华北电力大学

专业硕士学位论文

基于机器学习的氢燃料电池电堆电压衰减预测与软件开发

Prediction and Software Development of Hydrogen Full Cell Stack Voltage Decay Based on Machine Learning

魏子涵

2025 年 5 月

国内图书分类号：TK91　　学校代码：10079
国际图书分类号：629　　密　　级：公开

专业硕士学位论文

基于机器学习的氢燃料电池电堆电压衰减预测与软件开发

硕士研究生：魏子涵
导　　师：高正阳教授
企业导师：段志洁高级工程师
申请学位：能源动力硕士
专业领域：动力工程
学习方式：全日制
所在学院：能源动力与机械工程学院
答辩日期：2025 年 5 月
授予学位单位：华北电力大学

Classified Index: TK91
U.D.C: 629

Thesis for the Professional Master's Degree

Prediction and Software Development of Hydrogen Full Cell Stack Voltage Decay Based on Machine Learning

Candidate: Zihan Wei

Supervisor: Prof. Zhengyang Gao

Enterprise mentor: Zhijie Duan

Professional Degree Applied for: Master of Energy and Power Engineering

Speciality: Power Engineering

Cultivation ways: Full-time

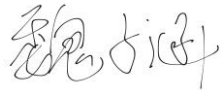
School: School of Energy Power and Mechanical Engineering

Date of Defence: May, 2025

Degree-Conferring-Institution: North China Electric Power University

华北电力大学学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于机器学习的氢燃料电池电堆电压衰减预测与软件开发》,是本人在导师指导下，在华北电力大学攻读学位期间独立进行研究工作所取得的成果。论文中除已注明部分外不包含他人已发表或完成的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。论文由本人撰写，未使用人工智能代写。本声明的法律结果将完全由本人承担。

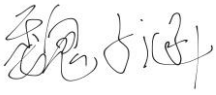
作者签名：  日期：2025 年 5 月 30 日

华北电力大学学位论文使用授权书

学位论文系研究生在华北电力大学攻读学位期间在导师指导下完成的成果，知识产权归属华北电力大学所有，学位论文的研究内容不得以其它单位的名义发表。研究生完全了解华北电力大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，同意学校将学位论文的全部或部分内容编入有关数据库进行检索，允许论文被查阅和借阅，学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。本人授权华北电力大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。
本人知悉学位论文的使用权限，并将遵守有关规定。

本学位论文属于（请在相应方框内打“ √ ”）
☑公开 ☐内部 ☐秘密 ☐机密 ☐绝密

作者签名：  日期：2025 年 5 月 30 日

导师签名：  日期：2025 年 5 月 30 日

摘要

氢燃料电池是一种可以将化学能转化为电能的电化学装置，其具有能量转化率高、噪音低及零排放等优点，但是其较短的使用寿命制约着它的大规模商业化进程。因此对氢燃料电池进行健康状态评估，实现准确的寿命预测，对延长氢燃料电池的使用寿命具有重要意义。但目前大多数研究只关注单一氢燃料电池电堆的寿命预测，难以满足实际应用中对多个不同氢燃料电池电堆预测的需求。对此，本文提出一种基于相似性分类的氢燃料电池寿命预测通用模型。本文主要研究内容如下：

首先，本文选取了三个具有代表性的氢燃料电池寿命衰减数据集，分别是法国实验室的氢燃料电池寿命衰减数据集、国家基础学科公共科学数据中心数据集和同济大学长期动态耐久性测试数据集。随后，提取它们的公用变量进行数据预处理，通过箱线图分析、观察数据的分布，同时剔除异常值。接着，采用主成分分析（PCA）方法对处理后的数据进行降维，得到相似性图和两个主成分 PC1 和 PC2。这两个主成分累计解释了 87.11%的总方差。其中，PC1 代表最大方差方向，反映了数据中最主要的变化趋势；PC2 则是第二大方差方向，用于捕捉剩余的重要信息。进一步分析发现，PC1 主要反映了能量转换与输出相关因素，PC2 则侧重于热管理效率相关因素。最终，基于主成分分析的结果，将相似性图划分为三个区域：高温低压区、低温高压区和低热管理效率区，为后续研究提供了数据支持。

然后，基于划分的三个区域，选取电堆电压作为衰退指标，构建了预测模型。首先对数据进行重建和平滑处理，以减少数据量并提升数据质量。接着，建立长短时记忆神经网络模型（LSTM）对电堆电压进行预测，结果显示模型在训练集上的拟合效果非常好，决定系数 R^2 分别为 0.972、0.981 和 0.951。然而，在测试集上，尽管拟合效果仍然较好，但决定系数 R^2 下降至 0.932、0.920 和 0.947，其他的评价指标也随之变差。这表明模型在训练集上表现优异，但在测试集上的泛化能力需要改进。为提升模型在测试集上的泛化能力，在长短时记忆网络后面引入了 Kolmogorov-Arnold Networks（KAN）层，构建出 LSTM-KAN 模型。经过对比分析，LSTM-KAN 模型在训练集和测试集上的表现都更为出色，尤其是在转折点处，预测值与实际值贴合度更高了。具体来说，测试集上的决定系数 R^2 分别从 0.932、0.920、0.947 提升到了 0.959、0.953、0.955，其他的评价指标也得到显著的改善。这些结果表明，LSTM-KAN 模型相较于 LSTM 模型对寿命预测效果更好，因此将其作为基于相似性分类的预测通用模

型的预测部分模型。最后，为验证通用模型的适用性，使用生成对抗网络在三个原始数据集的基础上生成出三组新的数据集来进行验证。结果表明，基于相似性分类的氢燃料电池寿命预测通用模型能很好的预测多组氢燃料电池电堆老化数据。

最后，利用 Python 语言和 Streamlit 库设计了 GUI 界面，开发了氢燃料电池寿命预测软件。软件的主要功能包括箱线图分析、PCA 降维相似性分析、模型预测结果分析。随后通过输入测试数据验证了软件的可靠性，每个分析部分均可正常运行，相关图和结论都可以正常显示出来，可以实现对氢燃料电池的寿命预测。

关键词：氢燃料电池；相似性分类；寿命预测；GUI 设计

Abstract

Hydrogen fuel cells are an electrochemical device that can convert chemical energy into electrical energy. They have the advantages of high energy conversion rate, low noise, and zero emissions. However, its short service life restricts its large-scale commercialization process. Therefore, assessing the health status of hydrogen fuel cells and achieving accurate life prediction are of great significance for extending their operational lifespan. However, most current research only focuses on predicting the lifespan of a single hydrogen fuel cell stack, which is difficult to meet the demand for predicting multiple different hydrogen fuel cell stacks in practical applications. This article proposes a universal model for predicting the lifespan of hydrogen fuel cells based on similarity classification. The main research content of this article is as follows:

Firstly, this article selected three representative datasets of hydrogen fuel cell life decay, namely the French laboratory's hydrogen fuel cell life decay dataset, the National Basic Discipline Public Science Data Center dataset, and the Tongji University long-term dynamic durability testing dataset. Subsequently, extract their common variables for data preprocessing, analyze and observe the distribution of data through box plots, and remove outliers. Next, principal component analysis (PCA) was used to reduce the dimensionality of the processed data, which resulted in a similarity map and two principal components PC1 and PC2. These two principal components cumulatively explained 87.11% of the total variance. Among them, PC1 represents the direction of maximum variance, reflecting the main trend of change in the data; PC2 is the second largest variance direction, used to capture the remaining important information. Further analysis showed that PC1 mainly reflects factors related to energy conversion and output, while PC2 focuses on factors related to thermal management efficiency. Finally, based on the results of principal component analysis, the similarity map was divided into three regions: high temperature and low-pressure region, low temperature and high-pressure region, and low heat management efficiency region, providing data support for subsequent research.

Secondly, based on the three divided regions, the stack voltage was selected as the decay index to construct a prediction model. Firstly, reconstruct and smooth the data to reduce its volume and improve its quality. Next, a Long Short-Term Memory (LSTM) neural network model was established to predict the voltage of the fuel cell stack. The results showed that the model had a very good fitting effect on the training set, with determination coefficients R^2 of 0.972, 0.981, and 0.951,

respectively. However, on the test set, although the fitting effect was still good, the coefficient of determination R^2 decreased to 0.932, 0.920, and 0.947, and other evaluation indicators also deteriorated accordingly. This indicates that the model performs well on the training set, but its generalization ability on the testing set needs improvement. To improve the generalization ability of the model on the test set, a Kolmogorov-Arnold Networks (KAN) layer was introduced after the long short-term memory network to construct the LSTM-KAN model. After comparative analysis, the LSTM-KAN model performs better on both the training and testing sets, especially at turning points where the predicted values are more closely aligned with the actual values. Specifically, the coefficient of determination R^2 on the test set increased from 0.932, 0.920, and 0.947 to 0.959, 0.953, and 0.955, respectively, and other evaluation metrics also showed significant improvement. These results indicate that the LSTM-KAN model performs better in predicting lifespan compared to the LSTM model, therefore it is considered as the predictive part model of the general prediction model based on similarity classification. Finally, to validate the applicability of the universal model, a generative adversarial network was used to generate three new datasets based on the three original datasets for verification. The results indicate that the universal model for predicting the lifespan of hydrogen fuel cells based on similarity classification can effectively predict the aging data of multiple sets of hydrogen fuel cell stacks.

Finally, a GUI interface was designed using Python language and Streamlit library, and a hydrogen fuel cell life prediction software was developed. The main functions of the software include box plot analysis, PCA dimensionality reduction similarity analysis, and model prediction result analysis. Subsequently, the reliability of the software was verified by inputting test data, and each analysis section could run normally. Relevant graphs and conclusions could be displayed normally, enabling the prediction of the lifespan of hydrogen fuel cells.

Keywords: Hydrogen fuel cell; Similarity classification; lifespan prediction; GUI design

目录

第 1 章 绪 论

1.1 研究背景及意义

能源是二十一世纪非常关键的资源，对国家的经济发展和人民的生活质量有着重要的影响。随着我国经济的持续快速增长，对于能源的需要求量也在不断增加。目前，我国的能源结构仍以传统能源为主，这种结构带来了不少环境问题，如雾霾、全球变暖和酸雨等等。同时，也对国家能源安全也构成了威胁。为此，我国近年来不断加快能源结构调整的步伐，并在 2020 年提出了“双碳目标”，目的就是为了通过发展和应用清洁能源技术来推动能源转型。

交通运输是我国碳排放增长最快的领域，近几年全国交通运输碳排放平均增速保持在 5%以上。随着我国经济发展水平的不断提升，交通运输需求仍将保持较快增长态势，未来我国交通碳排放将面临上升压力[1]。氢能作为一种快速可再生、零排放的替代燃料，在众多新型清洁能源中备受瞩目，被认为是未来最有发展前途的新型清洁能源。发挥氢能在交通能源替代方面的重要作用，可助力交通领域实现深度脱碳。

燃料电池技术是氢能在交通领域的重要应用之一，氢燃料电池（Hydrogen Fuel Cell，HFC）是一种可以将氢气和空气中氧气反应的化学能转化成电能并存储起来的装置。它突破了传统热机效率的限制，能量转化率高，燃料来源广泛[2]。同时它的产物仅为水和少量的二氧化碳，这对助力交通领域深度脱碳，以及解决环境污染问题具有积极意义。

氢燃料电池虽然具有能量转化率高、运行噪音低且无污染排放等特点，但其在实际应用面临寿命短等问题，制约了其大规模的商业化进程[3-5]。事实上，人们普遍认为，氢燃料电池的寿命达到 2000 h 以上，才有望实现广泛应用，但是在交通运输领域，其寿命需求更高，需超过 6000 h[6]。美国能源部于 2007 年对此进行了评估，认为到 2015 的寿命应超过 5000 h[7]。然而不管是氢燃料电池的寿命，还是其它的性能，均受多种失效机制的影响[8, 9]。本田汽车的研究表明，氢燃料电池堆两两电池之间若存在温度、电压等参数不一致，将显著降低电池组整体性能，大幅缩短电池使用寿命[10]。若是无法有效解决这种不一致性问题，提高整体性能，其电池的耐久性将难以达到商业化应用的竞争力要求。此外，一些常见问题如膜干燥、水淹以及空气供应不足等，也会导致氢燃料电池性能衰退和使用寿命缩短[11]。相关研究表明，氢燃料电池若闲置超过一年，容易发生膜干燥现象[12]；若出现使用不纯的反应气体或缺乏适当的过滤器等情况，将可能导致氢燃料电池组的使用寿命急剧下降至数小时[13]。

因此，面对如此多复杂的影响氢燃料电池寿命的因素，如何延长氢燃料电池寿命技术及其应用成为了目前大家的研究热点与难点问题。特别是在国内，延长氢燃料电池寿命技术发展的相对滞后，已然成为阻碍我国氢燃料电池技术前行的主要瓶颈之一。要突破这一瓶颈，除了在材料、设计、制备等技术层面实现突破外，对于氢燃料电池的剩余寿命做出精确预测也成为氢燃料电池研究领域的一个重要方向[15, 16]。毕竟，在众多因素的影响下，能否实现精准的剩余寿命预测在很大程度上决定了氢燃料电池能否投入到实际应用中去。这是因为该预测不仅可以提前发现当前氢燃料电池存在的问题，而且可以提前让用户知道剩余使用时长，再根据时长来维修或更换器件，以此来提高氢燃料电池的寿命降低成本，为氢燃料电池技术的实际应用与推广提供坚实保障。

1.2 国内外研究现状

故障预测与健康管理（Prognostics and Health Management，PHM）是一种利用传感器采集系统关键数据，结合数据处理与智能算法评估设备健康状态，从而实现故障检查测、隔离、预测以及健康维护的技术[17]。PHM 能够在故障发生前准确的预判，并且根据可用的资源信息给出维修决策，其核心包括故障诊断、故障预测以及剩余使用寿命预测。本文主要围绕寿命预测方向展开研究。

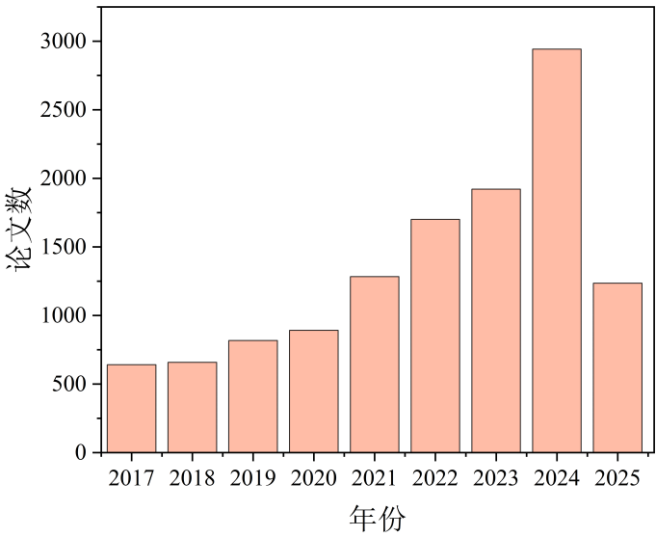


图 1-1 近十年文献发表量

通过调研发现，近几年“寿命预测”引起了学术界和工业界的高度关注相关研究论文的数量年均增长率保持在 15%以上，如图 1-1 所示。

目前，研究者们已经系统性的总结了氢燃料电池寿命预测的发展历程，分析了各类方法的优缺点，并对其未来的发展方向进行了展望[18-20]。同时，他们还提出了很多新的预测算法，并通过对比分析来提升模型的预测精度和可靠性[21]。部分研究者在此基础进一步深入，系统研究了燃料电池的测试条件、方法以及健康评

估指标，并对多种运行条件下的测试数据进行了对比分析[22]。值得注意的是，寿命预测技术已经被认为是氢燃料电池能否在汽车领域被推广的关键技术之一。当前常被用来评估燃料电池老化程度的指标有电堆电压、功率和内阻。通过实时监测这些参数，能帮助用户准确掌握燃料电池的老化情况，然后根据情况及时采取措施，延长使用寿命。他们的研究方法总结起来主要可以分为三类：基于模型的方法、基于数据驱动的方法及混合方法，如图 1-2 所示。

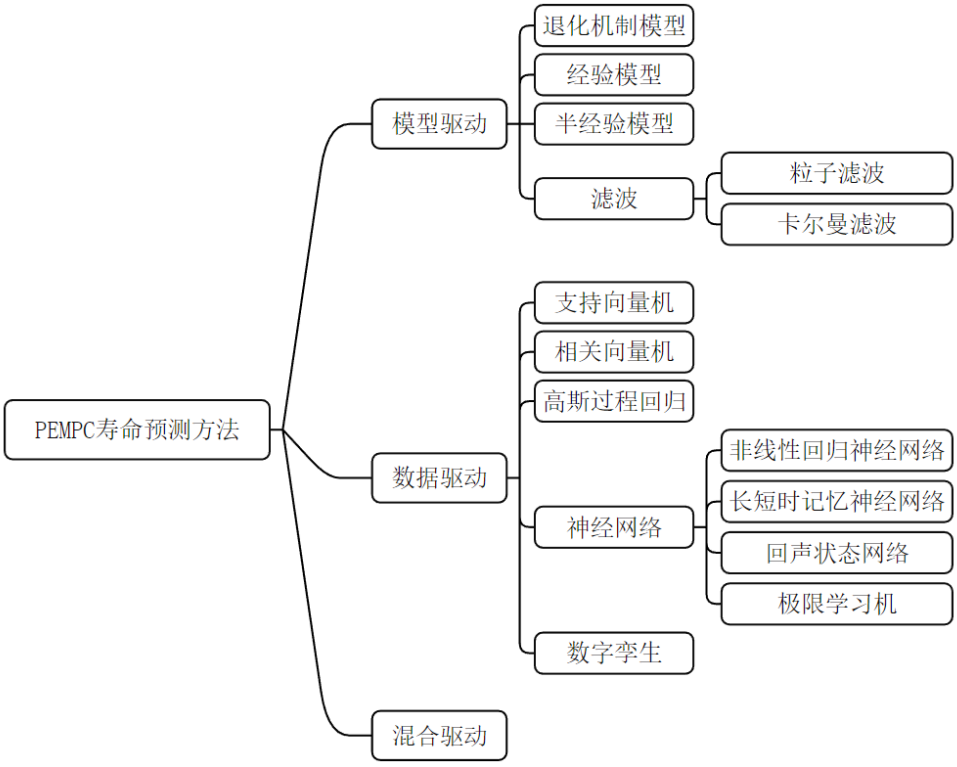


图 1-2 燃料电池寿命预测方法总结

1.2.1 模型驱动

基于模型驱动的研究方法是指通过建立数学模型来模拟燃料电池内部的反应机制，模型可以根据燃料电池结构或反应方式来建立，也可以依据经验来建立，其准确性取决于建立的模型与实际情况之间的契合度[23, 24]。模型驱动又分为衰退机理模型、经验模型和半经验模型[25]。

衰退机理模型是通过深入剖析燃料电池内部衰退过程及原理搭建起来的数学模型，这种模型对数据的依赖性小，但普适性强、准确性高。如 Robin 等[26]采用多尺度建模方法研究，将催化剂溶解模型与单体质子交换膜燃料电池（Proton Exchange Membrane Fuel Cell，PEMFC）的动态仿真模型耦合起来，得到了 Pt 颗粒粒子半径增加等效 PEMFC 运行期间 ECSA 的衰减，并且进行了两次在动态工况下 2000 h 的耐久性试验，验证了模型的准确性；Futter 等[27]通过建立了 PEMFC 的

二维数学模型用于模拟质子交换膜的化学降解过程，并通过加速应力试验验证了降解模型的正确性，最后认为该模型可以有效的预测质子交换膜在不同工况下的衰退速率。

燃料电池内部衰退机理的复杂性要求研究者能深入理解其内部的物理化学反应过程，因此构建出精确的衰退机理非常具有挑战性。而为了降低这一难度，大部分的研究者都会采用经验模型或者半经验模型来进行研究分析。如 Bressel 等[28]提出了一种基于模型的质子交换膜燃料电池预测算法，将降解模型与扩展卡尔曼滤波（EKF）相结合，用于估计电池健康状态和预测剩余使用寿命，并通过实验数据验证了该方法的有效性；Ou 等[29]提出一种基于 PEMFC 极化特性的半经验模型，该模型通过引入恢复因子，结合 ECSA 衰退模型和等效电阻模型，用来分析燃料电池的衰减规律。然后，他们通过两个不同运行条件下的试验，证明了该模型的准确性。然而，尽管采用经验或者半经验模型能够在一定程度上简化建模的过程，但是这种方式预测的精度很低，适用差，同时对于实验数据的依赖性很高。

1.2.2 数据驱动

基于数据驱动的方法是指通过统计学和数据挖掘等方法，从现有的燃料电池衰减数据中学习燃料电池的老化行为，然后利用学习出来的结果对电池的健康状态进行监测或者寿命预测。这种方法不需要研究者深入理解其内部的物理化学反应过程，而是通过对大量的衰减试验数据学习，往往就能达很好的预测效果。正因如此，数据驱动方法已被广泛应用，成为燃料电池寿命预测的首选方法。

Chen 等[30]提出了一种结合粒子群优化（PSO）、移动窗口法的灰色网络模型（GNNM），用于对 PEMFC 在不同工况下的老化情况进行预测，并通过 3 个不同的老化实验进行了验证，同时也发现 PSO-GNNM 在数据量较少的预测中表现更为出色；Jouin 等[31]提出一种混合预测的方法，用于预测恒定电流工况下质子交换膜燃料电池堆功率老化情况。当数据充足时，该方法可以实现准确预测，剩余有用寿命估计值的误差小于 5%，这足以用于决策；Vichard 等[32]提出了一种基于回声状态网络的 PEMFC 寿命预测方法，研究结果发现电堆电压的大小和环境温度有一定的关系，并且证明了该方法在不同的温度下，都能较为准确的预测出系统老化的情况。

随着研究的不断深入与技术的持续发展，机器学习领域的神经网络技术逐渐在质子交换膜燃料电池老化预测中崭露头角。学者们开始着眼于利用神经网络强大的非线性映射能力和自学习能力，挖掘复杂运行数据背后隐藏的规律，进而实现更为精准的寿命预测。在这样的研究趋势下，Laghrouche 等[33]基于 ANN 神经网络构建了一个寿命预测模型，输入变量为电堆电流、堆温度、空气压力、氢气压

力和空气湿度。对数据集的前 400 个小时的数据学习之后，模型预测的相对误差小于 5%。

Liu 等[34]提出一种长短时记忆循环神经网络（Long short-Term Memory，LSTM）的方法来解决燃料电池的寿命预测问题，他们使用该方法学习了 1154 小时燃料电池老化数据之后，发现使用 LSTM 进行预测的精度能够高达 99.23%。这个预测的精度已经非常的高了，但是在质子交换膜电池老化预测中，神经网络技术仍面临诸多挑战。因为不同的运行工况和环境因素都会使燃料电池的老化机制更加复杂，但是单一的神经网络模型很难捕捉到这些变化。

为解决单一神经网络的局限性，研究人员持续探索更有效的神经网络架构。发现在序列数据预测领域，基于 LSTM 网络模型和其变体因其能够有效处理时间序列中的长期依赖关系，逐渐成为研究热点，得到了广泛应用[35-37]。

Wang 等[38]提出一种堆叠长短时记忆（S-LSTM）模型，用于预测动态条件下燃料电池的短期老化趋势。该模型就是 LSTM 模型的变体，它在原模型的基础上加入了差分进化算法进行优化参数，让其精度大概提高了 20%；Ma 等[39]提出一种基于长短时记忆网络的 G-LSTM 循环神经网络深度学习模型，用于质子交换膜燃料电池老化预测，用记录下 1 kW、1.2 kW 和 25 kW 三种工况下的老化数据进行训练和验证，发现预测精度比较理想，均方根误差为 0.0040，平均绝对百分比误差为 0.0013，优于 RVM 等方法。

上述的数据驱动方法都没有考虑燃料电池内部复杂的物理和化学反应过程，而数字孪生作为一种 AI 技术，它可以持续的学习燃料电池内部复杂变化来不断完善自己。所以它能快速的适应燃料电池环境和运行条件的变化，非常适合用于在线预测，具有很好的应用前景[40]。

1.2.3 混合驱动

模型驱动的方法通常预测精度都很高，但是燃料电池内部衰退机理的复杂，难以深入了解，极具挑战性；而数据驱动的方法则不需要考虑复杂的反应机理，但是却非常依赖用来学习历史数据的质量。因此，一些研究者尝试着将这两种方法的优点结合起来，提出了一种混合驱动的方法[41]。该方法通常先将实验数据放入模型中进行处理，让其更贴近实际数据，然后再利用数据驱动的方法对处理后的数据进行学习并预测。如 Xie 等[42]就提出一种融合了模型驱动和数据驱动的燃料电池寿命预测的方法。他们先采用奇异谱分析（SSA）对老化数据进行处理，然后利用深度高斯过程（DGP）对处理后的老化进行学习，同时提供置信区间来保证预测结果的可靠性。通过验证发现该方法能够很好预测出燃料电池的剩余使用寿命；Wang 等[43]使用退化行为模型和滑动窗口模型识别方法提取燃料电池的健康指

标。随后，使用基于符号的长短时记忆网络来预测健康指标退化趋势并估计剩余寿命。该方法提供了接近燃料电池生命周期 50%的预测范围，估计的剩余使用寿命的平均相对准确性接近 90%。

对于当前阶段的混合驱动方法来说，虽然结合了另外两种方法的优势，但是依然要求研究者对燃料电池内部的物理化学反应过程有较深理解，同时计算量也偏大。最主要的问题就是它对于燃料电池模型的构建依赖性很大。所以，混合驱动方法目前的应用得并不多[44]。

1.2.4 目前研究方法的局限性

前文主要介绍了目前国内外氢燃料电池寿命预测领域的三种研究方法的现状。模型驱动的方法是通过建立数据模型来模拟燃料电池内部复杂的物理和化学反应，其中衰退机理模型虽然精度很高但是建模难度很大，而经验模型和半经验模型虽简化了建模的难度但是精度很低；数据驱动的方法则是借助数据挖掘和统计方法从老化数据学习衰退行为，是当前研究的主流方法。在数据驱动方法中，神经网络技术因其在时间序列分析中的强大能力而被广泛应用，尤其是LSTM及其变体。然而，神经网络技术在面对不同的电堆老化数据时，预测精度非常的不稳定。混合驱动的方法虽然结合了两者的优点，但是仍需要考虑复杂的反应机理。

尽管现有的方法在氢燃料电池寿命预测和研究中得到了广泛应用，但目前的研究都存在一个普遍的局限性，即大多数研究只关注单一氢燃料电池电堆的寿命预测，而缺乏对氢燃料电池寿命预测通用模型的探讨和研究。因此，现在没有一个通用的模型能够同时预测很多不同电堆的老化数据。然而，在实际应用场景中，每个氢燃料电池电堆的情况都是不一样的，相关企业考虑到成本问题是不可能对每个电堆都构建出一个预测模型。因此，开发出一个适用于所有燃料电池电堆的通用模型是非常必要的。

