilde{\mathbf{w}}_{cx}\tilde{\mathbf{x}}_t + \tilde{\mathbf{w}}_{ch}\tilde{\mathbf{h}}_{t-1} + \tilde{\mathbf{b}}_c), \quad (12)$$

$$\tilde{\mathbf{h}}_t = \tau(\tilde{\mathbf{c}}_t) \odot \tilde{\mathbf{o}}_t. \quad (13)$$

最终, NLSTM神经元的输出变为

$$\mathbf{c}_t = \tilde{\mathbf{h}}_t, \quad (14)$$

$$\mathbf{h}_t = \tau(\mathbf{c}_t) \odot \mathbf{o}_t. \quad (15)$$

3 基于NLSTM网络的RUL预测

NLSTM网络模型结构如图3所示, 包含了输入层、NLSTM层(含Dropout层)、全连阶层和输出层. 假设 $\mathbf{X}_t = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, \dots, \mathbf{x}_N\}^T$ 表示第 t 个输入数据样本, 其中 $\mathbf{x}_t \in \mathbb{R}^{s \times 1}$ 是 t 时刻来自 s 个所选传感器的信号数据组成的向量. 如图3所示, 输入的多传感监测数据首先被输入到NLSTM层中, 获取输出 $\mathbf{h}_t$, 具体表示为

图 3 (网络版彩图) NLSTM网络
Figure 3 NLSTM network.

$$\mathbf{h}_t = \mathbb{H}(\mathbf{X}_t), \quad (16)$$

其中, $\mathbb{H}$ 表示NLSTM层对数据的非线性变换, 具体可以参考公式(2)~(15). 其次, NLSTM层的输出被输入到全连接层中进行线性转换, 以对数据进行进一步变换和降维, 表示为

$$\mathbf{F}_t = \mathbf{W}_1 \mathbf{X}_t + \mathbf{b}_1, \quad (17)$$

其中 $\mathbf{W}_1$ 和 $\mathbf{b}_1$ 分别是权重矩阵和偏置. $\mathbf{X}_t$ 表示 t 时刻输入的数据张量, $\mathbf{F}_t$ 表示第 t 输出的数据张量. 接着, 数据被输入到输出层进一步降维, 并输出预测的RUL结果, 表示为

$$\hat{y}_t = \mathbf{W}_2 \mathbf{F}_t + b_2, \quad (18)$$

其中 $\mathbf{W}_2$ 和 b_2 分别是权重矩阵和偏置. 值得说明的是, 由于输出的RUL为标量, 所以输出层中的神经元个数设置为1.

在模型训练阶段, 将输入数据输入到NLSTM模型中, 模型按照公式(16)~(18)的过程前向传递. 为了实现NLSTM的反向传播, 需要为NLSTM模型设计一个恰当的模型损失函数. 基于模型输出的预测值 $\hat{y}_t$ 和真实的数据值 y_t , 引入均方误差函数, 设计如下的损失函数

$$L = \frac{1}{K} \sum_{t=1}^K (y_t - \hat{y}_t)^2, \quad (19)$$

其中, K 表示训练样本数量.

在测试过程中, 先将在线测试数据预处理, 再实时输入到训练好的NLSTM模型中, 即可得到预测的RUL.

4 案例验证1

4.1 数据集介绍

为了验证提出方法的有效性, 本文采用了NASA涡扇发动机数据集^[24]. 本数据集通过商业化的航空涡扇发动机软件(CMAPSS)来模拟发动机在不同飞行高度、马赫数等参数的情况下各关键部件的故障和性能退化过程. 在试验中, 涡扇发动机的初始状态均为正常状态(拥有随机的磨损程度), 随着运行过程逐渐产生故障, 导致性能下降, 直至系统发生故障. C-MAPSS通过改变输入参数(包括燃油速度和压力)来模拟涡扇发动机旋转部件的不同故障和退化过程.

NASA涡扇发动机退化数据集一共包含了4个子集, 分别为FD001~FD004. 不同的数据子集拥有不同的运行工况数量和故障数量. 在四个子集中, 所有的训练发动机都具有全寿命数据, 即从初始状态到最终失效状态的数据. 测试发动机的仅包含从初始状态到最终失效前的任意时刻的传感器数据, 以供数据使用者对测试发动机的RUL进行预测. 涡扇发动机数据集四个子集的信息如表1所示.

