

哈尔滨理工大学

毕 业 设 计

题 目： 基于知识蒸馏框架的锂电池健康

状况评估方法研究

院、系： 理学院、应用数学系

姓 名： 曾昱榆

指导教师： 谢巍

系 主 任： 毕卉

2024 年 06 月 13 日

基于知识蒸馏框架的锂电池健康状况评估方法研究

摘要

对锂电池的健康状况（State of Health, SOH）进行评估，对于电动汽车、可再生能源储存系统以及便携式电子技术等多个领域而言是极为关键的。对电池 SOH 进行准确的预测不仅有助于电池的寿命延长，还能提升能源使用效率和保障运行的安全性。然而，锂电池在工作过程中会经历不可挽回的化学变化和材料降解，这会降低其性能，因此开发出高效的 SOH 预测方法成为了研究的关键。

本文提出了三种不同的锂电池 SOH 评估方法。首先，本文提出了一个基于 SOH 退化原则的预测方法，这种方法利用随机微分方程 (SDE) 生成时间序列，以及使用双支 ResNet 模型使得模型在具有优秀表现的同时能一定程度上以数学形式解释模型工作原理。紧接着，本文探讨了基于片段数据的 LSTM-Attention 模型预测技术。最后，本研究采用了一个跨模态知识蒸馏的框架来融合这两种不同的模型。通过采纳异构模型间的知识传递方法，这一创新技术进一步优化了 SOH 预测性能。

此外，本文还详细推导了提出的随机微分方程-双支 ResNet 模型和 LSTM-Attention 模型的相关理论误差，填补了当前大部分神经网络模型在支撑理论方面的不足，并提升了该模型的可解释性与稳定性。

实验数据显示，在 Oxford 电池退化数据集中进行的测试里，知识蒸馏框架在 SOH 预测上的表现十分优秀，所有测试的电池的均方根误差 (RMSE) 均低于 0.01。特别是在 Cell2 和 Cell5 上，RMSE 数值分别为 0.0006 和 0.0009，这些数据明显优于传统的预测手段。该计算模型不仅在捕捉电池 SOH 的变化趋势方面表现突出，同时也在数值模拟上展示出优越的泛化性能。

关键词 锂电池，健康状况，知识蒸馏

Research on Lithium Battery State of Health

Estimation Methods based on Knowledge

Distillation Framework

Abstract

Evaluating the State of Health (SOH) of lithium batteries is critical in a number of areas, including electric vehicles, renewable energy storage systems, and portable electronics. Accurate prediction of the SOH of a battery not only helps to extend the life of the battery, but also improves energy efficiency and ensures operational safety. However, Li-ion batteries undergo irreversible chemical changes and material degradation during operation, which can degrade their performance, making the development of efficient SOH prediction methods critical to research.

In this paper, three different methods for SOH assessment of Li-ion batteries are proposed. Firstly, a prediction method based on the principle of SOH degradation is proposed, which utilizes Stochastic Differential Equations (SDEs) to generate the time series, as well as the use of a two-branched ResNet model that allows for excellent performance while explaining the model workings mathematically to a certain extent. Immediately following, this paper will explore the LSTM-Attention model's prediction techniques based on fragmented data. Finally, this study employs a cross-modal knowledge distillation framework to fuse these two different models. By adopting a knowledge transfer method between heterogeneous models, this innovative technique further optimizes the SOH prediction performance.

In addition, this paper also derives in detail the relevant theoretical errors of the proposed stochastic differential equation-dual-branch ResNet model and the LSTM-Attention model, which fills in the deficiencies of most of the current neural network models in terms of supporting theories and improves the interpretability and stability of the model.

The experimental data show that the knowledge distillation framework performs very well in SOH prediction in the tests conducted on the Oxford battery degradation dataset, and the root mean square error (RMSE) of all the tested cells is lower than 0.01. Particularly, the RMSE values of Cell2 and Cell5 are 0.0006 and 0.0009, respectively, which are significantly better than those of traditional prediction means. The computational model not only excels in capturing the trend

of cell SOH, but also demonstrates superior generalization performance in the application area of numerical simulation.

Keywords Lithium-ion Batteries, State of Health, Knowledge Distillation

目 录

摘要.....	I
Abstract	II
第 1 章 绪论	1
1.1 课题背景及研究意义.....	1
1.2 国内外研究现状分析.....	1
1.3 论文内容与结构安排.....	3
第 2 章 基于退化机制的 SOH 预测方法	4
2.1 SOH 的相关定义.....	4
2.2 基于随机微分方程的时间序列生成器.....	4
2.2.1 伊藤过程.....	5
2.2.2 深度欧拉表达.....	5
2.3 双支 ResNet.....	6
2.4 损失函数.....	8
2.5 误差分析.....	9
2.5.1 泛化误差分析.....	9
2.5.2 模型训练稳定性分析.....	10
2.5.3 随机微分方程估计误差分析.....	12
2.6 实验验证.....	13
2.6.1 数据来源.....	13
2.6.2 优化器.....	14
2.6.3 训练策略.....	15
2.6.4 评价指标.....	16
2.6.5 实验结果与分析.....	16
2.7 本章小结.....	17
第 3 章 基于片段数据的 SOH 预测方法	19
3.1 模型结构.....	19
3.1.1 LSTM.....	19
3.1.2 Attention 机制.....	20
3.1.3 LSTM-Attention.....	21
3.2 损失函数.....	23
3.3 误差分析.....	23
3.4 实验验证.....	24
3.4.1 训练策略.....	24

3.4.2 实验结果与分析.....	24
3.5 本章小结.....	25
第 4 章 基于知识蒸馏框架的 SOH 预测方法.....	26
4.1 知识蒸馏框架.....	26
4.2 模型结构.....	27
4.2.1 教师网络.....	27
4.2.2 学生网络.....	28
4.3 损失函数.....	28
4.3.1 软损失.....	29
4.3.2 硬损失.....	29
4.3.3 总损失.....	29
4.4 实验验证.....	30
4.4.1 训练策略.....	30
4.4.2 评价指标.....	31
4.4.3 实验结果与分析.....	33
4.5 本章小结.....	35
结论.....	36
致谢.....	38
参考文献.....	39
附录 A.....	42
附录 B.....	50

第1章 绪论

1.1 课题背景及研究意义

锂电池 (LIBs) 现如今是所有现代化技术中的核心元素, 无论是智能手机还是电动汽车(EVs)、再生能源的存储方式或者是手持电子设备, 锂电池都扮演着至关重要的角色, 因为它们为我们提供了几乎全部的能量来源。随着这些设备被广泛用于各种场合, 准确预测电池的健康状况 (State of Health SOH) 变得尤为关键。这不仅有助于更有效地进行电池健康管理, 延长电池的使用寿命, 还可以提高整体能效, 并确保能源供应的稳定性。

但是, 考虑到电池在实际操作中会遭遇一系列无法逆转的化学变化、电极使用物质的退化以及电极表面的副产品生成, 这都不可避免地导致电池的性能和安全隐患逐步减少。因此, 对电池的 SOH 进行精确预测, 在保证电池可靠性、延长其耐用寿命以及在节约成本和确保环境可持续性方面, 都是至关重要的。

如今, 针对 LIBs SOH 的预测手法可以主要归类为由模型驱动和由数据驱动的两类。模型驱动型方法主要涉及经验模型、物理电化学模型以及融合模型等。这些方法都运用了过去的数据和实验研究来构建关系模型, 虽然这些模型在某种程度上有一定的适应性, 但由于其对数据以及研究人员的严苛要求, 因此在实践中具有一定局限性。除此之外, 近些年来数据驱动方法已经获得了令人瞩目的进展, 这是通过大量的数据集合和机器学习算法来为电池 SOH 研究提供宝贵的洞见和理解的。然而, 这些技术也面临着特征筛选和提取的问题。传统的机器学习方法大多依赖于手工特征提取, 深度学习方法可以自动从初级数据中提取高级特征, 但当处理具有复杂时间序列特点的数据时, 这种方法还具有一定局限性。

1.2 国内外研究现状分析

如上文所述, 目前 SOH 预测方法主要分为模型驱动和数据驱动。模型驱动方法主要利用经验模型、物理-电化学模型和融合模型, 并结合先验知识或自适应滤波器, 以达到精准预测 SOH 目的^[1]。经验模型是基于历史数据和实验结果构建和预测关系, 包括指数模型^[2]、多项式模型^[3]和集成模型^[4]。这些模型表现出较好的适应性, 且需要的领域特定知识较少。韩槿萱 (2023) 提出了一种基于分数阶模型的锂电池荷电状态 (SOC) 和 SOH 协同估算方法。通过分数阶模型描述电池的 electrochemical 行为, 利用协同估算算

法提高 SOC 和 SOH 的预测精度^[5]。朱浩然等（2023）提出了一种结合差分热伏安法和 GPR 的锂电池 SOH 估算方法。通过差分热伏安法进行特征提取，然后利用 GPR 模型进行 SOH 估算^[6]。李煜等（2024）出了一种基于扩展卡尔曼滤波（EKF）算法的纯电动汽车锂电池 SOC 与 SOH 联合估算方法。通过构建二阶 RC 等效电路模型，并进行参数辨识，进行 SOC 估算^[7]。然而，由于它们本质上只是描述数据关系而非解释内在原理，因此它们的理论基础并不完善。此外，经验模型的效果极大地依赖于输入数据的完整性和模型质量。数据中的噪声、错误、缺失值或不适当的建模假设都可能显著降低其性能。然而，现实世界中的 SOH 数据往往存在噪声且可能不连续，同时很难找到合适的模型假设，使得经验模型不适合实际的 SOH 预测。

