



储能科学与技术

Energy Storage Science and Technology

ISSN 2095-4239, CN 10-1076/TK

《储能科学与技术》网络首发论文

题目: 一种基于模态分解和机器学习的锂电池寿命预测方法
作者: 肖浩逸, 何晓霞, 梁佳佳, 李春丽
DOI: 10.19799/j.cnki.2095-4239.2022.0341
收稿日期: 2022-06-20
网络首发日期: 2022-07-13
引用格式: 肖浩逸, 何晓霞, 梁佳佳, 李春丽. 一种基于模态分解和机器学习的锂电池寿命预测方法[J/OL]. 储能科学与技术.
<https://doi.org/10.19799/j.cnki.2095-4239.2022.0341>



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。



收稿日期：2022-06-20；修改稿日期：2022-06-27。

基金项目：湖北省教育厅基金（B2020013）资助

第一作者：肖浩逸（1999—），硕士研究生（在读），主要研究方向：工业大数据分析，E-mail：2218403061@qq.com；

通讯作者：何晓霞（1979—），博士研究生，副教授，主要研究方向：高维数据的统计推断，E-mail：hexiaoxia@wust.edu.cn。

一种基于模态分解和机器学习的锂电池寿命预测方法

肖浩逸，何晓霞，梁佳佳，李春丽
（武汉科技大学理学院，湖北 武汉 430065）

摘 要：锂离子电池剩余使用寿命（RUL）是电池健康管理的一个重要指标。本文采用电池容量作为健康状况的指标，使用模态分解和机器学习算法，提出了一种 CEEMDAN-RF-SED-LSTM 方法去预测锂电池 RUL。首先采用 CEEMDAN 分解电池容量数据，为了避免波动分量里的噪音对模型预测能力的影响，且又不完全抛弃波动分量里的特征信息，本文提出使用随机森林（RF）算法得到每个波动分量的重要性排序和数值，以此作为每个分量对原始数据解释能力的权重。然后将权重值和不同波动分量构建的神经网络模型得到的预测结果进行加权重构，进而得到锂离子电池的 RUL 预测。文章对比了单一模型和组合模型预测精度，加入了 RF 的组合模型预测精度让五种神经网络的表现都有进一步的提升。最后，对表现较好的两种网络——LSTM 和 GRU 引入了简单编码解码（SED）的机制，让其更好地学习到序列数据全局时间上的特征和远程的依赖关系。以 NASA 数据集作为研究对象进行该方法的性能测试。实验结果表明，CEEMDAN-RF-SED-LSTM 模型在电池 RUL 预测表现效果好，预测结果相比单一模型具有更低的误差。

关键词：锂离子电池；寿命预测；自适应白噪声完整集成经验模态分解；随机森林；神经网络

doi: 10.19799/j.cnki.2095-4239.2022.0341

中图分类号：TM912

文献标志码：A

文章编号：2095-4239（2022）XX-1-12

A Lithium Battery Life Prediction Method Based on Mode Decomposition and Machine Learning

XIAO Haoyi, HE Xiaoxia, LIANG Jiajia, LI Chunli

(College of Science, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

Abstract: Lithium-ion battery remaining useful life (RUL) is an important indicator of battery health management. In this paper, using battery capacity as an indicator of health status, using modal decomposition and machine learning algorithms, a CEEMDAN-RF-SED-LSTM method is proposed to predict lithium battery RUL. Firstly, adaptive white noise full ensemble empirical mode decomposition (CEEMDAN) is used to decompose the battery capacity data. In order to avoid the influence of the noise in the fluctuation component on the prediction ability of the model, and not completely discard the characteristic information in the fluctuation component, this paper proposes to use the random forest (RF) algorithm to obtain the important value of each fluctuation component. Sexual ranking and numerical values are used as weights for each component's ability to explain the original data. Then, the weight value and the prediction result obtained by the neural network model constructed by different fluctuation components are weighted and reconstructed, and then the RUL prediction of the lithium-ion battery is obtained. The article compares the prediction accuracy of the single model and the combined model, and the combined model prediction accuracy of RF is added to further improve the performance of the five neural networks. Finally, the Simple Encoder-Decoder (SED) mechanism is introduced for the two networks with better performance, LSTM and GRU, to better learn the global temporal features and long-range dependencies of sequence data.

The performance of the method is tested with the NASA dataset as the research object. The experimental results show that the CEEMDAN-RF-SED-LSTM model performs well in battery RUL prediction, and the prediction results have lower errors than the single model.

Key words: Li-ion battery; life prediction; adaptive white noise full ensemble empirical mode decomposition; random forest; Neural Network

在环境污染, 新冠疫情, 国际局势等一系列不确定性因素的背景下, 伴随着石油价格的上涨, 能源和储能系统变得至关重要。锂离子电池由于具有能量密度高、污染低、充电快、重量轻、性能好等优点, 已经广泛应用到一系列电子设备中, 在航空航天, 移动通信和新能源电动汽车等各个领域都是主要的能源来源^[1-2]。但是伴随着电池使用次数和充放电循环次数的增加, 电池性能会下降, 容量通常会降低, 安全性也大幅下降, 并且甚至可能会导致一些灾难^[3]。因此为了确保电子设备的安全性和可靠性, 有必要去对锂电池的健康状况(State of Health, SOH)进行有效的控制和管理^[3-4]。其中电池剩余使用寿命(Remaining Useful Life, RUL)是电池健康管理的重要指标, 剩余使用寿命的预测能提高电源性能, 对于优化电池管理系统设计和终生校准保护至关重要^[1,5]。

对于 RUL 的预测方法, 目前多数文献采取数据驱动方法。基于数据驱动方法可以直接从电池的监控测量指标中去分析与挖掘电池性能退化趋势的影响因素^[6], 该方法能回避一些电化学反应的复杂和专业性问题, 具有较强的普适性^[7]。数据驱动的方法的文献层出不穷^[8-10], 其中有使用传统统计学方法, 如指数模型, 多项式模型甚至是混合模型去拟合容量衰减的轨迹^[11]。文献[12]采用融合了经验指数和多项式回归, 并加入粒子滤波(Particle Filtering, PF)等算法融合, 还有高斯过程回归(Gaussian Process Regression, GPR)^[7,13], 自回归移动平均(Auto-Regressive Moving Average, ARMA)^[14]等等方法做出电池剩余寿命的预测。

