

# 结合 LSTM 和改进 Transformer 联合特征提取的 PEMFC 剩余使用寿命预测方法

韩云飞<sup>1</sup>, 高锋阳<sup>1</sup>, 张建刚<sup>2</sup>, 黄娇娇<sup>3</sup>

(1.兰州交通大学 自动化与电气工程学院, 甘肃 兰州 730070; 2.兰州交通大学 数理学院, 甘肃 兰州 730070; 3.兰州城市学院 培黎机械工程学院, 甘肃 兰州 730070)

**摘要:**质子交换膜燃料电池(proton exchange membrane fuel cell, PEMFC)性能退化的精确预测是提升 PEMFC 系统性能的关键。传统性能退化预测方法一般通过机理模型和预测算法实现,并通过优化模型参数和算法结构以提高预测精度,但是对老化数据长时间尺度所隐含的细节特征考虑不够充分。因此提出一种长短期记忆网络(long short-term memory, LSTM)结合改进 Transformer 联合特征提取模型以实现 PEMFC 输出电压的精确预测。基于传统 Transformer 结构进行优化设计,构建适用于 PEMFC RUL 预测的改进型 Transformer 模型。在传统 LSTM 的基础上嵌入改进 Transformer 模型,构造 LSTM 结合改进 Transformer 联合特征提取模型。在稳态、动态和伪动态工况下分别对 LSTM、卷积神经网络(convolutional neural networks, CNN)、CNN-LSTM、改进型 Transformer 和联合特征提取模型进行输出电压预测及量化对比。结果表明,改进 Transformer 和联合预测模型在均方根误差  $E_{RMS}$ 、平均绝对误差  $E_{MA}$  和  $R^2$  等评估指标上相较于其他对比模型均有显著提升,验证了所提预测模型能够在一定程度上提高 PEMFC 的 RUL 预测精度。

**关键词:**质子交换膜燃料电池;寿命预测;长短期记忆网络;联合特征提取

中图分类号:TM911.4

文献标志码:A

文章编号:1000-2758(2025)06-1224-11

质子交换膜燃料电池(proton exchange membrane fuel cell, PEMFC)在军事、交通、分布式发电以及便携式电源等多个领域展现了广泛的应用前景<sup>[1]</sup>。然而,PEMFC 的实际应用中随着运行时间的增长,组件在应力和腐蚀作用下逐渐老化,导致性能下降<sup>[2]</sup>,输出电压不再能满足负载需求,限制了 PEMFC 的商业化进程和大规模应用,也直接影响到系统的可靠性和安全性<sup>[3]</sup>。为此通过基于历史数据和实时监测数据对 PEMFC 的健康状态进行评估<sup>[4]</sup>,可以实现对其性能退化趋势的有效预测,进而采取适当的维护措施延长 PEMFC 的服役寿命。

关于 PEMFC 寿命预测研究主要有模型驱动方

法、数据驱动方法以及结合二者的混合方法<sup>[1]</sup>。模型驱动方法基于对 PEMFC 内部物理化学过程的深入理解和建模,通过建立数学模型来描述 PEMFC 的工作原理和老化机制,从而预测其性能退化和 RUL。Futter 等<sup>[5]</sup>考虑了 PEMFC 组件层面的老化机理,建立了 PEMFC 不同运行工况下老化模型,分别对催化剂层和质子交换膜的老化机理进行了解释,但是对温度依赖性、反应速率、模型参数选择及长期行为的预测缺乏实验验证。Ma 等<sup>[6]</sup>研究发现,燃料电池内部的电化学反应会引起内部温度、气体流量以及溶液条件发生变化,这些变化会导致质子交换膜、电极、催化剂层和气体扩散层产生不同程度的老化,但是未充分讨论模型在不同工况下的适应性和泛化能力。模型驱动方法优点是能够在描述 PEMFC 内部各组件物理现象的基础上提供机理解释,但其内部各组件之间复杂的多尺度特性导致建模需要精确的参数和复杂的计算<sup>[7]</sup>。

数据驱动预测方法利用大量已知实验数据对模

收稿日期:2025-03-27

基金项目:甘肃省科技计划项目联合科研基金一般项目(25JRRA1165)

与中车“十四五”科技重大专项计划(2021CXZ021)资助

作者简介:韩云飞(1996—),博士研究生

通信作者:张建刚(1978—),教授 e-mail:zhangjg7715776@126.com

型进行训练,从而实现了对电池性能退化和 RUL 的精确预测。这种方法优势在于不需要完全理解或模拟电池内部复杂的退化机理,而是依赖于机器学习从历史数据中自动学习模式和趋势<sup>[8]</sup>。常见的数据驱动方法有最小二乘支持向量机(least squares support vector machine, LSSVM)<sup>[9]</sup>、高斯过程状态空间模型(Gaussian process state space, GPSS)<sup>[10]</sup>和各种神经网络<sup>[11-15]</sup>。此外,为了提高预测精度和泛化能力,一些研究者还提出了将不同算法组合使用的方法,例如卷积神经网络(convolutional neural networks, CNN)与 LSTM 相结合的模型(CNN-LSTM)<sup>[15]</sup>,不仅能够有效捕捉数据中的局部特征,还能处理长期依赖关系,在燃料电池性能退化和 RUL 预测中表现出色。尽管数据驱动方法在无需详细了解老化机制的情况下能够实现高性能预测,但在面对显著扰动的数据时,可能会更关注于短期变化而忽略了长期趋势<sup>[16]</sup>,往往在预测过程中会忽略数据中的频繁扰动特征而出现预测不可信或控制决策出错。混合方法旨在综合利用模型驱动和数据驱动方法的优势,通过结合 PEMFC 的物理知识与实测数据信息,实现更准确且可靠的寿命预测<sup>[12]</sup>,不仅提高了 RUL 预测的精度和可靠性,还降低了对大量高质量历史数据的依赖,因为模型驱动部分可以通过理论分析补充数据不足问题。然而,混合方法其算法复杂度和计算量相对较高,需要进一步优化以平衡性能与计算资源之间的关系;此外,在预测过程中难以实时修正老化趋势,使得迭代精度较难控制<sup>[17-18]</sup>。