近年来，数据驱动的方法在预测电池 SOH 上也得到了广泛应用，其避免了对复杂物理现象进行深入理解的必要性。这种方法显著提高了模型在各种现实条件下的泛化能力。这些方法旨在表征 SOH^[8]。数据驱动的方法大致分为传统的机器学习（ML）方法和深度学习（DL）技术。传统的 ML 方法包括但不限于支持向量机（SVM）^[9]、随机森林（RF）^[10]和高斯过程回归（GPR）^[11]，在表征非线性关系和识别内在模式方面表现出强大的能力。然而，这些技术高度依赖于适当的特征选择；不相关或冗余的特征都会显著影响模型性能。由于这些模型常常需要手动选择和提取特征，可能会耗费时间并需要使用者对数据和任务有深入的理解。相比之下，深度学习方法，如深度神经网络（DNN）^[12]、卷积神经网络（CNNs）^[13]、长短时记忆网络（LSTMs）^[14]和 Transformer^[15]一定程度上缓解了这一问题。方斯顿和他的团队为 SOH 估计提出了一种融合双向长短时记忆网络（Bi-LSTM）的新技术^[16]。这些技术能够自主学习和从原始数据中提取高级特征，从而显著减少了手动特征提取的负担。此外，这些深度学习模型在 SOH 预测中都有着独特的优势。例如，CNNs 具有强大的多尺度特征提取能力，其强大的能力在图像处理 and 充电状态（SOC）预测领域得到了证明^{[17][18]}。然而，由于 CNNs 对输入序列顺序不敏感，限制了其在电池健康估计中的应用，无法辨别数据序列中的时间依赖性^[19]。相反，时间序列模型如 LSTMs 和循环神经网络（RNNs）天生适合处理时间数据，其在石油价格预测、能源负载预测和脑电图（EEG）数据异常检测等许多领域的效率得到了验证^[19]。SOH 数据本质上就是时间序列数据，如电压、电流和温度等特征会随着时间的推移而变化，同时 SOH 预测任务已被证明适合使用时间序列模型进行分析^[18]。为了结合各种模型的优势，许多研究人员探索了混合模型，如 CNN-LSTM^[20]、卷积双向门控循环网络^[21]和 BiLSTM-Attention^[22]，将多种架构融合在一起以增强模型能力。为了增强 SOH 预测的准确度，潘子良和他的同事提出了结合了贝叶斯优化和 LSTM 的混合预测方法。在这之中，贝叶斯优化算法被引入 LSTM 的超参数调整中，有效

地解决了因主观经验引发的参数调整问题^[23]。辛付宇及其团队提出了一个融合 CNN 和门控循环单元 (GRU) 的综合模型用于锂电池的 SOH 预估。在其中, CNN 被应用于特征抽取中, GRU 处理则针对长时间的依赖性进行处理^[24]。莫易敏及其团队提出了一种融合迁移学习和 GRU 神经网络的新方法。此方法通过对电池相似度进行分类和构建相应的迁移学习战略, 从而解决了锂电池 SOH 估值中的难题^[25]。雷孟飞等 (2024) 提出了一种结合自注意力机制和 GRU 神经网络的方法, 其中, 自注意力机制用于增强模型表征能力, GRU 用于处理时间序列数据^[26]。然而, 这些组合模型通常类似于几个模型的简单叠加, 并不一定能充分利用子模型的各自优势。且对于大部分数据驱动模型而言, 缺少相关的理论基础使其最大的短板, 因此如何开发出同时具有优异表现和一定理论基础的模型是目前研究的关键。

1.3 论文内容与结构安排

本文基于锂电池高速发展的背景下, 对锂电池 SOH 估计进行了详细研究, 并提出了基于数据驱动深度学习的三种预测方法。本文的结构安排如下:

- 1) 第 1 章阐述了课题背景与研究意义, 并对国内外研究现状进行了分析。
- 2) 第 2 章提出了基于退化机制的 SOH 预测方法, 详细介绍了基于随机微分方程的时间序列生成器以及双支 ResNet, 并对损失函数进行了说明。同时对模型进行了误差分析, 并通过实验验证模型的有效性。
- 3) 第 3 章提出了基于片段数据的 SOH 预测方法, 介绍了基于 LSTM-Attention 的模型, 并对模型进行了误差分析和实验验证。
- 4) 第 4 章介绍了基于跨模态知识蒸馏框架的 SOH 预测方法, 从知识蒸馏框架介绍, 模型构成, 损失函数, 实验验证四个方面进行了阐述了分析。

第2章 基于退化机制的 SOH 预测方法

2.1 SOH 的相关定义

首先要清晰地了解 SOH 的相关定义。SOH 在电池管理系统里是一个普遍被采纳的专业词汇，用于描述电池与理想新电池进行的综合性能对比。在一般情况下，锂电池的 SOH 值会随着充放电周期的增长而逐步减小。电池运行的环境变量、工作温度、充电效率、循环充电频次以及放电的深度等，都有可能影响其 SOH 值。锂电池的 SOH 值能通过众多方式来评估和量化，其中，通过观察电池容量的减少来识别 SOH 是最普遍的方式。另外一些技术手段包括但不限于电压的降解分析、内部阻抗的测定以及利用电化学阻抗光谱分析法等。锂离子电池的 SOH 值可以使用各种技术来估量和测定。锂电池的 SOH 对电池的运用与维护起着至关重要的作用。这种技术为用户提供了了解电池的工作状态并进行维护的机会，同时也对于电池的稳定性和使用寿命的预估至关重要。当 SOH 值减少到某个特定水平，可能需要实施一些行动，例如更换新电池或者进行适当的维护工作。SOH 通常以百分比方式来标识，而当其达到 100% 时，意味着电池正处于崭新或最佳的状态。在操作实施过程中，常常通过设定 SOH 的阈值来确定电池的健康状态。这些阈值可以按照应用需求和电池类型的不同进行调整。比如说，在电动汽车的锂电池中，一旦 SOH 降低到大约 80%，通常人们会考虑考虑替换或者修复这块电池。本质上看，SOH 给出了一个评估电池现有性能的指标，这有助于预估电池的预期寿命，并决定何时需要替换或维护它。SOH 可由公式 2-1 得到，其中 $C_{current}$ 代表当前容量， $C_{initial}$ 代表当前循环的初始容量。

$$SOH = \frac{C_{current}}{C_{initial}} \quad (2-1)$$

2.2 基于随机微分方程的时间序列生成器

此前提到，在现存大部分深度学习方法中，都是单纯通过大量数据进行训练，缺乏理论指导，在 Remlinger 等 (2021) [27] 的启发下，本研究采用一种基于伊藤过程 (Itô process) 和深度欧拉卷积构建的时间序列生成器，以随机微分方程的形式构建预测方程。该生成器能够在准确估计时间边际分布的同时，正确捕捉时间动态。本研究设计的生成器设计足够简单以便于处理 (即输出可以被控制)，并且具有理论基础。为此，本研究将训练时间序列数据输入到本研究的算法中，并努力学习一个最能近似数据的经

验概率分布。这项任务可能很棘手，具体取决于序列长度、维度和数据分布的形状。

2.2.1 伊藤过程

假设给定一个时间序列的独立同分布样本，该时间序列被视为一个随机向量 $X = (X_{t_i})_{i=1\dots N}$ 在 $\sim^{d \times N}$ 上，从一个起始点 $X_0 \in \sim^d$ 开始，并在时间网格 $\mathbf{T} := \{0 = t_0 < t_1 < \dots < t_N = T\}$ 上观测。为简单起见，在下文中，本研究假设时间网格是均匀的，网格大小为 Δt 。离散时间样本被认为是从具有以下伊藤动态的连续时间基础过程 X 中抽取的：

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t \quad (2-2)$$

其中， $b_X : \sim^d \times \sim^d \rightarrow \sim^d$ 是漂移项， $\sigma_X : \sim^d \times \sim^d \rightarrow \mathbf{M}_{d \times d}$ 是扩散项， W 是一个 d 维布朗运动。假设参数 b_X 和 σ_X 满足 Lipschitz 条件，这些条件确保了方程 (2-2) 解的存在性和唯一性。

2.2.2 深度欧拉表达

$X = (X_{t_i})_{i=1\dots N}$ 的样本是从具有如动态方程 (2-2) 的连续时间抽取的，并且可以近似地看作是从以下欧拉离散化方案抽取的样本：

$$X_{t_{i+1}} \approx X_{t_i} + b_X(t_i, X_{t_i})\Delta t + \sigma_X(t_i, X_{t_i})\Delta W_{t_i} \quad (2-3)$$

其中 $(\Delta W_{t_i})_i$ 是一组独立同分布的 $\mathbf{N}(0, \Delta t)$ 随机变量。本研究依赖于这个近似，并引入时间序列的以下深度欧拉表示。从 $t_0 = 0$ 开始，从 $Y_0^\theta = X_0$ 生成时间序列，具体方案如下：

$$Y_{t_i+\Delta t}^\theta = Y_{t_i}^\theta + b_Y^\theta(t_i, Y_{t_i}^\theta)\Delta t + \sigma_Y^\theta(t_i, Y_{t_i}^\theta)Z_{t_i} \quad (2-4)$$