起源于人工智能领域的机器学习技术, 拥有着极其强大的预测能力, 在电池寿命预测领域也有广泛的应用。文献[15]使用电池早期循环数据, 构建 WOA-XGBoost 的混合模型进行寿命预测, 文献[16]使用相关相量机(Relevance Vector Machine, RVM)进行预测, 还有时间规整图和支持向量机回归(Support Vector Machine Regression, SVR)的结合^[17]等等。基于深度学习的高精度预测算法被认为是最适合的预测模型^[1]。神经网络系列算法由于具有强大的特征学习和非线性趋势拟合的能力, 同样也被广泛运用于 RUL 预测。文献[18]采用长短期记忆神经网络(Long Short-Term Memory, LSTM)进行电池寿命预测, 文献[2]采用四种神经网络对比研究, 发现长短期记忆网络预测效果最佳。因此很多学者都会基于改进的 LSTM 网络方法去预测电池剩余使用寿命^[19-20]。基于 LSTM 网络改进的门控单元循环神经网络(Gated Recurrent Unit, GRU)也经常在电池寿命预测中被采用^[21]。同样有将传统方法与神经网络结合起来去预测电池寿命^[22-23], 甚至是采用不同种类的神经网络融合的方法^[24-26]。

电池在充放电循环中会出现一定的容量再生的现象, 并且受到一些客观因素的影响(温度、电化学反应、测量设备误差等), 电池容量的原始数据总是充满噪音的, 高频动态和非线性容量曲线会影响神经网络去准确地学习到电池容量的衰退特征^[27]。针对该现象, 现有的大量文献都使用经验模态分解(EMD)对电池容量数据进行分解降噪处理, 然后再采用 ARMA^[28], GPR^[29]或神经网络^[30-32]等方法进行 RUL 预测。基于 EMD 改进的自适应白噪声完整集成经验模态分解(Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Analysis, CEEMDAN)广泛应用于股票投资, 生物医学等时间序列数据的领域^[33], 采用 CEEMDAN 进行降噪得到趋势项, 然后再构建神经网络进行预测会得到更好的预测效果。

然而大多数的文献对容量曲线进行 EMD 分解之后只保留了趋势项进行预测, 而对虽然充满噪音、却同样也可能含有一部分真实特征信息的高频波动分量进行了舍弃。为了更准确的预测效果, 分量里面的信息同样应给予一定的权重。文献[28]对每个波动分量进行白噪声检验, 然后利用皮尔逊相关系数进行加权重构。文献[30]采用灰色关联度分析(Grey Relation Analysis, GRA)进行了特征筛选, 文献[34]对电池原始特征提取和变换后, 采用随机森林(Random Forest, RF)筛选出重要性高的特征。

基于上述分析, 本文采用电池容量作为健康状况的指标。首先采用 CEEMDAN 分解电池容量的序列数据进行降噪, 为了准确度量波动分量对原始数据信息的解释程度, 本文提出了使用 RF 回归去得到每个分量对原始数据解释能力的权重。然后对每个不同频率的波动分量构建不同的神经网络模型得到预测结果, 再与 RF 得到的重要性的权重数值进行加权整合重构。这样既避免了波动分量中噪音对模型预测能力的影响, 且未完全抛弃波动分量里面的特征信息。最后本文对预测表现较好的两种网络——LSTM 和

GRU 引入了一种简单编码解码(Simple Encoder-Decoder)的机制, 让其更好地学习到序列数据全局时间上的特征和远程的依赖关系。使用美国国家航天局(NASA)的锂离子电池公开数据集验证该模型的有效性。实验结果表明, CEEMDAN-RF-SED-LSTM 在电池的 RUL 预测上表现效果好, 加入了 RF 和 SED 的组合模型预测结果上具有更高的精度。

1 相关理论基础

1.1 电池健康状况定义

电池的健康状况是表示电池当前性能退化程度的重要指标^[27]。随着电池使用次数增多, 最大可用容量会不断降低。电池剩余使用寿命由以下容量比定义:

$$SOH(t) = \frac{C_t}{C_0} \times 100\%, \quad (1)$$

其中, C_0 表示初始额定容量, C_t 表示 t 时刻测量的容量。一般认为电池寿命达到原始容量的 70%~80% 时, 则认为其寿命终止。而电池 RUL 就是指在电池在使用过程中容量达到初始容量 70%~80% 之前剩余还能使用的循环充放电次数。

1.2 自适应白噪声完整集成经验模态分解

CEEMDAN 是基于 EMD 的改进的算法^[35], 常用于处理非平稳非线性的数据。它将原始数据中加入白噪声, 可以放大各个模态之间的不相关程度, 便于将原始数据的模态分解开, 具有运算速度快, 屏蔽迭代次数少, 更好的模态分解效果等优点。其步骤主要分为如下几步:

(1) 向数据里加入振幅为 ε_{k-1} 的高斯白噪声 $\varepsilon_{k-1}E_{k-1}\omega_i(t)$, ($i = 1, 2, \dots, I$), $E_k(\cdot)$ 表示经过 EMD 分解后得到的第 k 个模态分量。

(2) 将每个加入了白噪声的数据 $R_{k-1}(t) + \varepsilon_{k-1}E_{k-1}(\omega_i(t))$, ($i = 1, 2, \dots, I$), 进行 EMD 分解, 得到每个第一个模态分量 $IMFk_i$, 然后取均值, 得到本轮分解出来的模态分量 $IMFk$ 。

(3) 更新本轮的残差 $R_k(t) = R_{k-1}(t) - IMFk$, 然后确定下一轮的白噪声幅值 ε_{k-1} , 重复步骤(1)和(2), 直到最后的残差无法分解停止。

(4) 得到最终的残差为 $R_n(t) = X(t) - \sum_{k=1}^n IMFk$, 其中 $X(t)$ 为原始数据, n 为模态总数。

1.3 随机森林

随机森林最早由 Breiman 提出^[36], 其思想来自于集成学习法。随机森林采用了随机特征选择的策略。使用重采样的样本, 但是在每棵决策树的一个节点分裂时, 仅随机从 p 个特征变量里选取 m 个变量进行分裂, 每棵决策树都如此。这样就尽可能降低了决策树之间的相关性。然而在每棵决策树分裂时未使用全部特征变量的信息, 无疑会增大偏差。但是会换取方差更大幅度的减小, 从而整体上降低总误差即均方误差。

在决策树进行节点分裂的时候, 每次使用了一个变量。可以考察该变量分裂使得损失函数即残差平方和下降的幅度。对于所有变量, 可以度量随机森林的每棵决策树由该变量分裂导致的损失函数的下降幅度, 再对每个下降幅度进行每棵决策树平均, 可以作为该变量重要性的度量。

随机森林算法在筛选变量方面具有优良的性质, 不易过拟合, 准确性高, 适用于高维数据, 并且不会受到自变量之间存在着多重共线性的影响。作为非线性非参数的方法, 很适合用来作为模态分解后的波动分量的重要性度量, 能准确衡量每个分量对于原始数据的解释能力, 并以此得到每个分量的重要性度量。

1.4 神经网络

本文采用了五种神经网络的基础模型, 分别为多层感知机(Multi-Layer Perceptron, MLP), 一维卷积神经网络(1D CNN), 循环神经网络(RNN), 长短期记忆神经网络(LSTM), 门控单元循环神经网络(GRU)。每种网络的基本架构简介如下。