为了精确地记录发动机健康退化过程, 21个传感器信号被实时记录以监测发动机的状态数据, 涡扇发动机的结构以及传感器信号的具体信息可以参考文献[24].

4.2 预测结果评价指标

为了评估RUL预测的效果, 本文采用均方根误差(root mean squared error, RMSE)和得分(Score)这两种常用的指标来评价预测性能.

RMSE是预测领域经典的评估指标, 公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N d_n^2}, \quad (20)$$

其中, $d_n = \widehat{R}_n - R_n$ 表示预测RUL $\widehat{R}_n$ 和实际RUL R_n 之

表 1 涡扇发动机数据集信息

Table 1 Information of turbofan engine dataset

数据子集	训练发动机数量	测试发动机数量	运行工况数量	故障数量
FD001	100	100	1	1
FD002	260	259	6	1
FD003	100	100	1	2
FD004	249	248	6	2

间的预测误差, n 表示测试发动机台数. 预测误差绝对值越小, RMSE越低.

Score是2012年预测和健康(PHM)挑战赛发布的一个官方评价指标^[24], 它通过一个得分函数来评估预测结果. 预测误差绝对值越小, Score越低. 得分函数表示为

$$\text{Score} = \sum_{n=1}^N s_n, s_n = \begin{cases} \exp\left(-\frac{d_n}{13}\right) - 1, & \text{if } d_n < 0, \\ \exp\left(\frac{d_n}{10}\right) - 1, & \text{if } d_n \geq 0. \end{cases} \quad (21)$$

图4显示了这两个预测性能评估指标之间的差异. 很明显, RMSE与预测误差是线性相关的, 而Score与预测误差是非线性相关的, 其对滞后预测的惩罚大于早期预测, 因为滞后预测可能会造成更加严重的后果.

此外, 为了比较NLSTM与LSTM性能提升水平, 使用如下公式衡量提升的程度:

图 4 (网络版彩图) RMSE与Score对比图

Figure 4 (Color online) Comparison between RMSE and Score.

$$\text{Imp} = \frac{I_{\text{LSTM}} - I_{\text{NLSTM}}}{I_{\text{LSTM}}} \times 100\%, I = \text{RMSE, Score}. \quad (22)$$

4.3 数据预处理

(1) 传感器信号筛选

尽管C-MAPSS数据集中记录了涡扇发动机的21种传感器信号数据, 但并非所有传感器信号数据都是有用的. 图5中展示了FD001中所有训练发动机的21种传感器信号. 显而易见的是, 这些传感器信号表现出不同的变化过程. 其中, 有些传感器信号值保持不变, 有

图 5 (网络版彩图) FD001中所有训练发动机的21种监测信号

Figure 5 (Color online) 21 monitoring signals of training engines in FD001.

些传感器信号值在不同的发动机上表现出了相反的趋势. 这些信号对于RUL预测来说都是无效信息. 在21种传感器信号中, 12种传感器信号展现出了航空发动机的健康变化, 包括传感器信号2, 3, 4, 7, 8, 9, 11, 12, 15, 17, 20和21. 因此, 选择这12个传感器的信号作为RUL预测方法的输入数据.

(2) 数据归一化

通常, 机械装备的监测信号分布在不同的数值范围内, 具有不同的尺度. 如果将这些信号直接输入到神经网络模型中, 则会出现数据淹没的问题, 即数据值大的信号数据和数据值小的信号数据融合时, 神经网络会对数据值大的信号数据的变化更敏感, 相反, 数据值小的信号数据的变化可能会被忽略掉, 导致每个传感器信号数据的贡献差异较大. 为了应对这个问题, 对所有监测信号数据进行归一化处理, 以避免不同传感器信号因分布数值范围不同而导致的数据贡献差异大的问题. 采用最小最大归一化方法对所有信号数据归一化, 公式如下:

$$x_{\text{norm}}^{i,j} = \frac{x^{i,j} - x_{\min}^{i,j}}{x_{\max}^{i,j} - x_{\min}^{i,j}}, \forall i, j, \quad (23)$$

其中 $x^{i,j}$ 表示在第 i 个工况条件下从第 j 个传感器获得的原始数据值. $x_{\max}^{i,j}$ 和 $x_{\min}^{i,j}$ 表示所有训练设备中第 i 个工况条件下第 j 个传感器信号的最大值和最小值.