1.3 本文研究内容

为了解决现有的模型在处理多个电堆老化数据时的局限性，本文提出了一种基于相似性分类的氢燃料电池寿命预测的通用模型构建方法。该方法的核心思想是通过降维技术将多组数据映射至低维空间，形成相似性图。然后通过分析数据之间的相似性和差异性对相似性图进行区域划分。随后，对划分的区域分别进行预测模型构建。当面对新的数据集时，该模型会先用降维技术将数据映射到相似性图上，快速判断它属于哪个区域，同时调用对应区域的模型进行预测。本文的研究思路如图 1-3 所示。

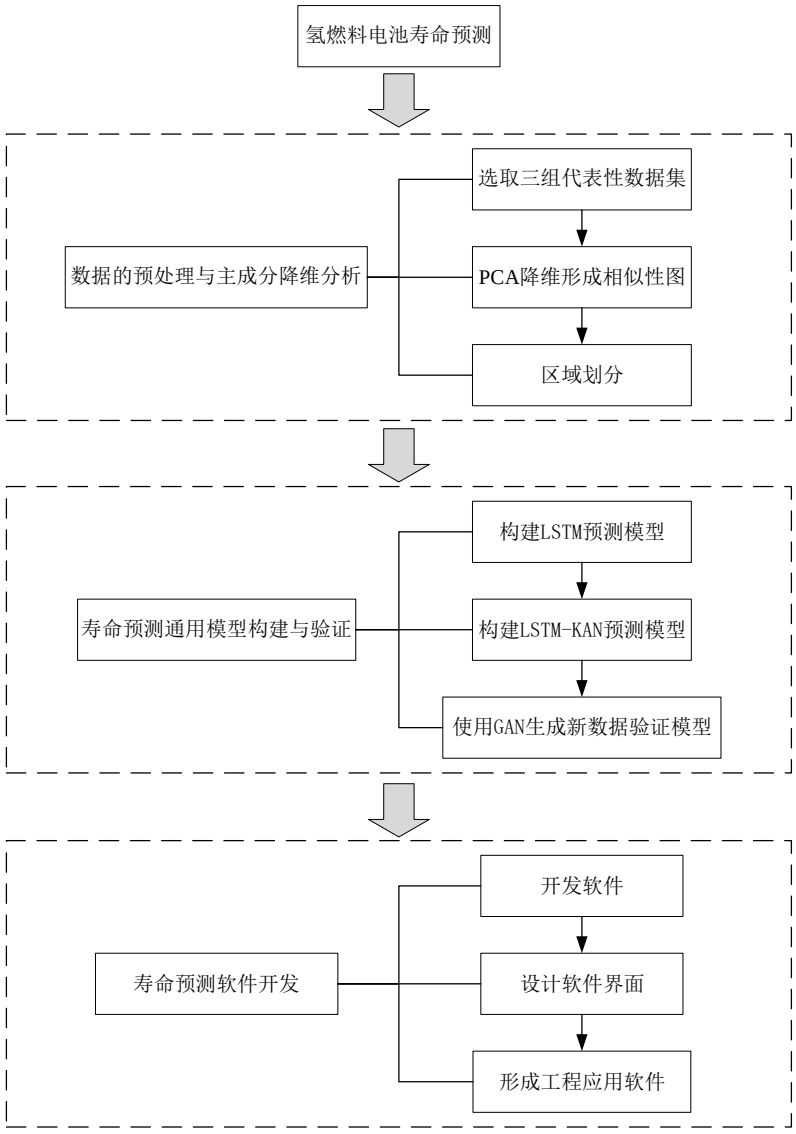


图 1-3 本文研究思路

本文具体章节安排如下：

（1）第一章为绪论。首先介绍了本文研究主体的背景信息以及研究的意义，然后分析目前燃料电池寿命预测的国内外研究成果，指出了现有研究的局限性，进而引出本文的研究工作以及意义。最后对本文的具体章节安排以及研究思路进行总结；

（2）第二章为氢燃料电池寿命预测的技术方法与模型构建。本章主要介绍了进行模型训练的三组数据集以及对数据集进行降维时采用的方法。然后介绍用于训练的模型 LSTM 和 KAN（Kolmogorov-Arnold Networks），以及相关评价指标。随后介绍用于新数据生成的生成对抗网络（Generative Adversarial Network，GAN），最后介绍用于软件开发相关内容。

（3）第三章为数据预处理与主成分降维分析。本章首先提取三个不同的氢燃

料电池电堆老化数据集的公共变量，对数据集进行箱线图分析，得到每个数据集的数据分布情况，同时剔除掉异常值。之后，使用主成分分析法（Principal Components Analysis，PCA）对三个数据集进行降维形成相似性图。最后分析两个主成分 PC1 和 PC2 的组成，明确两个降维的方向，并对相似性图进行区域划分。

（4）第四章为基于相似性分类的氢燃料电池寿命预测模型。本章主要在第三章基础上，对数据进行重建和平滑处理，减少数据量提升数据质量。然后搭建 LSTM 模型，预测氢燃料电池寿命，并分析预测效果。之后为改进效果搭建 LSTM-KAN 模型，预测寿命，并与 LSTM 模型的预测效果进行对比，验证了 LSTM-KAN 模型的可行性。最后，使用 GAN 生成的三组新数据，来验证基于相似性分类的氢燃料电池寿命预测的通用模型的可行性。

（5）第五章为氢燃料寿命预测软件开发。本章主要通过 Python 语言实现氢燃料电池数据分析以及结果数据，设计软件界面，完成氢燃料寿命预测软件开发。

（6）第六章为结论与展望。本章对全文的工作以及主要结论进行总结，并且指出本文不足之处以及提出改进方向。

第 2 章 氢燃料电池寿命预测的技术方法与模型构建

本文提出了一种基于相似性分类的氢燃料电池寿命预测的通用模型构建方法，主要是通过对多组数据进行降维、相似图区域划分和分别训练模型，来实现通用模型的构建。作为本文研究的基础章节，本章主要介绍数据集、训练模型以及评价指标的选用，以及用于新数据的生成和应用软件开发的工具，以期为氢燃料电池寿命预测领域带来新的突破和发展。

2.1 数据集介绍

本文提出了基于相似性分类的氢燃料电池寿命预测的通用模型构建方法，该方法需对多组数据进行降维，映射到低维空间形成相似性图，然后通过分析数据之间的差异对相似性图进行区域划分，之后再对不同区域分别进行模型训练。所以，本文的研究必须依赖多个氢燃料电池数据集。经过查阅相关文献，并对文献中公开的氢燃料电池研究数据库进行研究分析，最终确定了三个在氢燃料电池领域常用且具有代表性的数据集作为本文的研究对象，分别是法国实验室的氢燃料电池寿命衰减数据集、国家基础学科公共科学数据中心数据集、同济大学长期动态耐久性测试数据集，并分别命名为 FC1 数据集、FC2 数据集、FC3 数据集。

2.1.1 FC1 数据集

FC1 数据集是由法国燃料电池实验室提供的数据（FRCNRS3539，FCLAB，http://eng.fclab.fr/），该数据是 1 kW 的氢燃料电池在稳定环境下的运行数据，测试出的数据包含两个部分：一部分是在静态条件下采集的，而另一部分则是在动态条件下采集的[45]。本文采用其中动态条件下获取的数据，该数据包含多个测量参数，具体包括 5 个子电池的电压以及堆体的温度、电压、电流，还有反应气体进出口的温度、压力和湿度数等等，实验中监测参数如表 2-1 所示。

表 2-1 FC1 数据集监测参数

实验中监测参数	物理意义
Time	寿命时间（h）
U1-U5；Utot	单节电池和电堆电压（V）
I；J	电流（A）；电流密度（A/cm^2）
TinH2；ToutH2	进/出口氢气温度（℃）
TinAIR；ToutAIR	进/出口空气温度（℃）

接下表

续表 2-1

实验中监测参数	物理意义
TinWAT；ToutWAT	进/出口冷却水温度（℃）
PinH2；PoutH2	进/出口氢气压力（mbar）
PinAIR；PoutAIR	进/出口空气压力（mbar）
DinH2；DoutH2	进/出口氢气流量（L/mn）
DinAIR；DoutAIR	进/出口空气流量（L/mn）
DWAT	循环冷却水流量（L/mn）
HrAIRFC	电堆入口空气相对湿度（%）

2.1.2 FC2 数据集

FC2 数据集是由国家基础学科公共科学数据中心数据提供的 65 kW 两个电堆的燃料电池客车在江苏省张家港市的运行数据[46]，该客车搭载的氢燃料电池电堆是由两个相同的电堆串联而成的双堆，一共有 208 节电池[47]。该数据集包含了寿命预测的原始数据和相关的分析数据。原始数据主要包含了氢燃料电池电堆的采样数据（例如温度、压力、流量、电流、电压等信息）和操作时间的统计。分析数据主要包含了在学习阶段的老化参数变化曲线以及所预测的电堆平均电压等相关数据。本文采用该数据集的原始数据用于研究，其相关监测参数如表 2-2 所示。

表 2-2 FC2 数据集监测参数

实验中监测参数	实验中监测参数
时间	比例阀占空比
工作状态	A 堆氢气进堆压力
电池故障等级	B 堆氢气出堆压力
电池故障代码	空气流量
DCDC 输出电压	B 堆空气进堆压力
DCDC 输出电流	水进堆温度
DCDC 输出功率	水出 A 堆温度
燃料电池电压	B 堆水出堆温度
燃料电池电流	水电导率
氢燃料电池输出功率	A 堆氢气进堆温度
A 堆电压平均值	Purge 标志位
A 堆最低电压	氢气出 B 堆温度

接下表

续表 2-2

实验中监测参数	实验中监测参数
A 堆最低电压位置	B 堆空气进堆温度
B 堆电压平均值	电堆氢泄露
B 堆最低电压	环境温度
B 堆最低电压位置	密封腔体温度

2.1.3 FC3 数据集

FC3 数据集是同济大学提供的在 G20 实验平台进行的长期动态耐久性测试数据[48]。值得一提的是，该实验的动态负载循环是根据欧盟汽车质子交换膜燃料电池测试协议而设计的[49]，模拟了车辆的城市和郊区工况。该数据涉及 9 个电流负载，包括 0 A、1.76 A、4.42 A、9.48 A、10.37 A、14.81 A、20.7 A、29.58 A、35.53 A[50]。所测量的物理参数包括工作电流、输出电压、功率、以及各种反应物温度、阴阳极入（出）口压力以及入（出）口流速等，所测量的物理参数如表 2-3 所示。

表 2-3 FC3 数据集监测参数

实验中监测参数	实验中监测参数
秒数	阳极入口温度
小时数	阴极入口温度
循环位置	阳极出口温度
电流	阴极出口温度
电压	阳极氢气体积流量
功率	阴极氧气体积流量
阳极入口压力	阳极极板温度
阴极入口压力	阴极极板温度
阳极出口压力	阳极露水温度
阴极出口压力	阴极露水温度

2.2 PCA 数据降维

本文提出的模型需要将不同的数据集降维到一个相似性图中，因此需要选择合适的降维工具来保证降维结果的合理性。通过分析后采用了 PCA 进行降维[51]，它理解起来简单、计算简单方便，并且重构之后的数据表现优异、误差低[52]。因此，它成为了一种最常用、最基本的特征降维方法。在工程实际中，为了全面分析一个多因素影响问题，常常列出很多与输出有关的变量，虽然每个变量的影响

程度不同，但它们之间一般存在关联。较多的变量构成的高维数据组会使建模问题变得更加复杂。主成分分析方法通过坐标变换，可以找出各个变量的公共因素——主成分，最后得到的主成分之间相关性低，信息交叉率低，即数据冗余较少。

该方法的核心思想是，将原本高维度且信息复杂的数据进行映射到低维度的空间中，这样就能够达到提取到原本数据中最重要的元素的目的。同时得到这些变量就是原来数据中变量的线性组合，巧妙的解除了原本变量之间的相关性。

在主成分分析中，首先需要对样本进行采集，假设采集的矩阵是一个$n\times m$维的，即有 m 个变量，每个变量有 n 个数据值。然后就需要对此矩阵进行标准化处理，将每个变量的均值变为 0，同时要满足方差为 1。进行标准化之后，将原矩阵变成一个$n\times m$维的训练数据样本 $X_{n\times m}$。

随后对训练矩阵 $X_{n\times m}$ 进行分解，将其分解成 m 个向量外积的和，即：

$$X=t_1p_1^T+t_2p_2^T+\cdots+t_mp_m^T \tag{2-1}$$

式中，$t_i\in R^n$——得分向量，也可称为主成分量；

$p_i\in R^n$——负载变量。

可定义$T=[t_1t_2\cdots t_m]$为得分矩阵，$P=[p_1p_2\cdots p_m]$为负载矩阵。则式 (2-1)可以写成：

$$X=TP^T \tag{2-2}$$

每个得分向量是彼此相交的，即当$i\neq j$时则满足$t_i^Tt_j=0$。而负载向量不仅保持正交关系，还均为单位向量，即：

$$p_i^Tp_j=0\left(i\neq j\right) \tag{2-3}$$

$$p_i^Tp_j=1\left(i=j\right) \tag{2-4}$$

此时，对式 (2-1)两边同时右乘p_i，则可得：

$$Xp_i=t_1p_1^Tp_i+t_2p_2^Tp_i+\cdots+t_mp_m^Tp_i \tag{2-5}$$

随后，将式 (2-3)和式 (2-4)同时带入式 (2-5)，得：

$$t_i=Xp_i \tag{2-6}$$

式 (2-6)表明每个得分向量t_i是通过初始矩阵X在负载向量p_i方向上的投影，这个投影的长度就反映了初始矩阵X在该负载方向上的覆盖范围。那么如果t_i的向量长度越大，X在负载向量上的投影范围就越大。若将得分向量t_i按照向量长度排列：

$$\|t_1\|>\|t_2\|>\cdots>\|t_m\| \tag{2-7}$$

若如式 (2-7)，X的最大变化方向则有负载变量p_1表示，p_2则是X的第二大

变化方向，并且此向量在空间上会与 p_1 垂直。那么同理，随着负载向量数的增多其投影将会越来越少，m 如果足够大，则趋近于 p_m 的负载向量甚至可以忽略不计。若前 $r(r<m)$ 个主成分的方差贡献率超过经验阈值 0.85，则可把这 r 个主成分当作训练样本的代表，即原有的 m 维空间降维到 r 维。

训练数据样本 X 则分为主成分子空间和残差子空间的两部分之和，主成分子空间则是由前 r 个主成分的线性表示，剩余的 $m-r$ 个主成分则可线性表示出残差子空间。定义 $\hat{X}$ 为 X 的主成分子空间，E 为其残差子空间。于是，训练样本空间 X 则可表示为：

$$X=\hat{X}+E \tag{2-8}$$

$$\hat{X}=T_rP_r^T=\sum_{i=1}^{r}T_iP_i^T \tag{2-9}$$

$$E=T_eP_e^T=\sum_{i=r+1}^{r}T_iP_i^T \tag{2-10}$$

由式 (2-8)、(2-9)和 (2-10)可以确定样本的主成分子空间 $\hat{X}$ 以及残差子空间 E，从而实现对训练样本空间 X 的降维，用着两个空间的和就可以描述出原训练样本空间。

2.3 寿命预测模型及评价指标

寿命预测最主要的就是预测模型，完成数据降维以及区域划分之后，就需要对各个区域进行模型构建，为了评价模型的训练效果就得引入评价指标。本小节就介绍了本文选择的模型以及各种评价指标。

2.3.1 LSTM 模型

长短时记忆循环网络（Long Short-Term Memory，简称 LSTM）是由 Hochreiter 和 Schmidhuber 提出，他们引入了细胞的概念。具体而言，引入了“门”这个结构。通过“门”这个结构的信息会随着时间传播不断的保留或遗忘，以此来解决深度递归神经网络（RNN）中的梯度消失的问题[53]。它的原理其实非常的简单，就是通过控制“门”的开度大小，来对数据信息进行保存或者删除。这种方法有效的克服了 RNN 处理长序列数据时遇到的梯度消失、梯度爆炸以及长期记忆不足的问题，使得 LSTM 在长时间序列领域更具优势。这种优势使得 LSTM 被广泛应用到不同领域的时间序列的数据研究中，其中就包括道路交通和交通流量的相关预测[54]。因此对于氢燃料电池寿命这种长时间性能衰减数据的预测，LSTM 是一个很不错的选择。

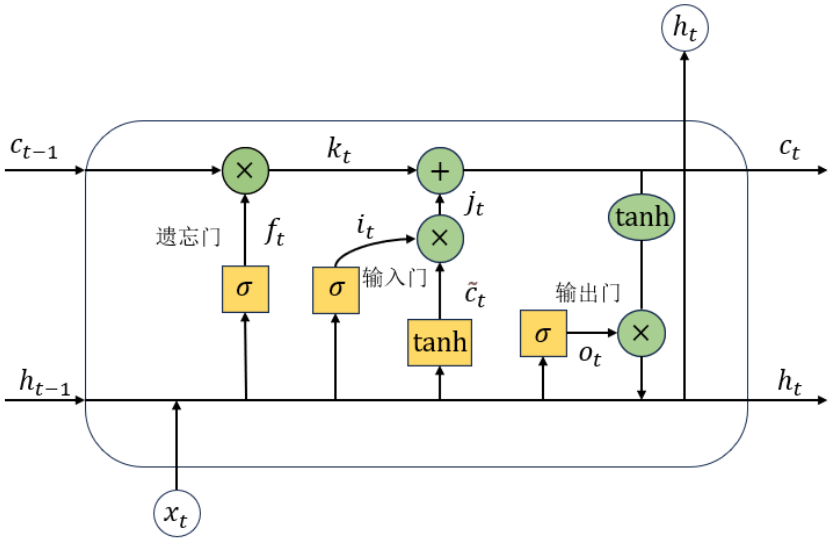


图 2-1 LSTM 网络结构

LSTM 的网络结构如图 2-1 所示，它主要由三个门来控制：遗忘门、输入门以及输出门。其中遗忘门主要用来控制数据中的哪些信息需要被遗忘掉，其计算公式如下：

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{2-11}$$

式中，σ 为激活函数 sigmoid；W_f 则是遗忘门中的权重矩阵；h_{t-1} 为上一个结构通过处理后的输出信息；x_t 为当前的输入；b_f 为遗忘门的偏置。

输入门决定了哪些信息应该在细胞中保留并更新。因此，输入门的公式应该分为两个部分，一部分决定保留哪些信息，一部分创建更新后的状态，如公式 (2-12)和 (2-13)所示。

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2-12}$$

$$e_t = \tanh\left(W_e \cdot \left[h_{t-1}, x_t\right] + b_e\right) \tag{2-13}$$

式中，W_i 和 W_e 为 i_t 和 e_t 的权重矩阵；b_i 和 b_e 为 i_t 和 e_t 的输入门偏置。

结合式 (2-12)和式 (2-13)，可以推导出更新细胞状态的表达式，如式 (2-14)。其中，C_t 为更新后的细胞状态。

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{2-14}$$

输入门则用来决定什么信息将要被输出到下一时刻，这一过程主要用到激活函数 sigmoid，然后使用 tanh 函数对这些信息进行处理。输出门的结果可以由下面两个式子得到：

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{2-15}$$

$$h_t = o_t \times \tanh\left(C_t\right) \tag{2-16}$$

最终，当前时刻的输出预测值如式 (2-17)所示。

$$\hat{y}_t = \sigma\left(Uh_t + b_y\right) \tag{2-17}$$

式中，$\hat{y}_t$ 输出的预测值；U 为输出的转换矩阵；b_y 为输出的偏置。

通过上述过程，LSTM 可以通过不同的方式对数据信息进行丢弃、维护以及更新，确保了在长序列训练过程中信息能够有效的传递。

2.3.2 KAN 模型

KAN 是一种基于 Kolmogorov-Arnold 表示定理的新型神经网络架构[55]。传统多层感知机（Multi-layer perceptron，MLP）模型中，通常将权重参数设置在网络的边缘，而神经元上则放置固定不动的激活函数。与 MLP 不同的是，KAN 模型将可学习的激活函数放置于网络边缘，使其随着网络训练自动调整，让网络权重参数被单变量函数替代[56]。这样能在保持网络灵活性的同时，能以较少参数实现对复杂网络的拟合。即将传统的“可学习权重+固定激活函数”的模型转变为“可学习激活函数+函数求和”的模型，所以 MLP 和 KAN 的网络结构图如 2-2 所示。

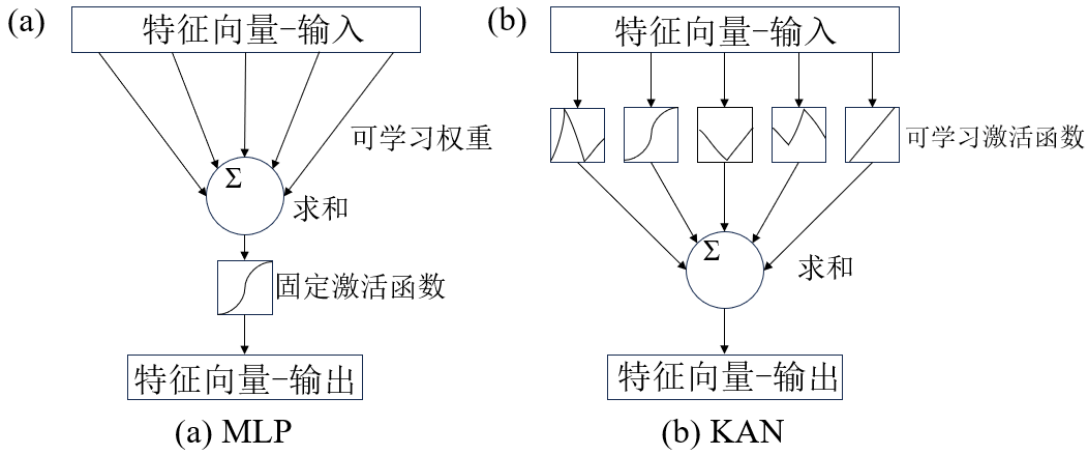


图 2-2 MLP 与 KAN 网络结构

KAN 模型是由外部函数和内部函数所构成的，如式 (2-18)所示。

$$f\left(x\right) = f\left(x_1,\cdots,x_n\right) = \sum_{q=1}^{2n+1}\phi_q\left(\sum_{p=1}^{n}\varphi_{q,p}\left(x_p\right)\right) \tag{2-18}$$

其中，x 为 n 维输入向量，$\varphi_{q,p}$ 表示内部函数定义域通常为[0,1]，其值域为实数集 R；$\varphi_{q,p}\left(x_p\right)$ 表示可训练的激活函数；ϕ_q 表示外部函数，其定义域和值域都为 R，用于内部函数 $\varphi_{q,p}$ 的输出组成的求和。

在 Kolmogorov-Arnold 定理中，内部函数形成了 n 输入和 $2n+1$ 的输出 KAN 层，内部函数则形成 $2n+1$ 输入和 n 输出的 KAN 层，因此本质上 $f\left(x\right)$ 是由两个 KAN 组成。为了使 KAN 更易于优化，Liu 等[56]提出了一种残差激活策略，通过引入一

种基函数$b(x)$，使可训练激活函数$\phi(x)$是由基函数$b(x)$和样条函数$spline(x)$的和：

$$\phi(x)=w_b b(x)+w_s spline(x) \tag{2-19}$$

式中，w_b和w_s实质上是冗余的，能够融入$b(x)$和$spline(x)$函数中。Liu 等人将基函数$b(x)$设置为 SiLU 函数，同时将样条函数$spline(x)$参数化为 B 样条曲线的线性组合：

$$b(x)=SiLU(x)=\frac{x}{1-e^{-x}} \tag{2-20}$$

$$spline(x)=\sum_i c_i B_i(x) \tag{2-21}$$

式中c_i表示可训练的参数，在训练过程中c_i可以持续优化，以调整样条形状，从而拟合训练数据。

2.3.3 评价指标

为了衡量模型的训练效果，本文选用了决定系数（R^2）、均方误差（MSE）、均方根误差（RMSE）、平均绝对误差（MAE）和平均绝对百分比误差（MAPE）这五个评价指标。

R^2 主要用来评估预测模型对于实际数据的预测效果。它在一定程度上能够反映模型对数据的学习程度，R^2 越接近于 1，则说明模型对于数据的学习程度就越好，其公式如下：

$$R^2=1-\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i-y_i)^2} \tag{2-22}$$

MSE 主要反映预测的结果和实际数据之间误差平方的平均值。这个值能够很好的反映预测的结果与实际值的偏差程度。RMSE 则是对误差平方后再开方，所以 RMSE 对于误差大小更为敏感。这两个评价指标都是用来评价预测值和实际值之间差值大小的指标，因此它们的值越小，越能说明模型的预测效果好。公式如下：

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{2-23}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{2-24}$$

MAE 主要反映预测值与真实值之间绝对误差的平均值，直接反映两个值之间的平均差异程度，能够更加直观体现他们之间偏离的平均幅度。同样也是值越小，说明模型的预测效果越好，其公式如下：

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i-y_i\right| \tag{2-25}$$

MAPE 以百分比的形式衡量预测值与实际值之间的平均误差，能更清晰的反映两者之间的误差比例，便于不同量级数据之间的比较。其值越小，模型的预测效果越好，公式如下：

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|\times 100\% \tag{2-26}$$

在上述公式中，n 是样本数量；y_i 是实际电压值；$\hat{y}_i$ 是预测电压值；$\bar{y}_i$ 是 n 个样本上的平均电压。

2.4 生成对抗网络

本文提出了一种基于相似性分类的氢燃料电池寿命预测通用模型，当完成了对三组氢燃料电池老化数据的降维、绘制出了相似图、根据数据的差异性和相似性划分出了区域，训练出对应区域的模型这些工作之后，就需要引入新的数据来验证该通用模型的正确性以及可靠性。但是考虑到获取真实且具有代表性的新数据条件有限，因此本文决定采用 GAN 这一先进的数据生成技术来生成新数据。

GAN 是由一个生成器和一个判别器组成[57]。生成器是用来根据现有数据尝试合成出新数据的工具，而判别器则是用来尝试将新数据与现有的真实数据区分的工具。GAN 使用方向传播算法和 Dropout 算法[58]来训练两个模型，其目的是为了让网络随着这两个算法的不断对抗而优化，这样网络生成的新数据质量就会变得更优，最终就会形成一个能够产生逼真数据的生成网络。

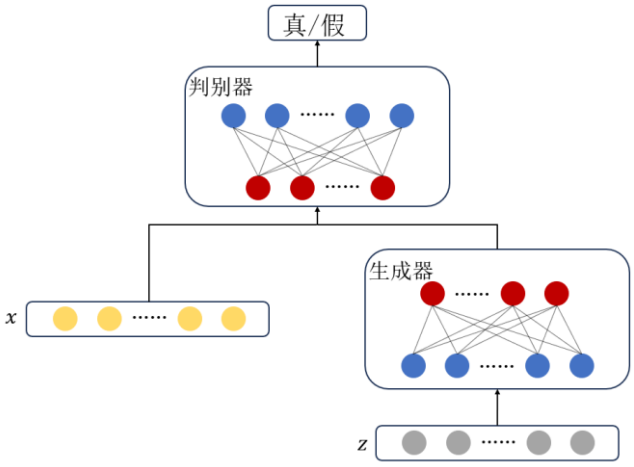


图 2-3 GAN 基本网络结构

基本的 GAN 网络结构如图 2-3 所示。从结构图中可以知道，生成器是利用随机噪声 z 生成一个类似于真实数据的样本。然后将样本和真实数据进行混合输入到判别器中，判别器则判断该样本的真实性。如果生成数据接近真实数据，则将其判别为“真”，反之则判别为“假”。