综合上述研究发现,现有的 PEMFC 寿命预测方法缺少对老化数据中细节特征的考虑,并且 PEMFC 的复杂工作机制导致其老化数据非线性特征提取困难,这给 RUL 预测方法的使用带来了挑战。因此,提出一种 LSTM 结合改进 Transformer 联合特征提取预测方法,通过 LSTM 时序预测的优异性能结合改进 Transformer 模型的多头注意力机制提取数据中的细节特征变化,以应对 PEMFC 复杂工况下数据频繁扰动导致的 RUL 预测精度受限问题,对比验证所提方法的精度和有效性。

## 1 PEMFC 老化测试

### 1.1 电堆参数

PEMFC 老化实验数据采用来自 IEEE 2014

PHM 数据挑战的燃料电池老化数据集<sup>[19]</sup>和 Greenlight 20 测试站采集到的数据集<sup>[20]</sup>。数据挑战所用 PEMFC 堆由 5 块单体 PEMFC 组成,各单体活化面积为 100 cm<sup>2</sup>,额定电流密度为 0.70 A/cm<sup>2</sup>,最大电流密度为 1 A/cm<sup>2</sup>。该老化实验总共监测收集了静态工况 FC1(0.70 A/cm<sup>2</sup> 恒流运行 1 154 h)和动态工况 FC2(FC1 基础上加入 5 kHz 的三角波干扰运行 1 020 h)2 组工况数据,且各工况共有如电流、电压、温度、气体出入口压力和流量等 24 维数据。Greenlight 20 测试站所用 PEMFC 活化面积为 25 cm<sup>2</sup>,设计满载电流为 35.6 A。该测试站每 50 h 周期性地地进行数据采集并且关闭测试其内部物理状态,每 100 h 进行 1 次 12 h 的停机休息。负载电流有 9 个等级,取用 10.4 A 电流工况数据结合每 50 h 关闭测试站导致的电压恢复现象<sup>[21-22]</sup>可构成伪动态 FC3 数据集。

### 1.2 数据预处理

根据已有研究发现 PEMFC 电压老化数据最能体现老化特性<sup>[1]</sup>,因此选择电压作为老化指标。

PEMFC 原始老化数据中存在噪声和尖峰,会对模型训练产生不利影响<sup>[23]</sup>。采用绝对中位数偏差法来辨识并剔除离群点,同时使用高斯移动平均滤波器消除高频干扰,从而净化数据集。完成滤波步骤后,为了优化计算效率,同时确保保留原始数据的退化特征信息和趋势,采用三次样条插值的方法,对数据进行 1 h 时间间隔重构。不仅解决了后续的数据训练计算繁琐问题,还保证了数据的时间连续性和代表性。图 1 为 FC1、FC2 和 FC3 电压数据经过

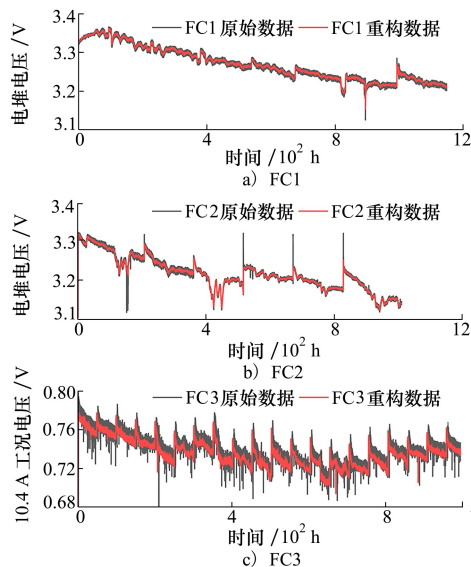


图 1 PEMFC 输出电压波形

上述处理后与原始数据的对比结果,可以直观地看到重构之后明显改善了数据质量。

## 2 LSTM 和改进 Transformer 原理

### 2.1 LSTM 算法原理

LSTM 神经网络可用来解决时间序列数据中的长期依赖问题,同时有效缓解循环神经网络(recurrent neural network, RNN)中存在的梯度消失和梯度爆炸问题<sup>[24]</sup>。如图 2 所示,在 LSTM 模型中,前一个时间步的输出  $H_{t-1}$  和单元状态  $C_{t-1}$  会作为输入传递到下一个时间步,即 LSTM 能够根据当前的输入以及之前的时间步信息调整其输出。这种机制使得 LSTM 具有“记忆性”,可以根据需要选择性地保留或遗忘历史信息,从而更好地捕捉长时间间隔内的依赖关系。在每个时间点  $t$ , LSTM 接收 3 个输入:当前的输入特征  $X_t$ 、上一时间步的隐藏状态  $H_{t-1}$  和单元状态  $C_{t-1}$ ,并产生 2 个输出:当前的隐藏状态  $H_t$  和更新后的单元状态  $C_t$ 。

LSTM 中信息的前向传播表示如下:首先使用当前输入的特征  $X_t$  和上一时刻隐藏状态  $H_{t-1}$  结合进行训练,相应表达式如(1)~(5)式所示。

$$F_t = \sigma[W_f X_t + W_f H_{t-1} + B_f] \quad (1)$$

$$I_t = \sigma[W_i X_t - W_i H_{t-1} + B_i] \quad (2)$$

$$C'_t = \tanh[W_c X_t + W_c H_{t-1} + B_c] \quad (3)$$

$$O_t = \sigma[W_o X_t + W_o H_{t-1} + B_o] \quad (4)$$

$$H_t = O_t \times \tanh(C_t) \quad (5)$$

式中:  $\sigma$  为激活函数;  $W_f, W_i, W_c$  和  $W_o$  为相应层权重矩阵;  $B_f, B_i, B_c$  和  $B_o$  为相应层输出产生的偏差向量。通过遗忘门对上一时间点输入的信息进行选择性地遗忘,随后通过 sigmoid 函数将输入门的值映射到  $[0, 1]$  区间内,紧接着利用 tanh 层创建新的状态变量  $C'_t$ ,将  $C'_t$  和输入门的乘积作为新加入的信息,此时新的状态变量  $C_t$  可被表示为

$$C_t = F_t \times C_{t-1} + I_t \times C'_t \quad (6)$$

最后的输出信息将由输出门  $O_t$  进行调控,并且通过一个 tanh 层对当前状态变量  $C_t$  进行缩放并与输出门  $O_t$  做乘积可得到当前的隐藏状态  $H_t$ 。至此 LSTM 中信息的前向传播算法全部完成。后续便于更新网络参数,需要得到相应层权重的梯度,然而梯度的计算需要通过时序反向传播算法完成,在这一过程中误差项在反向传播中几乎不发生衰减。