图6展示了FD001数据集中第1台训练发动机的12个归一化信号. 从图中可以看到所有归一化后的传感器信号值均处在[0,1]的区间范围内.

(3) 数据时间窗处理

在基于多传感器信号的RUL预测问题中, 向预测模型的输入数据中嵌入时间信息是一项重要的操作^[25,26]. 如果利用一个采样周期采集的信号作为预测模型的一个输入样本, 就不可避免地会忽略时间序列信号中的时间信息. 本文利用时间窗处理技术对多传感器数据进行处理, 采用固定大小的时间窗对连续采样周期采集的多传感器信号数据进行串接, 得到高维向量以作为NLSTM模型的输入. 图7展示了时间窗处理的示例. 图7中, 一个大小为25的时间窗被用于从监测信号中获取数据以构造高维输入向量. 一旦时间窗完成这次采集, 它将沿时间方向滑动一个采样周期, 以执行下一次采集, 直到信号结束.

许多已发表的论文认为分段线性标签技术适用于

图6 (网络版彩图) FD001数据集中第1台训练发动机的归一化信号

Figure 6 (Color online) Normalized signals of the first training engine in FD001.

图7 (网络版彩图)滑动时间窗示意图

Figure 7 (Color online) Schematic diagram of sliding time window.

C-MAPSS数据集^[27~29]. 分段线性标签技术假设发动机在初始阶段以不变的RUL正常工作, 在最后阶段线性退化. 参照文献^[27~29], 将分段线性标签值设为125. 图8显示了本文使用的分段线性标签技术. 从图中可以看到, 传统的真实RUL标签假设RUL随飞行周期的增加而线性减小. 本文采用分段线性标号法, 认为发动机RUL在初始阶段是固定的, 后期退化的标签值与传统线性标签相同.

4.4 预测结果分析与讨论

本节构造NLSTM网络预测模型, 自适应地挖掘多

图 8 (网络版彩图)分段线性标签图
Figure 8 (Color online) Piece wise label.

图 9 (网络版彩图)拥有不同时间窗尺寸的模型预测性能箱线图
Figure 9 (Color online) Boxplot of model prediction performance with different time window sizes.

传感器时间序列信号中的长期依赖特征, 并利用训练好的模型, 通过输入多传感器在线信号, 对实时RUL进行预测. 为了获得最优预测结果, 本节将对影响NLSTM模型预测性能的2个关键因素进行分析, 包括时间窗分析和NLSTM的网络结构分析. 其次将NLSTM的预测结果与LSTM模型及文献中的其他先进方法进行比较.

(1) 时间窗分析

输入信号序列长度是影响NLSTM模型预测性能的一个重要因素. 本小节讨论时间窗大小(即输入序列长度)对NLSTM模型性能的影响. 选择不同大小的时间窗(5~30)来为NLSTM模型构造输入向量以研究其

对预测性能的影响. 图9展示了拥有不同时间窗尺寸的NLSTM在FD001数据集中验证数据上的预测性能箱线图. 从图9可以明显看出, 随着时间窗尺寸的扩大, 预测结果的RMSE值逐渐减小, 这是由于时间窗尺寸较大, 使得输入向量可以包含更多的局部退化信息. 但是我们也应该知道, 输入序列的长度越长, 将导致计算过程越复杂, 模型训练时间也越长. 另外, 随着时间窗尺寸的增大, 性能提升的趋势变得不那么明显. 最终确定NLSTM模型的输入序列长度设定为25.