GAN 的目标函数的表达式如式 (2-27)所示：

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log\left(D(x)\right)\right] + E_{z\sim P_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \quad (2\text{-}27)$$

上述公式包含两个优化，可拆解为两部分：

优化 D，见式 (2-28)：

$$\max_D V(D,G) = E_{x\sim P_{data}(x)}\left[\log\left(D(x)\right)\right] + E_{z\sim P_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \quad (2\text{-}28)$$

当优化判别器 D 时，是不会对生成器进行操作的，式中 $G(z)$ 是生成器生成的新样本。对于式子的前半部分来说，由于输入是真实样本 x，那么输出则是越接近于 1 越好；而对于式子的后半部分，输入值是新样本 $G(z)$，那么输出值是越接近于 0 越好。

优化 G，见式 (2-29)：

$$\max_D V(D,G) = E_{z\sim P_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \quad (2\text{-}29)$$

同样当优化生成器 G 时，是不会对判别器进行操作的。相反的是该式子希望生成的样本 $G(z)$ 的输出值是 1，因此 $D(G(z))$ 越大越好。

根据目标函数可以得到判别器的损失函数如式 (2-30)：

$$-\left((1-y)\log\left(1-D\left(G(z)\right)\right) + y\log D(x)\right) \quad (2\text{-}30)$$

生成器的损失函数如式 (2-31)：

$$(1-y)\log\left(1-D\left(G(z)\right)\right) \quad (2\text{-}31)$$

总的来说，对于判别器 D 来说，需要解决的是一个二分类问题，即将数据分为真或假。而对于生成器 G 来说，需要做的就是不断调整生成数据 $G(z)$ 骗过判别器，让其认为它是真实数据。尽管该方法可以生成出很接近真实数据的样本，但是实际训练过程中是很难达到非合作博弈均衡的。所以它具有很强的不稳定性，也容易让数据梯度消失，同时也不能处理离散式数据。

2.5 软件开发

本文主要利用 Python 中的 streamlit 库设计寿命预测软件界面，实现寿命预测的工程化应用。上世纪末，Python 横空出世，经过这些年的发展，已经成为了目

前最流行和最受欢迎的编程语言。其中最主要的原因就是它的语言简洁明了、易上手、并且拥有很多的标准库，同时可以拓展很多第三方库。这些优点也让它在数据分析领域有着非常广泛的应用[59, 60]。利用 Python 进行数据分析主要依赖 NumPy 库和 Pandas 库，而实现数据结果分析的主要可视化主要借助 Matplotlib 库，最后通过 streamlit 库来进行交互界面的搭建[61]。下面对这四个库进行简单的介绍。

（1）NumPy 库

NumPy 库是 Python 的基础包之一，能够对任意一种类型的数据进行处理，同时仅占用很小内容就能实现高效运算。特别是对于高维数组，可以在整个数组上进行计算而不需通过循环，处理起来非常方便。

（2）Pandas 库

Pandas 是 Python 中主要用来读取、保存数据以及数据类型转换的一个标准库。它可以使用内置函数帮助用户读取各种格式的数据，然后可以根据读入的数据类型，采用对应的索引调用数据中的任意一行或一列数。处理分析数据之后，可以将结果保存成用户想要的文件类型，非常的高效、简单。

（3）Matplotlib 库

Matplotlib 是 Python 中主要用来进行数据分析结果绘图的一个标准库，它能让分析的结果更加直观的展现出来，方便用户进行分析，发现其中蕴藏的规律。它的功能非常强大，可以根据用户的需求绘制不同的图，基本可以满足大部分用于的需求。

（4）Streamlit 库

Streamlit 是一个开源 Python 库，专为数据科学家和机器学习工程师设计，能够快速将数据脚本转换为交互式 Web 应用程序。它通过少量代码即可实现传统复杂 Web 开发的功能，帮助用户快速构建高效的数据可视化工具[62]。

使用前三个库实现软件所需要功能后，就可以在原代码中接入 Streamlit 的代码，加入 Streamlit 的控件，同时将这些控件和相应功能代码链接在一起，形成一个完整的可使用的软件。

2.6 本章小结

本章主要介绍了本文提出的基于相似性分类的氢燃料电池寿命预测通用模型所采用的三个代表性数据集以及用于数据降维的主成分分析法，并在建模方面采用了 LSTM 和 KAN 神经网络，同时选取了决定系数等评价指标来对模型性能进行评价。为了验证该通用模型的正确性以及可靠性，引入了生成对抗网络生成新数据来验证模型有效性。最后利用 Python 及其相关库生成一个应用软件，实现通用模型的工程化应用。

第 3 章 数据预处理与主成分降维分析

氢能作为一种清洁、可再生的能源，具有广阔的应用前景，尤其是在氢燃料电池领域。氢燃料电池凭借其能量转换效率高和零排放的优势，逐渐成为清洁能源技术中的重要发展方向。然而，其较短的使用寿命却阻碍了其大规模的应用。这促使许多的研究者致力于氢燃料电池寿命预测的研究。然而，目前的研究面临着一个很大的问题——缺乏一个适用于不同电堆的通用预测模型。

为此，本文提出了一种基于相似性分类的氢燃料电池寿命预测通用模型。相似性分类作为其基础部分具有重要意义，因此本章重点探讨数据预处理和主成分降维分析方法，以实现数据的相似性分类。首先，对选出来的三组数据进行数据预处理，包括提取公共变量、统一变量和单位以及绘制出数据的箱线图，然后采用图基法处理异常值。接着，通过变量相关性分析发现数据间的线性相关关系，进而选择了 PCA 对数据进行降维处理。通过 PCA 降维得到了相似性图以及 PC1 和 PC2 这两个主成分。最后，通过分析 PC1 和 PC2 这两个主成分，明确数据间的差异性和相似性，借此来对相似性图进行区域划分。

3.1 数据预处理

3.1.1 公共变量提取

根据第二章可以知道三个数据集的变量并不统一，这是因为每个数据集都是根据自身需求来记录相关参数的。为了满足本文需求，需要提取出三个数据集公共变量，可以得到下表 3-1。

表 3-1 三个数据集公共变量

FC1 数据集	FC2 数据集	FC3 数据集
Time(h)	时间(h)	Time(s)
Utot(V)	B 堆电压平均值(V)	voltage(V)
I(A)	燃料电池电流(A)	current(A)
TinH2(℃)	A 堆氢气进堆温度(℃)	temp_anode_inlet(℃)
ToutH2(℃)	氢气出 B 堆温度(℃)	temp_anode_outlet(℃)
TinAIR(℃)	B 堆空气进堆温度(℃)	temp_cathode_inlet(℃)
PinH2(mbar)	A 堆氢气进堆压力(kPa)	pressure_anode_inlet(kPa)
PoutH2(mbar)	B 堆氢气出堆压力(kPa)	pressure_anode_outlet(kPa)
PinAIR(mbar)	B 堆空气进堆压力(kPa)	pressure_cathode_inlet(kPa)

提取 FC2 数据集的公共变量时，考虑到采集数据所用的电堆是双堆串联而成

的。理论上，因气体扩散、电极材料等差异，使得两堆在运行上可能存在细微差异，但鉴于这些细微差异对核心研究目标影响很小，故在本文研究中暂不考虑。

从表 3-1 中，可以发现三个数据集部分变量的单位并不统一。例如，压力数据有的以 mbar 为单位，有的则是 kPa。但在为了后续的分析的顺利进行以及合理，必须将这些单位进行统一。同时为了保证数据分析更加方便，本文将三个数据集的变量统一成 FC1 数据集变量。

此外，鉴于 FC1 数据集是在单电密条件下采集的数据，为确保数据的一致性，其他两个数据集也需采用单电密数据。因此，本文从这两个数据集中提取常用电密数据。

3.1.2 数据分析以及异常值处理

通过 3.1.1 的工作可以知道三个数据集存在 9 个公共变量，其中就包括时间变量这一关键变量，它在时间序列数据研究中具有重要的意义。确定好变量之后，首要任务就是全面剖析这些数据集的整体分布情况，重点关注是否存在异常值。这一步骤不仅可以提升数据的质量，更能有效的把握数据的整体规律，方便后面数据降维分析和模型构建工作的开展。

在众多数据分析工具中，箱线图就能很好的满足需求。它通过可视化的方式，能够简洁直观的呈现数据中的四分位数、中位数、最大值和最小值等关键统计量，同时，箱线图也能通过箱体和上下须线的组合直观的将异常值展现在图中。这些优点不仅能够帮助研究者观察到数据的整体分布情况，还能有效的识别数据中是否存在异常值，为后续的分析提供可靠的支持。FC1 数据集的箱线图如图 3-1 所示。

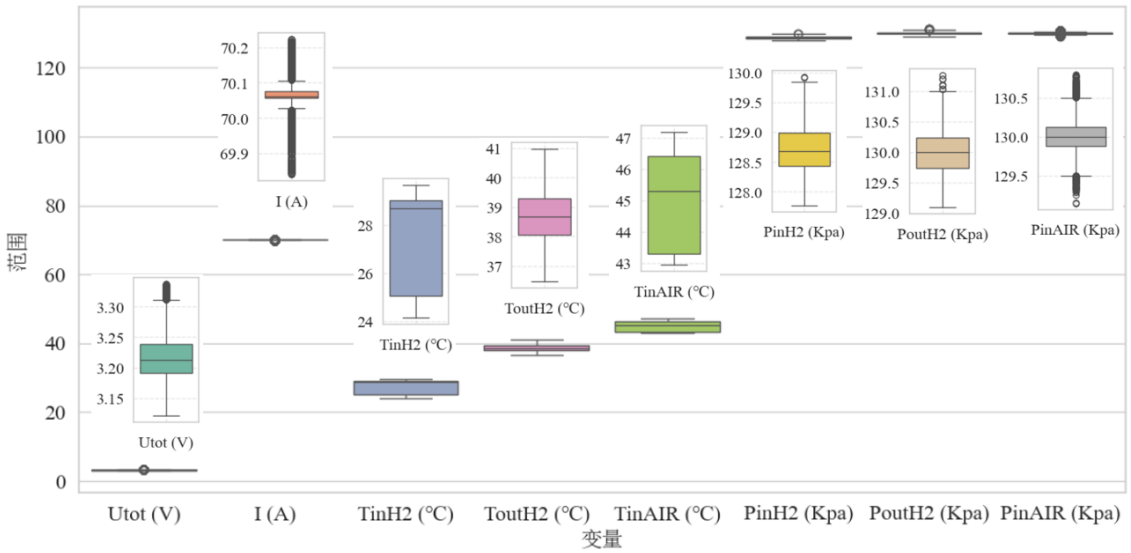


图 3-1 FC1 数据集箱线图

图 3-1 中仅展示了 FC1 数据集中除了时间变量以外的八个变量的分布范围以

及异常值情况。从图中可以看出每个变量的箱体位置和长度都是不同的，这表明各个变量的取值范围以及集中程度都是有区别的，如变量“Utot（V）”的取值范围是 3.1-3.3 V，分布得很集中，而变量“ToutH2（℃）”的取值范围是 36.5-41℃，相对而言数据就比较分散一些。同时，也可以观察到在变量“Utot（V）”和“I（A）”的箱线图的上下须线之外有很多黑色的点，这些就是该变量数据的异常值点。

以图 3-2 的“Utot（V）”箱线图为例来进一步剖析，从数据分布范围来看，数据跨度是从下边缘 3.10 V 到上边缘 3.30 V，其宽度仅为 0.20 V。再看绿色的箱体部分，其中位数大约在 3.22 V 的位置。这是该数据的中间值，具有重要意义，它将数据分为两部分，上下两侧的数据量相等。上四分位数大约在 3.24 V 处，下四分位数大约在3.19 V处，这意味着中间50%的数据就集中在这一相对狭窄区间内。从箱体的长度较短这一特点可以看出，大部分数据紧密地聚集在一起，数据的离散程度较低。

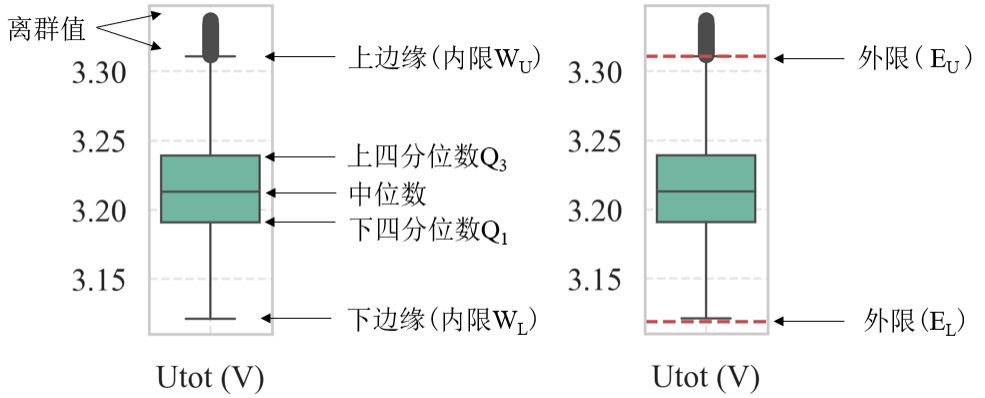


图 3-2 Utot（V）箱线图

通过箱线图可以很直观的找到异常值，这些值通常远离数据的主体部分，打破了整体数据的分布规律。从图 3-2 中就可以很清楚的看到在上边缘处的部分数据远离了箱线图的主体部分，因此就可以认为它是异常值。其实这些异常值形成原因有很多。一方面，可能是测量数据过程的误差造成的，如测点处仪器松动或者测试环境不稳定等原因；另一方面，某些异常值可能反映了某些特殊情况对数据产生的影响，这种异常值在数据分析中需要特别关注。

考虑上述这两个方面的原因，本文处理异常值时选择采用图基法，它将异常值分为温和异常值和极端异常值，并通过内外限来划分两者的位置。在图 3-2 的 Utot（V）的箱线图中上下外限用红色虚线标出，其具体的计算式如下：

$$E_U = Q_3 + 3IQR \tag{3-1}$$

$$E_L = Q_1 - 3IQR \tag{3-2}$$

上下内限为图中的上下边缘，计算式如下：

$$W_U = Q_3 + 1.5IQR \tag{3-3}$$

$$W_L = Q_3 - 1.5IQR \tag{3-4}$$

其中，Q_3 为上四分位数；Q_1 为下四分位数；IQR 为四分位距，即上下四分位数之差。

在图基法中，箱线图中内外限之间的部分是温和的异常值，而超出外限的部分则是极端异常值。考虑到一些异常值可能是因为某些特殊情况而造成的，并且有助于后续的电压衰减分析，因此本研究选择保留异常值中的温和异常值，同时剔除掉极端的异常值，以减少数据噪声对后续分析的影响。

分析完 FC1 数据集箱线图之后，对于 FC1 数据集的整体分布态势就有了清晰的了解，接下就是对其他两个数据集进行箱线图分析，如图 3-3 和 3-4 所示。

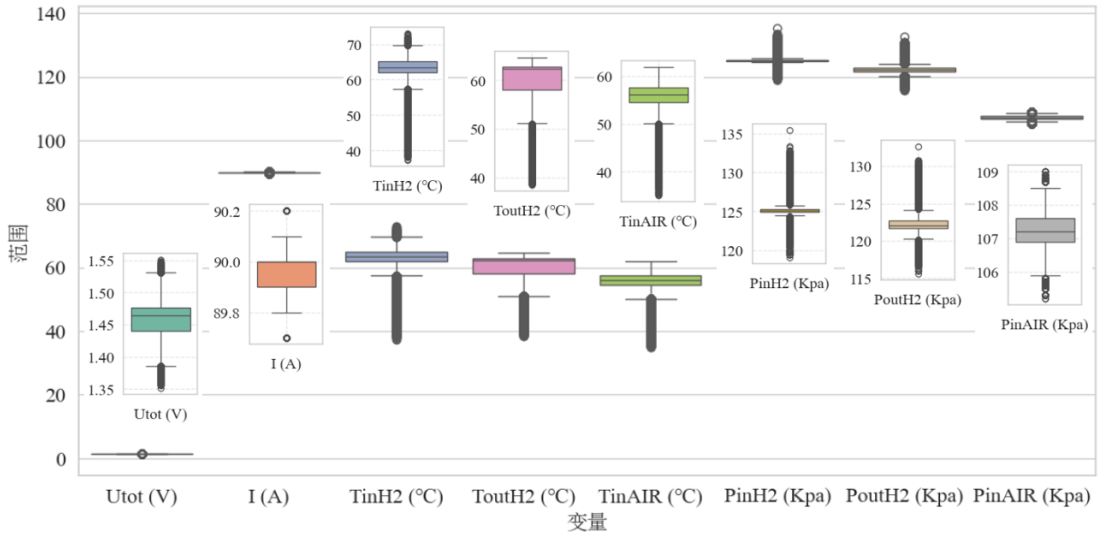


图 3-3 FC2 数据集箱线图

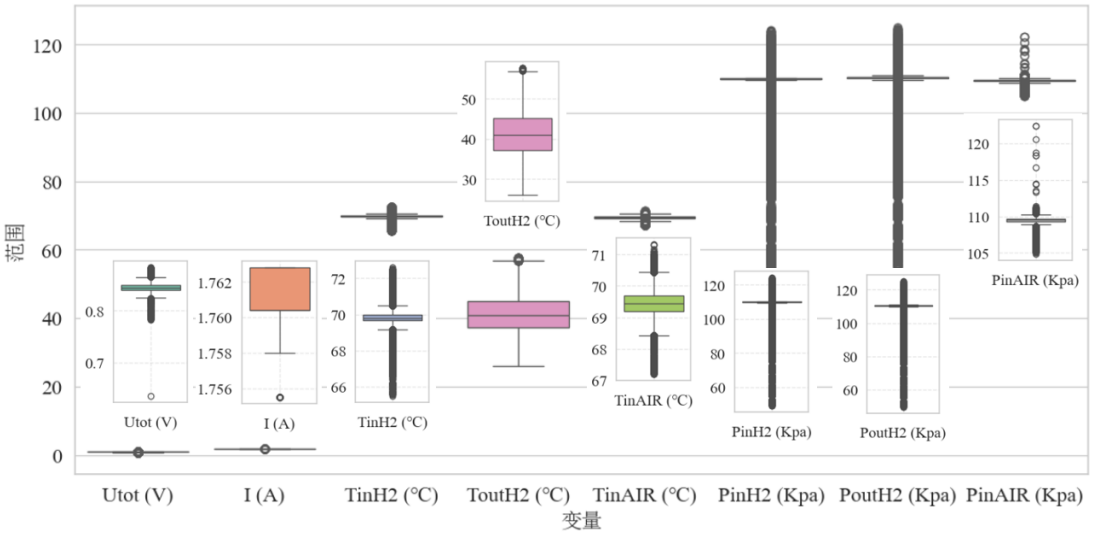


图 3-4 FC3 数据集箱线图

通过这两个数据集的箱线图和 FC1 数据集箱线图可以发现三个数据集在异常值存在一些差异。与 FC1 数据集相比，FC2 和 FC3 数据集不仅异常值数量更多，

而且异常值的离散程度更高了。具体来说，FC1 数据集的箱线图中只有部分变量存在异常值，而另外两个数据集的箱线图中每个变量都存在异常值；再以变量“PinH2（kPa）”为例，在 FC1 数据集中虽存在异常值，但可以看出异常值数据量少，且数据跨度不大；对应到 FC2 数据集的箱线图中的同一个变量，其异常值数量就很多了，且上下跨度在 10 kPa 左右；而在 FC3 数据集的箱线图中的变量“PinH2（KPa）”，其异常值数量就更多了，且跨度非常大，达到了 60 kPa 左右。这种差异可能是由于数据来源不同，采集设备、环境以及样本差异导致，因此，在后续分析中，鉴于这些特点需要采用合适的方法来进行异常值处理，如前文提到的图基法，从而降低不同的异常值对数据分析结果的干扰。

3.1.3 数据集对比分析

通过图基法分别对三个数据集进行异常值处理之后，接下来就需要对整体数据展开分析。其实，在前一小节的箱线图分析中就可以发现每个参数的分布范围也是存在一定差异的。因此，本小节打算将三个数据集放在一起进行对比分析，以便更好的了解数据之间的异同。

为了更直观的展现出这三个数据集分布范围之间的区别，本小节选用了雷达图作为可视化分析工具。通过雷达图，可以将三个数据集的各个变量统一到一个空间内，这样就能非常直观的分析出三个数据集分布范围之间的差异了。同时，由于变量“Utot”取值范围相较于其他变量小很多，于是对这个变量单独画了一个蜂群图，如下所示；

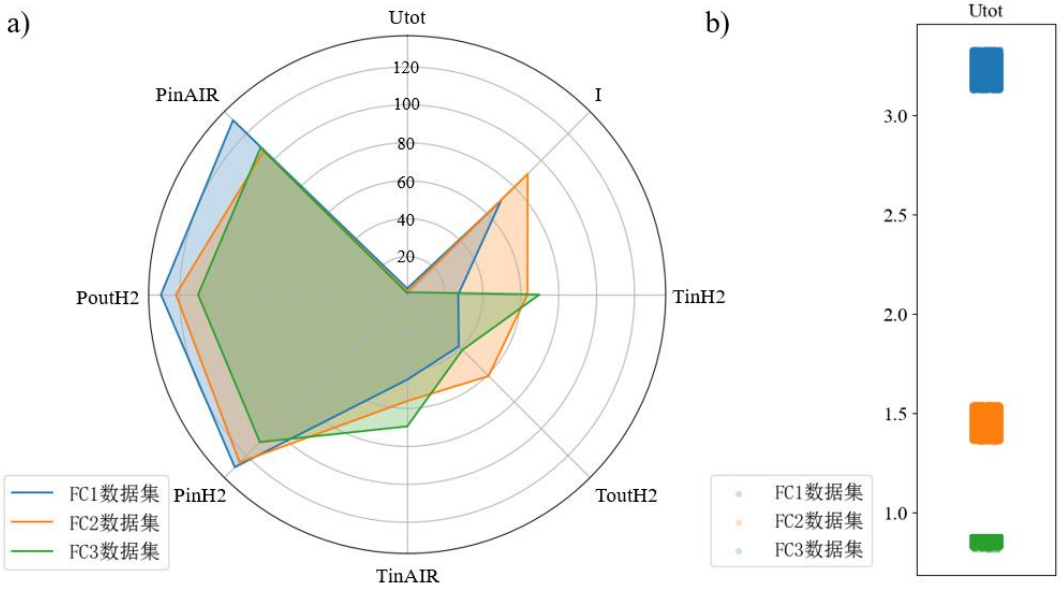


图 3-5 三个数据集分布对比分析图 a）三个数据集的雷达图；b）三个数据集的“Utot（V）”蜂群图

如果 3-5 a）所示，三个数据集的每个变量分布都不同，具有明显的差异性。

具体来说，某些变量的取值范围相对较窄，并且数据在该范围内呈现出高度集中的趋势，这表明这些变量在不同数据集之间的变化其实是很小的，波动性很小。其中，最为明显的就是变量“Utot”，在图 a）中三个数据集几乎和中心点重合，但从图 b）中还是可以发现它们存在一些差异，只不过数据范围波动很小。而另一部分变量取值范围则相对广泛，并且数据范围更加的分散变量“PoutH2”就是其中一个代表，三个数据集的数据在该变量轴上分布区间宽广，且分布范围不同，这就意味它们的取值范围存在较大的差异。

虽然这些变量分布各异、相互独立，但实际上这些变量彼此关联。从原理上来讲，这些变量最后都会影响到电堆电压（Utot）这个参数。这是因为这些变量取值范围的差异、分布的状态，最终会通过复杂的化学物理过程作用于电堆电压。在图 3-5 的 b）图中可以知道 FC1 数据集的电堆电压值最大，FC3 数据集的电堆电压值最小。对应到图 a）中，可以发现 FC1 数据集的压力参数均高于其他两个数据集，而温度参数则普遍低于另外两个数据集。但是对于 FC3 数据集来说，虽然压力参数是三个数据集最小的，温度参数却不是。这说明不能只从数据分布角度上来对这些数据集进行分类，需要采用更加深入全面的手段来进行分析。

3.2 数据降维分析

由于数据集中涉及到 9 个公共变量，这些变量之间具有复杂的相互作用，这使得直接分析具有一定的难度。为了简化这一分析过程，本文决定对数据集进行降维处理。降维不仅能够帮助更好的理解变量之间的关系，还能有效降低数据的复杂性，从而为后续的分析奠定基础。

目前，降维方法可以分为线性和非线性两大类。线性降维方法（如 PCA 等）通过寻找数据的线性组合来降低维度，适用于数据具有明显线性特征的情况；而非线性降维方法（如 t-SNE 等）则能够捕捉数据中的非线性结构，适用于处理复杂的数据分布。

3.2.1 变量相关性分析

为了确定数据是线性组合还是非线性组合，本小节采用变量相关性分析。首先将三个数据集进行合并，然后运用皮尔逊相关系数对各变量间的关联程度进行量化评估，得到结果如图 3-6 所示。

计算出来的皮尔逊相关系数的取值范围是[-1,1]，对应到图中就是蓝色和红色，颜色越深就表示这两个变量的相关性越强，不过蓝色表示的是负相关，而红色表示的是正相关。例如，变量“Utot（V）”就和很多变量之间的色块是深色的，这说明电堆电压这个变量和很多变量都是强相关的。其他的变量相互之间也存在一

定的相关性，像变量“I（A）”和变量“PinH2（kPa）”之间的相关性色块就是深红色的，这说明电流和氢气进气压力之间存在着很强正相关关系。在实际的物理过程中，氢气进气压力的上升确实会为电池内部的物理和化学反应提供更多的氢气，从而产生更多的电流，导致电流增大。这一现象验证了两个变量之间的强相关性。

从图中也可以看到时间变量“Time（h）”与其他变量相关性都非常小。这是因为时间变量本身在该数据体系中，就是一个独立的观测维度。它不像电堆电压、电流、温度、压力等变量，互相之间存在着物理化学上的相互作用关系。时间变量只是记录这些物理量变化的一个参考尺度，它不直接决定其他变量的数值变化，更多是反映这些变量随时间推移的演变情况，所以与其他变量的相关性较弱。