1.4.1 多层感知机

MLP 是指在神经网络中加入一个或者多个隐藏层去克服线性模型的限制, 能够更普遍的处理函数关系。原理公式为:

$$h = \sigma(W_1x + b_1), \quad (2)$$

$$o = W_2h + b_2, \quad (3)$$

式中, h 代表隐藏层输出, o 代表最终输出, W_1 、 W_2 、 b_1 、 b_2 为隐藏层和输出层的权重矩阵和偏置项, σ 代表非线性激活函数。

1.4.2 一维卷积神经网络

一维卷积神经网络可以用于处理序列数据问题，而且计算代价小。本文一维卷积网络如图 1 所示：

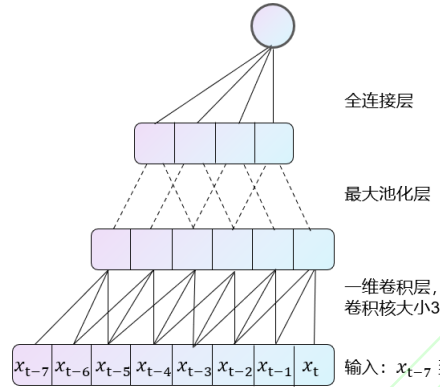


图 1 一维 CNN 结构图

Fig.1 1D CNN structure diagram

输入序列数据经过卷积核做互相关运算，得到局部一维序列子段，然后采用最大池化层提取每一小段的信息，保留主要的特征同时减少参数，最后采用全连接层得到 x_{t+1} 的估计量。

1.4.3 循环神经网络

循环神经网络(RNN)对具有序列特性的数据十分有效，它能挖掘数据中的时序信息。其原理见图 2：

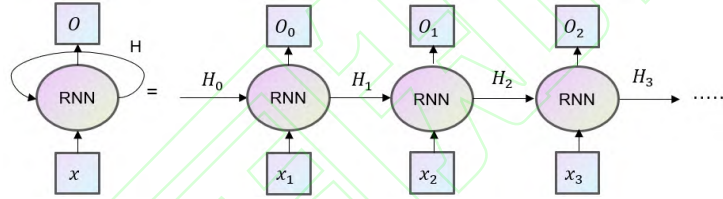


图 2 RNN 结构图

Fig.2 RNN structure diagram

可以看出每步的输出 O 不仅受到输入 x 的影响，还会受到上一时刻的隐状态 H 的影响。当前时间步隐藏变量由当前输入和前一时刻隐藏变量一起计算的得出：

$$H_t = \phi(X_t W_{xh} + H_{t-1} W_{hh} + b_h), \quad (4)$$

$$O_t = H_t W_{hq} + b_q, \quad (5)$$

其中， X_t 、 H_t 代表 t 时刻的输入和隐藏状态。 W_{xh} 、 W_{hh} 、 W_{hq} 分别为隐藏层、隐藏状态、输出层的权重矩阵， b_h 、 b_q 为隐藏层、输出层的偏置。 ϕ 为隐藏层激活函数。

1.4.4 门控循环单元神经网络

GRU 和 LSTM 改进了 RNN 梯度爆炸或是消失等缺点，并且通过“门”的机制可以控制隐藏状态。GRU 具备重置门、更新门和候选隐状态。重置门让 GRU 获取序列中的短期依赖关系，而更新门有助于学习序列中的长期依赖关系。其内部神经元结构如图 3 所示：

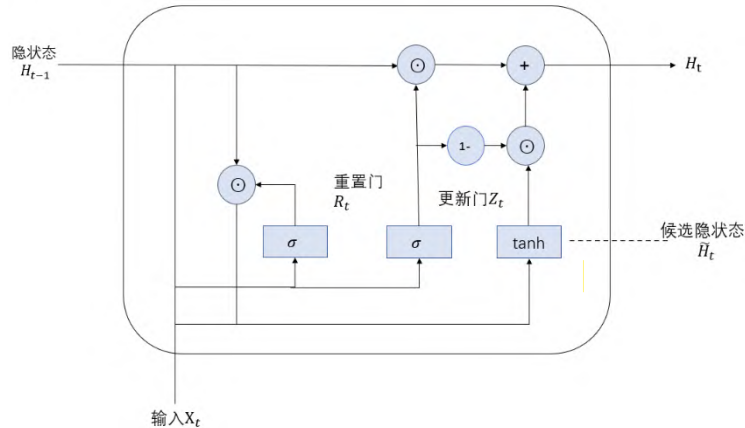


图 3 GRU 神经元内部结构图

Fig.3 Internal structure diagram of GRU neuron

其原理公式为:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r), \quad (6)$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z), \quad (7)$$

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h), \quad (8)$$

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t, \quad (9)$$

其中, R_t 代表重置门, Z_t 代表更新门, H_t 代表隐状态, \tilde{H}_t 代表候选隐状态。 W_{xr} 、 W_{xz} 、 W_{hr} 、 W_{hz} 、 W_{xh} 、 W_{hh} 分别为各个门和隐状态的权重矩阵, b_r 、 b_z 、 b_h 为其对应的偏置项。 σ 代表激活函数, \tanh 代表双曲正切激活函数。

1.4.5 长短期记忆神经网络

虽然相比 GRU, LSTM 的计算代价要大一些, 但许多文献表明 LSTM 是在序列数据预测中表现最为良好的循环神经网络^[1-2]。LSTM 同样具备学习序列数据之间的长期和短期的特征^[47]。它具备输入门、遗忘门、输出门和候选记忆元。其内部神经元结构如图 4 所示:

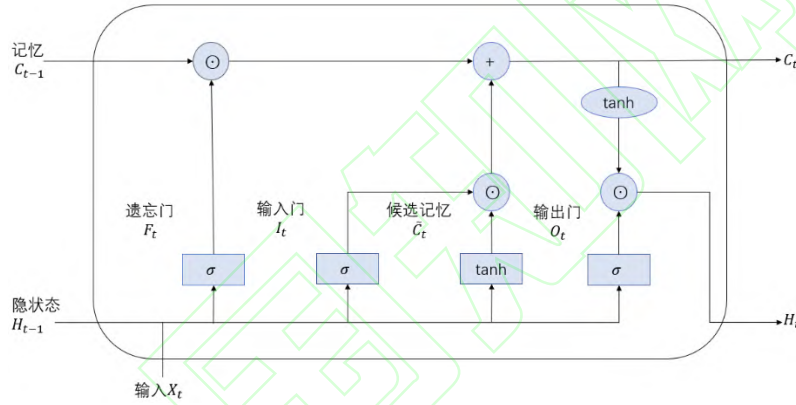


图 4 LSTM 神经元内部结构图
Fig.4 Internal structure diagram of LSTM neuron

其原理公式为:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i), \quad (10)$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f), \quad (11)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o), \quad (12)$$

