

DOI: 10.13382/j.jemi.B2104134

# 基于 XGBoost-RFECV 算法和 LSTM 神经网络的 PEMFC 剩余寿命预测\*

常家康<sup>1</sup> 吕宁<sup>2</sup> 詹跃东<sup>1</sup>

(1. 昆明理工大学信息工程与自动化学院 昆明 650500; 2. 昆明理工大学计算中心 昆明 650500)

**摘要:**针对质子交换膜燃料电池(PEMFC)寿命预测方法中 PEMFC 特征对其寿命的影响程度未知和模型预测精度低的问题,提出一种基于 XGBoost-RFECV 算法和长短期记忆(LSTM)神经网络的 PEMFC 剩余寿命预测方法。首先通过等间隔采样和 SG 卷积平滑法对 PEMFC 原始数据进行重构和平滑处理,有效提取 PEMFC 退化趋势。然后利用 XGBoost-RFECV 算法计算 PEMFC 不同特征的重要度,并选择平均交叉验证均方误差最小的 10 个 PEMFC 特征组成最优特征子集。最后将最优特征子集输入构建的双层 LSTM 神经网络实现 PEMFC 的剩余寿命预测。实验结果表明,该方法的平均绝对误差和均方根误差分别为 0.001 9 和 0.002 5,决定系数  $R^2$  为 0.974,与 XGBoost-RNN、XGBoost-LSTM 和 XGBoost-RFECV-RNN 方法相比预测精度更高,能够有效地预测 PEMFC 剩余寿命。

**关键词:** XGBoost-RFECV 算法; LSTM 神经网络; PEMFC; 剩余寿命

**中图分类号:** TM911.48 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Prediction of PEMFC remaining life based on XGBoost-RFECV algorithm and LSTM neural network

Chang Jiakang<sup>1</sup> Lyu Ning<sup>2</sup> Zhan Yuedong<sup>1</sup>

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;

2. Computer Center, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** Aiming at the problem that the influence of PEMFC characteristics on the life prediction method of the proton exchange membrane fuel cell (PEMFC) is unknown and the low prediction accuracy of the model, a PEMFC remaining life prediction method based on XGBoost-RFECV algorithm and LSTM neural network is proposed. First of all, the PEMFC original data is reconstructed and smoothed by equal interval sampling and SG convolution smoothing method, which effectively retains the original data degradation trend. Then the XGBoost-RFECV algorithm is used to calculate the importance of different PEMFC features, and the 10 PEMFC features with the smallest mean square error of average cross-validation are selected to form the optimal feature subset. Finally, the optimal feature subset is input into the constructed two-layer LSTM neural network to realize the remaining life prediction of PEMFC. The experimental results show that the average absolute error and root mean square error of the method are 0.001 9 and 0.002 5, respectively, and the coefficient of determination  $R^2$  is 0.974. Compared with the XGBoost-RNN, XGBoost-LSTM and XGBoost-RFECV-RNN model, the prediction accuracy is higher and it can effectively predict the remaining life of PEMFC.

**Keywords:** XGBoost-RFECV algorithm; LSTM neural network; proton exchange membrane fuel cell; remaining useful life

## 0 引言

燃料电池作为一种具有高能量密度、高能量转化率且无污染的新型发电设备拥有广阔的应用空间<sup>[1]</sup>。其中

质子交换膜燃料电池( proton exchange membrane fuel cell, PEMFC)与其他几种类型的燃料电池相比,具有电流大、启动迅速、模块化安装和噪声低等优点,被视为氢电动汽车、便携式电源以及中小型发电系统的最佳替代电源<sup>[2-4]</sup>。然而,当前 PEMFC 仍存在维护成本高和寿命

短等问题,这成为了其大规模商业化的瓶颈<sup>[5-6]</sup>。因此,深入开展 PEMFC 的剩余寿命(remaining useful life, RUL)预测研究对于降低维护成本和寿命结束前对其进行设备维护具有重大意义。

由于 PEMFC 是一种多输入多输出、变量间高度耦合的非线性系统,且内部退化机理尚无法完全掌握,因此采用模型方法很难建立精确的退化模型,用以描述燃料电池系统的退化机理和过程<sup>[7-8]</sup>。研究发现,采用数据驱动寿命预测方法,通过测量的历史数据来分析系统的退化过程,对于受许多因素影响的模型退化机理分析和模型推理分析困难时非常适用<sup>[9-10]</sup>。目前国内外有许多学者开展了相关方法研究。文献[11]采用回声状态网络(echo state network, ESN)模型将 PEMFC 的 RUL 预测误差降到 5% 以下。文献[12]利用多输入多输出回声状态网络(multiple inputs and multiple outputs echo state network, MIMO-ESN)进一步提高了寿命预测精度。文献[13]利用粒子群算法优化灰色神经网络(grey neural network, GNN)的权值和阈值从而提高模型的适应能力,并对不同运行条件下的 PEMFC 进行降解寿命预测。文献[14]采用核超限学习机(kernel extreme learning machine, KELM)方法提高了 PEMFC 剩余寿命预测的速度。文献[15]对 PEMFC 原始数据进行归一化后输入卷积神经网络(convolutional neural networks, CNN)预测 PEMFC 的堆栈电压,在原始数据含有大量噪声的情况下,预测电压的趋势和大小仍能够与实际电压基本一致。但是上述研究并未说明 PEMFC 具体特征对于其寿命的影响程度,并且在 PEMFC 剩余寿命预测精度方面仍有提升的空间。