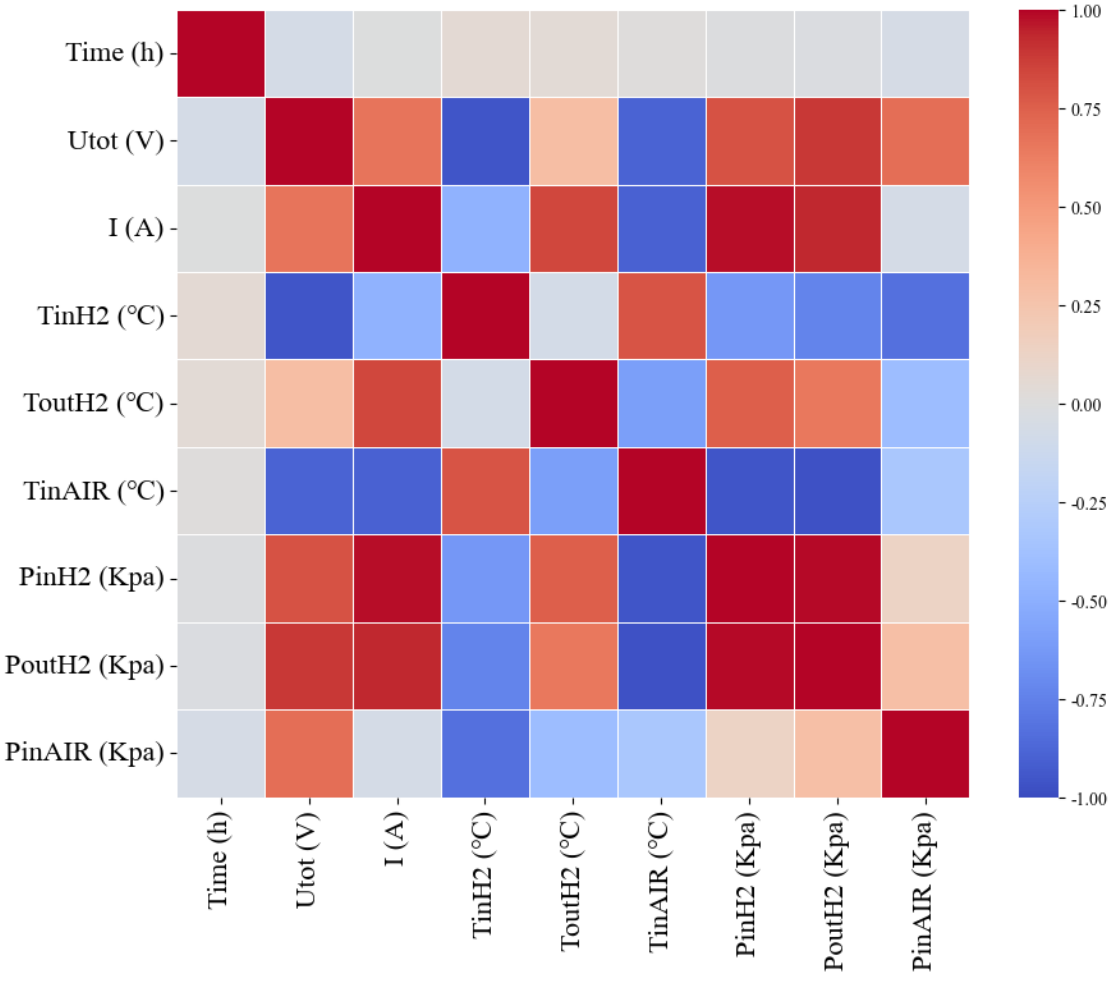


图 3-6 变量相关性热力图

总的来说，从变量相关性热力图中，可以明显看到大部分变量之间的相关性还是较强的，这就表明这三个数据集数据之间线性关系是占主导地位的。这种显著的线性关系，为本文在降维方法的选择上提供了明确的方向——PCA 降维。

3.2.2 PCA 降维

PCA 降维不仅依赖于变量之间的相关性，还要求数据具有相同的量纲。因此，首先需要做的就是先将三个数据集进行合并，然后将合并的数据进行标准化处理，最后才是进行 PCA 降维。其中，合并数据集是为了确保所有变量是在同一个分析框架下进行处理，标准化处理则能消除不同变量之间的量纲差异对分析结果的影响，使降维结果更加准确可靠。按照步骤处理之后就可以得到降维后的相似性图，如图 3-7 所示。

通过图 3-7 可以知道，三个数据集降维后映射到相似图上分为三个部分。其中，红色表示 FC1 数据集，蓝色表示 FC2 数据集，绿色表示 FC3 数据集。X 轴和 Y 轴分别表示两个主成分 PC1 和 PC2，它们分别解释了 63.55%和 23.56%的方差，合计解释了 87.11%的方差。通常来说，累计方差达到 85%以上，就可以认为主成分已经捕捉到了数据集中的大部分信息，而本文研究的累计方差高达 87.11%，大大超出了阈值。因此，可以认为这两个主成分就能很好的代表数据中的主要信息，为后续分析差异性和相似性提供了可靠的基础。

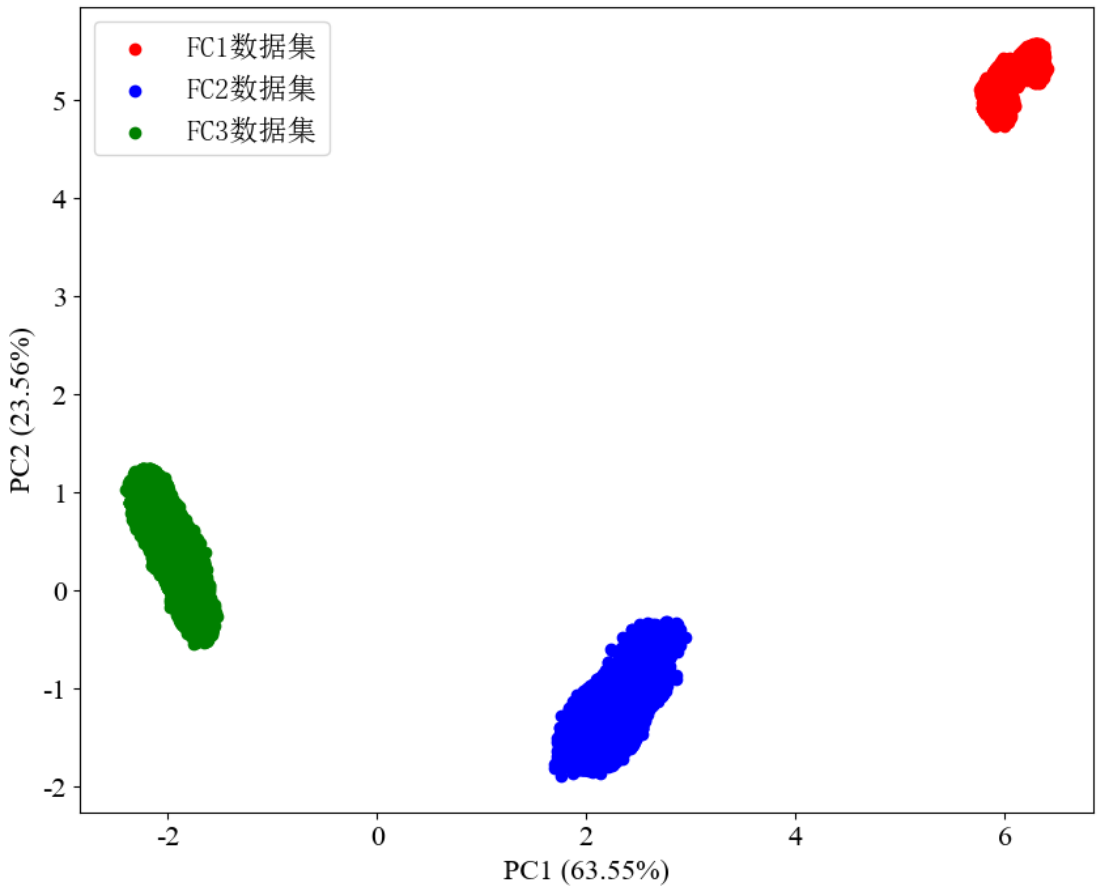


图 3-7 基于 PCA 的相似性图

确认了两个主成分的可靠之后，接下来就需要探讨这两个主成分 PC1 和 PC2

与初始变量之间的关系，明确它们两个所捕捉到的主要信息。于是，本节绘制出了主成分载荷图，如图 3-8 所示。图中每个变量的方向和长短都能很直观的反映这些变量在两个主成分中的贡献程度。

从主载荷图中可以看出每个变量的方向和长度都不一样，因此它们投影到两个主成分 PC1 和 PC2 的长度都不一样，并且投影后的数值也是有正有负的。这就分别对应着它们对于每个主成分的贡献度大小以及它们与主成分的正负相关性。基于上述分析，可以看出 PoutH2（kPa）和 PinH2（kPa）两个变量投影在 PC1 正方向的长度很长，而 TinAIR（℃）变量则投影在 PC1 的负方向上的长度很长，这就说明这三个变量是主成分 PC1 的主要组成变量；同理也可以看出 PinAIR（kPa）变量投影在 PC2 的正方向的长度很长，ToutH2（℃）和 TinH2（℃）两个变量投影在负方向的长度很长，也就说明这三个变量是主成分 PC2 的主要组成变量。

除了这些变量，在图中也可以看出 Utot（V）和 I（A）这两个变量对主成分也有一定的贡献度，只不过它们对于两个主成分的贡献度几乎一样，研究的价值并不高。

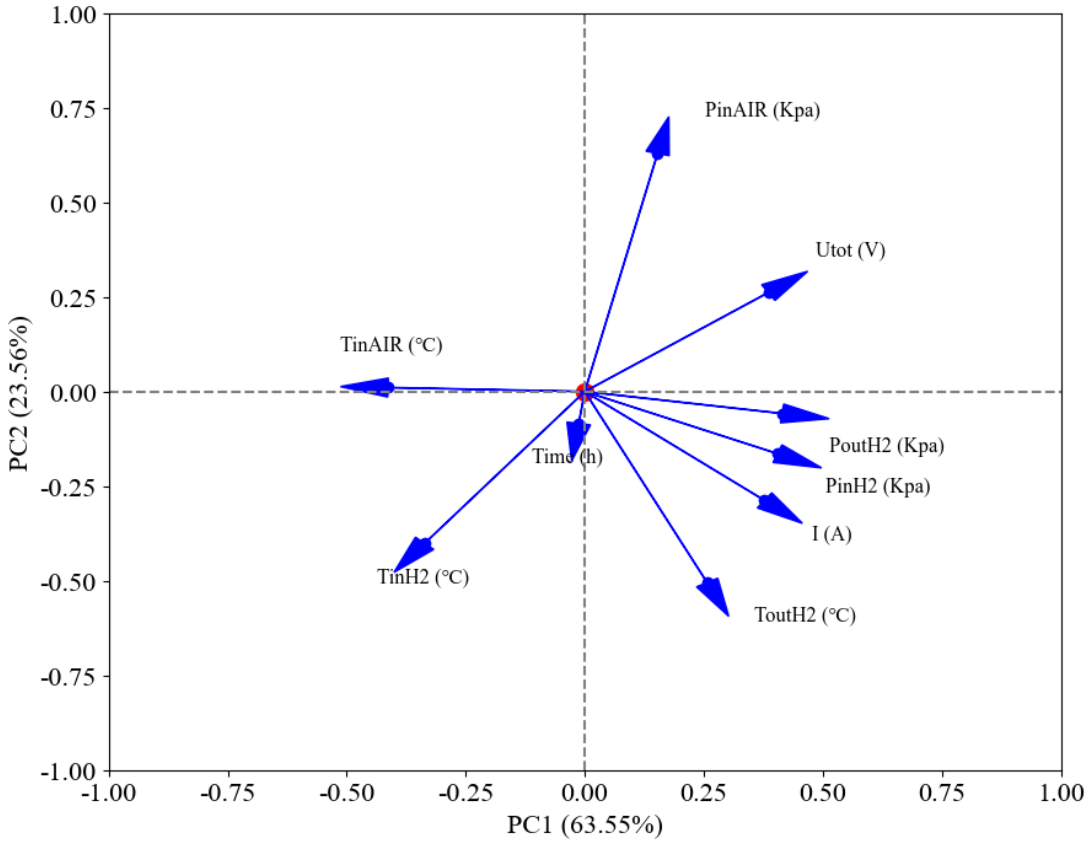


图 3-8 PCA 主成分载荷图

前文的分析，已经确定了哪些变量对于主成分 PC1 和 PC2 贡献度较高，但这些发现都只是粗略的表示了变量与主成分之间的关系。为了进一步了解每个主成分的详细情况，本节绘制出了 PC1 和 PC2 的独立载荷图来展示每个变量在主成分

中的权重大小，进而更好的理解主成分与变量之间的关系。

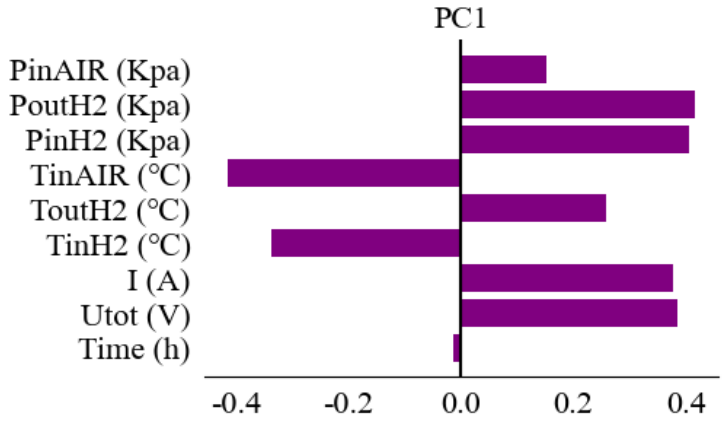


图 3-9 PC1 载荷图

如图 3-9 所示是 PC1 的载荷图，根据图中的各个变量在 PC1 方向上的载荷情况，可以推导出 PC1 的线性组合公式：

$$PC1 \approx 0.41 \times PoutH2 - 0.41 \times TinAIR + 0.40 \times PinH2 + 0.39 \times Utot + 0.38 \times I$$

由于载荷小的变量对主成分影响较小，因此式子中仅展示载荷最大的五个变量，并将单位省略掉。从给出的组合式中可以看出 PoutH2（氢气出口压力）、TinAIR（空气入口温度）、PinH2（氢气入口压力）、Utot（电压）和 I（电流）这五个变量的载荷绝对值是最大的。

其中，PoutH2 和 PinH2 的系数绝对值很大且为正，这反映出氢气进出口的压力变化对电池系统是具有一定影响的。从燃料电池系统的角度上来说，氢气压力的变化是会直接影响到其内部物理化学反应速率以及电池输出的电压、电流的大小。相关的文献也证明了这一点，例如，Celik 等[63]的研究就明确指出，氢气压力是影响电池性能的关键因素，合适的压力确保供氢充足，从而提高能量转换效率。

TinAIR 的系数也很大，但是为负值。这同样也反映出空气的入口温度变化对系统也是有一定影响的。理论上来说，空气的入口温度过高或过低都是会影响氧气分子的扩散速率和活性的，从而影响到电池内部物理和化学反应的速率的快慢。相关研究表明，这一现象对燃料电池的性能影响很大。例如，Yang 等[64]的研究就指出燃料电池堆工作温度对性能和寿命影响非常的显著，电堆温度过低就会降低化学反应活跃度与电池效率。这些发现表明空气入口温度是能量转换的关键调控因素。

综上所述，可以发现组成 PC1 的主要参数都是影响氢燃料电池能量转化与输出过程中的关键因素。这些参数的变化会直接影响氢燃料电池内部物理化学反应的快慢，进而改变电池的输出电压、电流的大小。因此，控制合适的氢气进出口压力，选择合适的空气入口温度，能帮助优化燃料电池性能并且提高其能效。

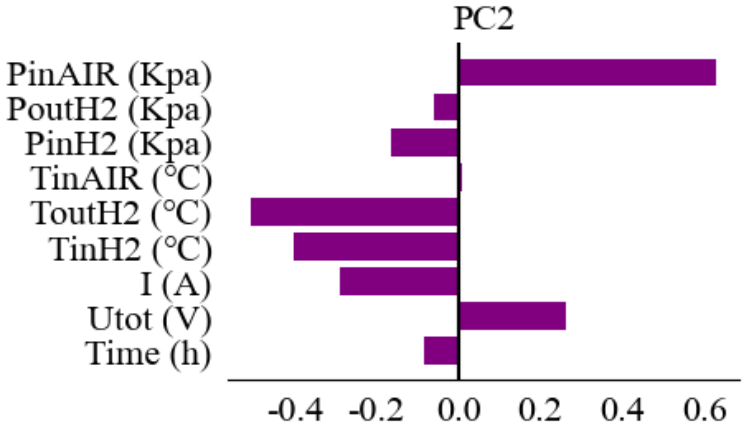


图 3-10 PC2 载荷图

同样可以根据图 3-10 中的 PC2 的载荷图得到 PC2 的线性组合公式，如下：

$$PC2 \approx 0.63 \times PinAIR - 0.50 \times ToutH2 - 0.40 \times TinH2 - 0.29 \times I + 0.26 \times Utot$$

从给出的 PC2 组合式中可以看出，PinAIR（空气入口压力）、ToutH2（氢气出口温度）、TinH2（氢气入口温度）、Utot（电压）和 I（电流）这五个变量的载荷绝对值是最大的。

其中，PinAIR 的系数绝对值最大且为正，这表明空气的入口压力变化对于电池系统的影响很大。理论上来说，合适的空气入口压力能保证氧气以恰当的流速和浓度进入到电池的内部，保证其物理化学反应能够高效进行。这一点也已经被 Wang 等[65]证明了，他们将分层控制法应用到质子交换膜燃料电池系统中，并通过优化空气供应控制，使系统运行效率提升 0.6%-2.6%，氢气消耗减少 5%以上。此外，若无法保证合适压力，不仅影响氧气的供应也会影响到电池的冷却效率，尤其在长期高负载运行时影响更为显著。

ToutH2 和 TinH2 的系数也很大但为负，这说明氢气进出口温度的变化对氢燃料电池系统有一定的影响。从氢燃料电池系统的角度上来说，氢气在电池内部发生反应的过程中往往伴随着热量的产生和交换。氢气的入口温度变化就会影响到反应的初始条件，进而影响到反应热量的产生和变化；出口的温度则是反映系统对于热量的利用率，当氢气出口温度比设定值低时，就可能是系统的热交换效率变高了，也有可能是反应放出的热量没有被充分利用。这一观点相关文献也已经证明了，例如，Salam 等[66]的研究就发现氢气入口温度升高就会提高反应速率和效率，但是过高就可能会导致质子交换膜脱水；出口温度降低则通常表示系统的热交换率升高了，但过低可能引发水蒸气凝结，影响质子交换膜的性能和输出功率。

综上所述，组成 PC2 的主要参数都是影响氢燃料电池热管理效率的关键因素。这些参数的变化会直接影响氢燃料电池系统内部的反应温度大小以及系统的冷却效率，进而影响整个系统的运行效率。因此，控制合适的空气入口压力以及氢气的进口温度，并检测好氢气的出口温度，对于整个氢燃料电池系统的热管理效率

具有重要意义。

总的来说，主成分 PC1 和 PC2 表示的方向是不一致的，并且各自对于氢燃料电池系统的影响是不同的。因此，可以认为主成分分析得到的两个主成分 PC1 和 PC2 是有意义的，同时通过分析也明确了这两个主成分和各个变量之间的关系，以及它们对于氢燃料电池系统的影响。

3.2.3 相似性图的区域划分

上一小节的分析表明，PC1 和 PC2 划分的主要原因在于，氢燃料电池系统中不同的变量对能量转换与输出和热管理效率这两个关键维度的贡献程度和作用方向的不同。这个结论是基于三组数据整体而言的，对于相似性图的区域划分作用不大。因此，本小节分别绘制两个主成分 PC1 和 PC2 的组成变量的雷达图，用于分析三组数据在每个主成分下各个变量之间的关系，根据此关系进行区域划分。

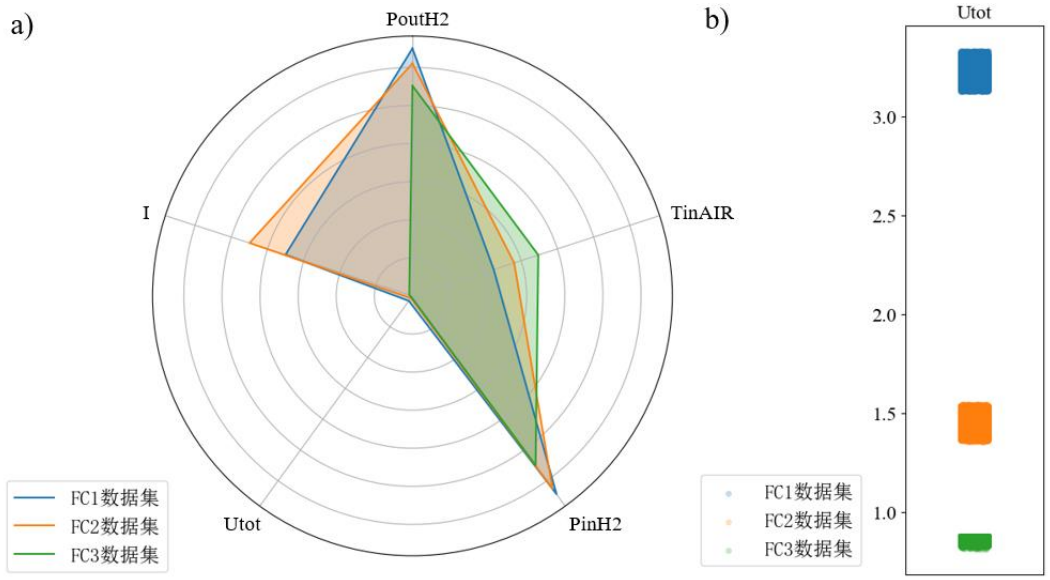


图 3-11 PC1 组成变量分布对比分析图 a）PC1 组成变量雷达图；b）三组数据集的“Utot（V）”蜂群图

结合图 3-11 和图 3-12 可以知道 FC1 数据集在 PoutH2、PinH2 和 PinAIR 三个压力指标上，均高于另外两个数据集的；而在 TinAIR、ToutH2、TinH2 三个温度指标上，则是弱于另外两个数据集的，这表明 FC 数据集在相似性图上相较于其他两组数据处于一个高压低温的区域。

进一步分析 FC2 数据集和 FC3 数据集，发现这两组数据在大多数变量的分布上是一样的。但是，它们在 ToutH2 和 TinH2 这两个温度指标的差值上差距很大，即 FC3 数据集的氢气进出口温差远大于 FC2 数据集，这可能是因为同济大学的燃料电池系统的热管理效率相对较低，无法有效地从燃料电池中转移掉多余的热量，导致氢气的进出口温度差距较大。

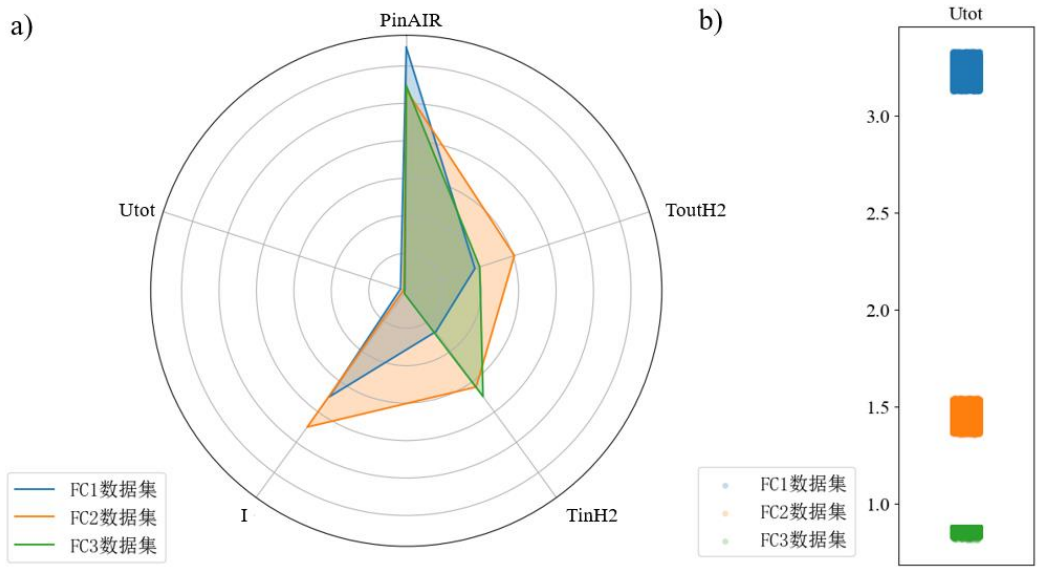


图 3-12 PC2 组成变量分布对比分析图 a）PC2 组成变量雷达图；b）三组数据集的“Utot（V）”蜂群图

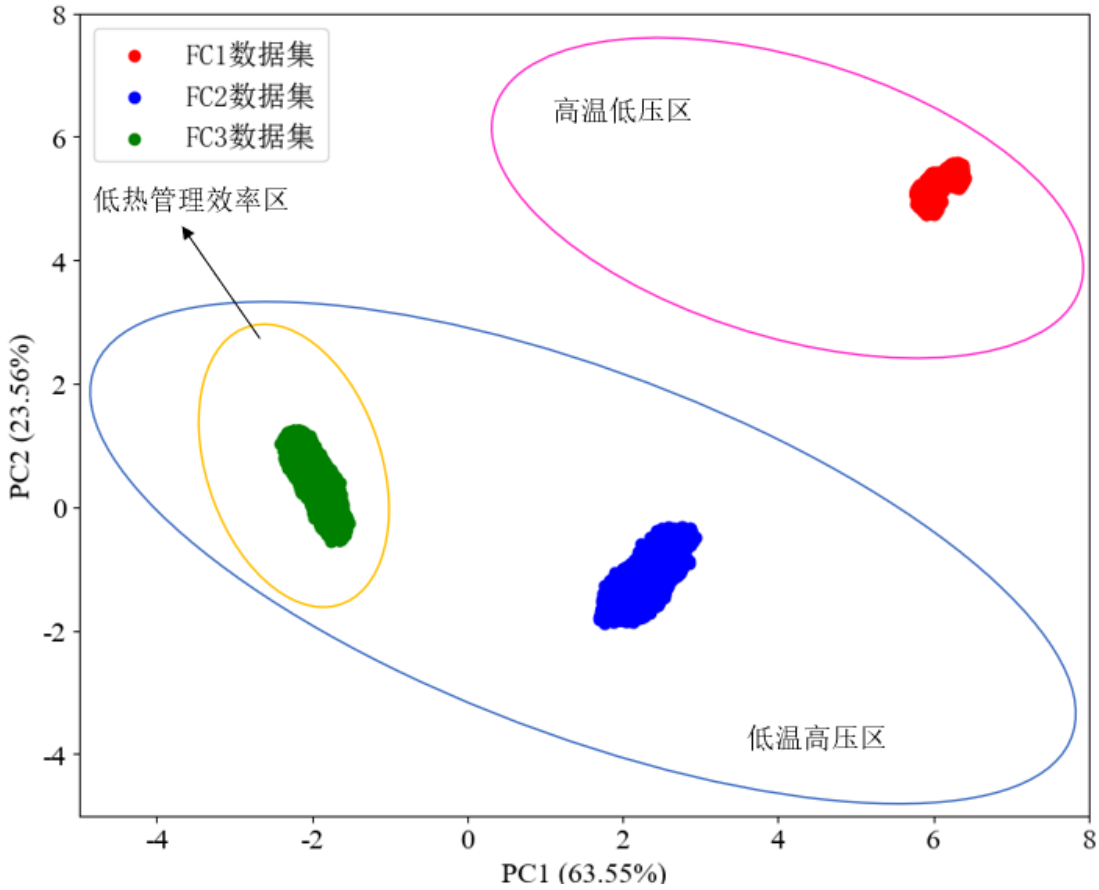


图 3-13 PCA 相似性图区域划分

基于以上的分析，本小节将图 3-7 的相似性图划分为三个不同区域：FC1 数据集划入高压低温区域，FC2 数据集和 FC3 数据集分别划入低温高压区域和低热管理效率区。这种划分不仅能够清晰展示各个数据集之间的差异，还能为后续的模