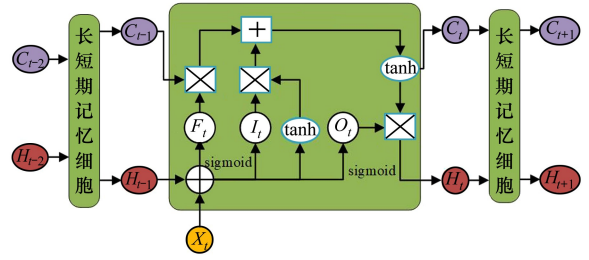


图 2 LSTM 结构

### 2.2 改进 Transformer 算法原理

Transformer 所具有的多头自注意力机制能够同时关注输入数据的不同部分,从而提取数据中更多层次主次特征信息,这种机制有效提高了时序预测模型的精准性和鲁棒性。传统 Transformer 由编码器和解码器两部分构成<sup>[25]</sup>,如图 3a)所示。由于 Transformer 最初设计是为处理自然语言任务,与该研究用 PEMFC RUL 预测存在较大差异,因此在原始 Transformer 基础上进行改进,设计了新的特征提取架构,丢弃了解码器结构,改进后的 Transformer 结构如图 3b)所示。该研究中将 PEMFC 退化指标电压数据分为训练集  $T_r$  和预测集  $F_o$  两部分,具体表示为

$$T_r = \{x_i^r, y_i^r\}_{i=1}^{n_r} \quad (7)$$

$$F_o = \{x_i^o\}_{i=1}^{n_o} \quad (8)$$

$$x_i^r = (x_{i1}^r, x_{i2}^r, \dots, x_{id}^r) \quad (9)$$

式中:  $x_i^r$  表示第  $i$  个训练集数据;  $n_r$  表示训练集样本数量;  $d$  表示数据维度;  $x_i^o$  表示第  $i$  个预测集数据;  $n_o$  表示预测集样本数量。数据通过前向计算获得训练集和预测集特征信息。前向计算具体流程如下。

首先将输入数据  $X_e$  进行层归一化将数据映射到  $[0, 1]$  区间内得到  $X_o$ ,这样可使网络训练收敛速度更快,归一化计算公式为

$$X = \frac{X_e - mX_e}{\sqrt{\text{var}(X_e)}}\lambda + \beta \quad (10)$$

式中:  $m$  表示平均值;  $\text{var}$  为方差;  $\lambda$  表示缩放因子;  $\beta$  表示平移参数。随后通过编码器将序列的所有变量信息嵌入到一个矩阵中,表达式为

$$S = U_b(X_{t,d}) \quad (11)$$

式中:  $S$  表示编码输出矩阵;  $U_b$  表示维度为  $b$  的编码函数;  $X_{t,d}$  表示行数为  $t$ ,列数为  $d$  的二维矩阵。

多头自注意力机制由不同的多个独立头部提取数据不同信息,使提取到的数据更加丰富。多头自

注意力机制的构建需要以下 3 步完成。

第 1 步 通过 3 个全连接层构建 3 个权重矩阵  $W_Q, W_K, W_V$ , 对  $X$  做线性映射, 从而形成 3 个新的矩阵向量  $Q, K, V$ , 其相应计算表达式为:

$$Q = L(X) = XW_Q \quad (12)$$

$$K = L(X) = XW_K \quad (13)$$

$$V = L(X) = XW_V \quad (14)$$

式中,  $L$  表示全连接层运算符。

第 2 步 将矩阵向量  $Q, K, V$  分别均分为  $h$  组, 接着组成不同的头部 ( $Q_i, K_i, V_i$ ),  $1 \leq i \leq h$ 。不同头部在不同的空间各自计算自注意力机制, 从而获得不同特征数据。矩阵均分过程表示为:

$$Q = [Q_1, Q_2, \dots, Q_h] \quad (15)$$

$$K = [K_1, K_2, \dots, K_h] \quad (16)$$

$$V = [V_1, V_2, \dots, V_h] \quad (17)$$

单头自注意力机制计算公式为:

$$A(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (18)$$

$$\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) = \frac{e^{Q_i K_i^T / \sqrt{d_k}}}{\sum_{i=1}^h e^{Q_i K_i^T / \sqrt{d_k}}} \quad (19)$$

式中:  $K_i^T$  为  $K_i$  的转置;  $d_k$  表示  $K_i$  的维度;  $Q_i$  与  $K_i^T$  的点乘为二者相关性计算, softmax 函数将  $Q_i$  与  $K_i^T$  的点乘矩阵映射到  $[0, 1]$  区间内。

第 3 步 为获得更多特征信息, 直接将归一化层和编码后的矩阵进行矩阵相加作为多头自注意力机制的输入。融合不同自注意力头部提取更加丰富的特征信息, 对提取的信息进行合并, 获得更加全面的特征数据, 有助于挖掘原始数据中更深处的隐含信息, 多头自注意力机制融合过程表示为

$$X_A = [A(Q_i, K_i, V_i), A(Q_{i+1}, K_{i+1}, V_{i+1}), \dots, A(Q_h, K_h, V_h)] W_0 \quad (20)$$

式中:  $A$  表示自注意力运算符;  $W_0$  的作用是将  $h$  个不同自注意力机制获得的特征信息进行融合。

最后, 将矩阵相加后的信息和多头自注意力机制融合后的信息进行残差与归一化连接, 残差连接可以保留更多有用信息, 也可以预防网络训练过程中的梯度消失和爆炸问题。为了进一步提高网络训练性能, 预防网络过拟合问题, 前馈神经网络之后添加有神经元丢弃层, 数据经过丢弃层之后将前馈神经网络输入数据结合再次进行残差与归一化处理, 以提高数据特征信息精度和鲁棒性。所构造的全连