针对以上分析,本文提出基于 XGBoost-RFECV 算法和长短期记忆(LSTM)神经网络的 PEMFC 剩余寿命预测方法。通过等间隔采样和 SG 卷积平滑法对 PEMFC 原始数据进行重构和平滑降噪处理,能够保留原始数据的退化趋势和降低计算复杂度。之后 XGBoost-RFECV 算法分析 PEMFC 特征重要度并给出最优特征子集。最后将最优特征子集输入构建好的 LSTM 模型预测 PEMFC 堆栈电压,并以此评估 PEMFC 剩余寿命。

## 1 相关理论

### 1.1 XGBoost 模型

XGBoost 是一种对梯度提升决策树(gradient boosting decision tree, GBDT)进行优化,通过不断生成新树以拟合前一颗树的残差来提高分类或预测精度的模型<sup>[16]</sup>。

对于训练样本  $D = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbf{R}^m, y_i \in \mathbf{R}, i = 1, 2, \dots, n\}$ , 一颗 CART 回归树的预测函数为:

$$\hat{y}_i = f(\mathbf{x}_i) \quad (1)$$

式中:  $\mathbf{x}_i$  为第  $i$  组样本组成的向量;  $\hat{y}_i$  为向量  $\mathbf{x}_i$  的预测值。

则 XGBoost 的目标函数表示为:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^t \Omega(f_i) \quad (2)$$

其中,

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3)$$

式中:  $t$  为回归树的数量;  $l$  为可微凸损失函数; 用来衡量预测值  $\hat{y}_i$  与真实值  $y_i$  之间的误差;  $\Omega(f_i)$  控制模型的复杂程度;  $\gamma$  为惩罚系数;  $\lambda$  为正则项系数;  $T$  为叶子节点的数目;  $w_j^2$  为叶子节点权值的平方。通过贪心算法求解目标函数并去掉常数项之后,最终目标函数表示为:

$$Obj = \sum_{j=1}^T \left[ G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \quad (4)$$

式中:  $G_j$  和  $H_j$  为第  $j$  个叶子节点中样本的一阶和二阶导数。将式(4)两边同时对  $w_j$  求导,取其最优解可得:

$$w_j^* = - \frac{G_j}{H_j + \lambda} \quad (5)$$

把  $w_j^*$  代入式(4)得:

$$Obj = - \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (6)$$

式(6)为 XGBoost 算法最小化目标函数。为使目标函数  $Obj$  最小化,采用贪婪算法,只考虑当前节点最优对当前节点进行分裂。一个叶子节点的分裂前后的信息增益定义为:

$$Gain = \frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (7)$$

一个叶子节点分裂时,计算该叶子节点所有特征相对应的  $Gain$ , 选择  $Gain$  最大特征进行分裂。所有叶子节点都分裂结束后, XGBoost 模型构建完成。特征  $p$  在单颗回归树的重要度计算公式为:

$$\hat{J}_p^2(T) = \sum_{j=1}^{L-1} \hat{I}_j^2(v_j = p) \quad (8)$$

式中:  $L$  为回归树叶子节点数目;  $v_j$  为与节点  $j$  相关的特征;  $\hat{I}_j^2$  为节点  $j$  分裂之后平方损失的减少值;  $I$  为 sign 函数, 当  $v_j = p$  时值为 1, 否则为 0。特征  $p$  在 XGBoost 下的整体重要度通过特征  $p$  在所有回归树重要度的平均值来衡量, 公式如下:

$$J_p = \frac{1}{t} \sum_{m=1}^t \hat{J}_p^2(T_m) \quad (9)$$

### 1.2 XGBoost-RFECV 算法

递归特征消除(recursive feature elimination, RFE)是一种搜索特征空间中最优特征子集的贪心算法。该方法

通过反复构建模型,逐次删除特征空间中模型评估重要度最低的特征以更新特征空间,直至得到所需数量的特征。交叉验证(cross-validation, CV)用于评估训练好的模型在新数据上的表现,可以在一定程度上减小过拟合,还可以从有限的数据中获取尽可能多的有效信息<sup>[17]</sup>。

XGBoost-RFECV 算法的伪代码表示为:

输入:样本集  $D = \{x_1, \dots, x_m, y\}$ ;

输出:最优特征子集  $F = \{x_1, \dots, x_i \mid i \leq m\}$ ;

Begin:

将  $D$  划分为  $k$  份样本子集  $D_1 \sim D_k$ ;

For  $j = 1$  to  $k$  do

  设定验证集为  $D_j$ ;

  训练集为  $S = \{D_i \mid i = 1, 2, \dots, k \text{ 且 } i \neq j\}$ ;

  For  $i = 1$  to  $m$  do

    训练集  $S$  输入 XGBoost 训练模型并同时计算特征  $x$  的重要度  $J$ ;