(2) 网络结构分析

网络结构是影响NLSTM模型预测性能的另一个重要因素. 本节研究了不同网络结构对所提出模型的性能的影响. 表2展示了待研究的8种网络结构. 不同结构NLSTM层和全连接层神经元数目不同. 每种结构采用随机交叉验证技术进行训练和校验. 所有网络结构的校验性能如表2所示. 从表中可以看出, 第5种网络结构的预测效果最优, 最终确定模型的结构为NLSTM层有64个神经元, 全连接层有32个神经元.

模型训练完成后, 将测试发动机数据输入到训练好的NLSTM模型中, 得到测试发动机的RUL. 图10展示了FD001中100台测试发动机的真实RUL与预测RUL对比结果. 从图中可以看到, 提出方法的预测RUL结果都与真实RUL结果有很高的相似度, 平均预测精度为0.79, 这证实了NLSTM模型预测的准确性.

(3) NLSTM与LSTM预测结果对比

本节对NLSTM和LSTM预测的性能进行了比较. 为了比较的公平性, LSTM网络与NLSTM拥有相同的网络结构和参数. 实验进行了十次, 以消除随机误差

表 2 FD001数据集中不同网络结构的预测性能

Table 2 Prediction performance of different network structures in FD001

序号	NLSTM层 神经元数目	全连接层 神经元数目	校验RMSE
1	16	8	22.49
2	16	16	19.39
3	32	16	18.54
4	32	32	17.28
5	64	32	16.94
6	64	64	17.11
7	128	64	17.98
8	128	128	18.91

的影响. 图11展示了NLSTM与LSTM模型在四个数据子集上的预测结果. 可以看出, NLSTM在所有数据子集上的RMSE均显著低于LSTM, 这证明了NLSTM模型相较于LSTM可以更好地实现对时间序列信号中的长期依赖性的深度捕捉.

(4) NLSTM与论文中其他先进方法对比

为了证明NLSTM模型的优越性, 本节采用一些在该数据集上成功应用的先进技术的预测结果来与提出的NLSTM模型预测结果进行比较, 包括多层感知器^[30]、支持向量回归^[30]、相关向量回归^[30]、卷积神经网络(convolutional neural networks, CNN)^[30]、卷积神经网络-门限循环单元(CNN-Gate recurrent unit, CNN-GRU)^[31], GRU和LSTM. 表3显示了NLSTM与其他先进方法的比较结果. 由表中可以看出, 该方法的RMSE和Score值均低于其他先进方法, 这证明了NLSTM具有较好的RUL预测性能. 此外, Imp指标值被用于衡量NLSTM相较于LSTM的性能提升. 从表中

可以看到, 性能提升最低为FD003的RMSE指标, 提升了1.62%, 最高为FD004的Score指标, 提升了22.37%. 在RMSE指标上, 四个数据子集的平均Imp为4.66%. 在Score指标上, 四个数据子集的平均Imp为15.18%.

5 案例验证2

5.1 数据集介绍

采用加工刀具数据验证所提出方法的有效性. 该加工刀具数据集源于2010年PHM数据挑战赛. 数据集包含六个加工刀具(C1~C6)的全生命周期测试数据, 每次加工刀具切削过程都可以监测到切削力信号(X, Y, Z 三个方向)、振动信号(X, Y, Z 三个方向)以及声发射信号. 关于该数据集的更多详细信息请参考文献^[32].

针对C1~C6, 采用交叉验证策略训练模型, 即每组刀具数据轮流作为测试数据, 其他五组刀具数据作为训练数据.

图 10 (网络版彩图) FD001中测试发动机的预测RUL与真实RUL对比

Figure 10 (Color online) Comparison between predicted RUL and real RUL of test engines in FD001.

图 11 (网络版彩图) NLSTM与LSTM预测结果对比. (a) FD001数据集; (b) FD002数据集; (c) FD003数据集; (d) FD004数据集

Figure 11 (Color online) Comparison of prediction results for NLSTM and LSTM. (a) FD001 dataset; (b) FD002 dataset; (c) FD003 dataset; (d) FD004 dataset.