接层共有 32 个隐含层。softmax 激活函数为神经网络带入了非线性因子, 可以使网络更容易贴近任意非线性函数。

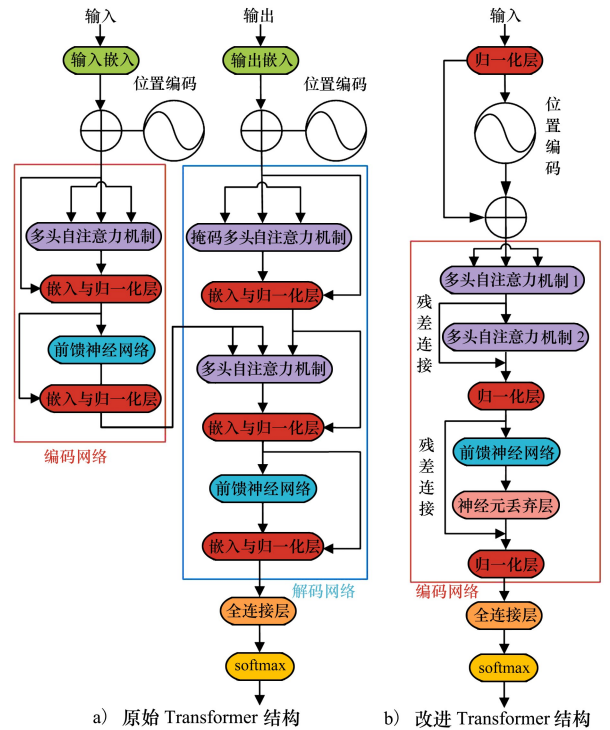


图 3 Transformer 结构

### 2.3 联合特征提取算法原理

在 LSTM 的基础上结合改进 Transformer 的联合特征提取模型, 用于 PEMFC RUL 预测。该模型通过同时提取电压数据的线性 and 非线性特征, 实现基于多维度信息的高精度 RUL 预测。

具体操作流程如下:

1) 对 PEMFC 测试所得老化数据进行降噪, 平滑以及重构处理, 提高数据质量。

2) 使用单元数为 32 的 LSTM 直接对输入数据进行非线性特征提取拟合迭代得到  $N$  个预测特征与  $M$  个残差序列, 使用 ReLU 激活函数对所得预测特征和残差序列进行非线性化, 使得网络更加贴近非线性函数, 全连接层进行预测特征和残差融合。

3) 对融合后信息进行 32 个输出通道的位置编码得到  $S$  个输出矩阵, 对全连接层融合信息和  $S$  个输出矩阵进行矩阵相加作为编码网络的输入, 数据通过 2 个 4 头, 且各自具有 32 个输出通道、128 个键、128 个查询通道、128 条值通道的自注意力机制层进行更深层次的自注意力特征提取。

4) 全连接层实现自注意力特征融合,使用 soft-max 激活函数进行非线性预测输出。

### 3 验证与讨论

#### 3.1 电压退化预测性能评价指标

为评价模型预测精度,选取均方根误差  $E_{RMS}$ 、平均绝对误差  $E_{MA}$  和决定系数  $R^2$  作为评价指标<sup>[26]</sup>。其相应计算公式为

$$E_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2} \quad (21)$$

$$E_{MA} = \frac{1}{N} \sum_{i=1}^N |y_i - y'_i| \quad (22)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - y'_i)^2}{\sum_{i=1}^N (\hat{y}_i - y'_i)^2} \quad (23)$$

式中:  $N$  为预测数量;  $y_i$  为真实值;  $y'_i$  为预测值;  $\hat{y}_i$  为电压平均值。

#### 3.2 电压退化预测

将重构后的 FC1、FC2 和 FC3 数据集划分为训练集和预测集分别输入 LSTM、CNN、CNN-LSTM、改进 Transformer 和联合特征提取预测模型进行变训练集占比输出电压预测。为增强对比公平性,所用模型参数设置一致:最大训练次数 100、初始学习率 0.005、学习率下降因子 0.1。训练集占比设置为 50%、60%、70% 和 80%。联合特征提取预测流程如图 4 所示。

首先利用 LSTM 优异的时序数据处理能力提取电压数据中的隐藏特征信息  $H_i$  和主要特征信息  $C_i$ , 经过融合处理作为改进 Transformer 的输入,利用改进 Transformer 强大的特征提取能力再次提取数据中的隐含信息和更深层次的特征信息,最后通过融合和非线性化作为预测电压数据的输出。预测量化结果中均方根误差和平均绝对误差均为  $10^{-4}$  量级,联合特征提取预测模型标记为 LSTM-TS。

##### 3.2.1 稳态条件下电压退化多步预测(FC1)

电压预测趋势结果对比如图 5 所示,碍于篇幅只展示训练集占比 50% 和 80% 的对比结果。联合特征提取预测模型在训练集达到 60% 时其精度有显著提升现象,LSTM、CNN、CNN-LSTM 预测模型训练集占比需要达到 70% 以上时才能有较好的电压跟随性。传统的时序预测网络需要大量的历史数据进行训练才能有较好的精确度,训练集占比较小导致模型训练阶段无法充分提取到数据隐含特征信息,所以提高训练集占比其预测精度才会有所提升。然而从改进 Transformer 单模型预测结果中可以发现训练集占比仅仅 50% 其预测结果就能和历史电压数据趋势基本保持一致,这主要得益于 Transformer 所具有的多头自注意力机制使其在训练阶段能够提取到丰富的历史数据特征,因此在预测阶段展现出了高度的精确性和鲁棒性。

为了进一步对比模型预测精度,对预测模型预测结果的量化统计如表 1 所示,从表中可发现改进 Transformer 模型的 3 个评价指标数值基本趋于平衡,其余 4 个预测模型的  $E_{RMS}$  和  $E_{MA}$  均随着训练集占比增大而有显著降低现象, $R^2$  随着训练集占比增大而有增大趋势,该结果和预测趋势图保持一致。联合预测模型累积均方根误差相较于 LSTM、CNN-LSTM 和 CNN 模型分别降低 20.16%、22.11% 和 40.73%, 累计平均绝对误差分别降低 16.59%、32.96% 和 47.58%, 累计决定系数分别提升 60.1%、77.52% 和 29.13%。改进 Transformer 预测模型累计均方根误差相较于 LSTM、CNN-LSTM 和 CNN 模型分别降低了 61.31%、62.56% 和 71.28%, 累计平均绝对误差分别降低 73.85%、78.98% 和 83.56%, 累计决定系数分别提升 145.1%、171.7% 和 97.65%。因此所提改进 Transformer 和联合预测模型均具有较高的预测精度,且改进 Transformer 预测模型对于静态工况其电压预测适用性更强。