    验证集  $D_j$  输入训练好的 XGBoost 模型计算均方误差 (MSE);

    删除训练集  $S$  和验证集  $D_j$  中最小特征重要度  $J_{\min}$  对应的特征  $x_{\min}$ ;

  End For  $i$

  保存依次删除最小重要度特征得到的  $m$  个特征子集对应 MSE

  End For  $j$

计算  $k$  折交叉验证中  $m$  个特征子集对应平均交叉验证均方误差  $MSE_{average}$ ;

在  $D$  中选择最小  $MSE_{average}$  对应特征子集为最优特征子集  $F$ ;

End Begin

### 1.3 LSTM 神经网络

RNN 因其单元结构的局限性,在处理长时间序列时很难将较早时间步的有效信息传递到后面的时间步。LSTM 神经网络能够很好的解决这个问题<sup>[18]</sup>。

LSTM 神经网络在 RNN 单元结构中添加了一个状态单元  $c$ , 并采用遗忘门  $f_t$ 、输入门  $i_t$ 、输出门  $o_t$  对状态单元  $c$  进行控制, LSTM 单元内部结构如图 1 所示。LSTM 单元在当前  $t$  时间步的 3 个输入分别为当前时间步 LSTM 的输入  $x_t$ 、上一时间步 LSTM 的输出  $h_{t-1}$  和上一时间步单元状态  $c_{t-1}$ 。LSTM 单元在当前  $t$  时间步的 2 个输出为输出  $h_t$  和单元状态  $c_t$ 。遗忘门的功能为丢弃上一时间步状态单元  $c_{t-1}$  中的部分信息。输入门的功能为将当前时间步的有效信息加入到上一时间步状态单元  $c_{t-1}$  中组成当前时间步状态单元  $c_t$ 。输出门的功能为计算向下一时间步传递的信息状态单元  $c_t$  和输出  $h_t$ 。

式 (10)~(15) 为  $t$  时刻 LSTM 层中内部单元状态的具体更新过程,即:

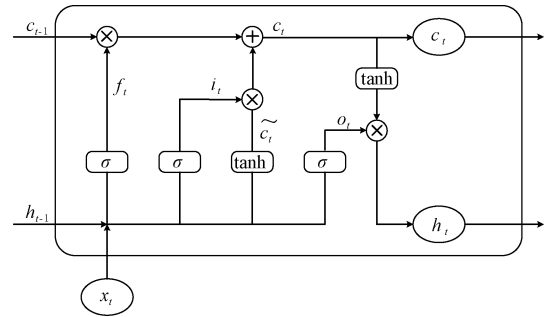


图 1 LSTM 单元内部结构

Fig. 1 Internal structure of LSTM unit

$$f_t = \sigma(W_{f_x}x_t + W_{f_h}h_{t-1} + b_f) \quad (10)$$

$$i_t = \sigma(W_{i_x}x_t + W_{i_h}h_{t-1} + b_i) \quad (11)$$

$$\tilde{c}_t = \tanh(W_{c_x}x_t + W_{c_h}h_{t-1} + b_c) \quad (12)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (13)$$

$$o_t = \sigma(W_{o_x}x_t + W_{o_h}h_{t-1} + b_o) \quad (14)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (15)$$

具体更新过程可分为如下 4 个步骤。

1) LSTM 首先应该判断出从上一时间步状态单元  $c_{t-1}$  中丢弃哪些无用信息。LSTM 将当前时间步输入  $x_t$  和上一时间步输出  $h_{t-1}$  传递给 sigmoid 函数得到当前时间步遗忘门的激活值  $f_t$ 。 $f_t$  的范围为 0~1。0 和 1 分别表示完全丢弃和保留上一时间步状态单元  $c_{t-1}$  信息,如式 (10)。

2) 提取当前时间步能够加入到状态单元  $c_t$  的有效信息。将当前时间步输入  $x_t$  和上一时间步输出  $h_{t-1}$  同步传递入 sigmoid 函数和 tanh 函数分别得到输入门的激活值  $i_t$  和当前时间步能够加入到状态单元的候选值向量  $\tilde{c}_t$ , 如式 (11) 和 (12)。

3) 计算当前时间步的状态单元  $c_t$ 。步骤 1) 保留上一时间步状态单元  $c_{t-1}$  有效信息;步骤 2) 利用输入门的激活值  $i_t$  和状态单元的候选值向量  $\tilde{c}_t$  得到当前时间步有效信息。使用两个步骤结果来对状态进行更新获得新的状态单元  $c_t$ , 如式 (13)。

4) 计算当前时间步的输出  $h_t$ 。输出  $h_t$  包含了先前输入的信息。首先,将上一时间步输出  $h_{t-1}$  和当前时间步输入  $x_t$  传递到 sigmoid 函数中,得到输出门的激活值  $o_t$ 。然后将新得到的状态单元  $c_t$  传递给 tanh 函数。最后将 tanh 函数的输出与输出门的激活值  $o_t$  相乘,以确定输出  $h_t$  应携带的信息,如式 (14) 和 (15)。