型建立以及验证奠定了基础。具体如图 3-13 所示。

图 3-13 中基于现有的数据集将相似性图划分为了三个区域，分别是：高温低压区、低温高压区以及低热管理效率区。其中，低热管理效率区是包含在低温高压区之内的，这一关系是通过对热管理效率与压力、温度之间内在联系的分析得出。

当有新的数据加入时，可以根据其大致的特性映射在相似图上。通过新数据映射在图上的位置，判断它们被划入哪一个区域。这一过程将有助于对新数据集的热管理效率大小和压力等特性进行初步分类。然后，通过得到的新数据在相似性图上的位置，并结合已有的区域特征，就能对新数据集的性能做出合理的解释和较为准确的预测。

目前，由于数据集的数量有限，基于相似性图所做的区域划分仅是初步尝试。虽然目前已经划分出了高温低压区、低温高压区以及低热管理效率区，但这些区域的划分存在一定的偏差，主要原因有两方面：一方面，有限的数据集使得某些关键的特征没有体现出来，从而影响到区域的划分。另一方面，实际应用中数据的多样性和复杂性远超现有的数据。因此，一些潜在的影响因素可能没有被考虑到，这就可能导致划分结果与真实情况存在偏差。

尽管如此，这种尝试也为后续的研究奠定了基础，同时也为该领域的数据分析提供了一条新的思路。随着研究的开展，若能收集的数据更多、更具代表性，相似性图会更加的完善。一方面，更多的电堆老化数据输入，可以进一步细化现有的区域，更好的勾勒出高温低压区、低温高压区以及低热管理效率区的边界。另一方面，随着数据的丰富，甚至有可能划分出更多的区域，例如“高热管理效率区”等等。这样一来，对于不同的新数据识别将会更加精确，后续的分析也会更加准确。

3.3 本章小结

本章主要围绕数据预处理和主成分降维分析展开。首先，通过提取三组数据的公共变量、统一变量名和单位，然后使用图基法处理异常值，来保障数据质量；随后，通过变量相关性分析，发现数据之间存在显著线性相关，于是选择了 PCA 方法来进行降维分析，得到了相似性图以及 PC1、PC2 这两个主成分，这两个主成分累计解释了 87.11%数据方差。进一步分析发现，PC1 主要反映能量转换与输出相关的因素，而 PC2 则侧重于系统热管理效率方面的因素。最后，基于主成分分析的结果，本章将相似性图划分为三个区域：高温低压区、低温高压区和低热管理效率区，为后续研究奠定了基础。

第 4 章 基于相似性分类的氢燃料电池寿命预测模型

上一章通过对三组数据集剔除了异常值，采用主成分分析进行降维，得到了相似性图和主成分 PC1、PC2。结果表明，主成分 PC1 和 PC2 累计解释了 87.11%的数据方差，其中 PC1 主要反映了能量转换与输出相关因素，PC2 则侧重于热管理效率方面的因素。然后，基于主成分分析的结果，将相似性图划分为三个区域：高温低压区、低温高压区和低热管理效率区。这些工作为构建基于相似性分类的氢燃料电池寿命预测通用模型奠定了基础，完成了相似性分类这一步骤。

因此，本章要在前一章的基础上，围绕氢燃料电池寿命预测模型的构建展开。研究内容主要分为三部分：首先，在前文数据预处理的基础上对数据进行重建和平滑处理，消除数据中的噪声和尖峰，提升数据质量；其次，建立 LSTM 神经网络，发现其预测效果有限。于是，在原模型框架的基础上引入了 KAN 层，构建了 LSTM-KAN 模型，提升模型的预测精度和泛化能力；最后，利用 GAN 生成的新数据来验证模型的正确性。

4.1 神经网络的构建过程概述

对于人工神经网络，不管是 LSTM 神经网络，还是其他神经网络，尽管它们的内部神经网络有所不同，但它们网络模型创建的流程都是一样的，可以分为下面几步，如图 4-1 所示。

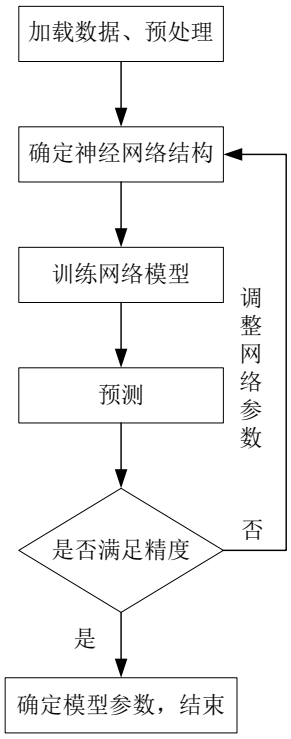


图 4-1 神经网络的搭建流程

在搭建预测模型的过程中，首先需要加载数据，然后进行数据预处理，包括数据清洗、数据重构、数据平滑处理以及归一化。其中，数据清洗部分的预处理工作已经在上一章完成了。接着，选择了 LSTM 模型以及在此基础上加入了 KAN 层的 LSTM-KAN 两种模型，以有效捕捉时序特征。

确定好模型之后就开始构建其对应的神经网络，同时将处理好的数据按比例划分成训练集和测试集作为模型的输入。然后，采用梯度下降法进行多次迭代，同时每次迭代都评估模型性能。最后，通过决定系数和均方误差等评价指标来对模型进行全面评估，并通过可视化分析展示预测效果，以便进一步进行调整参数优化模型。

4.2 数据重建和平滑处理

上一章已经进行了数据清洗，但选用的数据集数据量较大，且原始数据包含大量噪声和尖峰，这导致后续分析可能会耗费大量的计算资源。因此，需要对原电压数据进行数据重建和平滑处理。重建数据之后，本节采用局部散点平滑估计（Locally Estimated Scatterplot Smoothing，LOESS)方法对重建数据进行平滑处理。这种方法是一种非参数的平滑方法，它被广泛应用于数据分析和趋势识别，其核心在于对每个数据点进行局部加权线性回归。简单来说，LOESS 方法就是通过选择一定数量的邻近数据点，并为这些邻近点加上不同的权重，从而实现对这些数据的有效平滑和潜在趋势的捕捉。

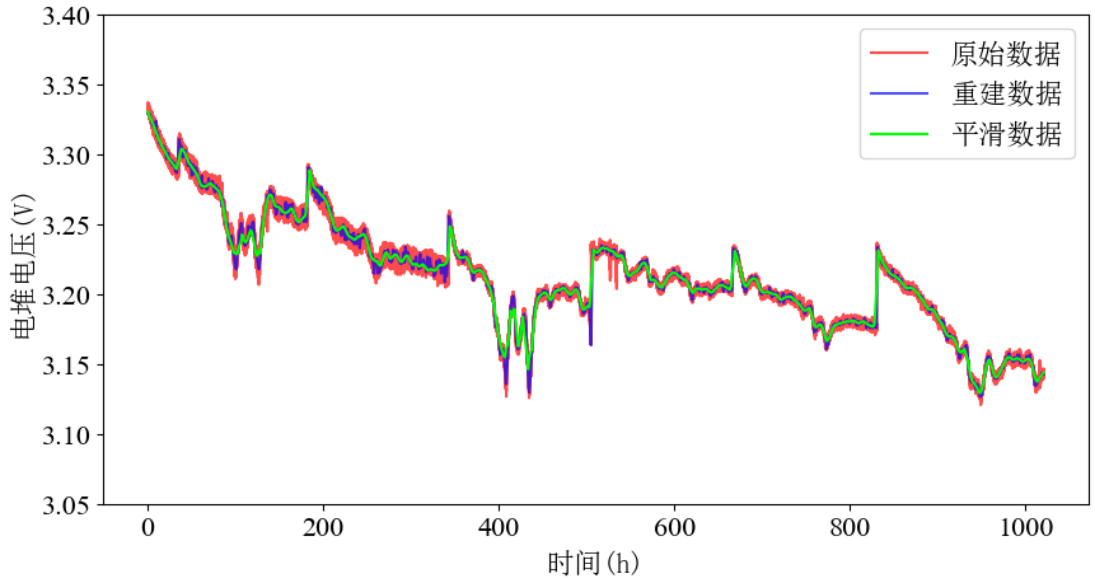


图 4-2 FC1 数据集电堆电压的衰减数据

重建数据是通过提取原始数据中的代表性的样本来降低数据量。本节采用了 1 小时的等距采样的方式，从原始数据中提取出 1021 组代表性的数据。随后，对重建后的数据采用 LOESS 方法来进行平滑处理，将平滑窗口的宽度设置为 10。经过

平滑处理后的数据不仅保留了原始数据的趋势，还消除了原始数据中的噪声和尖峰。现在以 FC1 数据集为例，其进行重建平滑处理后得到的结果如图 4-2 所示。

图中红色表示原始数据，蓝色表示重建数据，绿色表示平滑的数据。重建的数据基本和原始数据无差，但是平滑数据相较于原始数据在转折点处明显更加圆滑，整体上没有尖峰，同时它也保留了原始数据的主要趋势。保证了在提升数据质量的同时，也不会对后续的分析造成不利影响。

其他两个数据集也是按照相同的方式进行数据处理，保证所有数据集在处理流程上的一致性，尽可能减少因处理方式不同而带来的干扰。经过数据重建和平滑处理后，数据中的噪声和尖峰被有效去除，数据趋势更加明显，且满足了模型对数据格式和质量的要求，此时的数据经过归一化后就可直接作为模型的输入。

4.3 基于 LSTM 的寿命预测

在前文的研究中，提取了三个数据集的公共变量，也对每个数据集都做了相同的数据预处理，确保了变量的统一性。因此，在后续模型训练中无需再进行特征工程、变量剔除和变量重要性排序等工作。同时基于现有的数据处理框架，模型可以直接使用这些变量作为输入。

虽然在现有数据处理框架下不需要额外的数据再处理工作，但为了符合模型训练需求仍需按照 6:4 的比例将数据集划分为训练集和测试集，同时将电堆电压设置为目标变量。

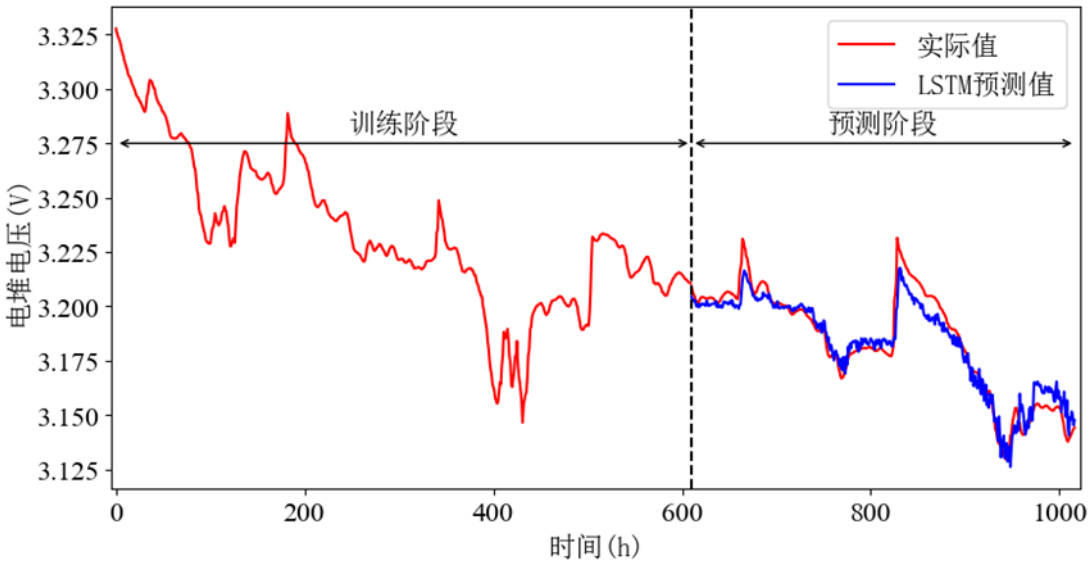


图 4-3 FC1 数据集的预测结果

接下来就是模型搭建，如下：首先通过三个 LSTM 层逐步提取数据特征，各层神经元数量分别为 35、35 和 30，激活函数均采用 relu。为了防止过拟合，在每个 LSTM 层后均添加 Dropout 层，Dropout 率分别为 0.2、0.2 和 0.1。然后，在 LSTM

层的基础上串联两个全连接层，神经元数量分别为 20、10，激活函数也采用 relu。在全连接层也后添加 Dropout 层，Dropout 率都设为 0.1。最后连接一个输出层，其神经元数量为 1。模型使用 Adam 优化器，以 MSE 为损失函数、MAPE 为评估指标对模型进行编译。模型训练的迭代次数是 1000、批次大小为 32，并利用测试数据进行验证，预测结果如图 4-3 所示。

图 4-3 中红色曲线代表实际数据，蓝色曲线代表预测数据，虚线将数据分为训练阶段和预测阶段。从整体上来看，电堆电压呈现出下降的趋势，其变化过程中也存在着明显的起伏。但是在预测阶段，可以观察到预测数据（蓝色曲线）与实际数据（红色曲线）走势表现出较高的一致性，这表明模型成功的捕捉到了训练集数据中的内在规律，并在测试集上实现了较好的预测效果。

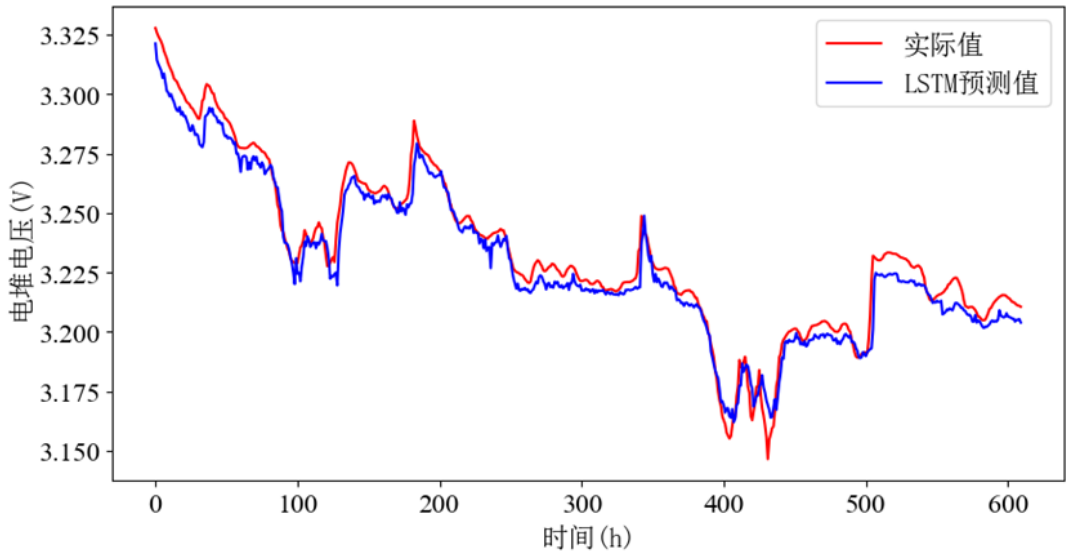


图 4-4 FC1 数据集的训练集结果

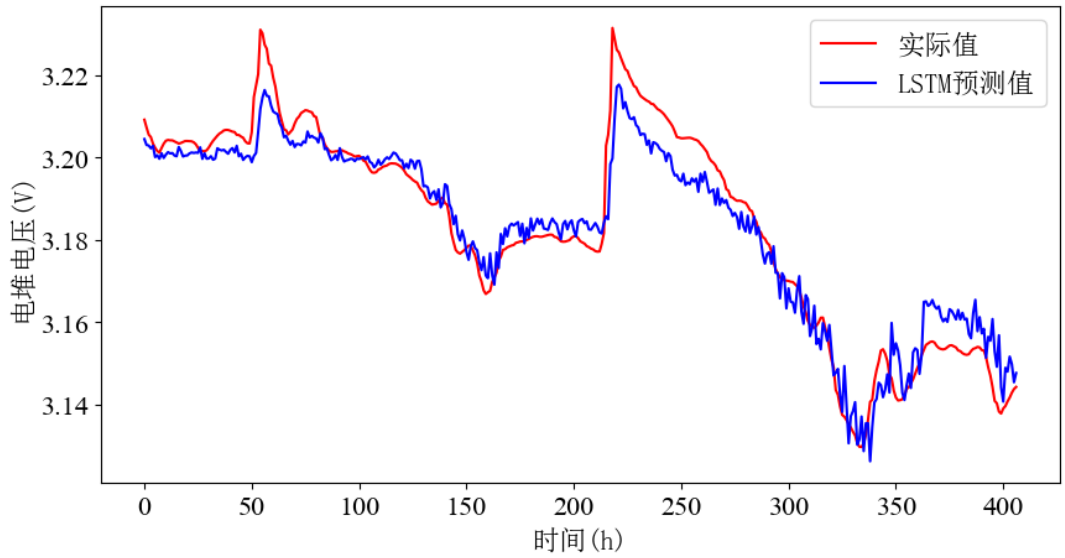


图 4-5 FC1 数据集的测试集结果

为了全面评估模型性能，需要进一步分析模型在训练集与测试集上的预测效果以及相关评价指标。于是，使用该模型分别对训练集和测试集的数据进行预测，预测结果如图 4-4 和图 4-5 所示，同时将它们的评价指标汇总到表 4-1。

表 4-1 FC 数据集的评价指标

	R^2	MSE	RMSE	MAE	MAPE
训练集	0.972205	0.000033	0.005751	0.004493	0.001390
测试集	0.932100	0.000041	0.006427	0.005109	0.001606

通过对训练集和测试集的预测结果和评价指标的分析，可以发现模型在训练集上拟合效果很好，实际值与预测值的贴合得很近，R^2 也达到了 0.972，MSE、RMSE、MAE 和 MAPE 等误差指标都处于较低水平，这表明该模型能够很好捕捉到数据特征与规律，具有较高的预测精度。然而，在测试集上，虽然实际值与预测值的曲线走势基本一致，但预测精度却下降了，R^2 值下降到了 0.932，各误差指标也相应变差了。这表明该模型在面对新数据时的泛化能力不足。因此，需要对该模型进行优化和改进，增强其在新数据上的适用性和预测的准确性，从而更好的满足实际应用需求。

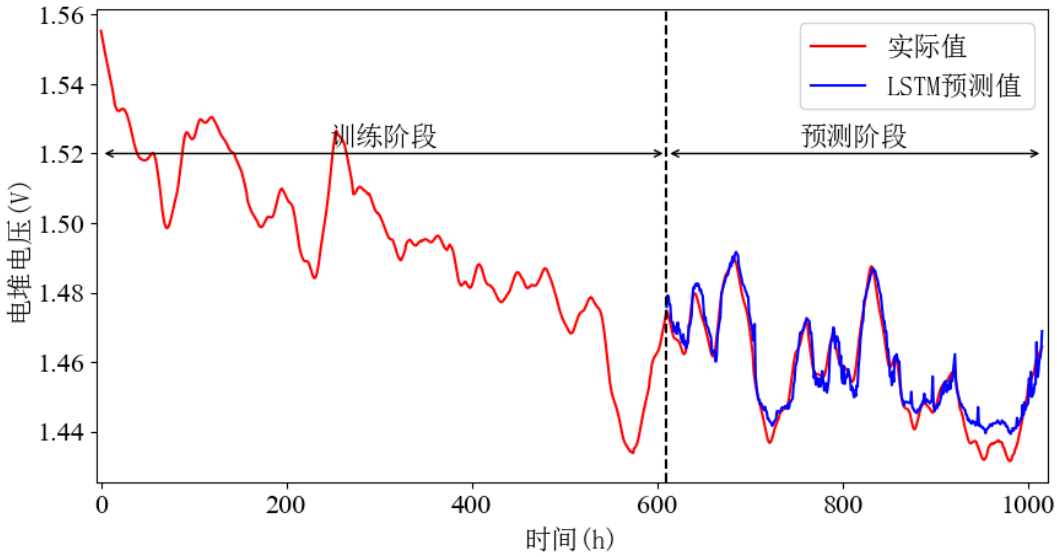


图 4-6 FC2 数据集的预测结果

此外，针对另外的两个数据集，也采用了相同的处理流程，将数据集按照 6:4 的比例划分为训练集和测试集，同时将电堆电压设定为目标变量。在模型搭建时，整体框架保持不变，仅对参数（如神经元数量、Dropout 率等）进行调整，使模型能够更好的适应不同数据集的特征，从而提高模型适用性和预测准确性。另外两个数据集模型的预测结果如图 4-6 和图 4-7 所示。

通过图 4-6 和图 4-7 可以发现，这两个数据集的模型同样也在训练集上捕捉到了数据的内在规律，在测试集上展现出较好的预测能力。然而，前文对于 FC1 数

据集模型的进一步分析发现该模型在测试集上的泛化能力不足。因此，也需要对这两个数据集构建出的模型在训练集和测试集上的预测结果进行深入分析，判断是否存在一样的问题。它们的相关评价系数汇总到了表 4-2 中，以便更全面的评估模型的表现并进行改进。

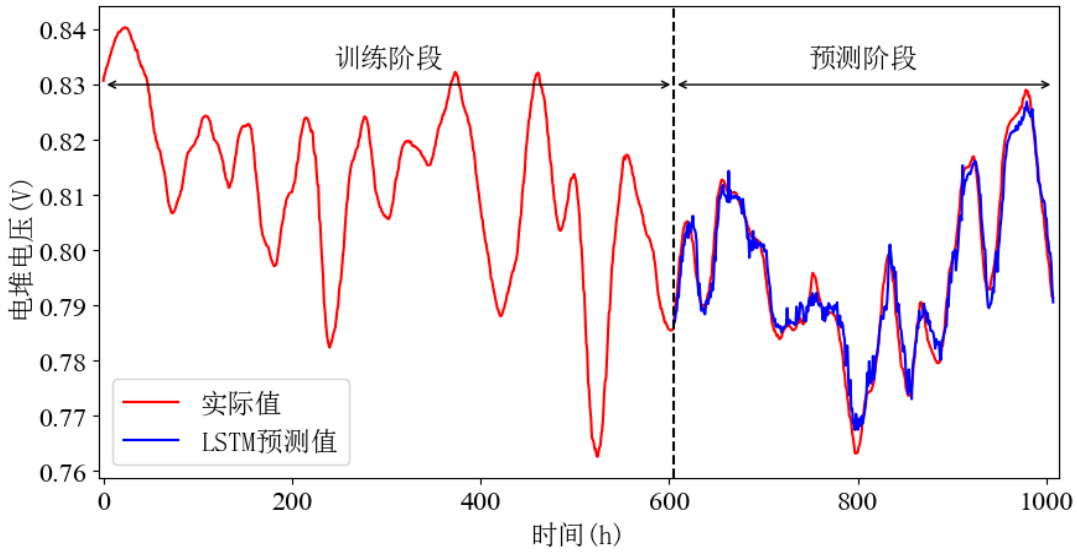


图 4-7 FC3 数据集的预测结果

表 4-2 FC2 和 FC3 数据集评价指标

		R^2	MSE	RMSE	MAE	MAPE
FC2 数据集	训练集	0.980889	0.000011	0.003360	0.002518	0.001679
	测试集	0.920255	0.000018	0.004245	0.003456	0.002378
FC3 数据集	训练集	0.950676	0.000012	0.003430	0.002709	0.003329
	测试集	0.947458	0.000012	0.003406	0.002766	0.003484

结合这两个数据集的预测结果和评价指标，可以发现这两个模型在训练集上的效果表现都很好，决定系数 R^2 都达到了 0.95 以上，MSE、RMSE、MAE 和 MAPE 等误差指标都非常的低。这一结果表明，这两个模型训练时都能很好的捕捉到数据的内在规律。

然而，在测试集上，两个模型的预测效果同样没有达到理想水平。相比较于训练阶段的良好表现，模型在面对新数据时，预测的准确性明显下降，预测值与实际值的误差也增大了。这一现象表明，当前模型在泛化能力方面存在一定的不足，对于新数据的预测能力还有待进一步提高。这也和前文分析的 FC1 数据集模型所面临的问题是一样的，训练集上预测效果很好，但测试集上的预测效果就下降了。因此，需要找到一个合适的方式来解决模型在测试集上泛化能力不行的问题。

4.4 基于 LSTM-KAN 的寿命预测

在前文所述中，三个数据集的模型在测试集表现上均暴露出在泛化能力不足的问题。经过分析，本文认为这一问题的根源可追溯到数据与模型本身的特性。因为本文选取的是三个数据集中的公共变量来用于模型训练与预测，这些变量虽具一定代表性，但难以全面、精准地描述整个氢燃料电池的复杂情况，所以用这些变量就无法很好的捕捉到目标值的趋势，从而导致训练集的预测效果变差，模型的泛化能力不足。这种数据输入层面的局限性，使得模型在数据处理过程中存在信息缺失的隐患。

于是，本节提出了一种方案来解决信息缺失的问题：在原来模型的基础上引入 KAN 层。它通过其独特的函数变换，将原始的输入信息映射到一个更高维的特征空间中，从而揭示数据中隐藏的复杂规律和非线性关系。这种方式有助于弥补仅使用公共变量带来的数据信息缺失问题，提升模型的表达能力。

经分析，本节决定在模型框架中将 KAN 层放在 LSTM 层之后，这一设计基于以下考虑：LSTM 层擅长处理时间序列数据，能从氢燃料电池的时间序列输入中提取和捕捉到时序相关特征，输出给下一层。这样 KAN 层就可以利用 LSTM 输出的特征，通过其独特的非线性变换，进一步挖掘深层次信息。这种组合方式不仅能够揭示 LSTM 层可能未充分捕捉的非线性关系，还符合深度学习模型中的层次化信息处理原则，方便参数调整和优化，使模型能够更加全面和深入地理解数据。