表 3 NLSTM网络与其他先进方法的比较结果

Table 3 Comparison results between NLSTM network and other advanced methods

方法	FD001		FD002		FD003		FD004	
	Score	RMSE	Score	RMSE	Score	RMSE	Score	RMSE
多层感知器 ^[30]	17972	37.56	780280	80.03	17409	37.39	5616600	77.37
支持向量回归 ^[30]	1382	20.96	589900	42.00	1598	21.05	371140	45.35
相关向量回归 ^[30]	1503	23.80	17423	31.30	1432	22.37	26509	34.34
CNN ^[30]	1287	18.45	13570	30.29	1596	19.82	7887	29.16
CNN-GRU ^[31]	761	18.67	8110	27.16	1400	18.98	7610	27.11
GRU	721	18.94	7556	25.91	891	19.14	7934	28.46
LSTM	694	18.21	6469	25.67	640	18.78	6144	28.19
NLSTM	607	17.37	5842	23.73	565	18.48	5021	27.10
Imp	14.33%	4.84%	10.73%	8.18%	13.27%	1.62%	22.37%	4.02%

5.2 预测结果分析与讨论

首先, 由于NLSTM无法从非过程数据(例如力信号和声发射信号)中直接捕捉数据特征, 因此需要利用统计特征从复杂非过程数据中提取特征. 本文所使用的特征如表4所示, 其中 x_i 表示每次采样信号中的第 i 数据点, $i = 1, 2, \dots, N$, N 表示每次采样信号数据点数量, $\bar{x}$ 表示每次采样所有信号数据点的均值.

表 4 数据特征

Table 4 Data features

序号	公式	描述
1	$F_1 = \frac{1}{N} \sum_{i=1}^N x_i $	绝对均值
2	$F_2 = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$	标准差
3	$F_3 = \max\{ x_i \}$	峰值
4	$F_4 = \max\{x_i\} - \min\{x_i\}$	峰峰值
5	$F_5 = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	均方根值
6	$F_6 = \left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right)^2$	方根幅值
7	$F_7 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$	峭度
8	$F_8 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$	歪度

接下来, 构造NLSTM网络预测模型, 利用多传感器监测信号提取的特征数据训练模型. 然后, 利用训练好的模型, 对在线数据进行RUL预测. NLSTM网络参数优化过程与案例1相同. 最终, NLSTM网络被设置为NLSTM层有128个神经元, 全连接层有64个神经元. 时间窗尺寸设置为15. 图12展示了刀具C1的预测RUL. 从图中可以看出, 预测误差在初始阶段相对较大, 整体趋势与实际RUL一致.

此外, 对LSTM与NLSTN的预测结果也进行了对比, 对比结果中如图13所示. 从图中可以看出, NLSTN在6个刀具上的预测结果均优于LSTM, 这也证明了NLSTM的嵌套结构相比于LSTM可以更好地挖掘时

图 12 (网络版彩图)刀具C1的预测RUL

Figure 12 (Color online) Predicted RUL of machining tool C1.

图 13 (网络版彩图)加工刀具数据集NLSTM与LSTM的预测结果对比

Figure 13 (Color online) Comparison of prediction results between NLSTM and LSTM in machining tool dataset.

间序列信号数据中的长期依赖特性, 获取更好的预测效果.

6 结论与展望

本文提出了一种新的基于NLSTM网络的机械装备RUL预测方法. 该方法利用多传感监测数据进行建模, 实现更准确、更稳健的预测. NLSTM模型对LSTM神经元中的存储空间进行改造, 将LSTM神经元嵌套在存储空间中, 实现对时间序列信号中的长期依赖性的深度捕捉. 为验证所提出方法的有效性和科学性, 本文分别使用了涡扇发动机和加工刀具实验案例来验证NLSTM的预测性能. 在涡扇发动机实验案例中, 开展了传感器信号选择、数据归一化和数据时间窗处理等数据预处理, 利用NLSTM模型对涡扇发动机RUL进行了预测. 实验结果显示, NLSTM相较于LSTM在Score和RMSE两个指标上分别提升了4.66%和15.18%, 此外, NLSTM的预测结果也优于文献中的其他先进方法. 在加工刀具实验案例中, 由于采集数据为非过程数据, 对多传感监测数据进行特征提取, 然后构建NLSTM实施RUL预测. 实验结果显示, NLSTM模型在六个刀具上的预测结果均优于LSTM.