### 1.4 PEMFC 剩余寿命预测模型框架

PEMFC 剩余寿命预测方法的框架由数据处理、特征选择和寿命预测 3 个模块组成。数据处理模块包括等间

隔采样、SG 卷积平滑法平滑滤波和归一化 3 部分。特征选择模块是利用 XGBoost-RFECV 剔除低重要性特征。寿命预测模块是将前两个模块处理完成的数据输入搭建 LSTM 神经网络模型进行训练,以 PEMFC 堆栈电压作为预测模型的输出,从而进一步计算得到 PEMFC 剩余寿命预测结果。

PEMFC 剩余寿命预测方法的流程如图 2 所示。

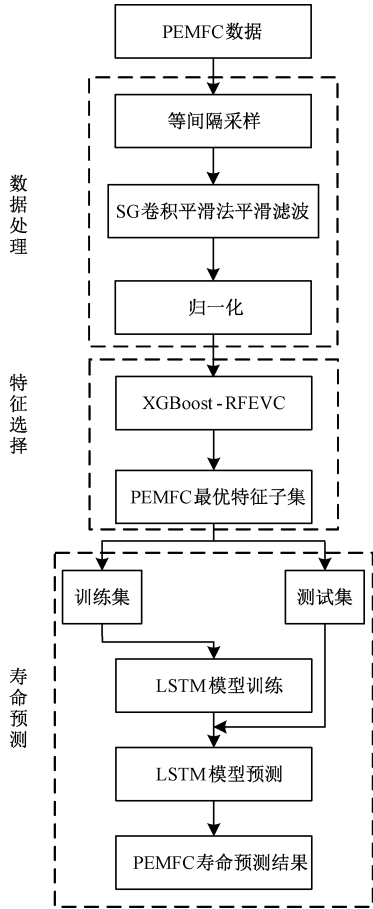


图 2 PEMFC 预测模型框架

Fig. 2 PEMFC prediction model framework

### 1.5 评估指标

为了评价本文提出方法对 PEMFC 剩余寿命预测的能力,将 MSE 作为评估 XGBoost 的预测性能指标,并选择平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean square error, RMSE) 和决定系数  $R^2$  3 种评价指标从不同方面评价 XGBoost-RFECV-LSTM 的预测性能。

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (19)$$

式中:  $y_i$ 、 $\hat{y}_i$  和  $\bar{y}_i$  分别代表 PEMFC 堆栈电压的实际值、预测值和平均值;  $n$  为测试集数据组数,  $t$  为预测点编号。

## 2 实验与分析

### 2.1 实验设备和数据介绍

实验设备为 IEEE PHM 2014 Data Challenge 提供的在标准工况下运行功率为 1 kW 的 PEMFC 电堆<sup>[19]</sup>。PEMFC 电堆由 5 片单电池组成,每片单电池的活化面积为 100 cm<sup>2</sup>,最大电流密度和标称电流密度分别为 1 和 0.70 A/cm<sup>2</sup>。PEMFC 电堆运行时,通过放置在气体分布系统中的两个独立加湿器实现空气和氢气送入电堆时的气体加湿。开启空气加湿器可得到需要的空气湿度。氢气加湿器在室温下可保证氢气的加湿程度。对主动式电子负载和冷却水系统进行操控以控制电堆内部的电流和温度。反应物的供给速率通过压力阀和流量阀调节。PEMFC 电堆原理如图 3 所示。

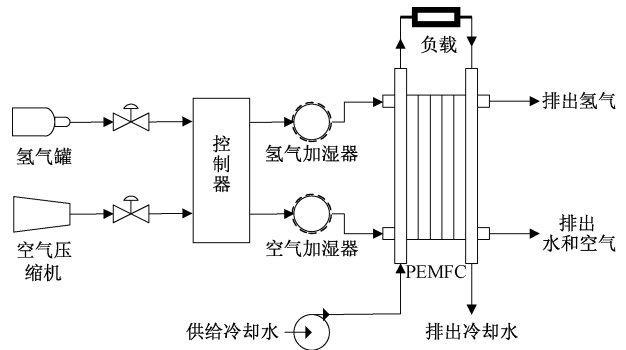


图 3 PEMFC 电堆原理

Fig. 3 Schematic diagram of PEMFC stack

对 PEMFC 电堆在下列时刻 (0、48、185、348、515、658、823、991 h) 进行一次极化曲线和电化学阻抗的测量,并利用传感器全程监测 PEMFC 电堆参数变化。极化曲线和电化学阻抗只能用于 PEMFC 的状态分析,不能用于 PEMFC 连续时间寿命预测。故本文实验数据为 137 000 组传感器时间序列监测值,记录了 PEMFC 电堆从健康到失效的 1 152 h 的全部运行状态。每一组的数据样本  $T_i$  由一个时间  $t$  和 23 个传感器监测值组成。传感器监测的 PEMFC 电堆具体参数如表 1 所示。