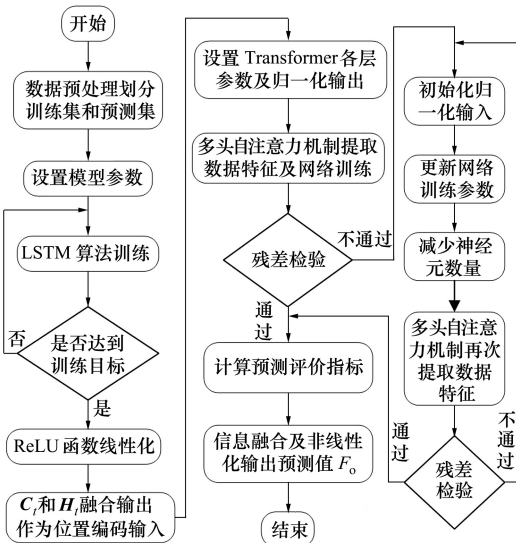


图 4 联合特征提取预测流程

表 1 静态工况不同训练集多步预测量化结果

训练集占比/%	预测方法	$E_{RMS}$	$E_{MA}$	$R^2$
50	LSTM	160.4	124.8	-0.02
	Transformer	37.52	19.44	0.998
	LSTM-TS	151.2	120.5	0.099
	CNN-LSTM	194.2	172.4	-0.49
	CNN	251.9	221	0.15
60	LSTM	145.9	107.1	-0.21
	Transformer	36.69	17.62	0.92
	LSTM-TS	76.42	54.52	0.67
	CNN-LSTM	116.1	104.3	0.24
	CNN	155.6	124.7	0.27
70	LSTM	48.14	28.7	0.83
	Transformer	41.5	18.7	0.88
	LSTM-TS	67.9	47.4	0.67
	CNN-LSTM	70.4	54.9	0.64
	CNN	101	77	0.48
80	LSTM	34.62	16.7	0.90
	Transformer	34.8	16.8	0.90
	LSTM-TS	15.1	8.91	0.98
	CNN-LSTM	18.1	13.5	0.98
	CNN	16	18.6	0.98

### 3.2.2 动态条件下电压退化多步预测(FC2)

动态工况多步预测趋势对比结果如图 6 所示,碍于篇幅只展示训练集占比 50% 和 80% 的对比结果。预测结果量化统计如表 2 所示。预测结果对比发现,联合预测模型预测精度较高,鲁棒性较好,并且对于波峰、波谷的预测性能同样表现优异,850 h 处的电压恢复现象也未能影响联合预测模型的预测精度。然而 LSTM、CNN 和 CNN-LSTM 模型对于动态工况下的预测性能则明显劣于改进型 Transformer 和联合预测模型,尤其对于波谷处的预测,LSTM 和 CNN 模型在波谷处的预测性能已经明显失去了电压跟随性,并且训练集占比越大波谷处的预测能力越弱。CNN-LSTM 模型随着训练集占比增大其预测结果有较小提升现象,说明传统时序预测模型需要对大量历史数据进行训练才能有较为理想的预测精度,并且电压突变处预测性能较低,主要因为模型训练阶段未能提取到数据中隐含的有效特征信息导致其无法输出精确的预测结果。以上结果也从侧面反映出改进 Transformer 和联合预测模型对于数据扰动频率大且宽的应用场合具有较高的数据预测能力和适用性,主要原因在于改进 Transformer 模型所具

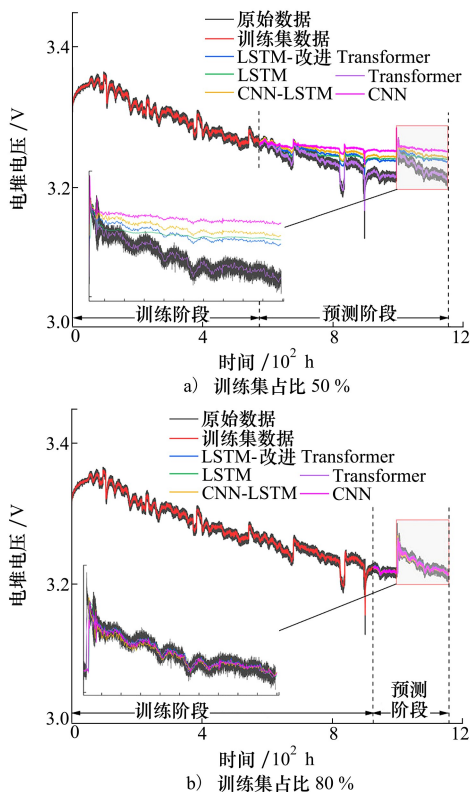


图 5 静态工况不同训练集多步预测结果

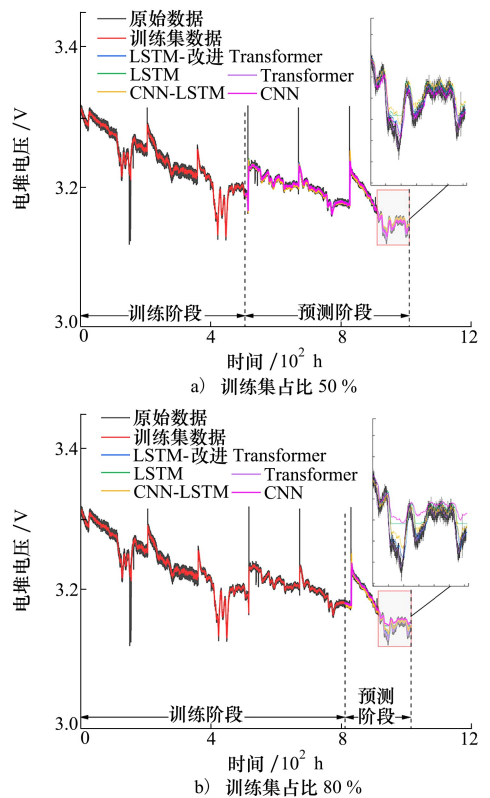


图 6 动态工况不同训练集多步预测结果

有的多头自注意力机制对于数据次要特征信息的提取能力较强,因此对于数据扰动频繁的场合适应性更高。从量化结果中可以发现 LSTM、CNN 和 CNN-LSTM 模型的  $E_{RMS}$  和  $E_{MA}$  有随着训练集占比增大而增大或先减小再增大的现象,说明数据的频繁扰动和电压恢复现象对于 LSTM、CNN 和 CNN-LSTM 模型的预测性能有较大影响,改进 Transformer 模型的 3 个评价指标基本趋于稳定,而联合预测模型的  $E_{RMS}$  和  $E_{MA}$  普遍低于其余模型,  $R^2$  普遍高于其余模型,这与预测趋势图保持一致。