其中， Z_{t_i} 是 $\mathbf{N}(0, \Delta t)$ 独立同分布的随机变量，并且函数 b_Y^θ 和 σ_Y^θ 是由神经网络近似的 θ 参数化函数。模型的目标是学习 b_Y^θ 和 σ_Y^θ ，使得过程 Y^θ 和 X 的分布接近。用于训练 b_Y^θ 和 σ_Y^θ 的双支 ResNet 模型将在第 2.3 节被介绍。

2.3 双支 ResNet

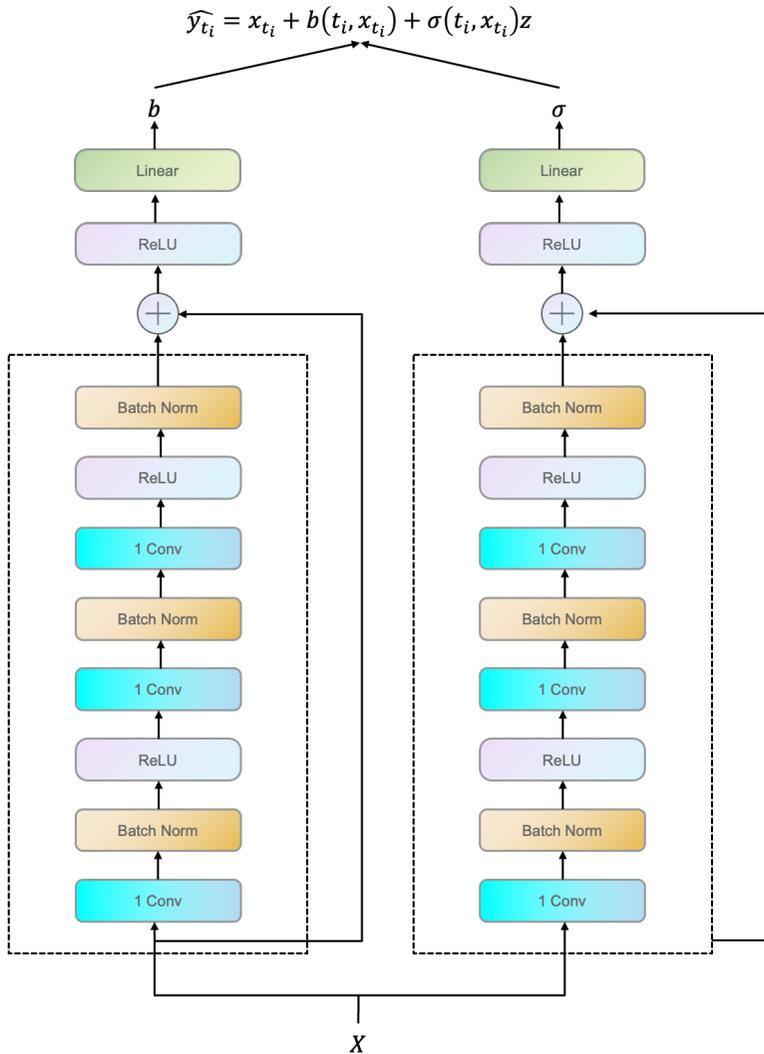


图 2-1 随机微分方程-双支 ResNet 模型结构图

ResNet 类模型（Residual Networks）是深度学习领域中一种极为成功且广泛应用的卷积神经网络架构，其引入的残差学习机制有效地缓解了传统深度神经网络训练中的退化问题，从而使得网络的深度可以大幅度增加而不导致模型性能的显著下降^[28]。ResNet 成功地借助残差块（Residual Block）来克服深度神经网络在学习阶段遇到的退化问题。传统的深度网络在其层数逐步增长时，由于梯度可能会在反向传播过程中逐步消散，从而使得整个网络的训练面临巨大挑战。残差块的特点在于它采用了 Skip Connection，这意味着可以在一定的层面上直接传递梯度，从而有力地解决了梯度消失以及梯度爆炸的问题。该架构设计保证了，即使网络的层数

有所增加，模型的训练也能够保持稳定，成功避免了训练偏差随着网络深度的加深而增加的情况。其次，由于采用了模块化设计方案，残差网络展现出极高的灵活性。ResNet 里的残块可以被认为是不同功能的单元，通过将这些功能单元堆叠在一起，能便捷地构造出各种不同深度层次的网络结构。比如说，ResNet-50、ResNet-101 以及 ResNet-152 这些模型的实现方式，就是通过简洁地堆叠不同数量的残差块来完成。这样的模块化设计不仅便于网络的构建和扩展，也使得研究人员能够更方便地进行网络结构的探索和优化。残差网络的引入使得深度学习模型在更大数据集上的训练成为可能，同时在保持高效特征提取能力的同时，提高了模型的泛化性能。ResNet 所具有的类似于 CNN 的共享权重架构和固有的平移不变性确保它不会对输入数据中特征的具体位置过于敏感。这一特性对于 SOH 预测非常有利，因为具体的退化模式可能会在电池的不同生命周期阶段表现出来。此外，在现实世界的电池使用数据中噪声非常常见，而 ResNet 对噪声和异常值有较好的鲁棒性，使其在提供一致的预测性能方面更加可靠。此外，Xiong 等(2023)验证了 ResNet 在锂电池 SOH 验证任务上的有效性^[29]。

基于上述原因，本研究提出了如图 2-1 所示的双支 ResNet (Double Branch ResNet) 对第 2.1 节中设计的随机微分方程模型进行训练，过两个并行分支分别模拟随机微分方程中的漂移项 b_Y^θ 和扩散项 σ_Y^θ 。每个 Branch 由三层卷积层以及 ReLU 激活函数层和一层线性层组成，每一个 Branch 的卷积层-ReLU 层-线性层可看做一个 1-Lipschitz 函数。Branch1 负责处理确定性部分，也称作漂移项。而 Branch2 在处理随机性部分时，会乘以随机噪声，这也被称为扩散项，从而更为有效地捕获了时间序列数据中的随机性波动与发展轨迹。本研究提出的双支 ResNet 的组成结构可表示为以下形式：

- 输入层： $h_0 = \phi(\mathbf{A}x)$;
- 残差层： $h_l = \phi(h_{l-1} + \tau \mathbf{W}_l h_{l-1})$;
- 全连接层： $h_f = \phi(\mathbf{W}_l h_{f-1})$;
- Branch1 输出： $b_Y^\theta = h_{f_{B1}}$;
- Branch2 输出： $\sigma_Y^\theta = h_{f_{B2}}$;
- 输出层： $Y_{t_i+\Delta t}^\theta = Y_{t_i}^\theta + b_Y^\theta(t_i, Y_{t_i}^\theta)\Delta t + \sigma_Y^\theta(t_i, Y_{t_i}^\theta)Z_{t_i}$;
- 初始化： $\mathbf{A} \in \sim^{m \times 1}$ 和 $\mathbf{W}_l \in \sim^{m \times m}$ 对于 $l=[L]$ ，其中 \mathbf{A} 和 \mathbf{W}_l 的各个元素分别独立采样自 $N\left(0, \frac{2}{m}\right)$ 和 $N\left(0, \frac{2}{m}\right)$ 。

其中 $\phi(\cdot)$ 为 ReLU 激活函数 $\phi(\cdot) := \max\{0, \cdot\}$ 。具体来说，假设输入维度为 1，因此 $x \in \sim^1$ ；中间层具有相同的宽度 m ，因此 $h_l \in \sim^m$ ，其中 $l=0,1,\dots,L$ ；

输出维度为 1，因此 $y \in \mathbb{R}^1$ 。令激活前的值为 $g_0 = \mathbf{A}x, g_l = h_{l-1} + \tau \mathbf{W}_l h_{l-1}$ ，其中 $l=1,2,\dots,L-1$ 。用 $h_{i,l}$ 表示当输入向量为 x_i 时的 h_l 的值

2.4 损失函数

在构思模型的损耗函数时，寻找一个平衡是关键，这需要在边缘的分布拟合与时间结构的精确描述之间实现。一方面，我们不应该只关注于边界分布问题，因为即使所有 $t_i \in \mathbf{T}$ 满足 $L(X_{t_i}) \sim L(Y_{t_i}^o)$ ，也并不意味着 $b_X = b_{Y^o}$ 或 $\sigma_X = \sigma_{Y^o}$ 。例如，考虑一个时间网格 $\{t_0, t_1, t_2\}$ ，其中 $\Delta t = t_{i+1} - t_i$ ，定义 $b_X : (t, x) \rightarrow \frac{-2x}{\Delta t}$ ， $b_Y : (t, x) \rightarrow 0$ ，且所有 $\sigma_X(t, x) = \sigma_Y(t, x) = 1$ 对所有 $x \in \mathbb{R}$ 都成立，这些 b 和 σ 是 Lipschitz 连续的：

$$\|b_X(t, x) - b_X(t, y)\|_2 \leq K \|x - y\|_2 \quad (2-5)$$

$$\|b_Y(t, x) - b_Y(t, y)\|_2 \leq K \|x - y\|_2 \quad (2-6)$$

对于一维情况，有：

$$\begin{aligned} X_{t_0} &= 0, & X_{t_1} &= \varepsilon_{t_1}^X, & X_{t_2} &= -X_{t_1} + \varepsilon_{t_2}^X \\ Y_{t_0} &= 0, & Y_{t_1} &= \varepsilon_{t_1}, & Y_{t_2} &= Y_{t_1} + \varepsilon_{t_2} \end{aligned}$$