表 1 传感器监测参数

Table 1 Sensor monitoring parameters

参数	PEMFC 参数	符号
基本参数	U1 单电池电压/V	V1
	U2 单电池电压/V	V2
	U3 单电池电压/V	V3
	U4 单电池电压/V	V4
	U5 单电池电压/V	V5
	堆栈总电压/V	U <sub>tot</sub>
	电流(A)	I
	氢气入口温度/°C	T <sub>inH2</sub>
	氢气出口温度/°C	T <sub>outH2</sub>
	空气入口温度/°C	T <sub>inAIR</sub>
运行参数	空气出口温度/°C	T <sub>outAIR</sub>
	冷却水入口温度/°C	T <sub>inWAT</sub>
	冷却水出口温度/°C	T <sub>outWAT</sub>
	空气入口压力/mbara	P <sub>inAIR</sub>
	空气出口压力/mbara	P <sub>outAIR</sub>
	氢气出口压力/mbara	P <sub>outH2</sub>
	氢气入口压力/mbara	P <sub>inH2</sub>
	氢气入口流速/(L/mn)	D <sub>inH2</sub>
	氢气出口流速/(L/mn)	D <sub>outH2</sub>
	空气入口流速/(L/mn)	D <sub>inAIR</sub>
	空气出口流速/(L/mn)	D <sub>outAIR</sub>
	冷却水流速/(L/mn)	DWAT
空气入口湿度/%	HrAIRFC	

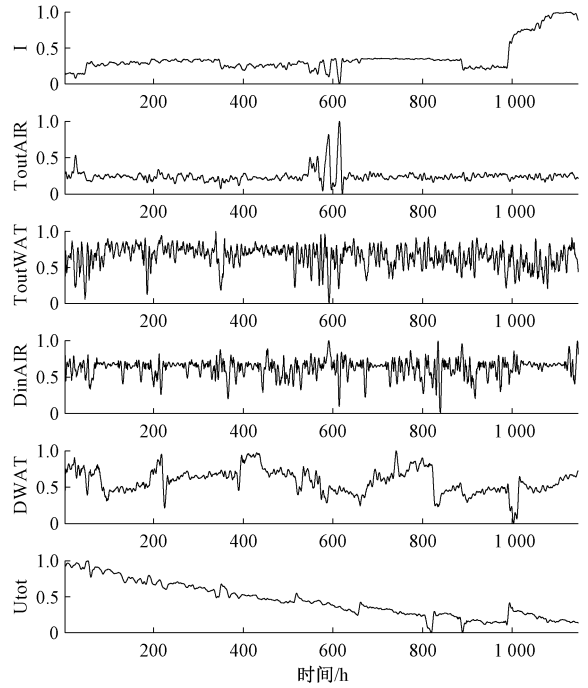


图 4 PEMFC 部分运行参数特征和堆栈电压波形  
Fig. 4 Part running parameters characteristics and stack voltage waveform of PEMFC

2.2 数据处理

PEMFC 电堆的降解过程非常缓慢,通过传感器监测的原始老化数据在较短时间内变化很小,并且原始数据中具有大量噪声和尖峰,为使预测算法更好地捕捉信号的趋势变化需要对数据进行处理。本文首先对 PEMFC 原始数据采取每 50 组提取一组数据的等间隔采样,共提取 2 740 组具有代表性的数据样本。之后使用 SG 卷积平滑法对重构数据进行平滑处理,其中移动平滑窗口宽度  $n = 51$ ,多项式次数  $k = 9$ 。最后将平滑数据归一化到  $[0, 1]$ 使其具有相同的量纲,以减少变量之间因量纲差异较大对模型产生的影响。PEMFC 部分参数特征和 PEMFC 堆栈电压经数据处理后的波形如图 4 所示。

2.3 特征选择

本文基于 Python 机器学习库 Sklearn 中的 XGBRegressor 和 RFECV 模块进行 PEMFC 的特征重要度计算以及特征选择。XGBRegressor 中参数均设定为默认值。递归特征消除从排序后的训练样本中系统的删除重要度最低的特征更新特征子集,之后对 XGBoost 模型进行递归重新训练和评估。递归消除过程中的第  $i(i = 1, 2, \dots, 22)$ 个特征子集由重要度前  $i$  的 PEMFC 特征组成,且每个特征子集都训练一个相应的 XGBoost 模型。通过 15 折交叉验证法对每个特征子集相对应的 XGBoost 预测模

型重复运行 15 次,取 15 次预测结果均方误差的平均值评价特征子集的优劣。

PEMFC 最优特征子集的平均交叉验证均方误差变化如图 5 所示。平均交叉验证均方误差越小代表特征子集对应模型的预测准确率越高。当特征子集规模小于 10 时,随着所选特征数目的增加平均交叉验证均方误差在不断变化中逐渐减小,预测准确率总体呈上升趋势。当特征数为 10 时,平均交叉验证均方误差达到最小,预测准确率最高。故本文选择平均交叉验证均方误差最小的 10 个特征组成最优特征子集。