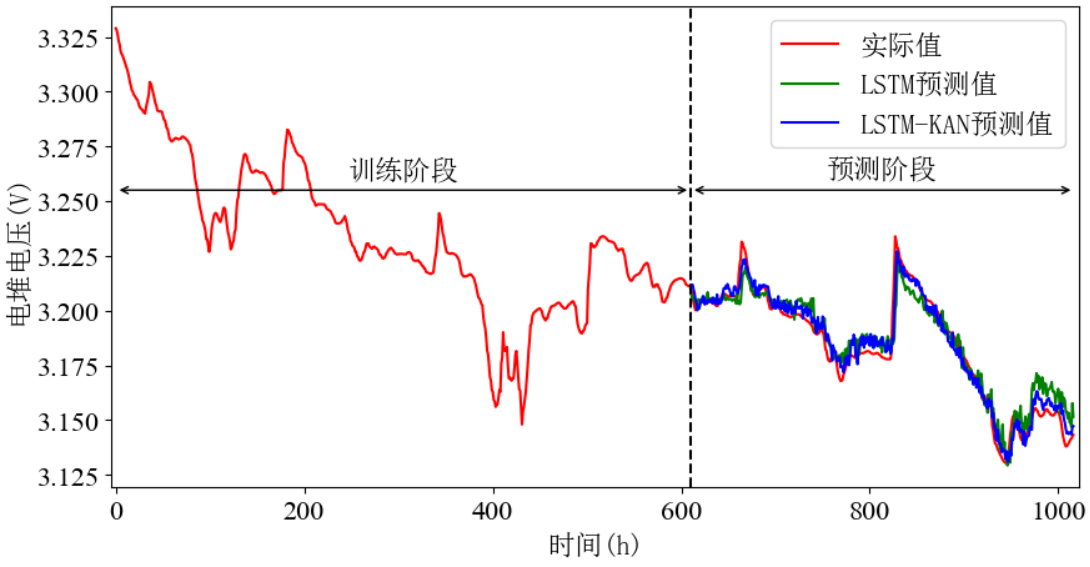


图 4-8 FC1 数据集的预测结果对比

对于 KAN 层的设置，本节主要通过自定义 kan_transform 函数对输入数据进行简单的平方，并在此基础乘 0.45 的缩放因子。该函数通过 Lambda 层对三个 LSTM 层输出的数据进行变换，生成 KAN 层的输出结果。随后，将 LSTM 层和 KAN 层

的输出进行融合，直接输入全连接层和 Dropout 层中，形成一个完整的 LSTM-KAN 神经网络模型。其他参数都按照前一节的设置，预测结果如图 4-8 所示。

图中的 LSTM 模型预测结果是绿色曲线，LSTM-KAN 模型预测结果是蓝色曲线，可以发现 LSTM 模型和 LSTM-KAN 模型的预测结果存在明显差异。在预测阶段部分，LSTM-KAN 模型的预测数据与实际数据的拟合程度明显高于 LSTM 模型。这表明所引入的改进措施有效提升了模型的性能，使模型能够更好地捕捉实际数据的变化趋势和特征，从而提高了预测的准确性和可靠性。同时也说明 KAN 层的引入一定程度上弥补了数据信息的缺失造成的影响，解决了单一的 LSTM 模型在测试集上泛化能力不足的问题。

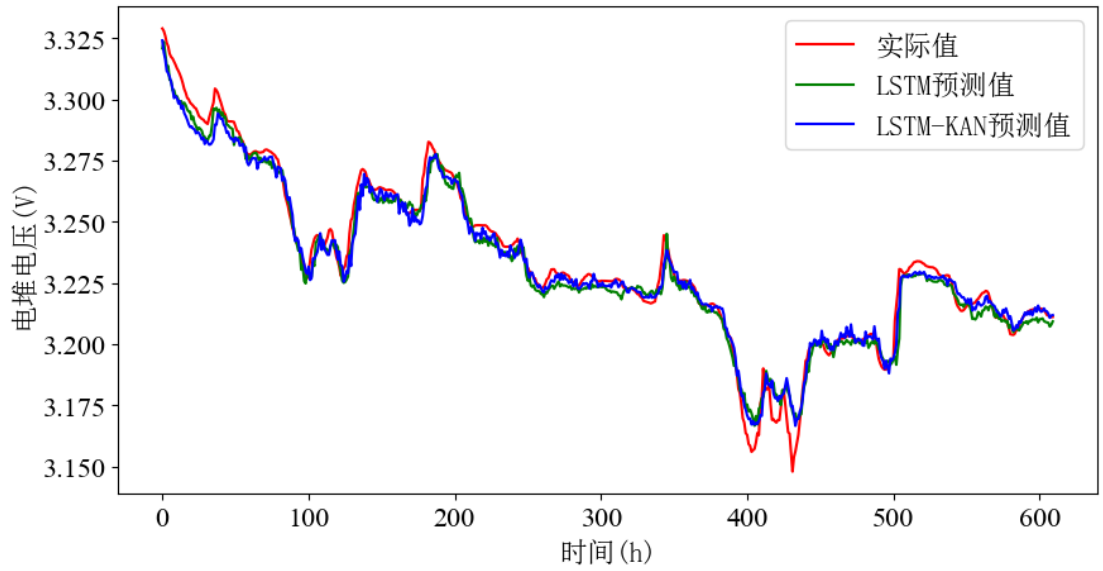


图 4-9 FC1 数据集的训练集结果对比

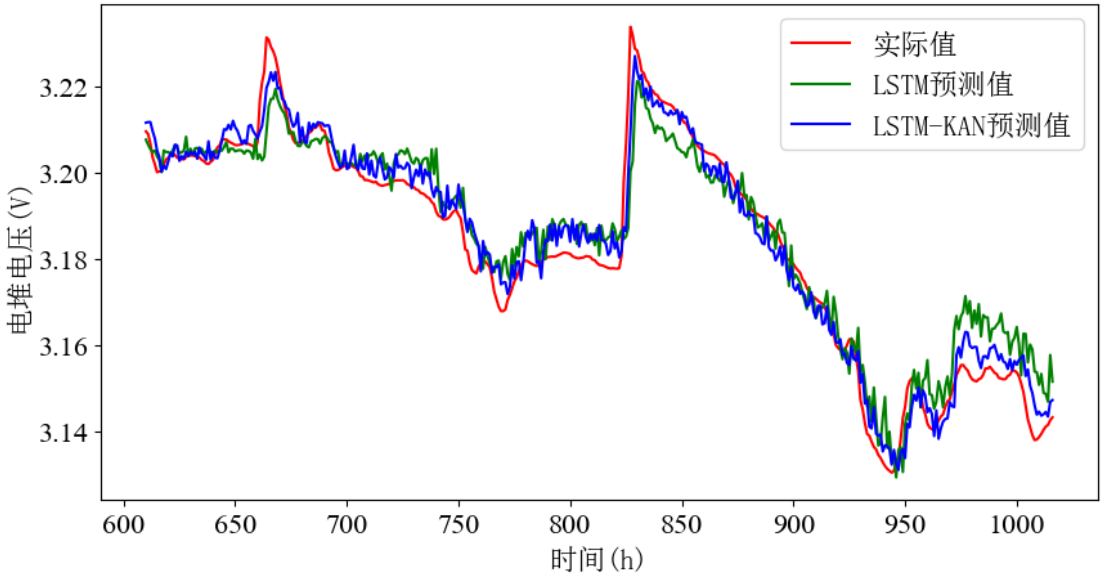


图 4-10 FC1 数据集的测试集结果对比

表 4-3 FC1 数据集评价系数比较

		R^2	MSE	RMSE	MAE	MAPE
LSTM	训练集	0.972205	0.000033	0.005751	0.004493	0.001390
	测试集	0.932100	0.000041	0.006427	0.005109	0.001606
LSTM-KAN	训练集	0.976563	0.000028	0.005313	0.003852	0.001189
	测试集	0.958556	0.000026	0.005054	0.004012	0.001260

同样，可以借助模型在训练集和测试集上的拟合程度以及评价系数，绘制了相应的对比图，如图 4-9 和 4-10，进一步验证了改进方案的有效性。同时也将训练集和测试集上的评价系数比较汇总到表 4-3。

在训练集的预测曲线中，氢燃料电池电堆的电压波动较大。LSTM 模型的预测结果能大致跟随其波动的趋势，但细节处的偏差非常明显，尤其在转折点处；而 LSTM-KAN 模型的预测结果与实际值贴合度非常的高，尤其在波动较大的区域，其预测更加精准，拟合能力更强。这表明 LSTM-KAN 模型在训练过程中能够更好地学习数据的特征。

在测试集上，实际值的波动同样较大。LSTM 模型的预测效果与训练集类似，偏差很明显，尤其在 820 h 到 950 h 这段电压急剧下降的区域，预测结果与实际值差距较大，这表明其泛化能力较差。相比之下，LSTM-KAN 模型在测试集上的表现更为出色，预测结果与实际值曲线更为贴近，甚至在电压急剧下降的区域也能较好的拟合，这表明新模型能够更好的适应新数据的变化趋势，也说明 LSTM-KAN 模型的泛化能力更强。从评价系数来看，无论是训练集还是测试集，LSTM-KAN 模型的 R^2 等评价指标均优于 LSTM 模型，进一步验证了其优势。

通过分析 FC1 数据集的 LSTM 模型和 LSTM-KAN 模型的对比结果来看，LSTM-KAN 模型无论是在训练集和还是测试集上的表现都要优于 LSTM 模型。在训练阶段，LSTM-KAN 模型展现出了更好的拟合能力，能够把数据特征学习得更好；在测试阶段，LSTM-KAN 模型则具有更好的泛化能力，能够更准确的预测新数据。这种结果表明，引入 KAN 层之后，模型的整体性能得到了有效提升，使其在学习氢燃料电池电压衰退等相关数据时，能够更精准的捕捉复杂的数据变化规律。

同样，对于另外两个数据集，本节也在它们原有的 LSTM 框架上加入了 KAN 层。然后，通过对 LSTM 层的各种参数以及 KAN 层的缩放因子进行调整，使 LSTM-KAN 模型能够更好的适应当前数据集的数据特征，从而提升模型的适用性和准确性。最后，分别绘制模型在两个数据集上的预测结果，如图 4-11 和 4-12 所示。同时，将模型在训练集和测试集上的评价系数进行汇总，如表 4-4 所示。

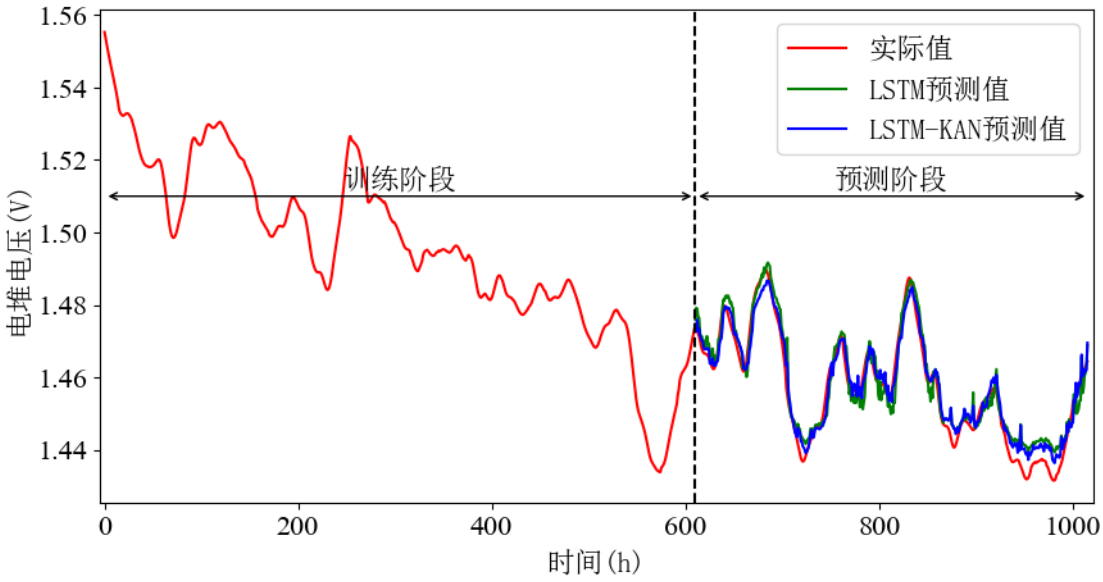


图 4-11 FC2 数据集的预测结果对比

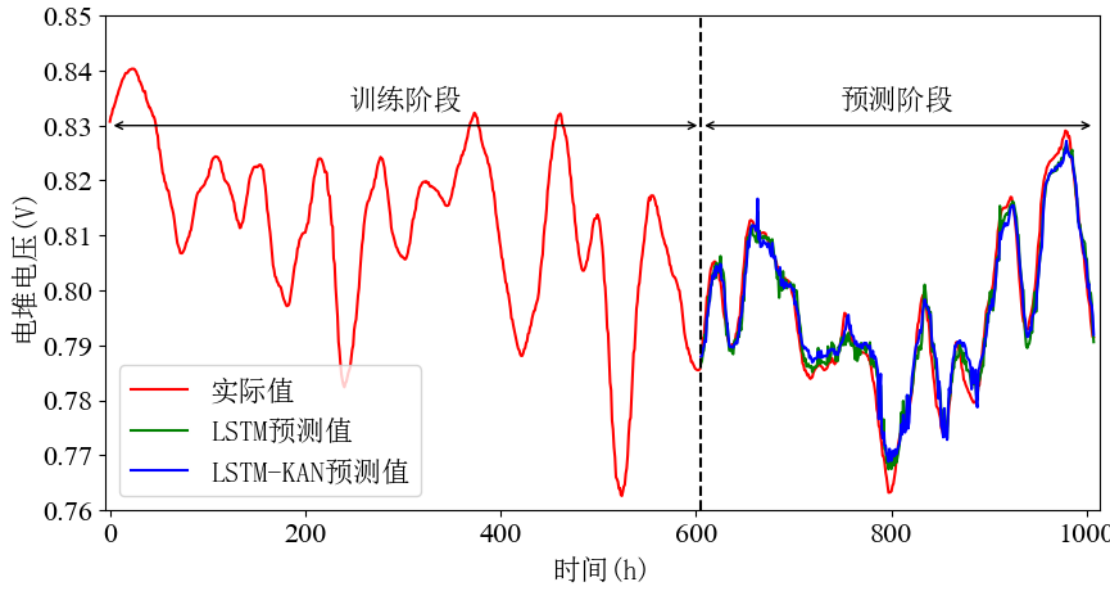


图 4-12 FC3 数据集的预测结果对比

通过对图 4-12 和图 4-13 的 FC2 数据集和 FC3 数据集的预测结果分析可知，在预测阶段 LSTM-KAN 模型相较于 LSTM 模型与实际值的拟合程度更高，说明两个数据的 LSTM-KAN 模型在性能上均优于 LSTM 模型。例如，在 FC2 数据集的工作时间约 800 h-1000 h 区间，LSTM-KAN 模型能够更紧密地跟随实际值波动；而在 FC3 数据集的工作时间约 900 h-1000 h 区间，LSTM-KAN 模型能够更准确地捕捉电压上升趋势。

从表 4-4 的评价指标来看，FC2 数据集的训练集上，LSTM-KAN 模型的误差指标与 LSTM 模型相差不大，但在测试集上，LSTM-KAN 模型的 R^2 为 0.953 相较于 LSTM 模型的 R^2 的 0.92 有了很大的提高，并且其他各项误差指标都更低了。FC3

数据集的训练集和测试集上，LSTM-KAN 模型的 R^2 值均高于 LSTM 模型，且误差指标也更低。这表明，LSTM-KAN 模型不仅在训练集上对数据拟合效果更好，而且在测试集上的预测准确性和泛化能力也更强，能够更好的适应新数据和实际应用场景。所以，本文选择将 LSTM-KAN 模型作为基于相似性分类的氢燃料电池寿命预测的通用模型的预测部分模型。

表 4-4 FC2 和 FC3 数据集评价指标对比

			R^2	MSE	RMSE	MAE	MAPE
FC2 数据集	LSTM	训练集	0.980889	0.000011	0.003360	0.002518	0.001679
		测试集	0.920255	0.000018	0.004245	0.003456	0.002378
	LSTM-KAN	训练集	0.980058	0.000012	0.003432	0.002902	0.001935
		测试集	0.952777	0.000011	0.003266	0.002688	0.001848
FC3 数据集	LSTM	训练集	0.950676	0.000012	0.003430	0.002709	0.003329
		测试集	0.947458	0.000012	0.003406	0.002766	0.003484
	LSTM-KAN	训练集	0.968529	0.000008	0.002740	0.002112	0.002610
		测试集	0.954849	0.000010	0.003157	0.002541	0.003201

4.5 基于相似性分类的通用模型的验证

本文提出了一种基于相似性分类的燃料寿命预测的通用模型。该模型的核心思想是通过 PCA 降维，将多组数据映射到低维空间中形成相似性图。然后，通过分析降维后的结果，明确数据之间的差异性和相似性，同时根据这些性质对相似性图进行区域划分。随后，对每个区域的数据分别进行模型训练。当模型读入新数据时，首先会通过降维将数据映射到已有的相似性图上，然后判断该数据与哪个区域的数据相似，最后调用该区域的模型进行预测。

目前，模型的相似图以及相关模型已经完成。现在，就需要输入一批新数据来验证该模型的适用性如何。然而，获取真实且有效的新数据存在一定的困难，于是就想到使用相关工具在已有数据的基础上生成新数据。GAN 就是一种先进的数据生成工具，可以用它来生成高质量的虚拟数据用于模型验证。

4.5.1 对抗生成网络（GAN）生成新数据

GAN 由生成器和判别器两部分组成，使用时需要分别搭建各自的神经网络。本节在搭建过程中，生成器采用深度递归神经网络（RNN）架构，其中包含了三个门控循环单元（GRU）层，且每层 GRU 之后都连接批归一化层和激活函数 LeakyReLU，最后有一层全连接层。设计成这种结构就是为了能够有效的捕捉时

间序列中的长期依赖关系，同时也能加速模型的收敛速度，并且增强模型的稳定性。激活函数则能够让模型有更强的非线性表达能力，帮助模型有效的避免梯度消失问题。

当噪声向量输入生成器之后，经过上述网络就能在全连接层输出符合要求的时间序列数据。该数据就会被送到判别器中，通过其网络来判别数据是真实数据还是生成数据。它的网络内部包含一个 GRU 层和两个全连接层，其中，GRU 层用来初步提取数据的时序特征，全连接层则会对时序特征进行进一步筛选和分类。最后，通过激活函数 Sigmoid 处理，输出一个区间在[0,1]之间的标量值，直观的表示数据为真实数据的概率。

完成网络搭建之后，就需要准备输入生成器中的噪声向量了。本文为了能让 GAN 生成的数据具有一定的代表性，选择将现有的三组数据合并成一组数据作为输入。这样就可以确保生成器能学习到三组数据的特征，生成出区别于当前数据的新数据，从而更好的验证模型。

之后就对模型进行初始化。本节定义了输入维度、隐藏维度、输出维度以及序列长度等关键参数。训练时，选用 Adam 作为模型的优化器，并将学习率设定为 0.00001。在这里，本小节增设了学习率调度器，每经过 20 轮训练，都会在前一个学习率的基础上乘以 0.9。这种做法能保证模型训练初期快速收敛参数，随着不断训练，学习率减小，再对模型参数进行更加精细的调整，达到提升训练速度的效果。

训练过程中，为提高数据的真实性，需要增强判别器判别功能。于是，每轮训练都要对判别器进行 3 次更新，让其不断优化，试图让其欺骗判别器，让它误判生成数据为真实数据。这种对抗式的训练方式就会促使生成器和判别器不断提升性能，共同进化。完成训练之后，生成器就可以生成新的数据，并通过反归一化让数据恢复到原始尺度。

4.5.2 新数据的相似性分类

通过上一小节的方法生成三组新数据之后，就对这三组数据进行 PCA 降维处理，并将其映射到图 3-7 中形成图 4-13。

通过分析图 4-13 可以直接判断新数据集所属的区域及其靠近的数据集：新数据集 1 落在高温低压区且与 FC1 数据集重合，表明两者在数据特征上具有一定的相似性；新数据集 2 和新数据集 3 则分别落在低热管理效率区和低温高压区，并分别靠近 FC2 数据集和 FC3 数据集。因此，后续可根据数据集所属区域，调用对应的模型进行预测。

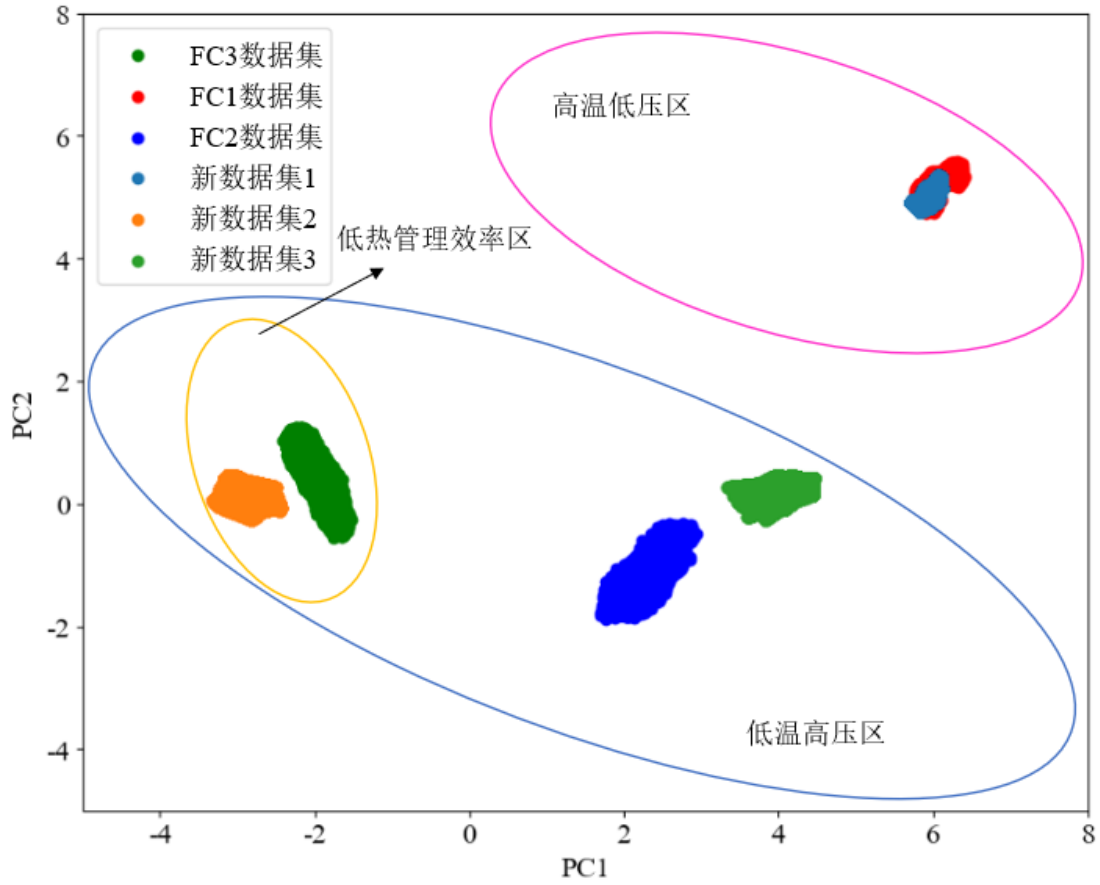


图 4-13 新数据映射后的相似性图

4.5.3 新数据预测效果

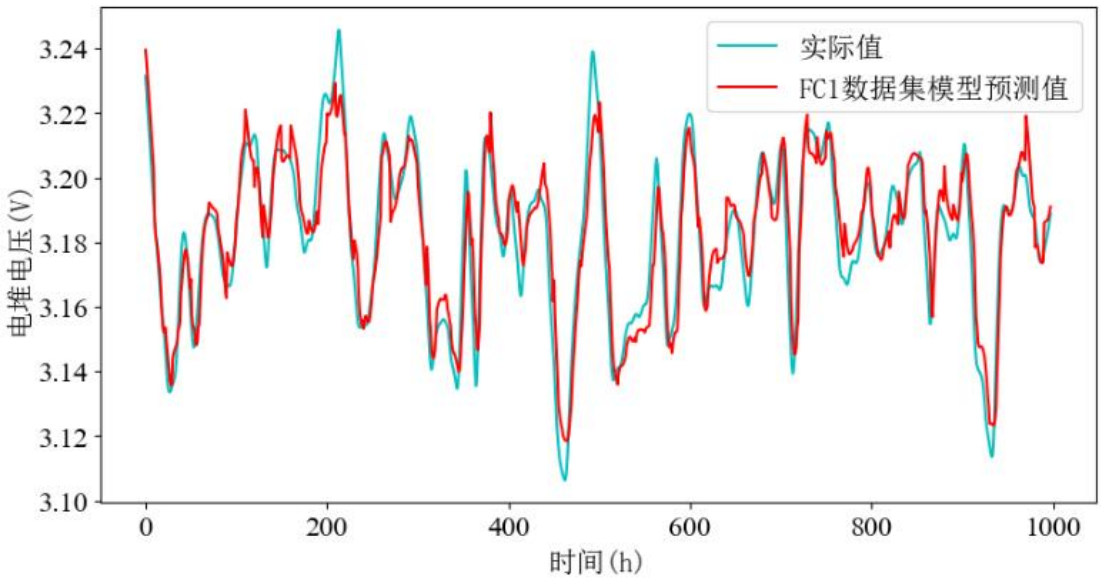


图 4-14 FC1 数据集模型预测效果

上一小节通过对新数据的 PCA 降维映射到相似性图上的结果分析，得到了每

个新数据集对应的区域。本节将根据这些结果分别对每组新数据集进行预测，通过预测的效果来验证生成数据的适用性和基于相似性分类的氢燃料电池寿命预测的通用模型的预测能力。

本节还是先以与 FC1 数据集重合的新数据集 1 为例，按照本文的研究的思路，选用训练好的 FC1 数据集的 LSTM-KAN 模型来对新数据 1 的数据进行预测，效果如图 4-14 所示。

从图中可以看出，FC1 数据集模型能够很好的预测出新数据集 1 的电堆电压变化趋势，且决定系数 R^2 也高达 0.87，这表明模型的整体预测效果很好。但是，仔细观察图中的细节可以发现，模型在一些峰值和谷值上的预测效果很差，经分析认为这可能是因为两组数据在某些变量上存在差异而造成的。于是，为进一步明确原因，本节画出了这两组数据的变量折线图进行分析，如图 4-15 所示。

从图中可以看出，这两组数据虽然在分布范围上重合度很高，但是数据分布的特征却存在很大的差异。例如，变量 TinH2 在 FC1 数据集中先是平缓变化，但是到 400 h 左右突然急剧上升，而新数据 1 却没有类似的特征。这种数据集之间的差异也就导致了 FC1 数据集模型对于新数据 1 的预测效果有所欠缺。但从最后的预测结果上来说，还是达到了预期，能够有效的证明本文的提出模型的正确性。