其中 $\varepsilon_{t_i}^X, \varepsilon_{t_i}^Y \sim N(0, \Delta t)$ 是 *iid* 的。那么对于 $i \in \{0, 1, 2\}$ ，有 $L(X_{t_i}) = L(Y_{t_i})$ ，但是 $E(X_{t_2} | X_{t_1} = z) = -z \neq z = E(Y_{t_2} | Y_{t_1} = z)$ 。

另外，在处理其边缘分布的情况下，可以考虑将时间序列视作定义在 \mathbb{R}^{n+1} 上的向量样本，但这种方法未必带来更为理想的成果。正 Yoon et al. 在文章中所探讨的，哪怕是基于记忆机制的网络，采用学习联合分布 $L(X_{t_0}, \dots, X_{t_n})$ 也未必能充分确保网络对时间动态有准确的捕获^[30]。在处理时间序列的损失分析时，不能单纯基于其边缘或联合分布特性来进行。为了更好地探讨这一问题，本研究中采用了 Gelbrich 等中引入的 Wasserstein-2 距离作为损失函数^[31]。同时，利用这一函数，本研究能对模型理论估计误差进行更为深入的评估和分析。Wasserstein-2 距离的计算公式如下：

$$W_2^2(L(X), L(Y)) = \|E[X] - E[Y]\|_2^2 + B^2(\text{Var}(X), \text{Var}(Y)) \quad (2-7)$$

其中， B 是 Bures 度量，定义为 $B^2(A, B) = \text{Tr}(A) + \text{Tr}(B) - 2 \text{Tr} \left(A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{1/2}$ ，

适用于正定矩阵 A 和 B ^[32]。如果 X 和 Y 是高斯分布，那么 Wasserstein-2 距

离可以被定义为 W_2 。公式(2-7)捕捉了分布之间有意义的几何特征。此外，由于 W_2 的计算方法对异常值反应异常敏锐，这增强了对分布估计的准确率。

2.5 误差分析

2.5.1 泛化误差分析

对神经网络模型进行全面误差的分析成为理解其在实际应用场景中性能表现的关键要素。尽管优化算法能够大幅度地减小训练过程中的误差，但泛化误差却揭示了在未见数据中该模型的具体表现，它是评价该模型在实际应用中价值的核心指标。通过深入分析泛化误差，能够更加明确模型在过度拟合与欠缺拟合方面的表现，这进一步有助于指导模型的构建和选择过程。本项研究使用了 Rademacher 复杂度这一衡量模型泛化误差的量化方法，它是一种用来衡量机器学习模型复杂性的评估工具，并用于确定模型在随机数据分析中的性能。复杂性更高代表假设的类型变得更为灵活，更容易适应随机噪声，然而其泛化表现可能并不理想；如果复杂性降低，这意味着模型的假设类更为简洁，因此更适合在处理新数据时展现出出色性能。鉴于这个观点，本文给出以下推论：

推论 2.1： 假设给定一个时间序列 *i.i.d.* 样本，表示为一个随机向量 $X = (X_{t_i})_{i=1 \dots N} \in \mathbb{R}^{1 \times N}$ ，从点 $X_0 \in \mathbb{R}^1$ 开始，在时间网格 $T := \{0 = t_0 < t_1 < \dots < t_N = T\}$ 上观察。同时假设给定随机微分方程-双支 ResNet 模型 F 是稳定可训练的，同时损失函数满足 *K-Lipschitz* 条件，且每一层的权重矩阵的 Frobenius 范数 $\|W_l\|_F$ 是有界的，则给定模型的泛化误差边界可以表示为：

$$R_N(F) = R_N(F_{B1}) + R_N(F_{B2}) \leq \frac{L \times \log(d)}{\sqrt{N}} \times (C_{B1} + C_{B2}) \quad (2-8)$$

其中， F_{B1} ， F_{B2} 表示 F 中两支的函数类， L 为模型层数， d 为模型参数总数， C_{B1} ， C_{B2} 为常数。

证明： 对于数据集 $\{(x_i, y_i)\}_{i=1}^N$ 和函数类 F ，Rademacher 复杂度定义为：

$$R_N(F) = E_{\sigma} \left[\sup_{f \in F} \frac{1}{N} \sum_{i=1}^N \sigma_i f(x_i) \right] \quad (2-9)$$

其中， σ_i 是 Rademacher 独立随机变量，取值为 ± 1 且概率相等。根据深度

神经网络的理论，对于双支 ResNet 中的每一个分支，每层的 Rademacher 复杂度可以表示为：

$$R_N(F_l) \leq \frac{1}{\sqrt{N}} \|W_l\|_F \quad (2-10)$$

其中 $\|W_l\|_F$ 为该层权重矩阵 W_l 的 Frobenius 范数。假设 $\|W_l\|_F$ 是有界的，即 $\|W_l\|_F \leq C_l$ ， C_l 为常数。那么对于整个 Branch，同时将参数数量 d 考虑进去，Rademacher 复杂度可以表示为：

$$R_N(F_B) \leq \frac{\log(d)}{\sqrt{N}} \sum_{l=1}^L \|W_l\|_F \leq \frac{L \times \log(d) \times C}{\sqrt{N}} \quad (2-11)$$

其中， $C = \max\{C_1, C_2, \dots, C_L\}$ 。那么对于整个网络的 Rademacher 复杂度则可进一步表示为：

$$R_N(F) = R_N(F_{B1}) + R_N(F_{B2}) \leq \frac{L \times \log(d)}{\sqrt{N}} \times (C_{B1} + C_{B2})$$

基于方程 2-8，可以明确地看到，随着网络的进一步加深，模型的泛化误差将会逐渐增加，这进一步增强了过拟合的危险，并导致了模型的稳定性及其训练能力的降低。依据 Zhang 等的研究表明，在对 ResNet 进行控制 $\tau \leq \frac{1}{\sqrt{L}}$ 的情况下，能够保证网络输出不会出现爆炸性增长，从而确保模型具有稳定和可训练的特性，进而帮助减少模型过度拟合的风险。本文将在下一小节对 $\tau \leq \frac{1}{\sqrt{L}}$ 时对本文提出的双支 ResNet 的训练稳定性进行分析以进一步支持推论 2.1。

2.5.2 模型训练稳定性分析

推论 2.2： 假设给定一个时间序列 *i.i.d.* 样本，表示为一个随机向量 $X = (X_t)_{t=1 \dots N} \in \sim^{1 \times N}$ ，且 $X_0 \in \sim^1$ 的误差的 Frobenius 范数 $\|\delta h_0\|_F$ 是有界的。同时给定随机微分方程-双支 ResNet 模型满足推论 2.1 中的条件，且模型中的 SDE 满足方程 (2-4) 的形式，且每一层残差层中满足 $\tau \leq \frac{1}{\sqrt{L}}$ ，则给定模型的输出边界可以表示为：

$$\|\delta h_L\|_F \leq \left(e^{c_1 \sqrt{L}} + \sqrt{\frac{2}{\pi}} e^{c_2 \sqrt{L}} \right) \|\delta x\|_F \tilde{N}O\left(\frac{1}{\sqrt{L}}\right) \quad (2-12)$$

证明： 首先对常规 ResNet 模型进行分析。假设输入的误差为 δh_1 ，

则经过残差块后的输出误差可以表示为:

$$\delta h_{l+1} = \delta h_l + \tau W_l \delta h_l \quad (2-13)$$

由于 $\|W_l\|_F$ 有界, 可以得到:

$$\|\delta h_{l+1}\|_F \leq \|\delta h_l\|_F + \tau \|W_l \delta h_l\|_F \quad (2-14)$$

假设权重矩阵的谱范数有界, 即 $\|W_l\| \leq c_l$ 其中 c 是一个常数, 可以得到:

$$\|\delta h_{l+1}\|_F \leq (1 + \tau c) \|\delta h_l\|_F \quad (2-15)$$

对于第 L 层的推导, 递归不等式 (2-15) 可以得到:

$$\|\delta h_L\|_F \leq (1 + \tau c)^L \|\delta h_0\|_F \quad (2-16)$$

由于 $\tau \leq \frac{1}{\sqrt{L}}$, 根据渐进不等式, 可以得到:

$$\left(1 + \frac{c}{\sqrt{L}}\right)^L \approx e^{c\sqrt{L}} \quad (2-17)$$

因此, 输出层的误差可以表示为:

$$\|\delta h_L\|_F \leq e^{c\sqrt{L}} \|\delta h_0\|_F \quad (2-18)$$

对于本文提出的双支 ResNet 模型, 假设 Branch1 的输出误差是 $\delta h_{1,l}$, 则通过残差块后的输出误差可以表示为:

$$\delta h_{1,l+1} = \delta h_{1,l} + \tau_1 W_{1,l} \delta h_{1,l} \quad (2-19)$$

同理易得, Branch2 的输出误差可以表示为:

$$\delta h_{2,l+1} = \delta h_{2,l} + \tau_2 W_{2,l} \delta h_{2,l} \quad (2-20)$$