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c), \quad (13)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t, \quad (14)$$

$$H_t = O_t \odot \tanh(C_t). \quad (15)$$

其中, I_t 代表输入门, F_t 代表遗忘门, O_t 代表输出门, \tilde{C}_t 代表候选记忆元, C_t 代表记忆元, H_t 代表隐状态。 W_{xi} 、 W_{hi} 、 W_{xf} 、 W_{hf} 、 W_{xo} 、 W_{ho} 、 W_{xc} 、 W_{hc} 分别为各个门和隐状态的权重矩阵, b_i 、 b_f 、 b_o 、 b_c 为其对应的偏置项。 σ 代表激活函数, \tanh 代表双曲正切激活函数。

1.4.6 简单编码解码

本文引入了一种简单编码解码(SED)的结构以便让神经网络能在预测时间步长上通过解码器的隐藏状态更好地学习序列数据全局时间上特征和远程依赖关系。编码器中的输出作为隐藏状态神经元放入解码器再次进行循环神经网络的隐藏层运算。其结构如图 5 所示:

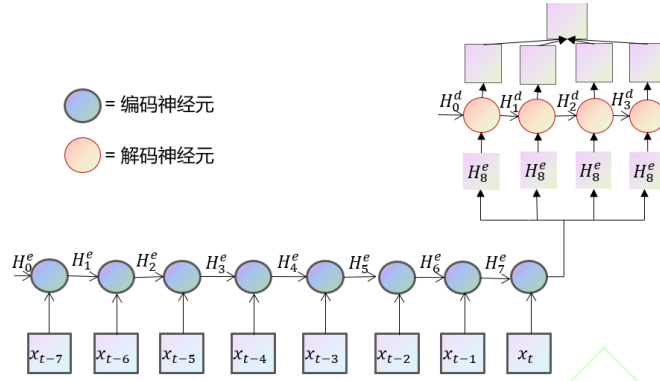


图 5 简单编码解码结构图
Fig.5 Simple encoding and decoding structure diagram

2 预测流程及评价指标

2.1 预测流程

本文提出的 RUL 预测的流程如图 6 所示，其主要步骤如下：

(1)选取一个电池的所有电容量数据进行 CEEMDAN 分解为 N 个分量 $IMFi$ ，然后将每条分量进行神经网络拟合，得到每条分量的神经网络模型。

(2)将每条分量作为特征变量，原始数据作为响应变量，进行随机森林回归，得到每个分量的变量重要性的数值，作为权重。

(3)将每条分量的神经网络模型进行预测，得到所有分量的预测值，然后利用随机森林的变量重要性权重进行加权重构，得到原始数据的估计值。最后和真实数据进行评价指标的计算，作为预测模型性能的评价准则。

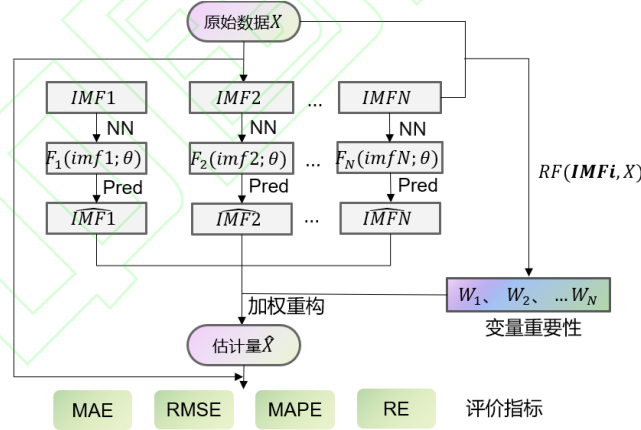


图 6 整体预测方法框架
Fig.6 Overall Forecasting Methodology Framework

2.2 模型评价指标

对于模型的预测结果采用如下四个指标进行评价，分别是平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RSME)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)和相对误差(Relative Error, RE)。其计算公式定义见式(16)-(19)：

$$MAE = \frac{1}{n} \sum_{t=1}^n \|x_t - \hat{x}_t\|, \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2}, \quad (17)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left\| \frac{x_t - \hat{x}_t}{x_t} \right\|, \quad (18)$$

$$RE = \frac{|RUL^{pred} - RUL^{true}|}{RUL^{true}}. \quad (19)$$

其中， n 为序列的长度， x_t 为整个序列的取值， RUL^{pred} 和 RUL^{true} 代表预测的剩余使用寿命和真实的剩余使用寿命。

3 RUL 预测结果和分析

3.1 数据集介绍

本文使用的是美国国家航天局(NASA)的锂离子电池公开数据集。选取第一组四个电池作为研究对象,编号为 B0005、B0006、B0007、B0018。其测试的温度为 24 度,充电过程是以 1.5A 的恒定电流进行充电,直到电池电压达到 4.2V,然后以恒定电压充电,直至充电电流降至 20mA。放电过程是以 2A 的恒定电流模式进行放电,直到电池 5、6、7 和 18 的电压降到 2.7V, 2.5V, 2.2V 和 2.5V 时停止。电池达到寿命的情况为额定容量下降到初始容量的 70%,即四块电池的额定容量从 2Ahr 到 1.4Ahr 时,可认为其寿命终止。四块电池的容量随充放电循环次数的变化见图 7:

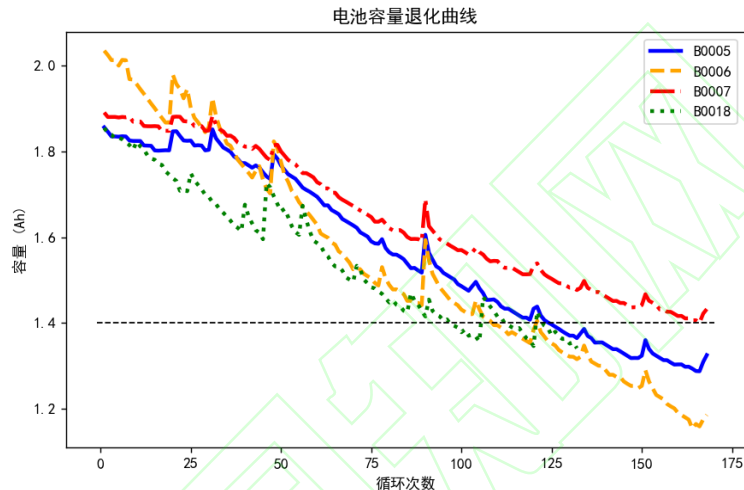


图 7 NASA 数据集电池容量退化曲线

Fig.7 NASA dataset battery capacity degradation curve

3.2 CEEMDAN 分解

使用 CEEMDAN 对四组电池的容量数据进行分解, B0005 和 B0006 电池容量分解结果如图 8 所示:

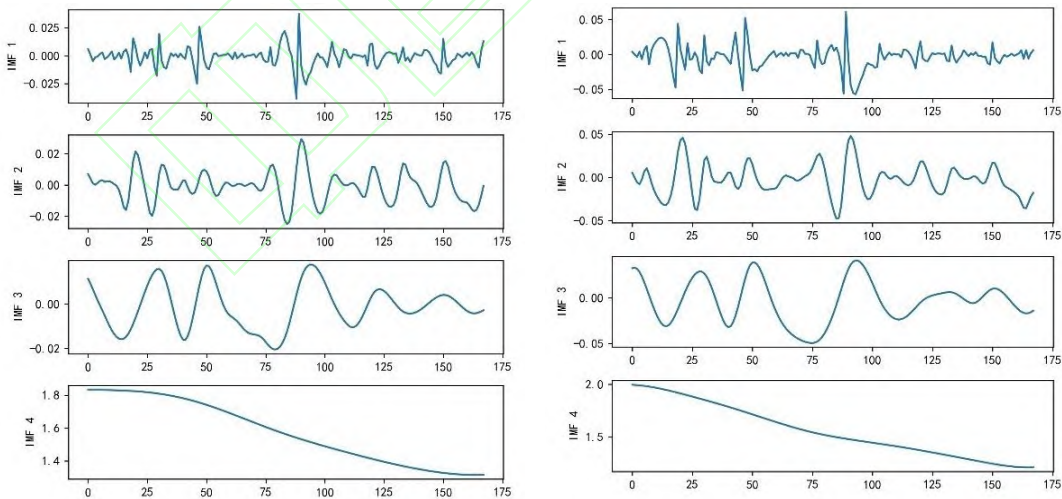


图 8 B0005 和 B0006 电池容量分解的模式分量

Fig.8 Modal components of B0005 and B0006 battery capacity decomposition

如图 8 所示,实验过程中每个电池都分解出了四条分量,其中 IMF4 为趋势项,其他 IMF1 到 IMF3 为高频分量,这些分量包含噪音,但也可能包含一部分真实信息。分量依次按照频率的高低进行排列。并且每条分量的最大值和最小值,即波动幅度都不完全相同。下面通过随机森林对分量进行权重优化。

3.3 随机森林回归

将每个电池的 IMF1 至 IMF4 作为特征变量,原始数据作为响应变量进行随机森林回归,然后进行变量重要性计算和排序,得到数值结果见表 1:

表 1 随机森林回归结果
Table 1 Random forest regression results

电池编号	拟合优度	IMF1	IMF2	IMF3
B0005	0.999284	0.023469	0.064610	0.109861
B0006	0.999052	0.072530	0.035768	0.124490
B0007	0.998983	0.027500	0.088808	0.076621
B0018	0.998613	0.025708	0.041343	0.174602

所有的电池的随机森林回归的拟合优度是 99%以上，意味着分解出来的分量对于原始数据有非常好的解释能力。IMF4 是趋势项，因此权重固定为 1，由此计算得到的每个高频波动分量的权重系数如表 1 所示。从表中可以得出每条分量对于原始数据的解释能力是不一样的，而且不一定频率越低的分量对原始数据的解释能力越高，例如 B0006 电池的 IMF1 的重要性系数比 IMF2 高，B0007 电池的 IMF2 的重要性系数比 IMF3 要高。因此对于每个分量的权重系数调整是有必要的。

3.4 神经网络预测

本文首先采用单独神经网络，CEEMDAN+神经网络，CEEMDAN+RF+神经网络三种方法在 B0005 号和 B0006 号电池上进行测试，比较五种不同网络的预测性能，以及加入了 CEEMDAN 和 RF 进行优化后的模型的性能对比。计算每个模型的预测四个评价指标以及运行时间如表 2 和表 3 所示：

表 2 B0005 预测评价指标结果
Table 2 B0005 Prediction and evaluation index results

电池编号	神经网络模型	模型方法	运行时间(s)	MAE	RMSE	MAPE	RE
B0005	LSTM	LSTM	18.77	0.058592	0.070959	0.043809	0.040650
		CEEMDAN-LSTM	70.65	0.043490	0.057525	0.026992	0.026260
		CEEMDAN-RF-LSTM	74.93	0.040182	0.053559	0.024962	0.024390
	RNN	RNN	38.07	0.280588	0.360674	0.180698	0.335772
		CEEMDAN-RNN	87.56	0.245580	0.332067	0.169977	0.390244
		CEEMDAN-RF-RNN	91.95	0.071269	0.087691	0.047033	0.195122
	GRU	GRU	17.84	0.060329	0.072215	0.044820	0.073171
		CEEMDAN-GRU	61.86	0.045586	0.060776	0.028049	0.028130
		CEEMDAN-RF-GRU	66.39	0.041444	0.055729	0.025512	0.024390
	CNN	CNN	13.09	0.068268	0.081806	0.039025	0.440650
		CEEMDAN-CNN	47.96	0.062904	0.073718	0.040348	0.304553
		CEEMDAN-RF-CNN	52.11	0.062832	0.073593	0.040437	0.280813
	MLP	MLP	13.02	0.070968	0.081720	0.047639	0.325203
		CEEMDAN-MLP	45.63	0.068850	0.080262	0.045182	0.329496
		CEEMDAN-RF-MLP	49.97	0.068734	0.080241	0.044737	0.305152

表 3 B0006 预测评价指标结果
Table 3 B0006 Prediction and evaluation index results

电池编号	神经网络模型	模型方法	运行时间(s)	MAE	RMSE	MAPE	RE
B0006	LSTM	LSTM	20.72	0.054751	0.069080	0.037688	0.068037
		CEEMDAN-LSTM	64.83	0.044735	0.058687	0.028165	0.056075
		CEEMDAN-RF-LSTM	69.37	0.039161	0.049791	0.024893	0.054112
	RNN	RNN	38.66	0.207933	0.325623	0.177686	0.439252
		CEEMDAN-RNN	87.25	0.249714	0.345865	0.180759	0.355140
		CEEMDAN-RF-RNN	91.36	0.086660	0.110239	0.061186	0.196262
	GRU	GRU	17.84	0.064288	0.082282	0.044646	0.046729

	CEEMDAN-GRU	58.64	0.042328	0.052737	0.027216	0.065421
	CEEMDAN-RF-GRU	62.91	0.041394	0.051566	0.026716	0.065421
	CNN	13.19	0.083262	0.095125	0.054467	0.233645
CNN	CEEMDAN-CNN	47.86	0.046205	0.062880	0.030333	0.029597
	CEEMDAN-RF-CNN	51.37	0.041894	0.054645	0.027848	0.028037
	MLP	13.06	0.085937	0.096994	0.064232	0.088037
MLP	CEEMDAN-MLP	39.78	0.051744	0.066709	0.034459	0.084112
	CEEMDAN-RF-MLP	44.08	0.048342	0.063234	0.032276	0.037383