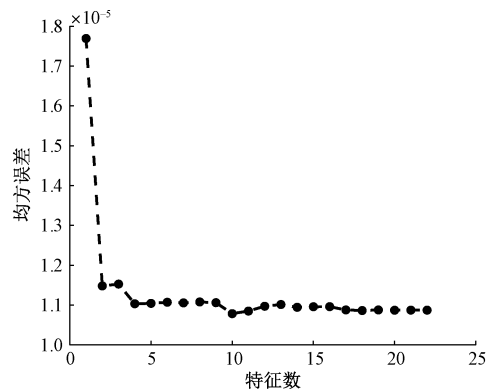


图 5 不同 PEMFC 特征子集的平均交叉验证均方误差  
Fig. 5 Average cross-validation scores of different PEMFC feature subsets

PEMFC 最优特征子集的平均交叉验证均方差变化如图 5 所示。平均交叉验证均方差越小代表特征子集对应模型的预测准确率越高。当特征子集规模小于 10 时,随着所选特征数目的增加平均交叉验证均方差在不断变化中逐渐减小,预测准确率总体呈上升趋势。当特征数为 10 时,平均交叉验证均方差达到最小,预测准确率最高。故本文选择平均交叉验证均方差最小的 10 个特征组成最优特征子集。

XGBoost 特征重要度和 XGBoost-RFECV 最优特征子

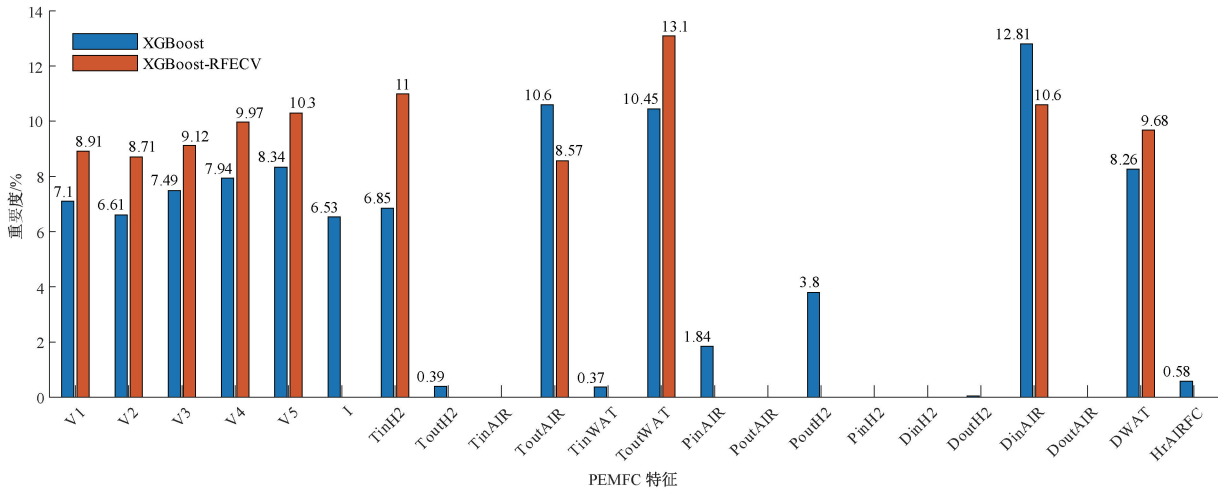


图 6 XGBoost 和 XGBoost-RFECV 特征重要度

Fig. 6 Feature importance of XGBoost and XGBoost-RFECV

### 2.4 剩余寿命预测

本文构建双层 LSTM 神经网络作为 PEMFC 剩余寿命预测模型。将处理完成的 2 740 组 PEMFC 数据的前 1 310 组数据 [0 h, 550 h] 设定为训练集,并设置后 1 430 组数据 [551 h, 1 152 h] 为测试集。前 1310 组训练集数据输入 LSTM 神经网络模型训练模型参数。然后将后 1 430 组测试集数据送入训练完毕的 LSTM 神经网络模型进行预测,模型输出结果为 PEMFC 堆栈电压。LSTM 模型参数如表 2 所示。当 PEMFC 堆栈电压衰退至初始堆栈电压的 96% 时,确定 PEMFC 电堆达到寿命,定义其失效<sup>[20]</sup>。PEMFC 初始堆栈电压为 3.349 3 V,则 PEMFC 的失效电压为 3.215 3 V。时间  $t=808.9$  h 的 PEMFC 实际堆栈电压为 3.218 V,与实际失效电压最为接近。所以设定 PEMFC 在 808.9 h 时失效。由于 LSTM 神经网络模型在 [0 h, 550 h] 为训练阶段,所以 PEMFC 的实际 RUL 为 258.9 h。

基于 XGBoost-RFECV-LSTM 方法的预测结果如图 7 所示。当时间  $t=809.35$  h 时,PEMFC 堆栈电压预测值为 3.217 8 V,与实际失效电压 3.215 3 V 最接近。此时,PEMFC 预测 RUL 为 259.35 h。

集中特征的重要度如图 6 所示。基本参数重要度较为均衡,对 PEMFC 堆栈电压有直接影响。运行参数中  $T_{inH2}$ 、 $T_{outAIR}$ 、 $T_{outWAT}$  和  $DWAT$  的重要度表明 PEMFC 堆栈电压对温度的变化较敏感。运行参数中有关压力的特征重要度均较小。PEMFC 电堆在标准工况下易控制压力在较小范围内变化,而受 PEMFC 内部电化学反应和外部条件的综合影响温度控制较难。因此标准工况下 PEMFC 温度变化对其使用寿命较为重要。