结合两个分支, 总的输出误差可以表示为:

$$\delta y_{l+1} = \delta h_{1,l+1} + z \cdot \delta h_{2,l+1} \quad (2-21)$$

根据方程 (2-16) 递归每个分支的误差传播公式:

$$\|\delta h_{Branch1,L}\|_F \leq (1 + \tau_1 c_1)^L \|\delta h_0\|_F \quad (2-22)$$

$$\|\delta h_{Branch2,L}\|_F \leq (1 + \tau_2 c_2)^L \|\delta h_0\|_F \quad (2-23)$$

由于 $\tau_1 \leq \frac{1}{\sqrt{L}}$, $\tau_2 \leq \frac{1}{\sqrt{L}}$, 因此有:

$$\|\delta h_{Branch1,L}\|_F \leq e^{c_1\sqrt{L}} \|\delta h_0\|_F \quad (2-24)$$

$$\|\delta h_{Branch2,L}\|_F \leq e^{c_2\sqrt{L}} \|\delta h_0\|_F \quad (2-25)$$

最终的输出误差上界可以表示为:

$$\|\delta h_L\|_F \leq e^{c_1\sqrt{L}} \|\delta h_0\|_F + |z| \cdot e^{c_2\sqrt{L}} \|\delta h_0\|_F \quad (2-26)$$

考虑到最终输出层中 Z 是正态分布，可以对其取期望 $\mathbb{E}\left[\left|Z_{t_i}\right|\right] = \sqrt{\frac{2}{\pi}}$ ，因此输出误差上界可进一步表示为：

$$\|\delta h_L\|_F \leq \left(e^{c_1\sqrt{L}} + \sqrt{\frac{2}{\pi}} e^{c_2\sqrt{L}} \right) \|\delta x\|_F \tilde{N}\mathcal{O}\left(\frac{1}{\sqrt{L}}\right)$$

该结论确保了随着网络深度的增加，模型输出不会出现爆炸性增长，从而保证了模型的稳定性和训练过程的可控性，一定程度上避免了过拟合的风险。

2.5.3 随机微分方程估计误差分析

由于基于随机微分方程的时间序列生成器为本研究模型最核心的组成部分，因此本研究认为若要对模型进行误差理论分析，也应从随机微分方程的角度对估计误差进行分析。

推论 2.3: 假设 SDE 满足方程 (2-4) 的形式，且 $\sigma_{Y^0}^2(t_i, Y_{t_i}^\theta)$ 严格为正，且 $b_Y^\theta(t_i, Y_{t_i}^\theta)$ 中 $Y_{t_i}^\theta$ 满足 K -Lipschitz 条件，同时假设漂移项 b_Y 和扩散项 σ_Y 在 t_i 时刻的误差 $e(b_{t_i})$ 以及 $e(\sigma_{t_i})$ 是有界的，那么 SDE 的估计误差可以表示为：

$$\dot{U}_{total} \leq \sqrt{\frac{\left(\frac{\dot{U} + \Delta x}{\Delta t} + 2\Delta x\right) \log(2/\delta_b)}{2N}} + \sqrt{\frac{\left(\sqrt{\frac{\dot{U}}{\Delta t}} + 2\Delta x\right) \log(2/\delta_\sigma)}{2N}} \cdot \sqrt{\frac{2}{\pi}} \quad (2-27)$$

证明: 首先定义 SDE 中漂移项 b_Y 和扩散项 σ_Y 在 t_i 时刻的误差 $e(b_{t_i})$ 以及 $e(\sigma_{t_i})$ ：

$$e_b(t_i) = b_Y^\theta(t_i, Y_{t_i}^\theta) - \hat{b}_Y^\theta(t_i, Y_{t_i}^\theta) \quad (2-28)$$

$$e_\sigma(t_i) = \sigma_Y^\theta(t_i, Y_{t_i}^\theta) - \hat{\sigma}_Y^\theta(t_i, Y_{t_i}^\theta) \quad (2-29)$$

由于 $e(b_{t_i})$ 以及 $e(\sigma_{t_i})$ 是有界的，即 $|e(b_{t_i})| \leq M_b$ ， $|e(\sigma_{t_i})| \leq M_\sigma$ ，那么根据 Hoeffding 不等式，对于这两个有界独立误差，可以给出误差的集中性：

$$P\left(\left|\frac{1}{N} \sum_{i=1}^N e_b(t_i)\right| \geq \dot{U}_b\right) \leq 2 \exp\left(-\frac{2N\dot{U}_b^2}{M_b^2}\right) \quad (2-30)$$

$$P\left(\left|\frac{1}{N}\sum_{i=1}^N e_{\sigma}(t_i)\right| \geq \dot{U}_{\sigma}\right) \leq 2 \exp\left(-\frac{2N\dot{U}_{\sigma}^2}{M_{\sigma}^2}\right) \quad (2-31)$$

设 b 的误差概率 $\delta_b = 2 \exp\left(-\frac{2N\dot{U}_b^2}{M_b^2}\right)$ ，等式两边取对数：

$$\log(\delta_b) = \log(2) - \frac{2N\dot{U}_b^2}{M_b^2} \quad (2-32)$$

所以有：

$$-\frac{2N\dot{U}_b^2}{M_b^2} = \log 2 - \log(\delta_b) = \log\left(\frac{2}{\delta_b}\right) \quad (2-33)$$

即可求得 $\dot{U}_b = \sqrt{\frac{M_b^2 \log(2/\delta_b)}{2N}}$ 。同理，设 σ 的误差概率 $\delta_{\sigma} = 2 \exp\left(-\frac{2N\dot{U}_{\sigma}^2}{M_{\sigma}^2}\right)$ ，

可求得 $\dot{U}_{\sigma} = \sqrt{\frac{M_{\sigma}^2 \log(2/\delta_{\sigma})}{2N}}$ 。由于 Z_{t_i} 满足标准正态分布，因此有

$E[Z_{t_i}] = \sqrt{\frac{2}{\pi}}$ ，因此基于随机微分方程的时间序列生成器估计误差为：

$$\begin{aligned} \dot{U}_{total} &= \sum_{i=1}^N (e_b(t_i, Y_{t_i}) + e_{\sigma}(t_i, Y_{t_i}) Z_{t_i}) \\ &\tilde{N} \sqrt{\frac{M_b^2 \log(2/\delta_b)}{2N}} + \sqrt{\frac{M_{\sigma}^2 \log(2/\delta_{\sigma})}{2N}} \cdot \sqrt{\frac{2}{\pi}} \end{aligned} \quad (2-34)$$

根据 Remlinger 等（2021）给出的结论， $M_b^2 = \frac{\dot{U} + \Delta x}{\Delta t} + 2\Delta x$ ，

$M_{\sigma}^2 = \frac{\dot{U}}{\Delta t} + 2\Delta x$ ，因此总体的估计误差可以进一步写做：

$$\dot{U}_{total} \leq \sqrt{\frac{\left(\frac{\dot{U} + \Delta x}{\Delta t} + 2\Delta x\right) \log(2/\delta_b)}{2N}} + \sqrt{\frac{\left(\frac{\dot{U}}{\Delta t} + 2\Delta x\right) \log(2/\delta_{\sigma})}{2N}} \cdot \sqrt{\frac{2}{\pi}}$$

2.6 实验验证

2.6.1 数据来源

表 2-1 Oxford 数据集测试环境

Dataset	Cell No.	Charging Profile	Discharging Profile	Cycles
	Cell1	CC-CV	Artemis Urban driving profile	82
	Cell2	CC-CV	Artemis Urban driving profile	77
	Cell3	CC-CV	Artemis Urban driving profile	81
	Cell4	CC-CV	Artemis Urban driving profile	51
	Cell5	CC-CV	Artemis Urban driving profile	50
	Cell6	CC-CV	Artemis Urban driving profile	50
	Cell7	CC-CV	Artemis Urban driving profile	81
	Cell8	CC-CV	Artemis Urban driving profile	81

本研究所选用的数据集为由牛津大学收集的 Oxford Battery Degradation Dataset。如表 2-1 所示，不同的 Cell 采集时使用的充电/放电策略以及测试循环次数不同，具体如下：

牛津大学使用了八个 Kokam (SLPB533459H4) 锂离子软包电池 (OX-x)，每个电池的标称容量为 0.74 Ah (其中 x 表示电池编号)，进行长期电池退化测试。为了模拟车辆的城市使用情况，这八个电池在 40°C 的加热房间内定期以 2C 速率充电，并通过动态曲线进行放电。为了确定电池的保持容量，电池在每 100 次充放电循环后，采用恒定电流循环技术进行测试。

2.6.2 优化器

接下来将讨论优化器，它是通过最小化损失函数来提高神经网络性能的关键部分。一旦确定了最终的优化目标 (损失函数)，模型的参数可以使用 BP 算法和优化器进行更新。本文中采用了 Adam 优化器。与其他优化器如随机梯度下降 (SGD) 和均方根传播 (RMSProp) 相比，Adam 结合了 RMSprop 和动量的优点，计算梯度的指数移动平均值以及平方梯度的指数移动平均值。这种方法允许对学习率进行更细微的调整，通常能更快收敛。