参考文献

- 1 Lei Y, Li N, Guo L, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech Syst Signal Process*, 2018, 104: 799–834
- 2 Wu J, Wu C, Cao S, et al. Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines. *IEEE Trans Ind Electron*, 2019, 66: 529–539
- 3 Qin Q B, Wu S C, Hu Y N, et al. Fatigue strength and residual life assessment of high-speed railway vehicle used S38C hollow axles (in Chinese). *Sci Sin Tech*, 2019, 49: 840–850 [秦庆斌, 吴圣川, 胡雅楠, 等. 高速动车组S38C车轴疲劳强度及剩余寿命评价. 中国科学: 技术科学, 2019, 49: 840–850]
- 4 Xiang S, Qin Y, Zhu C, et al. Long short-term memory neural network with weight amplification and its application into gear remaining useful life prediction. *Eng Appl Artificial Intell*, 2020, 91: 103587
- 5 Li L L, Zhang X B, Tseng M L, et al. Optimal scale Gaussian process regression model in Insulated Gate Bipolar Transistor remaining life prediction. *Appl Soft Comput*, 2019, 78: 261–273
- 6 Wang Y, Deng C, Wu J, et al. A corrective maintenance scheme for engineering equipment. *Eng Failure Anal*, 2014, 36: 269–283
- 7 Liao L, Köttig F. A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Appl Soft Comput*, 2016, 44: 191–199
- 8 Ramasso E, Rombaut M, Zerhouni N. Joint prediction of continuous and discrete states in time-series based on belief functions. *IEEE Trans Cybern*, 2013, 43: 37–50
- 9 Gao Y L, Pan J Y, Ji G L, et al. Time series modeling method based on boosting gradient descent theory (in Chinese). *Sci Sin Tech*, 2011, 41: 929–943 [高云龙, 潘金艳, 吉国力, 等. 基于Boosting梯度下降理论的时间序列建模方法. 中国科学: 技术科学, 2011, 41: 929–943]
- 10 Ompusunggu A P, Papy J M, Vandenplas S. Kalman-filtering-based prognostics for automatic transmission clutches. *IEEE/ASME Trans Mech*, 2016, 21: 419–430
- 11 Wu J R, Xu J X, Huang X L. An indirect prediction method of remaining life based on glowworm swarm optimization and extreme learning machine for lithium battery. In: *Proceedings of the 2017 36th Chinese Control Conference*. Dalian, 2017. 7259–7264