从表 2 和表 3 中，首先对比五种不同神经网络在 B0005 和 B0006 的预测表现，无论是单独使用神经网络或是采用不同的组合方法，五种神经网络的性能均为 LSTM 和 GRU 最佳，其中 LSTM 在精度上略优于 GRU，但同时计算时间也略有增多。CNN 和 MLP 在预测精度表现中等，但其计算时间要显著低于其他三种网络。表现最差的网络是 RNN，无论从精度还是运算时间上来看，都不如其他四种网络。这可能是电池数据中的高频动态和非线性容量曲线影响了其性能。

再从方法组合的角度进行对比，采取了 CEEMDAN 分解之后再去做神经网络的预测基本上会比单独神经网络模型得到的结果具有更低的误差，并且模型加上 RF 进行分量的权重调整预测结果后，从 MAE、RMSE、MAPE、RE 四组误差指标上进行比较，全面优于只采用 CEEMDAN+神经网络方法的精度。此外，四组误差指标在每一个神经网络的模型上都是进一步下降，这说明采用 RF 进行分量权重调整的策略是有效的。

下面对表现较好的两种网络——LSTM 和 GRU 引入了简单编码解码(SED)的机制，让其更好地学习到序列数据全局时间上的特征和远程的依赖关系。图 9 比较所有的组合方法的在 B0005 和 B0006 上的预测性能表现。

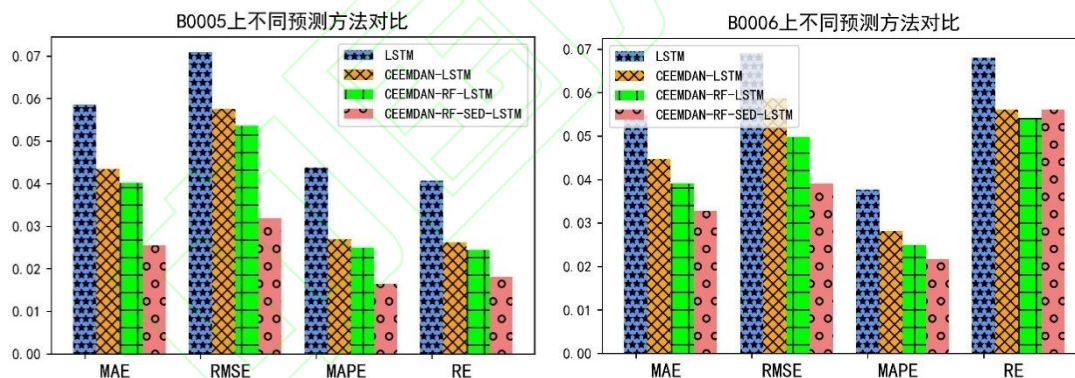


图 9 所有组合方法的评价指标对比

Fig.9 Comparison of evaluation indicators for all combination methods

由于 LSTM 网络表现最佳，图 9 中只展示了 LSTM 网络在不同方法组合中的四组评价指标，其他网络也类似图 9 中结果。可以看出在引入了简单编码解码(SED)的机制后，在电池 B0005 和 B0006 上进行训练与预测，所有误差指标都再次进一步下降，并且下降幅度较大，说明在进行简单编码解码后，神经网络模型能更好地学习到电池容量序列数据在全局时间上退化的特征。对四组电池使用组合模型得到的预测评价指标计算如表 4 所示，其 CEEMDAN-RF-SED-LSTM 拟合效果图见图 10。

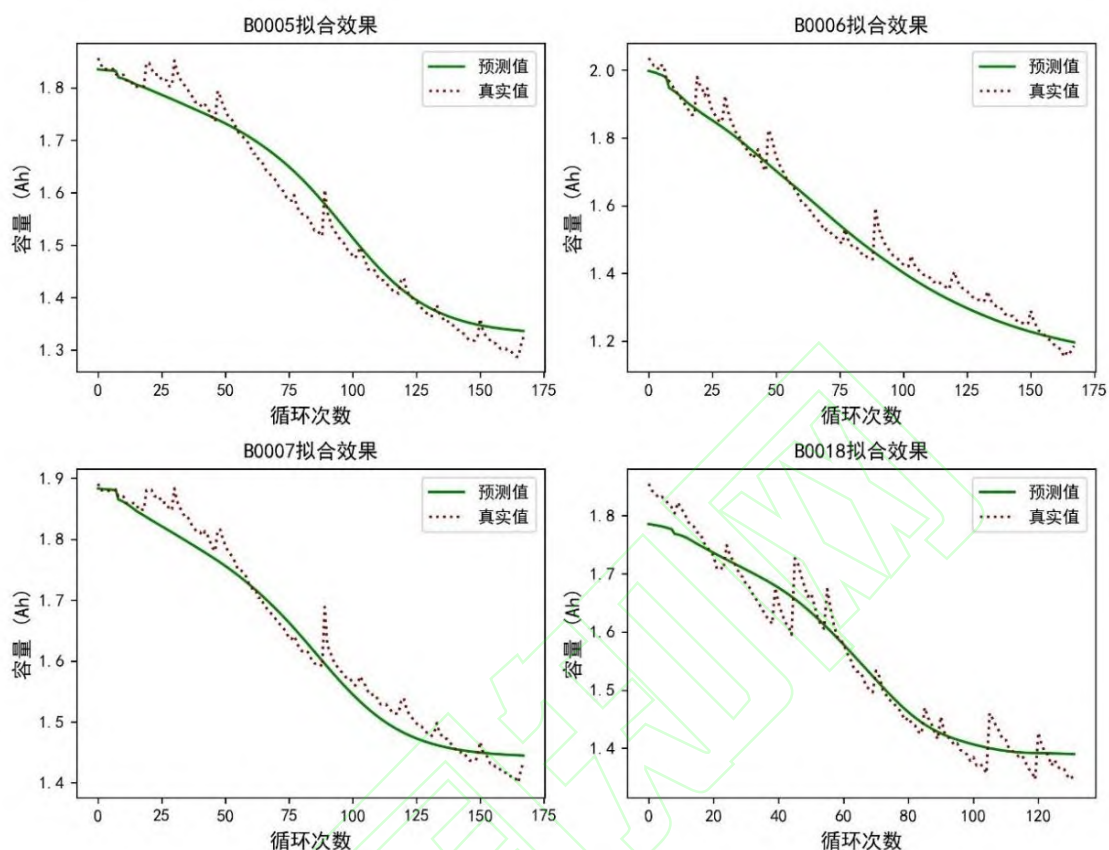


图 10 四组电池的 CEEMDAN-RF-SED-LSTM 模型拟合图

Fig.10 Fitting diagram of CEEMDAN-RF-SED-LSTM model of four groups of batteries

结合表 2、表 3 和表 4 可以得知，CEEMDAN-RF-SED-LSTM 方法相比单一的 LSTM 模型，误差指标平均下降了 40%~50%，相比 CEEMDAN-LSTM 方法，误差指标下降了 30%~40%。并且在其他神经网络上也具有类似的表现。四组电池的预测误差指标 MAE 和 RMSE 都控制在 5%以内，具有较高的预测精度。