在 Adam 算法中，学习率会在训练过程中自动变化，这可以避免在快速找到全局最优解时学习率不足的问题。关于参数 θ 的优化过程公式如方程 2-35 - 方程 2-40 所示：

$$g_q = \frac{\partial L_b}{\partial \theta_q} + \frac{\partial L_g}{\partial \theta_q} \quad (2-35)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2-36)$$

$$V_t = \beta_2 V_{t-1} + (1 - \beta_2) g_t^2 \quad (2-37)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2-38)$$

$$\hat{V}_t = \frac{v_t}{1 - \beta_2^t} \quad (2-39)$$

$$\theta_{t+1} = \theta_t - \gamma \frac{\hat{m}_t}{\sqrt{\hat{V}_t + \varepsilon}} \quad (2-40)$$

其中, g_q 是整个网络中 θ_q 的梯度, 包含学生模型和教师模型的梯度。 t 表示当前的训练周期。 m_t 和 V_t 分别代表动量和自适应学习率。 β_1 和 β_2 是两个超参数, 通常设置为 0.9 和 0.999。由于 β_1 和 β_2 的值, m_t 和 V_t 在初始训练过程中通常非常小并趋于 0, 导致估计的偏差。为了解决这个问题, 使用 \hat{m}_t 和 \hat{V}_t 进行偏差校正。 γ 是学习率, ε 是一个平滑系数, 用于避免分母为 0, 在本文中设置为 $1e-8$ 。

2.6.3 训练策略

为了更好地表现本研究所提出模型的强劲表现, 不同于其他有关时间序列数据研究所采用的数据预处理策略, 即滑动窗口, 本研究采用单特征单数据单输出的方式构建训练策略。具体来说, 本研究采用上一时刻的 SOH 数据作为输入, 用下一时刻的 SOH 数据作为标签, 具体表现形式如下:

$$y_{ii} = x_{ii+1} \quad (2-41)$$

同时为了检验模型的泛化能力, 即在预测不同于训练数据充放电策略的电池 SOH 时的表现, 本文使用其他 Cell 作为训练数据, 选择目标 Cell 作为测试数据 (如当测试 Cell1 时, 采用 Cell2-Cell8 作为训练数据)。

在完成数据预处理以及训练集测试集分割后, 随机微分方程-双支 ResNet 通过训练集 (SOH 数据) 进行离线训练。所有随机微分方程-双支 ResNet 的参数通过 Adam 优化器迭代更新, 以 Wasserstein-2 距离作为优化目标。离线训练算法总结如算法 1 所示。

算法 1: 基于随机微分方程-双支 ResNet 的离线训练策略

算法 1: 随机微分方程-双支 ResNet 的离线训练策略

输入: 样本 X , m batch size, γ 学习率

初始化: θ (随机选择)

1: **while** Not converged **do**

```

2:   for  $t_i = 0 \dots T$  do
3:       采样  $z \sim N(0, \Delta t)$ 
4:        $\hat{y}_t^{(k)} \leftarrow x_t + g_\theta^b(t_i, x_t) \Delta t + g_\theta^\sigma(t_i, x_t) z$ 
5:        $\ell \leftarrow W_2^2(L(y^{(k)}), L(\hat{y}^{(k)}))$ 
6:        $\theta = \theta - \gamma \nabla_\theta \ell$ 
7:   end while
输出:  $\hat{y}$ 
    
```

2.6.4 评价指标

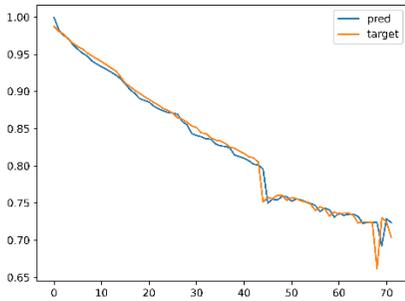
MAE（平均绝对误差）和 RMSE（均方根误差）用于评估估计的 SOH 的误差。在第 k 次循环下，预测的 SOH 表示为 $\tilde{y}^{(k)}$ ，真实的 SOH 表示为 $y^{(k)}$ ，RMSE 和 MAE 计算公式如下：

$$RMSE = \frac{1}{N} \sqrt{\sum_{k=1}^N (\tilde{y}^{(k)} - y^{(k)})^2} \quad (2-42)$$

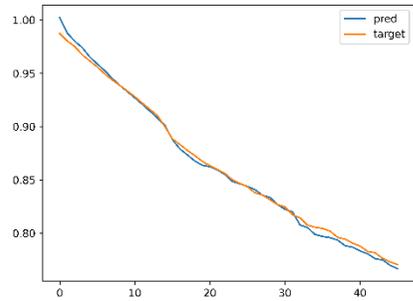
$$MAE = \frac{1}{N} \sum_{k=1}^N |\tilde{y}^{(k)} - y^{(k)}| \quad (2-43)$$

RMSE 用于计算预测值与实际值之间的平方差的平方根平均值，该方法对不寻常的数据值高度敏感，并常用于回归分析中，RMSE 水平愈低，则表示该模型在性能方面表现更出色。利用 MAE 方法来计算预测值和实际值的绝对差异的平均值，这样可以将所有的误差视为相同，且异常值的影响较轻。MAE 值偏低，这同样意味着准确度有所增加。RMSE 与 MAE 都是用于评价模型准确度的关键参数，它们都能助力模型的挑选和校验过程。

2.6.5 实验结果与分析



(a) Cell2



(b) Cell4

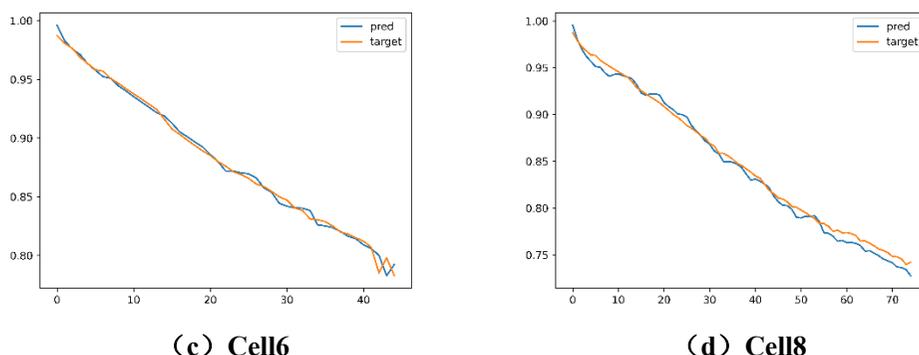


图 2-2 随机微分方程-双支 ResNet 在 Oxford 数据集上测试结果

本研究选取了 Oxford 数据集中的 4 个电池单元 (Cell2、Cell4、Cell6、Cell8) 作为测试对象, 通过实验结果可以看出, 模型在 SOH 预测任务上表现出色。首先, 从趋势上看, 随机微分方程-双支 ResNet 模型的拟合曲线与真实曲线高度重合, 在趋势上几乎完全一致, 充分展示了模型在捕捉 SOH 变化趋势方面的卓越能力。具体来说, 数值结果如表 2-2 所示, 除了 Cell2 外, 所有测试对象的 RMSE 均在 0.01 以下, 而所有测试对象的 MAE 也都小于 0.01, 这表明模型不仅在趋势学习上表现优异, 在数值拟合上同样具备高度的准确性。此外, 模型在不同测试对象上的一致优异表现进一步证明了其良好的泛化能力和适用性。不管是 Cell2、Cell4、Cell6, 亦或是 Cell8, 这些模型都能为 SOH 提供非常高度的准确性预测, 展现在各种电池单元上的出色稳定表现。这样的鲁棒性确保了随机微分方程-双支 ResNet 模型在真实场景中表现出广阔的适宜性和高度的稳定性。从一个更有创新性的视角出发, 结合随机微分方程与双支 ResNet 的模型设计, 为 SOH 预测任务呈现了一个创新的解决途径。此种技术方法除了在有理论上的深厚基础之外, 在实际数据集的检验中也指出了它在捕获电池 SOH 变化中的错综复杂动态上所展现的独到长处。

表 2-2 随机微分方程-双支 ResNet 在 Oxford 数据集上数值测试结果

Cell No.	RMSE	MAE
Cell2	0.0118	0.0096
Cell4	0.0043	0.0025
Cell6	0.0047	0.0054
Cell8	0.0065	0.0076

2.7 本章小结

本章提出了一种依赖于退化过程的锂电池 SOH 的预测技术。首先, 本文描述了一个基于随机微分方程的时间序列生成器。通过使用伊藤过程

（Itô process）与深度欧拉卷积模型，这一时间序列生成器成功建立了预测方程，并模拟了电池 SOH 的变化情况。接着，文中进行了详尽的描述，介绍了双支 ResNet 模型，该模型被用作时间序列数据的特性抽取和预测模型。之后，这一章也对随机微分方程，即双支 ResNet 模型，进行了详细的理论误差探讨，并在 Oxford 数据集上对这一模型进行了实际的实验检验。