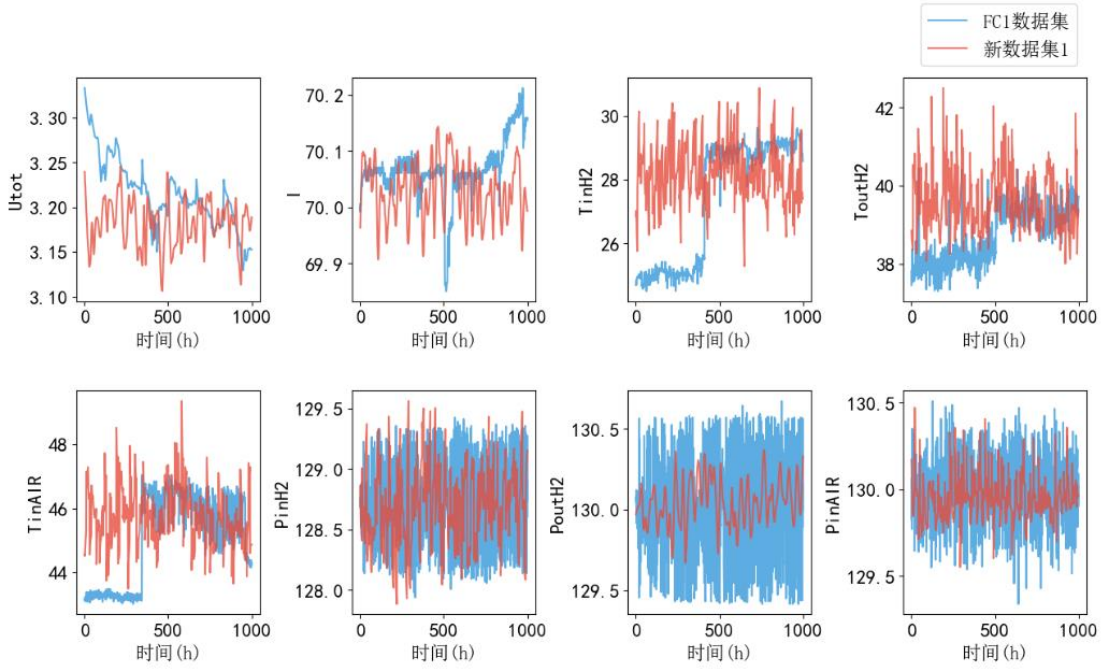


图 4-15 FC1 数据集与新数据集 1 的变量特征比较图

为了进一步说明数据集 1 在落点出的 FC1 数据集预测效果最好，本节将该数据集在三个模型上的预测结果综合在一张图上，如图 4-16 所示。

通过此图可以看出，新数据集 1 使用 FC1 数据集模型的预测效果最佳，而使用其他两个数据集模型的预测的结果与实际数据存在较大的偏差。这也更好的说

明了本文的研究思路的正确性，即通过数据集的区域归属选择对应的模型能够有效提高预测精度。

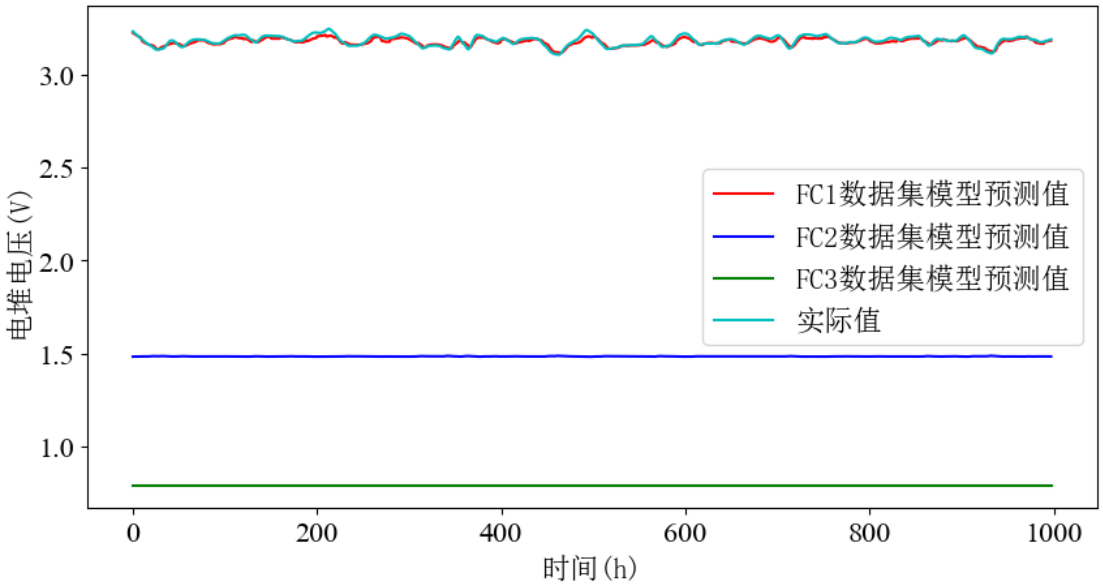


图 4-16 新数据集 1 在三个模型上的预测结果

为了全面评估模型的性能，本节还对新数据集 2 和新数据集 3 的预测效果进行了分析。鉴于新数据集 2 和新数据集 3 分别呈现出与 FC3 数据集和 FC2 数据集的相似性，于是采用这两个数据集的模型对它们进行了预测。其预测结果如图 4-17 和图 4-18 所示。

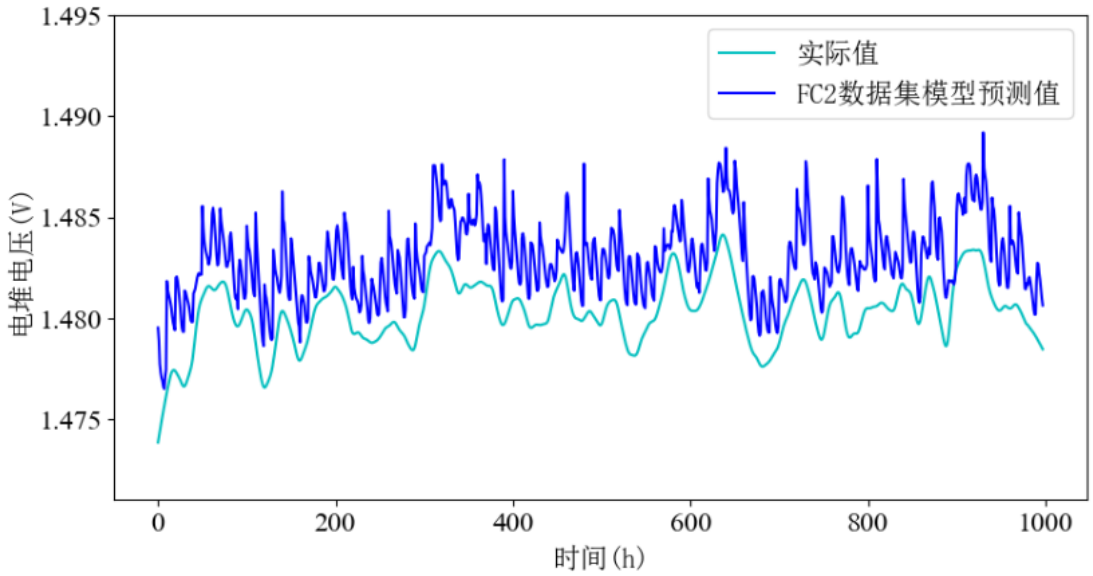


图 4-17 FC2 数据集模型预测效果

通过这两张图的可以看出，这两个模型在对新数据集进行预测时，都能够很好的预测数据集的趋势。尽管预测数据出的数值范围和实际数据存在一定的偏差，但是整体的趋势吻合度非常的高。这一结果可以通过分析图 4-13 的相似性图来解

释，这两个新数据集虽然分别接近用于预测的两个数据集，但是它们之间相距较远，这说明两个数据集之间存在较大的差异。

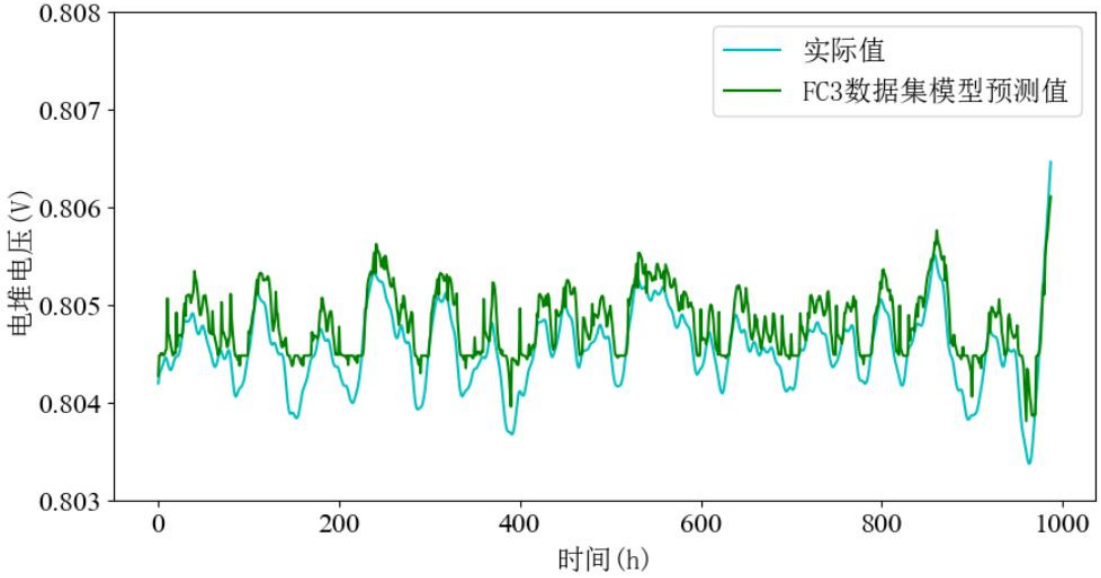


图 4-18 FC3 数据集模型预测效果

总的来说，本文提出的基于相似性分类的氢燃料电池寿命预测通用模型在预测效果上达到预期的目标。对于新数据集 1 的预测结果的 R^2 达到了 0.872，并且整体的贴合效果非常的好。其他的两个新数据集模型也能很好的捕捉到数据整体的变化趋势，表现出很强的泛化能力。这充分的表明当前模型已经具备了一定的预测能力。

当然，该通用模型还存在一定的局限性，尤其是在数据量方面。有限的数据集使得相似性图显示得很稀疏，不同的电堆老化数据之间在图上相距较大，这就到导致无法准确找到最优预测模型，进而影响到最后预测效果。随着未来数据集规模的扩大、老化情况的丰富以及新的预测模型的引入，该模型的预测精度和可靠性会得到很大提升，从而满足实际应用的需求。

4.6 本章小结

本章主要围绕构建基于相似性分类的氢燃料电池寿命预测通用模型的预测模型部分展开，通过一系列的优化和改进，成功提升模型的预测能力和泛化能力。首先是在异常值处理后的数据基础上进行数据重建和平滑处理，去除掉数据的噪声和尖峰，使数据的趋势更加明显，更符合预测模型的数据要求。

处理完数据之后，就开始建立 LSTM 模型，通过分析三个数据集的 LSTM 模型的预测结果发现，该模型因为使用三个数据集的公用变量而导致数据信息缺失，从而造成模型在测试集上的泛化能力较弱。为解决这个问题，本章在 LSTM 层之后加入了能够挖掘现有数据中隐藏规律和非线性关系的 KAN 层，力求能够提升模

型的泛化能力。最后，经过分析发现加入 KAN 层之后形成的 LSTM-KAN 模型预测能力和泛化能力有了明显的提升，并决定将其作为基于相似性分类的氢燃料电池寿命预测通用模型的预测部分模型。

构建了完整的基于相似性分类的氢燃料电池寿命预测通用模型之后，就需要对该模型的应用能力进行分析。于是，本章利用对抗生成网络生成三组特色各异的数据，然后使用主成分分析法将这些数据降维映射到相似性图中，并分析出每个新数据集对应的区域，随后调用对应区域的预测模型进行预测，最后发现对应的模型能够很好的预测出新数据的趋势，展现出较强的泛化能力。这一结果表明，本文构建的基于相似性分类的氢燃料电池寿命预测模型已经具备了一定的预测能力，能用于实际应用。

第 5 章 基于 Python 的相似性分类的氢燃料电池寿命预测软件开发

第三章通过箱线图分析将数据中的异常值剔除，利用主成分分析将三组数据降维形成相似性图，同时对相似性图进行区域划分；第四章则是在划分区域后的相似性图的基础上构建出三个区域的预测模型，形成了完整的基于相似性分类的氢燃料电池寿命预测通用模型。然后，通过使用对抗生成网络生成的新数据验证该模型的可行性。然而，这些研究结果尚未实现工程化应用，缺乏一个便于操作和实际应用的界面。

因此，本章将在前两章的基础上，利用 Python 中的 Streamlit 框架开发出基于相似性分类的氢燃料电池寿命预测通用模型的软件界面，实现寿命预测的工程化应用。该软件界面命名为 Hydrogen Cell LifeX，其主要功能包括箱线图分析、PCA 降维相似性分析、模型预测结果分析。最后通过输入数据对软件进行测试，判断各个功能是否正常运行、结果是否准确。

5.1 Hydrogen Cell LifeX 各功能模块的程序设计

开发的氢燃料电池寿命预测系统软件 Hydrogen Cell LifeX 主要包括三个功能模块，分别为：箱线图分析、PCA 降维相似性分析、模型预测结果分析。这三个功能板块主要对应新数据读入后从处理到预测的三个流程。

（1）箱线图分析

箱线图分析主要用于对新数据进行前期的数据清洗，对异常值进行剔除，其分析程序的设计流程如图 5-1 所示。此部分主要是通过箱线图找出数据中的异常值，然后通过 3.1.2 节中提到的图基法进行异常值剔除。此外，系统还将输出数异常值的统计表格，同时向用户提示当前分析功能已执行完毕。

（2）PCA 降维相似性分析

PCA 降维相似性分析主要用于对剔除异常值后的新数据进行降维映射到相似性图上，然后系统将会分别计算新数据与三个数据集与质心之间的欧氏距离，通过比较距离的长短，来自动判断该新数据更接近与哪一类数据，从而确定其在相似性图中的落点位置以及所属区域。同时也会输出相似图以及系统判断的结果，供用户参考判断。其分析程序的设计流程如图 5-2 所示。

（3）模型预测结果分析

模型预测结果分析主要根据 PCA 降维相似性分类的结果，调用其对应数据集的模型对新数据集进行预测，生成寿命预测的结果和相关的评价指标，让用户判

断预测效果的好坏。其分析程序的设计流程如图 5-3 所示。

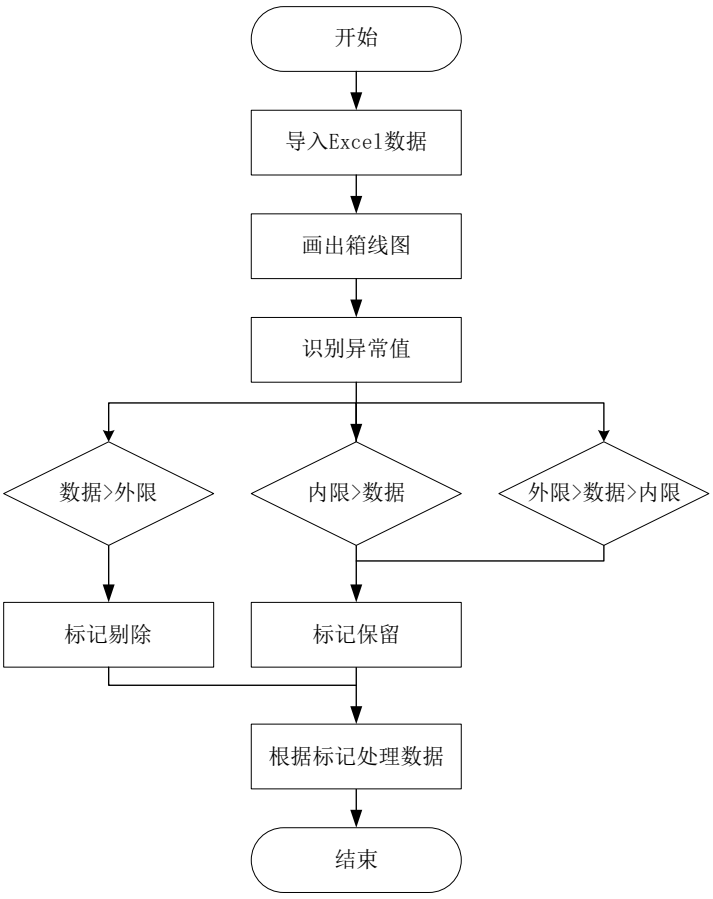


图 5-1 箱线图程序设计流程图

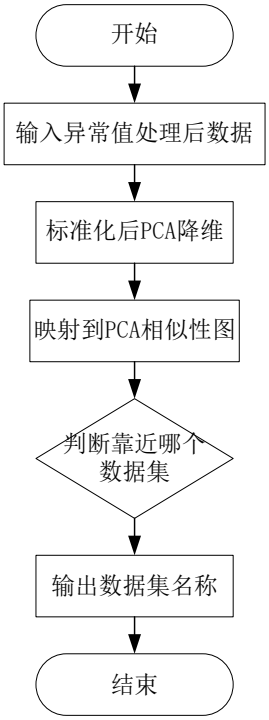


图 5-2 PCA 降维相似性分析程序设计流程图

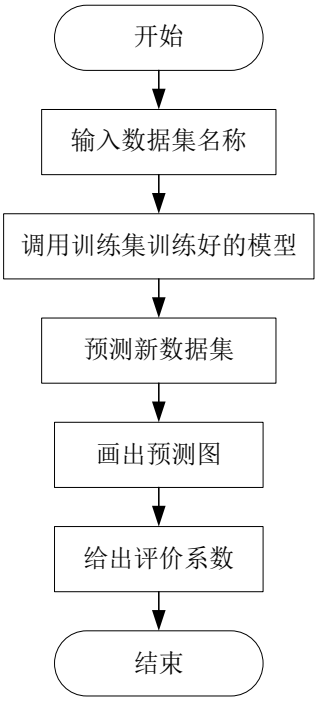


图 5-3 模型预测结果分析程序设计流程图

5.2 Hydrogen Cell LifeX 的界面设计

完成所有的功能模块的程序开发后，就需要一个工具来将代码和其他资源整合起来，生成一个可以简易操作的软件界面。这样不仅可以管理和保护代码，还可以确保软件在不同的环境下稳定运行，从而实现通用模型的工程化应用。基于快速开发的考虑，本章采用 Python 中的 Streamlit 库来进行界面设计开发。

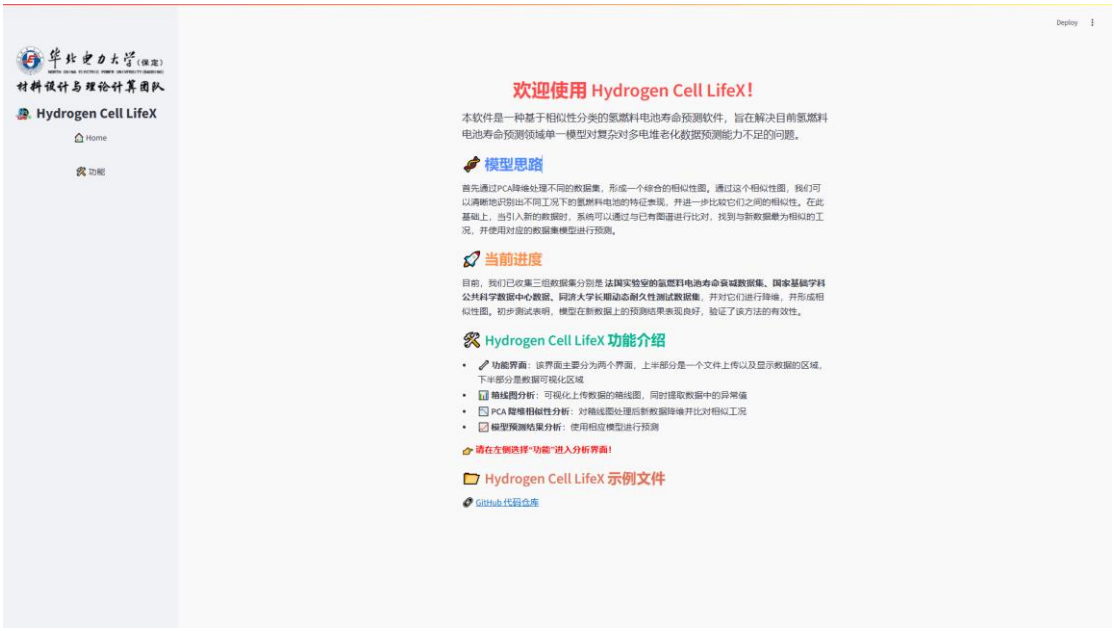


图 5-4 软件的 Home 界面

该软件界面主要分为两个界面，一个初始界面 Home，另一个是功能界面。初始界面 Home 是用户进入软件后的看到的第一个页面，主要用于介绍软件的背景信息、功能以及当前的开发进度。该界面也简要的阐述了基于相似性分类的氢燃料电池寿命预测通用模型的思路，并展示了该软件的核心功能。同时，在界面的最下方，提供了一个示例文件的下载链接。用户可以通过点击链接直接跳转到数据的下载界面下载数据。然后按照本章的功能介绍顺序进行操作使用。初始界面 Home 如图 5-4 所示。

功能界面则是需要点击侧边栏的“功能”按钮才会出现的功能界面。功能界面主要包括五个功能。第一个是读取文件，点击后就会弹出选择框，让用户选择要进行处理的数据文件，并且要求文件大小不能超过 200 MB，且必须为 csv 或者 xlsx 文件；其中三个则是该软件的核心功能，分别是：箱线图分析、PCA 降维相似性分析、模型预测结果分析；第五个则是三个核心功能实现之后会提供图片下载功能，方便用户保存当前的分析结果。同时该界面分为上下两部分，上部分为文件读取，下部分则是数据可视化区域，即核心功能画图区域。功能界面如图 5-5 所示。

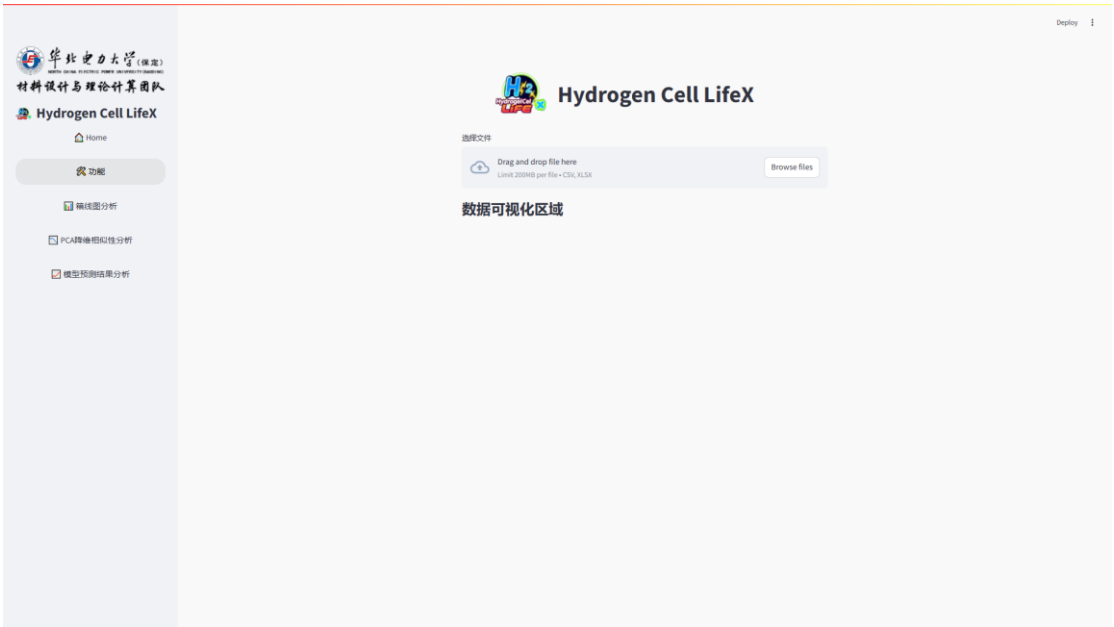


图 5-5 软件的功能界面

设计完后的软件需要测试每个功能是否能够正常的运行和显示，这里先将点击侧边栏的“功能”按钮，进入功能界面。然后按照次序执行每一个功能板块，来测试界面交互和该软件的每个功能是否正常运行。以下是具体的测试步骤和流程说明：

首先测试的就是读入文件部分。用户需点击“选择文件”下的长条按钮，弹出文件选择对话框，选择需要导入的 Excel 数据文件进行上传。系统会自动读取文

件内容并解析数据。如图 5-6 所示，数据上传成功后，界面会清晰的展示上传文件的名称、占用内存的大小以及具体的数据预览内容，方便用户查看数据是否符合输入要求。

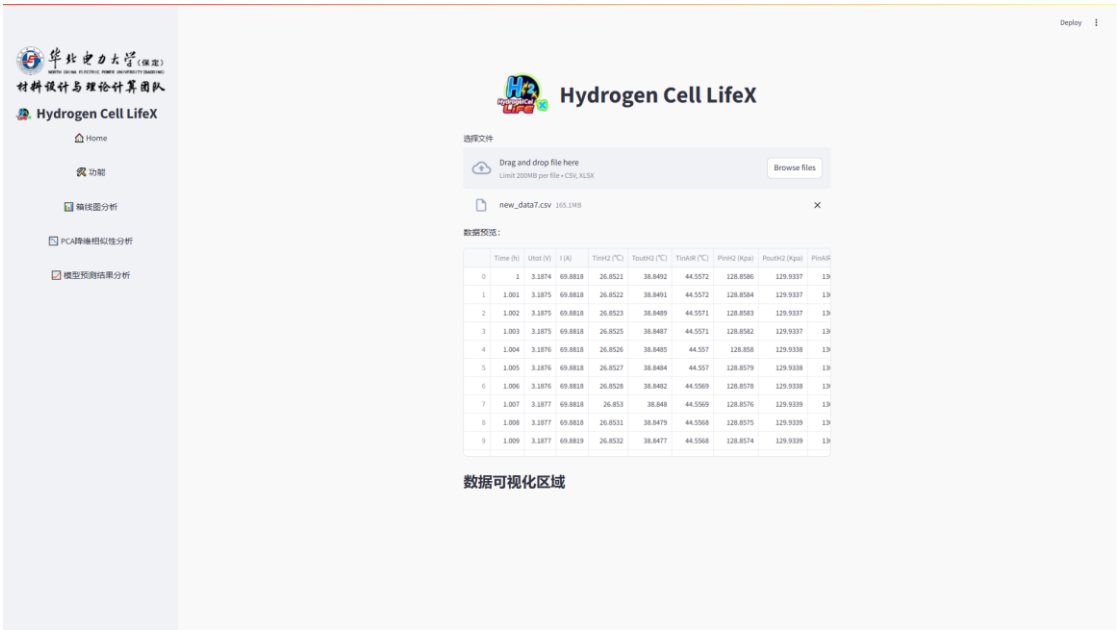


图 5-6 读入文件的运行结果

图 5-7 箱线图分析的运行结果

读入数据之后就需要对数据进行数据处理，这时就可以点击侧边栏的“箱线图分析”按钮。程序接收到指令后，就会将读入的数据传入到箱线图分析的流程中进行分析。然后，就会在数据可视化区域绘制出箱线图，同时提供异常值的统计表格以及提醒用户“异常值已处理，数据已更新！”，如图 5-7 所示。当显示出该

提示信息时，就说明后台数据已经完成了更新，后续将使用更新后的数据进行操作。

除此之外，最后会生成一个导出图片按钮，能够把输出的图保存到本地。点击按钮后，就会弹出另存为的窗口，用户可以选择要保存的文件夹。每个功能下绘制出的图都有自己默认的图片名字，当然，这个用户也可以根据自己的需求修改，具体下图 5-8 所示。