表 2 动态工况不同训练集多步预测量化结果

训练集占比/%	预测方法	$E_{RMS}$	$E_{MA}$	$R^2$
50	LSTM	42.24	20.07	0.975
	Transformer	35.77	14.93	0.982
	LSTM-TS	15.17	7.73	0.997
	CNN-LSTM	29.8	25.2	0.987
	CNN	14.8	7.79	0.997
60	LSTM	40.931	21.46	0.974
	Transformer	32.06	14.08	0.984
	LSTM-TS	15.43	8.39	0.996
	CNN-LSTM	65.8	45.3	0.932
	CNN	26.2	12.1	0.990
70	LSTM	51.27	27.15	0.955
	Transformer	34.6	14.99	0.980
	LSTM-TS	15.89	8.96	0.996
	CNN-LSTM	55.1	37	0.949
	CNN	48.3	28.4	0.961
80	LSTM	64.17	35.25	0.946
	Transformer	41	16.01	0.978
	LSTM-TS	20.68	10.89	0.994
	CNN-LSTM	33.6	28.3	0.985
	CNN	69.2	42.2	0.939

为了验证改进 Transformer 和联合预测模型在动态工况下对数据训练量的依赖性,将训练集占比分别设置为 40% 和 45% 进行训练预测,其中 45% 训练集占比预测结果如图 7 所示,量化结果如表 3 所示。训练集占比为 40% 时联合预测模型预测性能

较低,而训练集占比 45% 时改进 Transformer 和联合预测模型依然具有较高的电压跟随性和鲁棒性,其量化结果和上文基本一致。训练集占比 45% 时,联合预测模型  $E_{RMS}$  相较于 LSTM、改进 Transformer、CNN-LSTM 和 CNN 模型分别降低 62.29%, 51.86%, 66.23% 和 52.66%,  $E_{MA}$  分别降低 60.91%, 40.4%, 65.34% 和 60.25%,  $R^2$  分别提升 2.3%, 1.3%, 2.89% 和 1.22%。因此所提联合预测模型对于数据频繁扰动的动态工况电压预测能力更强、精确度更高,且数据依赖性较低。

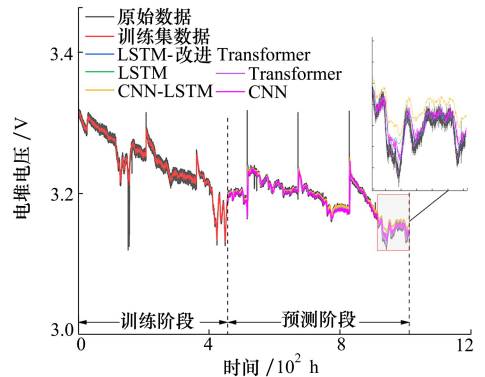


图 7 动态工况训练集占比 45% 的预测结果

表 3 训练集占比 40% 和 50% 的预测量化结果

训练集占比/%	预测方法	$E_{RMS}$	$E_{MA}$	$R^2$
40	LSTM	42.24	20.07	0.975
	Transformer	35.77	14.93	0.982
	LSTM-TS	15.17	7.73	0.997
	CNN-LSTM	206.6	157.5	0.351
	CNN	196	144.2	0.02
45	LSTM	40.931	21.458	0.974
	Transformer	32.06	14.075	0.984
	LSTM-TS	15.434	8.388	0.996
	CNN-LSTM	45.7	24.2	0.968
	CNN	32.6	21.1	0.984

### 3.2.3 伪动态条件下电压退化多步预测(FC3)

不同方法的预测结果如图 8 所示。碍于篇幅,对于伪动态工况的电压趋势预测结果仅展示训练集占比 50% 的预测结果。其预测量化结果如表 4 所示。从电压预测趋势图中可以看到改进 Transformer 和联合预测模型的电压预测曲线和原始曲线保持高

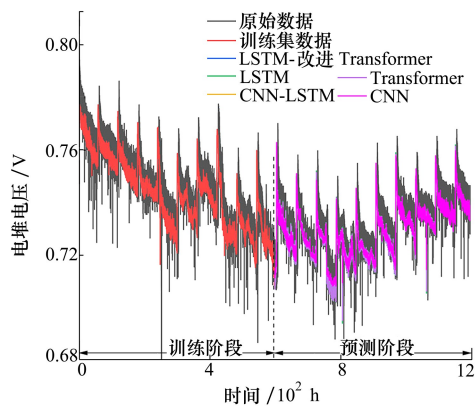


图8 伪动态工况训练集占比 50% 预测结果

表4 伪动态工况不同训练集多步预测量化结果

训练集占比/%	预测方法	$E_{RMS}$	$E_{MA}$	$R^2$
50	LSTM	58.4	35.9	0.679
	Transformer	1.11	0.728	0.999
	LSTM-TS	7.98	2.85	0.994
	CNN-LSTM	25.5	15.7	0.939
	CNN	14.7	8.29	0.98
60	LSTM	39.9	23.9	0.863
	Transformer	1.24	0.834	0.999
	LSTM-TS	9.4	2.74	0.992
	CNN-LSTM	15.1	12.6	0.98
	CNN	22.3	10	0.957
70	LSTM	32.9	18.7	0.877
	Transformer	1.85	1.31	0.999
	LSTM-TS	1.89	1.13	0.999
	CNN-LSTM	5.03	3.27	0.997
	CNN	9.11	7.05	0.991
80	LSTM	33	19.2	0.798
	Transformer	2.07	1.32	0.999
	LSTM-TS	2.01	1.34	0.999
	CNN-LSTM	6.8	5.09	0.991
	CNN	7.83	5.27	0.989

度吻合状态,能够高精度地反映出原始电压数据的变化特征和波动趋势,该工况下大约每隔 50 h 停机监测 PEMFC 内部物理状态,导致出现高频率的电

压恢复现象,但并未影响改进 Transformer 和联合预测模型的预测性能。而其他 3 种预测模型的预测结果与实际曲线均存在明显偏差,特别是电压发生突变或恢复节点处。从预测量化结果中可以发现改进 Transformer 模型的 3 个评价指标基本趋于平衡,联合预测模型的  $E_{RMS}$  和  $E_{MA}$  随着训练集占比增大有降低趋势, $R^2$  有增大的现象,主要由于前期 LSTM 对于数据特征提取能力有限导致后续改进 Transformer 没有充足的数据特征供其提取;LSTM、CNN-LSTM 和 CNN 预测模型的  $E_{RMS}$  和  $E_{MA}$  随着训练集占比增大有降低趋势,并且数值波动较大, $R^2$  呈现先增大后降低的趋势。说明传统的电压预测模型需要大量的历史数据进行训练,并且训练集占比较小时 LSTM、CNN-LSTM 和 CNN 预测模型未能充分提取到电压数据内部的变化特征,导致对于波峰、波谷处的预测性能较低,误差较大,随着训练集占比增大,LSTM、CNN-LSTM 和 CNN 预测模型预测误差减小,所以在训练阶段大量的历史数据使得模型能够有效提取数据中的细节特征,预测性能逐渐提高。