表 4 四组电池预测评价指标结果

Table 4 Prediction and evaluation index results of four groups of batteries

电池编号	模型方法	MAE	RMSE	MAPE	RE
B0005	CEEMDAN-RF-SED-LSTM	0.025569	0.031899	0.016515	0.018130
	CEEMDAN-RF-SED-GRU	0.037491	0.044218	0.024056	0.089431
B0006	CEEMDAN-RF-SED-LSTM	0.032727	0.039157	0.021635	0.056075
	CEEMDAN-RF-SED-GRU	0.041408	0.049697	0.028029	0.093458
B0007	CEEMDAN-RF-SED-LSTM	0.021852	0.026331	0.013319	0.059524
	CEEMDAN-RF-SED-GRU	0.028275	0.036218	0.017499	0.061905
B0018	CEEMDAN-RF-SED-LSTM	0.025835	0.031901	0.016380	0.094737
	CEEMDAN-RF-SED-GRU	0.030818	0.040062	0.019301	0.081053

实验结果表明，本文提出的 CEEMDAN-RF-SED-LSTM 在电池的 RUL 预测上表现效果好，预测结果相比单一模型具有较高的精度。

4 结论

本文采用模态分解和机器学习算法，提出了一种 CEEMDAN-RF-SED-LSTM 方法框架去预测 RUL。选取电池容量作为健康因子，使用随机森林回归对自适应白噪声完整集成经验模态分解(CEEMDAN)方法

分解出来的电池容量的序列数据每条波动分量计算它们对原始数据的解释程度,得到每个分量的重要性排序和数值,然后与不同频率的波动分量的神经网络模型得到预测进行加权整合重构。进而得到锂离子电池剩余使用寿命的预测。选取 NASA 数据集验证该模型的精度,得到结论如下:

(1) 电池数据由于容量再生和一些客观因素的影响,原始数据总是充满噪音的,高频动态和非线性容量曲线会影响模型去准确地学习到电池容量的衰退特征。而采用 CEEMDAN 分解对电池数据进行降噪处理是有效的,能进一步帮助模型提高精度。

(2) 分解出来的高频分量虽然充满噪音,却同样可能含有部分数据真实特征信息。实验结果表明,本文使用随机森林回归去调整每条波动分量权重的方法是有效的,五种神经网络预测结果的误差都进一步下降。

(3) 在时间序列数据的预测的问题上,五种神经网络中 LSTM 和 GRU 两种网络表现最为良好,一维 CNN 由于其计算代价小、速度快,综合考虑同样也具有不错的效果。

(4) 对神经网络引入简单编码解码(SED)的机制,可以让其更好地学习到序列数据全局时间上的特征和远程的依赖关系。结果表明使用 SED 机制可以让 LSTM 和 GRU 神经网络的预测误差指标进一步下降。

实验结果表明,在相同的预测条件下,CEEMDAN-RF-SED-LSTM 模型在电池 RUL 预测上表现最为优良,预测误差最低。该模型为现有电池 RUL 预测研究提供了参考。

参考文献

- [1] WANG S L, JIN S Y, BAI D K, et al. A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries[J]. Energy Reports, 2021, 7:5562-5574.
- [2] ZRAIBI B, MANSOURI M, LOUKILI S E. Comparing deep learning methods to predict the remaining useful life of lithium-ion batteries[J]. Materials Today: Proceedings, 2022.
- [3] 胡振恺,雷博,李勇琦,等.储能用锂离子电池安全性测试与评估方法比较[J].储能科学与技术,2022,11(5):1650-1656.
- [4] HU Z K, LEI B, LI Y Q, et al. Comparison of safety testing and evaluation methods for lithium-ion batteries for energy storage[J]. Energy Storage Science and Technology, 2022,11(5):1650-1656.
- [5] 李建林,肖珩.锂离子电池建模现状综述[J].储能科学与技术,2022,11(2):697-703.
- [6] LI J L, XIAO Y. A review of the current state of lithium-ion battery modeling[J]. Energy Storage Science and Technology, 2022,11(2):697-703.
- [7] 李练兵,李思佳,李洁,等.基于差分电压和 Elman 神经网络的锂离子电池 RUL 预测方法[J].储能科学与技术,2021,10(6):2373-2384.
- [8] LI L B, LI S J, LI J, et al. RUL prediction method of lithium-ion battery based on differential voltage and Elman neural network[J]. Energy Storage Science and Technology, 2021,10(6):2373-2384.
- [9] 姚芳,张楠,黄凯.锂离子电池状态估算与寿命预测综述[J].电源学报, 2020, 18(3):9.
- [10] YAO Fang, ZHANG Nan, HUANG Kai. A review of lithium-ion battery status estimation and life prediction[J]. Journal of Power Sources, 2020,18(3):9.
- [11] 韩云飞,谢佳,蔡涛,等.结合高斯过程回归与特征选择的锂离子电池容量估计方法[J].储能科学与技术,2021,10(4):1432-1438.
- [12] HAN Y F, XIE J, CAI T, et al. A Li-ion Battery Capacity Estimation Method Combining Gaussian Process Regression and Feature Selection[J]. Energy Storage Science and Technology, 2021,10(4):1432-1438.
- [13] 赵显赫,耿光超,林达,等.基于数据驱动的锂离子电池健康状态评估综述[J].浙江电力, 2021, 40(7): 65-73.
- [14] ZHAO X H, GENG G C, LIN D, et al. A review of data-driven health assessment of lithium-ion batteries[J]. Zhejiang Electric Power, 2021, 40(7):65-73.
- [15] 黎冲,王成辉,王高,等.基于数据驱动的锂离子电池健康状态技术研究[J].中国电力, 2022:1-18.
- [16] LI C, WANG C H, WANG G, et al. Research on data-driven lithium-ion battery health status technology[J]. China Electric Power,2022:1-18.
- [17] 李建林,王哲,许德智,等.基于数据驱动法的锂离子电池寿命预测关键技术综述[J].电器与能效管理技术, 2021(9):10-18.
- [18] LI J L, WANG Z, XU D Z, et al. A review of key technologies for lithium-ion battery life prediction based on data-driven method[J]. Appliances and energy efficiency management technology,2021(9):10-18.
- [19] 张若可,郭永芳,余湘媛,等.基于数据驱动的锂离子电池 RUL 预测综述[J].电源学报,2022:1-15.
- [20] ZHANG R K, GUO Y F, YU X Y, et al. A review of data-driven Lithium-ion battery RUL predictions[J]. Journal of Power Sources,2022:1-15.
- [21] XING Y J, MA E W M, TSUI K L, et al. An ensemble model for predicting the remaining useful performance of lithium-ion batteries[J]. Microelectronics Reliability,2013,53(6):811-820.
- [22] WANG J W, DENG Z W, YU T, et al. State of health estimation based on modified Gaussian process regression for lithium-ion batteries[J]. Journal of Energy Storage,2022,51:104512.
- [23] 周亚鹏,郭彪,王一纯,等.基于 PF-ARIMA 的锂离子电池剩余寿命预测[J].电池工业,2022,26(1):19-22.
- [24] ZHOU Y P, GUO B, WANG Y C, et al. Remaining life prediction of lithium-ion batteries based on PF-ARIMA[J]. Battery industry,2022,26(1):19-22.
- [25] 史永胜,李锦,任嘉睿,等.基于 WOA-XGBoost 的锂离子电池剩余使用寿命预测[J/OL].储能科学与技术:1-12[2022-06-28].
- [26] SHI Y S, LI J, REN J R, et al. Prediction of Remaining Service Life of Li-ion Battery Based on WOA-XGBoost[J/OL]. Energy Storage Science and Technology,2022:1-12.
- [27] Chen Z W, Shi N, Ji Y F, et al. Lithium-ion batteries remaining useful life prediction based on BLS-RVM[J]. Energy, 2021, 234(87):121269.
- [28] 韦荣阳,毛阆,高晗,等.储能科学与技术基于 TWP-SVR 的锂离子电池健康状态估计[J/OL].储能科学与技术:1-17[2022-06-28].DOI:10.19799/j.cnki.2095-4239.2022.0184.
- [29] WEI R Y, MAO Z, GAO H, et al. Energy Storage Science and Technology State-of-Health Estimation of Li-ion Batteries Based on TWP-SVR[J/OL]. Energy Storage Science and Technology,2022:1-17.
- [30] ZHENG S, RISTOVSKI K, FARAHAT A, et al. Long short-term memory network for remaining useful life estimation[C].2017 IEEE international conference on prognostics and health management (ICPHM),2017: 88-95.
- [31] 张少凤,张清勇,杨叶森,等.基于滑动窗口和 LSTM 神经网络的锂离子电池建模方法[J].储能科学与技术,2022,11(1):228-239.
- [32] ZHANG S F, ZHANG Q Y, SU Y X, et al. Lithium-ion battery modeling method based on sliding window and LSTM neural network[J]. Energy Storage Science and Technology, 2022,11(1):228-239.