表 2 LSTM 模型参数

Table 2 LSTM model parameters

参数名	设定值
第 1 层 LSTM 神经元	90
第 2 层 LSTM 神经元	63
优化器	Adam
学习率	0.001
损失函数	MAE
Dropout	0.1
时间步	1
输入特征维度	10
批大小	80
迭代次数	70

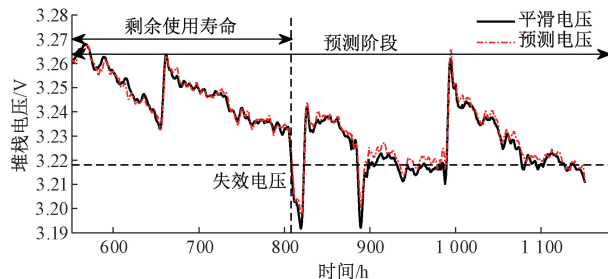


图 7 基于 XGBoost-RFECV-LSTM 的 PEMFC 剩余寿命预测结果

Fig. 7 Prognostic results of remaining life of PEMFC based on XGBoost-RFECV-LSTM

为了进一步验证 XGBoost-RFECV-LSTM 方法的预测性能,本文设计 3 组对比实验 XGBoost-RNN、XGBoost-LSTM 和 XGBoost-RFECV-RNN 的预测结果与其作对比。XGBoost 选择重要度大于 6 的 11 维特征构成特征子集输入 LSTM 模型。以 MAE、RMSE 和  $R^2$  作为预测模型准确度评价指标,结果如表 3 所示。由方法对比实验结果可得到 XGBoost-RFECV-LSTM 方法的 MAE 和 RMSE 分别为 0.001 9 和 0.002 5,决定系数  $R^2$  值为 0.974,3 个模型评价指标均好于其他 3 种对比方法,表明了 XGBoost-RFECV-LSTM 方法在 PEMFC 剩余寿命预测方面的精确性。

表 3 3 种方法的预测结果

Table 3 Forecast results of three algorithms

模型	MAE	RMSE	$R^2$
XGBoost-RNN	0.002 8	0.003 3	0.955
XGBoost-LSTM	0.002 7	0.003 1	0.96
XGBoost-RFECV-RNN	0.002 4	0.003	0.962
XGBoost-RFECV-LSTM	0.001 9	0.002 5	0.974

### 3 结 论

针对现有 PEMFC 剩余寿命预测方法的不足,提出了一种基于 XGBoost-RFECV 算法和 LSTM 神经网络的 PEMFC 剩余寿命预测方法。在 PEMFC 标准工况下,通过 XGBoost-RFECV 算法计算特征对 PEMFC 寿命影响重要度和降低特征维度,并且体现出 PEMFC 温度特征的变化对使用寿命的重要性。本文方法预测 PEMFC 寿命的 MAE 和 RMSE 分别为 0.001 9 和 0.002 5,  $R^2$  值为 0.974,并设计 XGBoost-RNN、XGBoost-LSTM 和 XGBoost-RFECV-RNN 3 组对比实验表明了该方法具有更高的预测准确度。本文所提出的剩余寿命预测模型没有考虑 PEMFC 在动态工况和极端工况下是否适用,这是下一步的工作方向。

### 参考文献

[ 1 ] 纪合超,陈涛,刘士华,等. 质子交换膜燃料电池温度监控系统的设计与开发[J]. 太阳能学报,2020,41(11):375-380.  
JI H CH, CHEN T, LIU SH H, et al. Design and development of temperature monitoring system for proton exchange membrane fuel cells[J]. Acta Energetica Sinica,2020,41(11):375-380.

[ 2 ] WANG G J, YU Y, LIU H, et al. Progress on design and development of polymer electrolyte membrane fuel cell systems for vehicle applications: A review [J]. Fuel Processing Technology,2018,179:203-228.

[ 3 ] HAN G, KWON Y K, KIM J B, et al. Development of a high-energy-density portable/mobile hydrogen energy

storage system incorporating an electrolyzer, a metal hydride and a fuel cell [J]. Applied Energy,2020,259:114175.