从表 4 中计算可得改进 Transformer 相较于 LSTM、CNN-LSTM 和 CNN 预测模型的  $E_{RMS}$  累计分别降低 96.18%,88.04%和 88.38%, $E_{MA}$  累计分别降低 95.71%,88.57%和 86.31%, $R^2$  累计分别提升 55.39%,27.95%和 27.62%;联合预测模型相较于 LSTM、CNN-LSTM 和 CNN 预测模型的  $E_{RMS}$  累计分别降低 87.1%,59.58%和 60.72%, $E_{MA}$  累计分别降低 91.56%,78.01%和 73.67%, $R^2$  累计分别提升 23.84%,1.97%和 1.71%。因此改进 Transformer 和联合预测模型对于电压波动较大和恢复频率较高的场合具有更高的预测性能和适用性。

### 3.3 PEMFC 预测电压与 RUL

根据 IEEE 2014 PHM 数据挑战中的《IEEE-Challenge-Details》<sup>[19]</sup> 文件,PEMFC 的 RUL 失效阈值可以选取初始电压的 96.5%,96%,95.5%,95%和 94.5%。美国能源部(United States Department of Energy,U.S.DoE)指出,PEMFC 系统可以用于固定式发电站、便携式设备以及各种车辆等不同场合中。PEMFC 的 RUL 指的是 PEMFC 的寿命结束点和预测起始点的时间差值,因此使用者可以根据 PEMFC 的具体使用工况选取电压失效阈值,再根据模型预测起始点和预测结果失效阈值对应时间进而估计 PEMFC 的 RUL。

## 4 结 论

针对 PEMFC 系统在静态、动态和伪动态工况下 RUL 预测,提出了一种 LSTM 结合改进 Transformer 联合特征提取预测模型,实现了 PEMFC 系统的 RUL 精确预测,主要结论如下:

1) 在原始 Transformer 结构的基础上进行改进重构,得到改进 Transformer 结构。结合 LSTM 优异性能嵌入改进 Transformer 结构,构造 LSTM 结合改进 Transformer 联合特征提取预测模型。以  $E_{\text{RMS}}$ ,  $E_{\text{MA}}$  和  $R^2$  作为预测评价指标,分别在稳态、动态和伪动态工况下进行多模型变训练集比例的多步预测,验证了改进 Transformer 和联合预测模型对

PEMFC 系统老化预测的适用性。

2) 改进 Transformer 和联合预测模型可精确反映电压老化趋势并提取历史数据中的细节特征信息,预测结果与原始数据均能保持高度一致,具有优异的电压预测准确性和稳定性。改进 Transformer 预测模型对于静态和伪动态工况具有更高的预测精度,联合预测模型对于动态工况具有更强的电压跟随性和鲁棒性,且对历史电压数据依赖性较低。

所提改进 Transformer 和联合预测模型在 PEMFC 系统的 RUL 预测过程中具有较高的预测准确度,可为后期在线预测做技术支持,并且可为后续动态工况长期预测技术的精度受限问题提供理论借鉴。

## 参考文献:

- [1] 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 Qi, LIU Jiawei, CHEN Weirong. 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 (in Chinese)
- [2] 马菁, 马强, 王俊杰, 等. 温度和阴极湿度对质子交换膜燃料电池的影响[J]. 西北工业大学学报, 2023, 41(6): 1162-1169  
MA Jing, MA Qiang, WANG Junjie, et al. Effects of temperature and cathode humidity on performance of PEM full cell[J]. Journal of Northwestern Polytechnical University, 2023, 41(6): 1162-1169 (in Chinese)
- [3] BO Z, RONGJIE W, DESONG J, et al. Parameter identification of proton exchange membrane fuel cell based on swarm intelligence algorithm[J]. Energy, 2023, 283: 128935
- [4] HUA Zhiguang, ZHENG Zhixue, PAHON E, et al. A review on lifetime prediction of proton exchange membrane fuel cells system[J]. Journal of Power Sources, 2022, 529: 231256
- [5] FUTTER G A, LATZ A, JAHNKE T. Physical modeling of chemical membrane degradation in polymer electrolyte membrane fuel cells: influence of pressure, relative humidity and cell voltage[J]. Journal of Power Sources, 2019, 410: 78-90
- [6] MA R, YANG T, BREAZ E, et al. Data-driven proton exchange membrane fuel cell degradation prediction through deep learning method[J]. Applied Energy, 2018, 231: 102-115
- [7] ZUO J, CADET C, LI Z, et al. A deterioration-aware energy management strategy for the lifetime improvement of a multi-stack fuel cell system subject to a random dynamic load[J]. Reliability Engineering & System Safety, 2024, 241: 109660
- [8] 李浩, 李浩, 杨扬, 等. 基于改进鲸鱼算法优化 GRU 的 PEMFC 老化预测[J]. 中国电机工程学报, 2024, 44(20): 8166-8178  
LI Hao, LI Hao, YANG Yang, et al. PEMFC aging prediction based on improved whale optimization algorithm optimized GRU[J]. Proceedings of the CSEE, 2024, 44(20): 8166-8178 (in Chinese)
- [9] CHENG Y, ZERHOUNI N, LU C. A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell[J]. International Journal of Hydrogen Energy, 2018, 43(27): 12314-12327
- [10] ZHU L, CHEN J. Prognostics of PEM fuel cells based on Gaussian process state space models[J]. Energy, 2018, 149: 63-73
- [11] 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
- [12] LIU H, CHEN J, HISSEL D, et al. Remaining useful life estimation for proton exchange membrane fuel cells using a hybrid