- [20] SHI Z Y, CHEHADE A. A dual-LSTM framework combining change point detection and remaining useful life prediction[J]. Reliability Engineering & System Safety, 2021:205.
- [21] XIAO B, LIU Y G, XIAO B. Accurate State-of-Charge Estimation Approach for Lithium-Ion Batteries by Gated Recurrent Unit With Ensemble Optimizer[J]. IEEE Access, 2019,7: 54192-54202.
- [22] WU, Y, LI W, WANG Y R, et al. Remaining Useful Life Prediction of Lithium-Ion Batteries Using Neural Network and Bat-Based Particle Filter[J]. IEEE Access, 2019,7:54843-54854.
- [23] 姚远,陈志聪,吴丽君,等.采用 GRU-MC 混合算法的锂离子电池 RUL 预测[J].福州大学学报(自然科学版), 2022, 50(2):169-174.
YAO Y, CHEN Z C, WU L J, et al. RUL prediction of lithium-ion batteries using GRU-MC hybrid algorithm[J]. Journal of Fuzhou University (Natural Science Edition), 2022,50(2):169-174.
- [24] 陈赐阳,陈德旺.基于 CNN-LSTM 的锂电池剩余寿命(RUL)间接预测研究[J].电源技术,2021,45(5):589-594.
CHEN C Y, CHEN D W. Indirect prediction study of residual life (RUL) of lithium batteries based on CNN-LSTM[J]. Power Sources technology, 2021,45(5):589-594.
- [25] 王英楷,张红,王星辉.基于 1DCNN-LSTM 的锂离子电池 SOH 预测[J].储能科学与技术,2022,11(1):240-245.
WANG Y K, ZHANG H, WANG X H. Lithium-ion battery SOH prediction based on DCNN-LSTM[J]. Energy Storage Science and Technology, 2022,11(1):240-245.
- [26] 梁海峰,袁凡,高亚静.基于 CNN-Bi-LSTM 网络的锂离子电池剩余使用寿命预测[J].电力自动化设备,2021,41(10):213-219.
LIANG H F, YUAN F, GAO Y J. Prediction of the remaining service life of lithium-ion batteries based on CNN-Bi-LSTM network[J]. Power automation equipment, 2021,41(10):213-219.
- [27] CHEN D, HONG W, ZHOU X. Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries[J]. IEEE Access, 2022, 10: 19621-19628.
- [28] 陈彦余,夏向阳,周文钊,等.基于 EMD-ARMA 的锂离子电池剩余寿命预测[J].电力学报,2021,36(1):43-50+59.
CHEN Y Y, XIA X Y, ZHOU W Z, et al. Prediction of residual life of lithium-ion batteries based on EMD-ARMA[J]. Journal of Electric Power, 2021,36(1):43-50+59.
- [29] 李洪,张营,邹博雨,等.基于 EMD-GPR 的锂离子电池剩余寿命预测方法研究[J].农业装备与车辆工程,2021,59(11):60-63.
LI Y, ZHANG Y, ZHOU B Y, et al. Research on residual life prediction method of lithium-ion battery based on EMD-GPR[J]. Agricultural Equipment and Vehicle Engineering, 2021,59(11):60-63.
- [30] JAMES C, CHEN T L, LIU W J, et al. Combining empirical mode decomposition and deep recurrent neural networks for predictive maintenance of lithium-ion battery[J]. Advanced Engineering Informatics, 2021,50:101405.
- [31] CHENG G, WANG X Z, HE Y R. Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network[J]. Energy, 2021,232:121022.
- [32] 赵沁峰,蔡艳平,王新军.锂电池在不同放电区间下的剩余寿命预测[J/OL].中国测试,2022:1-10.
<http://kns.cnki.net/kcms/detail/51.1714.TB.20211201.1038.006.html>.
ZHAO Q F, CAI Y P, WANG X J. Prediction of the remaining life of lithium batteries in different discharge intervals[J/OL]. Tested in China, 2022:1-10.
<http://kns.cnki.net/kcms/detail/51.1714.TB.20211201.1038.006.html>.
- [33] 黄凯,丁恒,郭永芳,田海建.基于数据预处理和长短期记忆神经网络的锂离子电池寿命预测[J/OL].电工技术学报,2022:1-14.
DOI:10.19595/j.cnki.1000-6753.tces.210860.
HUANG K, DING H, TIAN H J. Lithium-ion battery life prediction based on data preprocessing and long-term short-term memory neural network[J/OL]. Transactions of China Electrotechnical Society, 2022:1-14.DOI:10.19595/j.cnki.1000-6753.tces.210860.
- [34] YU Z, ZHEN P, YONG G, et al. Prognostics of battery cycle life in the early-cycle stage based on hybrid model[J]. Energy, 2021,221:119901.
- [35] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]//2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2011: 4144-4147.
- [36] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.