- [ 4 ] YAN H L, WANG G P, LU Z W, et al. Techno-economic evaluation and technology roadmap of the MW-scale SOFC-PEMFC hybrid fuel cell system for clean power generation [J]. Journal of Cleaner Production,2020,255:1202225.
- [ 5 ] 张雪霞,高雨璇,陈维荣. 基于数据驱动的质子交换膜燃料电池寿命预测[J]. 西南交通大学学报,2020,55(2):417-427.  
ZHANG X X, GAO Y X, CHEN W R. Data-driven based remaining useful life prediction for proton exchange membrane fuel cells [J]. Journal of Southwest Jiaotong University,2020,55(2):417-427.
- [ 6 ] 罗马吉,杨俊伟,赵岩,等. 不同衰退机理对 PEMFC 怠速工况性能衰退影响的模拟研究[J]. 太阳能学报,2021,42(3):414-421.  
LUO M J, YANG J W, ZHAO Y, et al. Simulation study on effect of different aging mechanisms on PEMFC performance degradation under idling condition [J]. Acta Energetica Sinica,2021,42(3):414-421.
- [ 7 ] 李奇,刘嘉蔚,陈维荣. 质子交换膜燃料电池剩余使用寿命预测方法综述及展望[J]. 中国电机工程学报,2019,39(8):2365-2375.  
LI Q, LIU J W, CHEN W R. Review and prospect of remaining useful life prediction methods for proton exchange membrane fuel cell [J]. Proceedings of the CSEE,2019,39(8):2365-2375.
- [ 8 ] RONG H L, XUE N X, PEI N W, et al. Review on hydrogen fuel cell condition monitoring and prediction methods [J]. International Journal of Hydrogen Energy,2019,44(11):5488-5498.
- [ 9 ] VICHARD L, HAREL F, RAVEY A, et al. Degradation prediction of PEM fuel cell based on artificial intelligence [J]. International Journal of Hydrogen Energy,2020,45(29):14953-14963.
- [ 10 ] LIU H, CHEN J, HISSEL D, et al. Prognostics methods and degradation indexes of proton exchange membrane fuel cells: A review [J]. Renewable and Sustainable Energy Reviews,2020,123:109721.
- [ 11 ] MORANDO S, JEMEI S, HISSEL D, et al. Proton exchange membrane fuel cell ageing forecasting algorithm based on echo state network [J]. International Journal of Hydrogen Energy,2017,42(2):1472-1480.
- [ 12 ] HUA Z G, ZHENG Z X, CÉCILE PÉRA M, et al. Remaining useful life prediction of PEMFC systems based on the multi-input echo state network [J]. Applied

- Energy, 2020, 265: 114791.
- [13] CHEN K, LAGHROUCHE S, DJERDIR A. Degradation prediction of proton exchange membrane fuel cell based on grey neural network model and particle swarm optimization [J]. Energy Conversion and Management, 2019, 195: 810-818.
- [14] 刘嘉蔚, 李奇, 陈维荣, 等. 基于核超限学习机和局部加权回归散点平滑法的 PEMFC 剩余使用寿命预测方法 [J]. 中国电机工程学报, 2019, 39(24): 7272-7279. LIU J W, LI Q, CHEN W R, et al. Remaining useful life prediction method of PEMFC based on kernel extreme learning machine and locally weighted scatterplot smoothing [J]. Proceedings of the CSEE, 2019, 39(24): 7272-7279.
- [15] 高雨璇. 质子交换膜燃料电池在线监测及寿命预测 [D]. 成都: 西南交通大学, 2019. GAO Y X. Online monitoring and remaining useful life prognostics of PEMFC [D]. Chengdu: Southwest Jiaotong University, 2019.
- [16] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system [C]. The 22nd ACM SIGKDD International Conference, August 13-17, 2016, San Francisco, America: 785-794.
- [17] WANG S H, CHEN S N. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling [J]. Journal of Petroleum Science and Engineering, 2019, 174: 682-695.
- [18] 魏昱洲, 许西宁. 基于 LSTM 长短期记忆网络的超短期风速预测 [J]. 电子测量与仪器学报, 2019, 33(2): 64-71. WEI Y ZH, XU X N. Ultra-short-term wind speed prediction model using LSTM networks [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(2): 64-71.
- [19] FCLAB Research. IEEE PHM 2014 DATA CHALLENGE [EB/OL]. 2014 [http://eng.fclab.fr/ieee-phm-2014-data-challenge/].

- [20] PAN R, YANG D, WANG Y J, et al. Performance degradation prediction of proton exchange membrane fuel cell using a hybrid prognostic approach [J]. International Journal of Hydrogen Energy, 2020, 45 (55): 30994-31008.

### 作者简介



**常家康**, 昆明理工大学硕士研究生, 主要研究方向为质子交换膜燃料电池的剩余寿命预测和故障诊断。

E-mail: 1936619806@qq.com

**Chang Jiakang** is a M. Sc. candidate at Kunming University of Science and Technology of China. His main research interests include remaining life prediction and fault diagnosis of proton exchange membrane fuel cell.



**吕宁**, 1984 年获得华中科技大学学士学位, 现为昆明理工大学高级实验师, 主要研究方向为数据建模与处理。

E-mail: lvning703@163.com

**Lyu Ning** received B. Sc. from Huazhong University of Science and Technology in 1984. Now she is a senior experimenter at Kunming University of Science and Technology. Her main research interest includes data modeling and processing.



**詹跃东** (通信作者), 2009 年获得昆明理工大学博士学位, 现为昆明理工大学教授、硕士生导师, 主要研究方向为燃料电池建模与优化以及故障检测与健康管理。

E-mail: ydzhan@163.com

**Zhan Yuedong** (Corresponding author) received Ph. D. from Kunming University of Science and Technology in 2009. Now he is professor and M. Sc. supervisor at Kunming University of Science and Technology. His main research interests include the fuel cell modeling and optimization, fault detection and health management.