- method[J]. *Applied Energy*, 2019, 237: 910-919
- [13] ZUO J, LV H, ZHOU D, et al. Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application[J]. *Applied Energy*, 2021, 281: 115937
- [14] MORANDO S, JEMEI S, HISSEL D, et al. ANOVA method applied to proton exchange membrane fuel cell ageing forecasting using an echo state network[J]. *Mathematics and Computers in Simulation*, 2017, 131: 283-294
- [15] FRANCISCO DA Costa Lopes, WATANABE E H, ROLIM L G B. A control-oriented model of a PEM fuel cell stack based on NARX and NOE neural networks[J]. *IEEE Trans on Industrial Electronics*, 2015, 62(8): 5155-5163
- [16] 杨洪, 陈景文, 华志广, 等. 基于集成型极限学习机的氢燃料电池寿命预测[J]. *电工技术学报*, 2025, 40(3): 964-974  
YANG Qi, CHEN Jingwen, HUA Zhiguang, et al. Research on PEMFC lifetime prediction based on ensemble extreme learning machine[J]. *Transactions of China Electrotechnical Society*, 2025, 40(3): 964-974 (in Chinese)
- [17] SUN Xilei, XIE Mingke, FU Jianqin, et al. An improved neural network model for predicting the remaining useful life of proton exchange membrane fuel cells[J]. *International Journal of Hydrogen Energy*, 2023, 48(65): 25499-25511
- [18] CHANG Ke, KAI Han, WANG Yongzhen. A hybrid method combining degradation mechanisms and deep learning for lifetime prediction of proton exchange membrane fuel cells under dynamic load cycle conditions[J]. *Journal of Power Sources*, 2025, 633: 236464
- [19] GOURIVEAU R, HILAIRET M, HISSEL D, et al. IEEE PHM 2014 data challenge: outline experiments scoring of results winners[C]//International Conference on Prognostics and Health Manage, Belfort, France, 2014
- [20] ZUO Jian, LV Hong, ZHOU Daming, et al. Long-term dynamic durability test datasets for single proton exchange membrane fuel cell[J]. *Data in Brief*, 2021, 35: 106775
- [21] CHU Tiankuo, WANG Qinpu, XIE Meng, et al. Investigation of the reversible performance degradation mechanism of the PEM-FC stack during long-term durability test[J]. *Energy*, 2022, 258: 124747
- [22] XIAO Biao, ZHAO Junjie, FAN Lixin, et al. Effects of moisture dehumidification on the performance and degradation of a proton exchange membrane fuel cell[J]. *Energy*, 2022, 245: 123298
- [23] 赵波, 张领先, 章雷其, 等. 基于 TCN 和 AUKF 联合迭代的 PEMFC 寿命融合预测方法[J]. *中国电机工程学报*, 2025, 45(9): 3609-3624  
ZHAO Bo, ZHANG Lingxian, ZHANG Leiqi, et al. PEMFC remaining useful life fusion method based on joint iteration of TCN and AUKF[J]. *Proceedings of the CSEE*, 2025, 45(9): 3609-3624 (in Chinese)
- [24] ZHENG Lu, HOU Yongping, ZHANG Tao, et al. Performance prediction of fuel cells using long short-term memory recurrent neural network[J]. *International Journal of Energy Research*, 2021, 45(6): 9141-9161
- [25] 汪敏, 郭鑫平, 唐洪明, 等. 深度 Transformer 迁移学习的页岩气储层核心参数预测案例[J]. *地球物理学报*, 2023, 66(6): 2592-2610  
WANG Min, GUO Xinping, TANG Hongming, et al. Prediction case of core parameters of shale gas reservoirs through deep Transformer transfer learning[J]. *Geophys*, 2023, 66(6): 2592-2610 (in Chinese)
- [26] 吴航宇, 王玮, 朱文超, 等. 基于 ARIMA-BiGRU 双数据驱动的燃料电池性能退化预测方法[J]. *中国电机工程学报*, 2025, 45(7): 2690-2699  
WU Hangyu, WANG Wei, ZHU Wenchao, et al. A fuel cell performance degradation prediction method based on ARIMA-BiGRU dual data-driven approach[J]. *Proceedings of the CSEE*, 2025, 45(7): 2690-2699 (in Chinese)

# Remaining useful life prediction method of PEMFC based on joint feature extraction of LSTM and improved Transformer

HAN Yunfei<sup>1</sup>, GAO Fengyang<sup>1</sup>, ZHANG Jiangang<sup>2</sup>, HUANG Jiaojiao<sup>3</sup>

(1.Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;  
2.College of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou 730070, China;  
3.Bailie Mechanical Engineering, Lanzhou City University, Lanzhou 730070, China)

**Abstract:** Proton exchange membrane fuel cells (PEMFC) are a crucial component of modern sustainable clean energy generation technology. Accurate prediction of performance degradation is key to enhancing the performance of PEMFC systems and is also an important step in promoting this clean energy technology for broader applications. Traditional methods for predicting performance degradation typically achieve their aims through mechanistic models and forecasting algorithms, refining model parameters and algorithm structures to improve accuracy. However, these methods often fall short in fully considering the detailed characteristics implied by aging data over long-time scales and the phenomenon of voltage recovery. Therefore, this paper proposes a predictive model that combines long short-term memory (LSTM) networks with an enhanced Transformer for joint feature extraction to achieve precise predictions of PEMFC output voltage. Initially, based on traditional Transformer architecture, an optimized design is performed to build an improved Transformer model suitable for PEMFC remaining useful life (RUL) prediction. Secondly, the improved Transformer model is embedded into the conventional LSTM framework, constructing a combined LSTM and improved Transformer joint feature extraction model. Under steady-state, dynamic and pseudo-dynamic datasets, finally, the LSTM, convolutional neural networks (CNN), CNN-LSTM, the improved Transformer, and the joint feature extraction model were evaluated for output voltage prediction and quantitatively compared. The results indicate that the improved Transformer and the joint prediction model show significant improvements over other comparative models in evaluation metrics such as RMSE, MAE, and  $R^2$ . This confirms that the proposed prediction model can enhance the RUL prediction accuracy of PEMFC to some extent.

**Keywords:** proton exchange membrane fuel cell; life prediction; long short-term memory; joint feature extraction

**引用格式:** 韩云飞, 高锋阳, 张建刚, 等. 结合 LSTM 和改进 Transformer 联合特征提取的 PEMFC 剩余使用寿命预测方法[J]. 西北工业大学学报, 2025, 43(6): 1224-1234

HAN Yunfei, GAO Fengyang, ZHANG Jiangang, et al. Remaining useful life prediction method of PEMFC based on joint feature extraction of LSTM and improved Transformer[J]. Journal of Northwestern Polytechnical University, 2025, 43(6): 1224-1234 (in Chinese)