- 12 Khelif R, Chebel-Morello B, Malinowski S, et al. Direct remaining useful life estimation based on support vector regression. *IEEE Trans Ind Electron*, 2017, 64: 2276–2285
- 13 Xi L F, Huang R Q, Li X L, et al. Remaining useful life prediction of ball bearing based on neural network (in Chinese). *J Mech Eng*, 2007, 10: 137–143 [奚立峰, 黄润青, 李兴林, 等. 基于神经网络的球轴承剩余寿命预测. *机械工程学报*, 2007, 10: 137–143]
- 14 Wang Q, Xu K, Kong X, et al. A linear mapping method for predicting accurately the RUL of rolling bearing. *Measurement*, 2021, 176: 109127
- 15 Sun B, Li Y, Wang Z, et al. An improved inverse Gaussian process with random effects and measurement errors for RUL prediction of hydraulic piston pump. *Measurement*, 2021, 173: 108604
- 16 Wang L, Zhang L, Wang X. Reliability estimation and remaining useful lifetime prediction for bearing based on proportional hazard model. *J Cent South Univ*, 2015, 22: 4625–4633
- 17 Chen D W, Wu J, Wang L P, et al. A method for predicting spindle rotation accuracy using vibration (in Chinese). *Sci Sin Tech*, 2020, 50: 819–828 [陈代伟, 吴军, 王立平, 等. 基于振动信号的电主轴回转精度预测方法. *中国科学: 技术科学*, 2020, 50: 819–828]
- 18 Guo L, Li N, Jia F, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 2017, 240: 98–109
- 19 Cheng Y, Zhu H, Wu J, et al. Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks. *IEEE Trans Ind Inf*, 2019, 15: 987–997
- 20 Qin Y, Xiang S, Chai Y, et al. Macroscopic-microscopic attention in LSTM networks based on fusion features for gear remaining life prediction. *IEEE Trans Ind Electron*, 2020, 67: 10865–10875
- 21 Xiang S, Qin Y, Zhu C, et al. LSTM networks based on attention ordered neurons for gear remaining life prediction. *ISA Trans*, 2020, 106: 343–354
- 22 Yu H M, Xu Y N, Chen X, et al. Fault diagnosis method of on-board ATC equipment based on Bi-LSTM model (in Chinese). *J Ordnance Eng*, 2021, 42: 147–151 [余辉敏, 徐永能, 陈新, 等. 基于Bi-LSTM模型的车载ATC设备故障诊断方法. *兵器装备工程学报*, 2021, 42: 147–151]
- 23 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9: 1735–1780
- 24 Saxena A, Goebel K, Simon D, et al. Damage propagation modeling for aircraft engine run-to-failure simulation. In: *Proceedings of the 2008 International Conference on Prognostics and Health Management*. Denver, 2008. 1–9
- 25 Li X, Ding Q, Sun J Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Eng Syst Saf*, 2018, 172: 1–11
- 26 Cheng Y, Hu K, Wu J, et al. A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings. *Adv Eng Inf*, 2021, 48: 101247
- 27 Zhao Z, Liang B, Wang X, et al. Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliability Eng Syst Saf*, 2017, 164: 74–83
- 28 Heimes F O. Recurrent neural networks for remaining useful life estimation. In: *Proceedings of the International Conference on Prognostics and Health Management*. Denver: IEEE, 2008. 1–6
- 29 Ramasso E. Investigating computational geometry for failure prognostics. *Int J Prognost Health Manage*, 2014, 5: 005
- 30 Babu G, Zhao P, Li X. Deep convolutional neural network based regression approach for estimation of remaining useful life. In: *Proceedings of the International Conference on Database Systems for Advanced Applications*. Taipei, 2016. 214–228
- 31 Wang W Q, Guo H, Fan Q F. Remaining useful life prediction of aero-engine based on CNN and GRU (in Chinese). In: *Proceedings of the 37th China Control Conference*. Wuhan, 2018. 9118–9123 [王文庆, 郭恒, 范启富. 基于CNN与GRU的航空发动机剩余寿命预测. 见: 第37届中国控制会议论文集. 武汉, 2018. 9118–9123]
- 32 Li X, Lim B, Zhou J, et al. Fuzzy neural network modelling for tool wear estimation in dry milling operation. In: *Proceedings of the Annual Conference of the Prognostics and Health Management Society (PHM)*, 2009. 1–11

A remaining useful life prediction method based on nested long short-term memory network for mechanical equipment

CHENG YiWei^{1,2}, ZHU HaiPing², WU Jun¹ & SHAO XinYu²

¹ School of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology, Wuhan 430074, China;

² School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Remaining useful life (RUL) prediction is an important technology to ensure the reliability, availability and safety of mechanical equipment. This paper proposes a RUL prediction method based on nested long short-term memory network (NLSTM) for mechanical equipment. This method realizes accurate prediction of RUL of mechanical equipment by fusing multi-sensor monitoring signals. Different from ordinary LSTM networks, NLSTM further deepens the storage unit and nests an LSTM neuron structure in the original LSTM storage space to achieve the deep capture of the long-term dependence in the multi-sensor time series signal. This paper uses two experimental cases of turbofan engine and machining tool to verify the prognostic performance of NLSTM. In the case of turbofan engine, it can be seen that compared with LSTM, the prediction performance of NLSTM has an improvement of 8.72% and 15.93%, respectively, on the two indicators, and the prediction results of NLSTM are also superior to other advanced methods in the literature. In the machining tool case, it can be seen that the prediction results of NLSTM are better than those of LSTM on all six machining tools.

recurrent neural network, nested long-short term memory network, remaining useful life prediction, multi-sensor monitoring data, mechanical equipment

doi: [10.1360/SST-2021-0332](https://doi.org/10.1360/SST-2021-0332)