第3章 基于片段数据的 SOH 预测方法

3.1 模型结构

3.1.1 LSTM

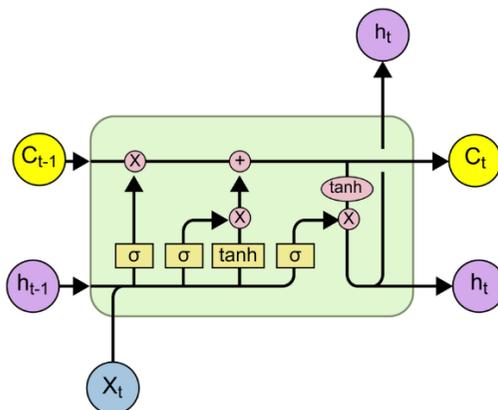


图 3-1 LSTM 结构示意图

LSTM 是一种特殊的递归神经网络 (RNN)，它能够处理和存储长时间序列数据中的依赖关系。LSTM 通过引入门控机制，可以有效地控制梯度的流动，使得梯度可以在较长的时间步内保持不变，解决了传统 RNN 中存在的梯度消失和梯度爆炸问题。除此之外，LSTM 的记忆单元允许信息长时间存储和传播，使得 LSTM 能够捕捉序列中的长时间依赖关系。这对于自然语言处理、时间序列预测等需要考虑长期上下文的信息尤为重要。更重要的是，LSTM 不需要对数据进行平稳化处理，适用于非线性和非平稳的时间序列数据。它能够自动学习时间序列中的复杂模式和依赖关系。LSTM 的结构如图 3-1 所示，其由三个主要的门 (gate) 组成：输入门 (input gate)、遗忘门 (forget gate) 和输出门 (output gate)，其工作原理可通过以下几个步骤描述：

(1) 遗忘门 (forget gate)

遗忘门决定了前一时刻的记忆细胞状态 C_{t-1} 中哪些信息需要被遗忘，它由一个 Sigmoid 函数来实现：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3-1)$$

其中， W_f 和 b_f 是权重矩阵和偏置向量， h_{t-1} 是前一时刻的隐藏状态， x_t 是当前输入 σ 是 sigmoid 激活函数。

(2) 遗忘门 (forget gate)

输入门决定了当前输入 x_t 中有多少信息需要写入记忆细胞。它包括两个部分：输入门 i_t 和新的候选记忆细胞状态 \tilde{C}_t 。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3-2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3-3)$$

其中， W_i ， W_c ， b_i ， b_c 是权重矩阵和偏置向量， \tanh 是激活函数。

(3) 记忆细胞状态更新

当前的记忆细胞状态 C_t 是由前一时刻的记忆细胞状态 C_{t-1} 和当前的候选记忆细胞状态 \tilde{C}_t 共同决定的，更新过程如下：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3-4)$$

(4) 输出门 (output gate)

输出门决定了记忆细胞状态 C_t 中有多少信息需要输出，并生成当前时刻的隐藏状态 h_t 。

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3-5)$$

$$h_t = o_t * \tanh(C_t) \quad (3-6)$$

其中， W_o 和 b_o 是权重矩阵和偏置向量。

3.1.2 Attention 机制

在深度学习领域，尤其是自然语言处理 (NLP) 领域以及时间序列预测领域，序列到序列 (seq2seq) 模型是一种广泛使用的架构，用于任务如机器翻译、文本摘要和对话生成。传统的 seq2seq 模型由编码器 (encoder) 和解码器 (decoder) 组成，其中编码器将输入序列编码为固定长度的上下文向量，而解码器根据该上下文向量生成输出序列。然而，固定长度的上下文向量限制了模型的能力，尤其是在处理长序列时。固定长度的上下文向量难以捕获输入序列中所有信息，导致信息丢失和性能下降。为了克服这些限制，Bahdanau 等人在 2014 年提出了 Attention 机制。在 Attention 机制中，解码器在生成每一个输出时，不再仅仅依赖于一个固定长度的上下文向量，而是通过计算输入序列中每个位置的注意力权重，动态地选择相关的输入信息。Attention 机制不仅提高了模型处理长序列的能力，还有效地解决了 seq2seq 模型中的偷看问题 (peek problem)。偷看问题指的是解码器在训练过程中可能直接看到目标序列，从而影响模型的训练效果。Attention 机制的计算过程可以用以下过程描述：

给定一个输入序列 $X = (x_1, x_2, \dots, x_n)$ ，其中 x_i 通常是一个向量表示。对于输入序列中的每个元素 x_i ，通过线性变换得到三个向量：查询 q_i ，键 k_i ，值 v_i 。线性变换用矩阵 W^q ， W^k ， W^v 表示，其中：

$$q_i = W^Q x_i, \quad k_i = W^K x_i, \quad v_i = W^V x_i \quad (3-7)$$

接下来通过查询 q_i 和所有键 k_j 进行点积得到注意力分数：

$$e_{ij} = q_i \cdot k_j \quad (3-8)$$

通常还会对这些分数进行缩放，以减小数值过大带来的不稳定性，特别是在高维空间中：

$$e_{ij} = \frac{q_i \cdot k_j}{\sqrt{d_k}} \quad (3-9)$$

其中 d_k 是键向量的维度。然后使用 **Softmax** 函数将注意力分数归一化为概率分布，得到注意力权重：

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})} \quad (3-10)$$

再利用注意力权重对值向量进行加权求和，得到最终的输出：

$$z_i = \sum_{j=1}^n \alpha_{ij} v_j \quad (3-11)$$

在自注意力机制（**Self Attention**）中，输入序列的每一个元素既作为查询，又作为键和值。上述的步骤应用于输入序列的每一个元素，形成矩阵运算形式：

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \quad (3-12)$$

$$E = QK^T / \sqrt{d_k} \quad (3-13)$$

$$A = \text{softmax}(E) \quad (3-14)$$

$$Z = A V \quad (3-15)$$

3.1.3 LSTM-Attention

在时间序列的预测中，SOH 预测被视为一个频繁研究的任务。此任务一般包括处理那些随时间变化、高度依赖的数据。通过使用 LSTM，能极有效地捕捉到数据集中存在的长期依赖性关联。多数 SOH 预测任务是基于跨越很长时段的数据来进行的，如涉及电压、电流等重要信息。这类数据一般都蕴含着丰富的过往资料，并展现了一个明确的长远变化方向。LSTM 利用了它独有的门控技术，可以处理和储存这些长时间的依赖性数据，进而更准确地预测电池的健康状态。

Attention 策略能够通过每一预测过程中灵活筛选相关的输入数据，从而进一步提升其执行效率。SOH 预测任务中的这点变得尤为关键，因为数据在不同的时间点可能对当前的健康状态带来各异的影响，且某些影响可能比其他因素更加显著。利用这种策略，模型能够更为灵活地为输入序列的各个部分分配恰当的比例（更明确地呈现关键信息区），进而增强了预测的准确率。传统的 LSTM 模型通过压缩输入序列来创建固定长度的上

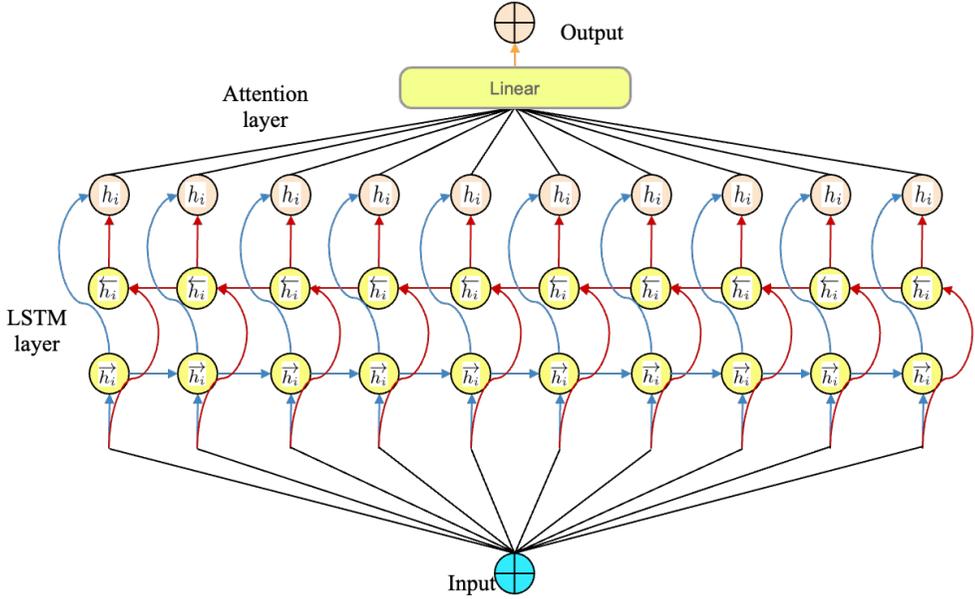


图 3-2: LSTM-Attention 结构示意图

下文向量，这种方式容易丢失有价值的信息。另一方面，通过根据重要性为各个时间步骤分配权重，Attention 使模型能够在每个预测步骤中对序列中的所有数据点有清晰的了解；从而确保没有信息瓶颈。通过这样做，模型能够捕获更多的历史数据，尤其是那些与当前预测直接相关的数据，从而提高预测性能。此外，由于 SOH 预测任务的复杂性，建立输入序列和输出序列之间的对齐是很复杂的任务。但 Attention 通过将每个预测步骤与来自不同时间步骤的输入的每个实例之间的相关性计算进行软对齐，极大地简化了这个过程。这种对齐机制的本质是更准确地掌握输入序列中对当前预测最重要的部分。综合以上原因，本文采用 LSTM-Attention 模型作为基于片段数据的 SOH 预测的解决方案，其计算过程如下：