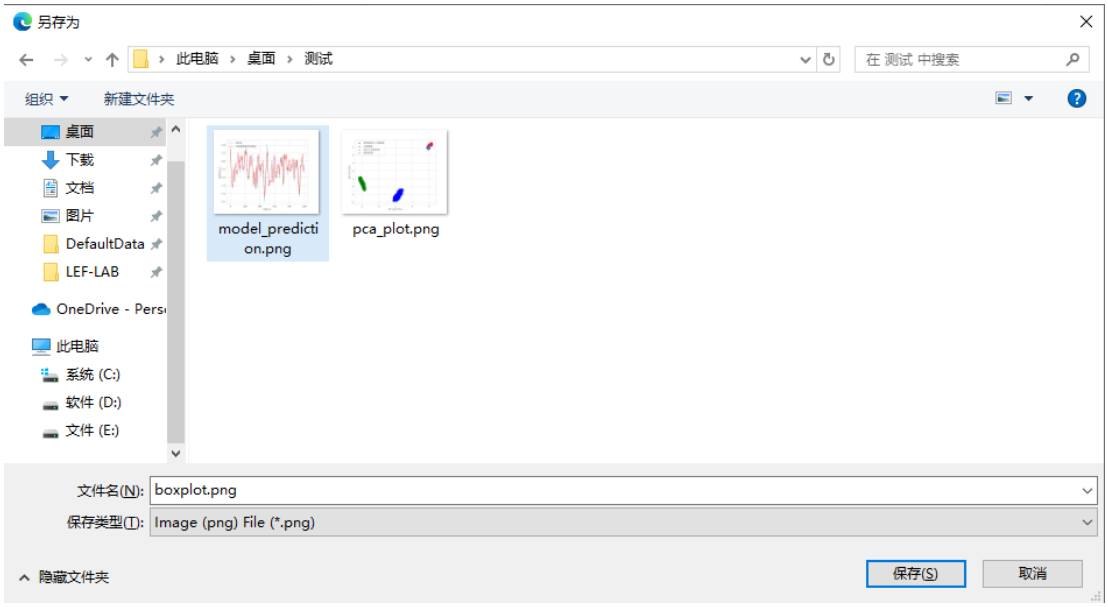


图 5-8 导出图片的运行结果

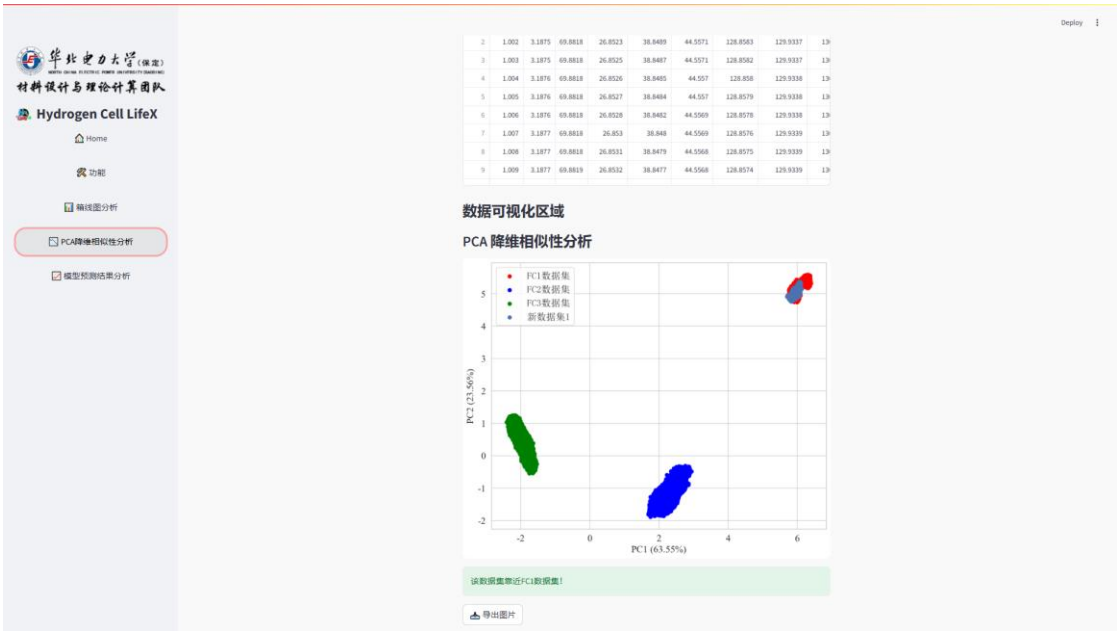


图 5-9 PCA 降维相似性分析的运行结果

完成箱线图分析处理完异常值之后，就需要进行 PCA 降维相似性分析。点击

侧边栏相应的按钮，就会将更新后的数据传入 PCA 降维分析的流程中，随后，就会数据可视化区域生成相似性图，同时系统会自动计算传入的数据和各个数据集之间的欧氏距离来判断新数据更为靠近哪个数据集，具体结果如图 5-9 所示。从图中可以看到，新数据的落点和 FC1 数据集相重合，下方也弹出提醒用户“该数据集靠近 FC1 数据集!”。同样的最后也生成了一个导出图片的按钮，方便用户导出分析图。

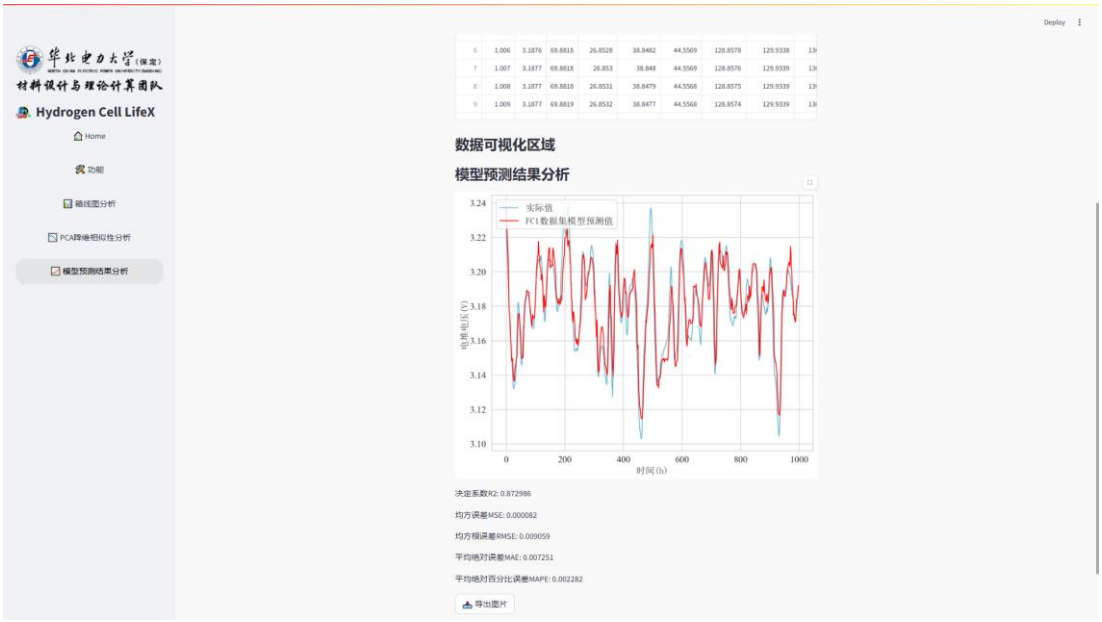


图 5-10 模型预测结果分析的运行结果

上一步骤得到新数据集靠近 FC1 据集的结果后，就可以点击模型预测结果分析的按钮了，这时就会调用模型内置的 FC1 据集模型对新数据进行预测，并生成出预测的结果图和相应的评价指标，具体结果如图 5-10 所示。从图中可以看出模型的预测值，与实际值贴合效果很好，并且相关评价指标都满足预测预期，这表明该软件实现的预测效果很出色。

从上面每个功能的测试效果来看，该软件的能够很好的实现每一个数据分析的功能，且界面交互非常的流畅，能够用于氢燃料电池系统的老化数据分析和寿命预测。

5.3 本章小结

本章主要利用 Python 中的 Streamlit 库来进行氢燃料电池寿命预测软件界面的开发。随后，通过输入示例数据来验证开发的软件是否能够正常运行。结果表明，软件的每个功能都能够正常运行，且生成出来的分析图和结论都符合预期，并且能正确的输出在指定区域。这一结果表明该软件可以正常运行实现对氢燃料电池

的寿命预测。该软件的界面设计得非常的简洁，界面的交互流畅，响应速度快，操作简便，具备良好的工程化应用潜力。

第 6 章 结论与展望

6.1 结论

随着我国能源结构调整和优化的步伐不断加快，以及在国家“双碳”战略的推动下，全社会对于新型清洁能源技术的关注度不断上升。然而，交通运输领域的碳排放问题依然严峻，急需找到有效的解决方法。在这种情况下，氢燃料电池凭借其清洁、高效的特性，被认为是取代传统能源的理想选择，但其较短的使用寿命却制约了它的大规模商业化进程。因此，如果能够实现对氢燃料电池寿命的精确预测，就可以提前发现电池性能的下降或潜在的故障，从而有效降低维护成本并延长其使用寿命。这不仅对推动氢燃料电池的技术发展具有重要意义，而且有助于其更广泛的市场推广。然而，目前大多数研究只关注单一氢燃料电池电堆的寿命预测，而缺乏对氢燃料电池寿命预测通用模型深入探讨，存在很大局限性。

为解决这种现状，本文提出了一种基于相似性分析的氢燃料电池寿命预测的通用模型。该模型首先通过降维技术将多组数据映射到低维空间，形成相似性图。然后通过分析数据之间的差异进行相似性图的区域划分。随后，对划分的区域分别训练预测模型，形成完整的通用模型。当新的电堆数据输入时，可将其降维后映射在相似性图上找到相似电堆数据区域，并调用对应的模型进行预测。本文主要研究内容与结论如下：

（1）本文采用了三组数据集，分别是法国实验室的氢燃料电池寿命衰减数据集、国家基础学科公共科学数据中心数据集、同济大学长期动态耐久性测试数据集，并分别命名为 FC1 数据集、FC2 数据集、FC3 数据集。首先对这三个数据集提取了公共变量，统一其变量名和单位。之后分别画出箱线图，观察每个变量的离散情况，再使用图基法处理异常值，保障数据质量。然后通过变量相关性分析，确定采用 PCA 降维。使用 PCA 降维对三组数据集进行降维得到相似图和 PC1、PC2 这两个主成分，分析发现它们合计能够解释 87.11%的总方差，同时也发现主成分 PC1 主要反映了能量转换与输出相关因素，主成分 PC2 则侧重于热管理效率方面因素。最后，根据分析的结果将得到的相似性图划分了三个区域，分别是：高温低压区、低温高压区以及低热管理效率区。

（2）利用数据重构降低数据量并抽取代表性的数据，即将所有测点用 1 小时为间隔做采用。之后通过局部散点平滑估计法对重建数据进行平滑处理来提升数据质量。平滑后的数据不仅保留了原数据的主要趋势，还消除了噪声和尖峰。随后将数据集按 6:4 的比例划分为训练集和测试集，并利用内置函数将数据进行归一化处理。

（3）然后，通过 LSTM 搭建寿命预测模型，对划分的这三个区域的数据进行了训练。结果显示，模型在训练集上的拟合效果非常好，决定系数 R^2 分别为 0.972、0.981 和 0.951。然而，在测试集上，尽管拟合效果仍然较好，但决定系数 R^2 下降至 0.932、0.920 和 0.947，其他的评价指标也随之变差。这表明模型在训练集上表现优异，但在测试集上的泛化能力尚需改进。

（4）为提升模型在测试集上的泛化能力，本文在 LSTM 层之后加入 KAN 层，构建成 LSTM-KAN 寿命预测模型，并将得到的预测结果与 LSTM 模型的结果进行对比。结果表明，无论是在训练集还是测试集上，LSTM-KAN 模型的表现都更为出色，尤其是在转折点处，预测值与实际值贴合度增高了。尤其是在测试集上的决定系数 R^2 分别从 0.932、0.920、0.947 提升到了 0.959、0.953、0.955，其它评价系数也得到显著的提升。这些结果表明，LSTM-KAN 模型能够更加准确地预测氢燃料电池的寿命。于是，把 LSTM-KAN 模型作为基于相似性分析的氢燃料电池寿命预测的通用模型预测部分的模型。

（5）为了验证基于相似性分析的氢燃料电池寿命预测的通用模型的适用性。本文采用了 GAN 生成了三组新数据，并按照对应的步骤进行了验证。结果表明模型在预测效果上达到预期的目标，对于新数据集 1 的预测结果的 R^2 达到了 0.872，并且整体的贴合效果非常的好。其他的两个新数据集模型也能很好的捕捉到数据整体的变化趋势，表现出很强的泛化能力。这充分的表明当前模型已经具备了一定的预测能力。

（6）为了便于工程化应用，利用 Python 中的 Streamlit 设计了 GUI 交互界面，开发出了氢燃料电池寿命预测软件，软件的主要功能包括箱线图分析、PCA 降维相似性分析、模型预测结果分析。通过输入数据对软件进行测试，发现软件的各个功能都能正常运行，且运行结果准确，可以用于氢燃料电池系统的数据分析和寿命预测。

6.2 展望

本文利用了三个数据集搭建了基于相似性分类的通用模型，并通过在 LSTM 模型中加入 KAN 层提升了通用模型的性能，但受限于研究条件与研究时间，加之自身的经验和能力不足，本文中还有部分问题需进一步展开研究，主要有：

（1）受限于自身经验与时间问题，本文仅采用了三个氢燃料电池数据集，因此降维后形成的相似性图存在很多的局限性，实际上这样的相似性图应该是由很多不同的电堆老化数据集形成，这样才能更好的反映出各电堆之间的差异性和相似性。因此后续应该采集更多的数据展开研究，同时也需要寻找更好的方法来揭示不同电堆之间的关系以及模型，进一步改进当前模型形成一个更好、更实用的

通用模型。

（2）本文开发出的氢燃料电池寿命预测软件，仅实现了从数据读入到预测的过程，这对于当今智能化的社会来说，实用性略显不足。因此后续应该在此基础上增加更多的功能，尤其是模型自更新的功能，能不断学习新读入的数据，不断完善自己模型，同时也提供给用户更多的模型选择。

参考文献

[1] 唐幸. “双碳”背景下应多措并举推进氢燃料电池汽车高质量发展[J]. 中国经贸导刊, 2025, (01): 49-51

[2] 付甜甜. 电动汽车用氢燃料电池发展综述[J]. 电源技术, 2017, 41(04): 651-653

[3] 邵志刚, 衣宝廉. 氢能与燃料电池发展现状及展望[J]. 中国科学院院刊, 2019, 34(04): 469-477

[4] 程一步. 氢燃料电池技术应用现状及发展趋势分析[J]. 石油石化绿色低碳, 2018, 3(02): 5-13

[5] Z. Li, J. Chen. Prognostics of PEM Fuel Cells Based on Gaussian Process State Space[J]. Energy, 2018, 149: 63-73

[6] FCLAB Research. IEEE PHM 2014 Data challenge. 2014

[7] Martin K E, Kopasz J P, McMurphy K W. Status of fuel cells and the challenges facing fuel cell technology today[J]. Fuel cell chemistry and operation, 2010, 1040: 1-13

[8] J. Wu, X. Z. Yuan, J. J. Martin, et al. A review of PEM fuel cell durability: Degradation mechanisms and mitigation strategies[J]. Journal of Power Sources, 2008, 184(1): 104-119

[9] M. W. Fowler, R. F. Mann, J. C. Amphlett, et al. Incorporation of voltage degradation into a generalised steady state electrochemical model for a PEM fuel cell[J]. Journal of Power Sources, 2002, 106(1): 274-283

[10] J. F. Wu, J. J. Martin, F. P. Orfino, et al. In Situ accelerated degradation of gas diffusion layer in proton exchange membrane fuel cell[J]. Journal of Power Sources, 2010, 195(7): 1888-1894

[11] Z. Y. Hu, L. Xu, J. Li, et al. A novel diagnostic methodology for fuel cell stack health: Performance, consistency and uniformity[J]. Energy Conversion and Management, 2019, 185: 611-621

[12] H. X. Lu, J. Chen, C. Z. Yan, et al. On-line fault diagnosis for proton exchange membrane fuel cells based on a fast electrochemical impedance spectroscopy measurement[J]. Journal of Power Sources, 2019, 430: 233-243

[13] N. H. Behling. Fuel Cells: Current Technology Challenges and Future Research Needs[M]. Elsevier Science, 2012

[14] D. M. Du-Plooy, J. Meyer. PEM fuel cells: Failure, mitigation and dormancy recovery understanding the factors affecting the efficient control of PEMFC[C]. 2017 IEEE AFRICON, 2017(9): 1119-1124

[15] Sutharssan T, Montalvao D, Chen Y K, et al. A review on prognostics and health monitoring of proton exchange membrane fuel cell[J]. Renewable and

Sustainable Energy Reviews, 2017, 75: 440-450

[16] Jouin M, Bressel M, Morando S, et al. Estimating the end-of-life of PEM fuel cells: Guidelines and metrics[J]. Applied Energy, 2016, 177: 87-97.m

[17] 曹宪林, 李服群, 张琨. 某机载雷达智能综合测试系统[J]. 计算机测量与控制, 2008, 16(5): 688-690

[18] Arias I K, Trinke P, Hanke-Rauschenbach R, et al. Understanding PEM fuel cell dynamics: The reversal curve[J]. International Journal of Hydrogen Energy, 2017, 42: 15818-15827

[19] Fan L, Tu Z, Chan S H. Recent development of hydrogen and fuel cell technologies: A review[J]. Energy Reports, 2021, 7: 8421-8446

[20] Sazali N, WanSalleh W N, Jamaludin A S, et al. New perspectives on fuel cell technology: A brief review[J]. Membranes, 2020, 10: 99

[21] Jia X, Liu X, Zhou Y. Performance Degradation and Life Prediction of Proton Exchange Membrane Fuel Cell. In Proceedings of the 2022 Prognostics and Health Management Conference (PHM-2022 London)[C], London, UK, 27-29 May 2022; pp. 433-437

[22] Wang, X. Remaining Useful Life Prediction of Proton Exchange Membrane Fuel Cell Based on Deep Learning. In Proceedings of the 2022 IEEE 5th International Conference on Electronics Technology (ICET)[C], Chengdu, China, 13-16 May 2022; pp. 290-296

[23] Wilhelm Y, Reimann P, Gauchel W, et al. Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge -based models[J]. Procedia CIRP, 2021, 99: 278-283

[24] Jin X B, RobertJeremiah R J, Su T L, et al. The new trend of state estimation: From model-driven to hybrid-driven methods[J]. Sensors, 2021, 21: 2085

[25] 刘浩. 质子交换膜燃料电池的寿命预测研究[D]. 浙江大学, 2019

[26] Robin C, Gérard M, Quinaud M, et al. Proton exchange membrane fuel cell model for aging predictions: Simulated equivalent active surface area loss and comparisons with durability tests[J]. Journal of Power Sources, 2016, 326: 417-427

[27] 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, 410-411: 78-90

[28] Bressel M, Hilairet M, Hissel D, et al. Extended Kalman filter for prognostic of proton exchange membrane fuel cell[J]. Applied Energy, 2016, 164: 220-227

[29] Ou M, Zhang R, Shao Z, et al. A novel approach based on semi-empirical model for degradation prediction of fuel cells[J]. Journal of Power Sources, 2021, 488:

229435

[30] 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

[31] Jouin M, Gouriveau R, Hissel D, et al. Joint Particle Filters Prognostics for Proton Exchange Membrane Fuel Cell Power Prediction at Constant Current Solicitation[J]. IEEE Transactions on Reliability, 2016, 65(1): 336-349

[32] 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

[33] Kui C, Laghrouche S, Djerdir A. Proton exchange membrane fuel cell degradation and remaining useful life prediction based on artificial neural network[C]//2018 7th International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, 2018: 407-411

[34] Liu J, Li Q, Chen W, et al. Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks[J]. International Journal of Hydrogen Energy, 2019, 44(11): 5470-5480

[35] Son H H. Toward a proposed framework for mood recognition using LSTM recurrent neuron network[J]. Procedia computer science, 2017, 109: 1028-1034

[36] Zhang Y, Xiong R, He H, et al. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries[J]. IEEE Transactions on Vehicular Technology, 2018, 67(7): 5695-5705

[37] Wu Y, Yuan M, Dong S, et al. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks[J]. Neurocomputing, 2018, 275: 167-179

[38] Wang F K, Cheng X B, Hsiao K C. Stacked long short-term memory model for proton exchange membrane fuel cell systems degradation[J]. Journal of Power Sources, 2020, 448: 227591

[39] Ma R, Yang T, Breaz E, et al. Data-driven proton exchange membrane fuel cell degradation predication through deep learning method[J]. Applied energy, 2018, 231: 102-115

[40] Meraghni S, Terrissa L S, Yue M, et al. A data-driven digital-twin prognostics method for proton exchange membrane fuel cell remaining useful life prediction[J]. International journal of hydrogen energy, 2021, 46(2): 2555-2564

[41] Sutharssan T, Montalvao D, Chen Y K, et al. A review on prognostics and health monitoring of proton exchange membrane fuel cell[J]. Renewable and Sustainable Energy Reviews, 2017, 75: 440-450

[42] Xie Y, Zou J, Peng C, et al. A novel PEM fuel cell remaining useful life prediction

method based on singular spectrum analysis and deep Gaussian processes[J]. International Journal of Hydrogen Energy, 2020, 45(55): 30942-30956

[43] Wang C, Li Z, Outbib R, et al. Symbolic deep learning based prognostics for dynamic operating proton exchange membrane fuel cells[J]. Applied Energy, 2022, 305: 117918

[44] 李鹏程. 基于循环神经网络氢燃料电池寿命预测技术[D]. 电子科技大学, 2021

[45] GouriveauR, Hilairet M, Hissel D, et al. IEEE PHM 2014 Data Challenge Details for participants, 2014

[46] 刘妹汝. 燃料电池寿命衰减模型数据集. (V1). 电子科技大学, 2023-03-07. 国家基础学科公共科学数据中心, https://cstr.cn/16666.11.nbsdc.vpbp27id

[47] 刘妹汝. 车用质子交换膜燃料电池电压衰退预测方法研究[D]. 电子科技大学, 2022

[48] Zuo J, Lv H, Zhou D, et al. Long-term dynamic durability test datasets for single proton exchange membrane fuel cell[J]. Data in Brief, 2021, 35: 106775

[49] Bloom I, Walker L K, Basco J K, et al. A comparison of Fuel Cell Testing protocols–A case study: Protocols used by the U.S. Department of Energy, European Union, International Electrotechnical Commission/Fuel Cell Testing and Standardization Network, and Fuel Cell Technical Team[J]. Journal of Power Sources, 2013, 243: 451-457

[50] 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

[51] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics and intelligent laboratory systems, 1987, 2(1-3): 37-52.

[52] Carreira-Perpinan M A. Continuous latent variable models for dimensionality reduction and sequential data reconstruction[D]. University of Sheffield, 2001

[53] Hochreiter S. Long Short-term Memory[J]. Neural Computation MIT-Press, 1997

[54] Liu J, Li Q, Chen W, et al. Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks[J]. International Journal of Hydrogen Energy, 2019, 44(11): 5470-5480

[55] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. Neural networks, 1989, 2(5): 359-366

[56] Liu Z, Wang Y, Vaidya S, et al. Kan: Kolmogorov-arnold networks[J]. arXiv preprint arXiv: 2404. 19756, 2024

[57] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144

[58] Hinton G E. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223

[59] 韩文煜. 基于python数据分析技术的数据整理与分析研究[J]. 科技创新与应用, 2020, (04): 157-158

[60] 张若愚. Python 科学计算: 第 2 版 [M]. 北京: 清华大学出版社, 2016

[61] Khorasani M, Abdou M, Fernández J H. Web application development with streamlit[J]. Software Development, 2022: 498-507

[62] Akkem Y, Kumar B S, Varanasi A. Streamlit application for advanced ensemble learning methods in crop recommendation systems–a review and implementation[J]. Indian J Sci Technol, 2023, 16(48): 4688-4702

[63] Celik N, Bayrak Z U, Tasar B, et al. Evaluation of the parameters of a PEM fuel cell system by using machine learning regression models[J]. Journal of Mechanical Science and Technology, 2025: 1-11

[64] Yang Y H, Lin Q, Zhang C F, et al. Study of the influence under different operating conditions on the performance of hydrogen fuel cell system for a bus[J]. Process Safety and Environmental Protection, 2023, 177: 1027-1034

[65] Wang R, Li K, Cao J, et al. Air supply subsystem efficiency optimization for fuel cell power system with layered control method[J]. Renewable Energy, 2024, 235: 121328

[66] Salam M A, Habib M S, Ahmed K, et al. Effect of temperature on the performance factors and durability of proton exchange membrane of hydrogen fuel cell: A narrative review[J]. Material Science Research India (Online), 2020, 17(2): 179-191

攻读硕士学位期间发表的论文及其它成果

（一）申请及已获得的专利

[1] 高正阳,卢焕,魏子涵,等. 一种电厂煤粉测试装置:中国, 22161316.6[P]. 2024-01-19.

[2] 高正阳,魏子涵,卢焕,等. 一种电厂煤粉在线检测装置:中国, 22161329.3[P]. 2024-03-01.

[3] 高正阳,魏子涵. 一种电力锅炉燃烧用双燃料燃烧系统:中国, 10386960.X[P]. 2024-06-18.

致谢

时光荏苒，岁月如梭。转眼间，三年的研究生学习生活即将画上句号。在这段时光里，我不仅学到了专业知识，更收获了宝贵的人生经验，在此，我衷心的感谢所有关心和支持我的人。

首先，我要感谢我的家人，你们一直以来都是我最坚强的后盾。无论我遇到什么困难，你们总能给予我最大的理解和支持。是你们的鼓励让我熬过来一个又一个难关，让我拥有和一切困难作斗争的勇气。

其次，我要感谢我的导师高正阳教授和杨维结副教授。从论文的选题到研究设计，再到论文修改，您们始终以严谨治学的态度和渊博的学识指引我。除了课题以外，还交给我很多横向课题，这不仅拓宽了我的知识视野，也提升了我的综合能力、实践能力和解决实际问题的能力。您们的悉心指导和无私帮助是我顺利完成学业的重要保障。

最后，我要感谢我的同学和朋友们。感谢刘昊、郑迪、刘凯龙、卢焕等师兄对我困难时的解惑，以及课题上的帮助；感谢师瑞阳、孙婷婷、孙逸啸、李翔、刘鑫塬、黄明烨、屈原正、刘飞洋、陈雨涛、甘志冉、葛晗、陈启勇等同窗三年以来的陪伴，感谢你们，有你们的每一天都充满了欢声笑语；感谢室友刘慧军、刘子瑞、宋文彬，感谢你们对我生活上的照顾。有你们的陪伴，我的研究生生活更加丰富多彩，感恩有你们的每一天！

承蒙诸位道友在修行路上慷慨相助，让我得以进阶。愿与诸君持剑并肩，共斩前路大妖！

华北电力大学学位论文答辩委员会情况

	姓名	职称	工作单位	学科专长
答辩主席	李杰	教授级高工	河北电力勘探设计院	氢能系统
答辩委员 1	吴兵	高级工程师	上海鲲华新能源科技有限公司	氢能系统
答辩委员 2	叶陈良	副教授	华北电力大学动力工程系	氢能系统
答辩委员 3	张炳东	副教授	华北电力大学动力工程系	氢能系统
答辩委员 4	董帅	副教授	华北电力大学动力工程系	氢能系统