首先创建 LSTM 编码器：

$$h_t, c_t = \text{LSTM}(x_t, h_{t-1}, c_{t-1}) \quad (3-16)$$

其中， x_t 是输入序列在时间步 t 的输入， h_t 是隐藏状态， c_t 是细胞状态。随后利用 Attention 机制对每个时间步进行解码，计算输入序列每个时间步的注意力权重：

$$e_{t,i} = \text{score}(h_t, h_i) \quad (3-17)$$

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{\sum_{j=1}^n \exp(e_{t,j})} \quad (3-18)$$

其中， $\text{score}(h_t, h_i)$ 表示解码器隐藏状态与编码器隐藏状态之间的相似性评

分。接下来根据注意力权重，计算上下文向量：

$$c_t = \sum_{i=1}^n \alpha_{t,i} h_i \quad (3-19)$$

最后将上下文向量与解码器隐藏状态结合，生成最终的预测输出：

$$y_t = \text{Output}(h_t, c_t) \quad (3-20)$$

3.2 损失函数

本研究选择了MSE（均方误差）作为LSTM-Attention模型在训练过程中所需的损失函数。这个函数是经常被用来评估预测值与实际值间偏差的回归损失函数，而其计算方法的公式是所有误差平方和的平均数值。MSE具备若干优势，首要之处是它的设计简洁明了，而计算方法也相当直观。接着，MSE对于大规模误差显示出更高的敏感性，主要因为误差平方会放大那些具有较强偏差值的错误，从而导致模型更加注重和纠正这些严重的误差。

在SOH的预测任务中，MSE发挥着不可或缺的角色。首先，MSE这一损失函数可以用来有效地评估模型预测的准确度。在MSE较低的情况下，模型预测的SOH值更加接近实际值，显示了模型在预测准确性方面的可靠性。接着，在模型训练阶段，MSE公司通过指导优化算法不断地调整参数，以便最小化预测误差，并提升模型的准确性和可信度。终究，MSE的对大误差的高度敏感性有助于模型在预测过程中更精准地区分和纠正大的误差，进而提高SOH预测的整体准确性和稳定性，确保模型能够准确地估算电池的剩余使用时长及总体表现。MSE的计算公式如方程3-21所示：

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3-21)$$

其中， y_i 是真实值， \hat{y}_i 代表预测值。

3.3 误差分析

本文接下来将对LSTM-Attention模型的泛化误差进行分析。本文依旧采用在第2章用到的Rademacher方法进行分析，具体定义如方程2-7所示。在LSTM-Attention模型中，其Rademacher复杂度主要与其参数数量相关，因此本文将基于其参数数量进行分析。LSTM-Attention模型由三部分组成，分别是LSTM层，Attention层和全连接层。对于LSTM层而言，LSTM的复杂度主要来自于其参数的数量。LSTM包含输入门、遗忘门、输出门和候选记忆的计算，这些计算通过线性层完成。假设输入维度为 m ，隐藏单元数为 h ，LSTM一层的参数数量大约为： $4(mh + h^2 + h)$ 对于多层LSTM，

参数数量会相应增加。而对于 Attention 层，Attention 机制的复杂度主要来自于其线性变换。假设隐藏层维度为 h ，线性层的参数数量为： h ，其次是 softmax 操作，其本身并不增加模型参数。最后是全连接层，最后的全连接层从隐藏层输出维度 h 映射到输出维度 o ，参数数量为： h_o 。综合考虑，可以通过计算网络中每一层参数的数量来估计复杂度：

- LSTM 层：假设有 L 层，每层有 $4(mh+h^2+h)$ 个参数，则总参数数量为 $4L(mh+h^2+h)$ ；
- Attention 层：参数数量为 h ；
- 全连接层：参数数量为 h_o 。

所以总的参数数量为： $4L(dh+h^2+h)+h+ho$ 。Rademacher 复杂度通常与参数的对数成正比，因此，LSTM-Attention 模型的 Rademacher 复杂度可以估计为：

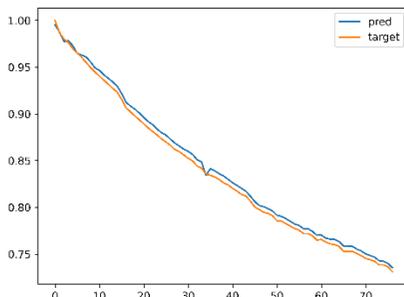
$$R_N(F_S) \propto \sqrt{\frac{1}{N} \log(4L(dh+h^2+h)+h+ho)} \quad (3-22)$$

3.4 实验验证

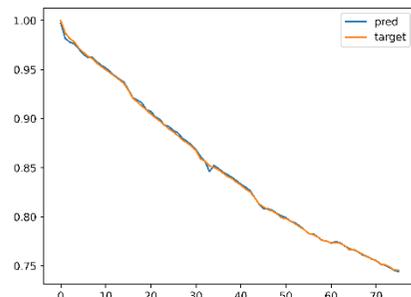
3.4.1 训练策略

与第 2 节相同，为了验证模型在真实工况下的实用性，针对 LSTM-Attention 模型的训练，如随机微分方程-ResNet 模型的训练一般，不对数据做过多的预处理。本研究首先选取如温度，电压，电流等能反映锂电池当前状况的数据作为训练数据，而后对其进行标准化处理以降低所需计算资源，后直接送入模型进行训练。评价指标同样采取与第 2 章相同的 RMSE 和 MAE，优化器也采用 Adam 优化器。

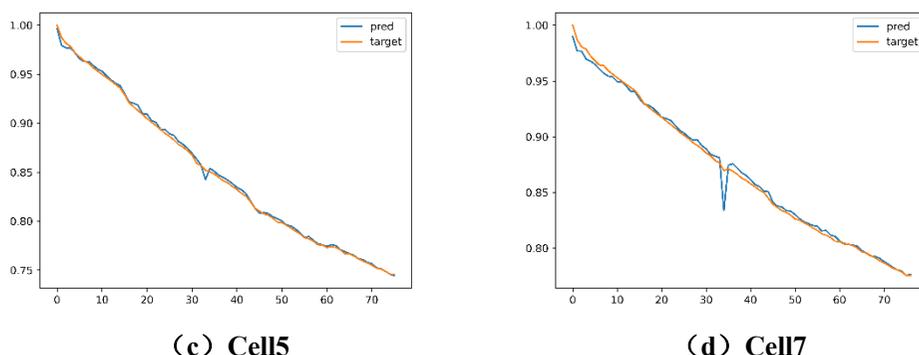
3.4.2 实验结果与分析



(a) Cell1



(b) Cell3



(c) Cell5 (d) Cell7

图 3-3 LSTM-Attention 在 Oxford 数据集上测试结果

在本研究中，为了进一步验证 LSTM-Attention 模型在 SOH 预测任务上的有效性，本文选取了 Oxford 数据集中另外四个电池单元（Cell1、Cell3、Cell5、Cell7）作为测试对象。实验结果如图 3-3 所示，从趋势上看，LSTM-Attention 模型学习到的拟合曲线与真实值高度一致，几乎完全重合，展示了该模型在趋势预测方面的卓越能力。数值结果如表 3-1 所示，无论从 RMSE 还是 MAE 的角度来看，所有测试单元在这两个评价指标上的数值均小于 0.01，其中 Cell3 的 RMSE 仅为 0.0017，MAE 仅为 0.0012，进一步验证了本文在第 3.1 节中对模型在 SOH 预测任务上强劲表现的假设。这种出色的表现不仅体现了 LSTM-Attention 模型在捕捉电池 SOH 变化趋势上的优越能力，也显示了其在数值拟合上的高度精确性。模型能够在不同电池单元上保持高预测准确性，说明了其良好的泛化能力。无论是 Cell1、Cell3、Cell5 还是 Cell7，LSTM-Attention 模型都能提供精确的 SOH 预测，说明了其强大的泛化能力。LSTM-Attention 模型在 SOH 预测任务上的优异表现归功于其结合了 LSTM 和 Attention 机制的优势。LSTM 有助于精确地捕获时间序列数据内的长期依赖关系，而 Attention 机制可以动态地调整模型在不同时间节点的关注度，这样做将有助于提升预测精度和模型解释性。

表 3-1 LSTM-Attention 在 Oxford 数据集上数值测试结果

Cell No.	RMSE	MAE
Cell1	0.0056	0.0054
Cell3	0.0017	0.0012
Cell5	0.0027	0.0021
Cell7	0.0053	0.0032

3.5 本章小结

本章提出了一种基于片段数据的锂电池 SOH 预测方法。首先详细介绍了 LSTM 和 Attention 机制的原理及其在时间序列数据处理中的应用，结合两者的优势，构建了 LSTM-Attention 模型。随后对 LSTM-Attention 模型进行了理论误差分析，并在 Oxford 数据集上进行了实验验证。

第4章 基于知识蒸馏框架的 SOH 预测方法

4.1 知识蒸馏框架

在本研究中，本文提出了一种全新的跨模态知识蒸馏框架，这与传统的师生网络结构通常在架构上完全一致的范式有所不同。与传统方法不同，本文的框架采用异构结构的网络，将随机微分方程-双支 ResNet 模型作为教师网络，LSTM-Attention 模型作为学生网络。通过整合两种不同的网络架构，本文的框架充分利用了每种模型的独特优势和能力。LSTM-Attention 以其在捕捉序列数据中的长距离依赖关系的能力而闻名。它使用 SOH 数据进行离线训练，并将其理解复杂模式的能力传授给随机微分方程-双支 ResNet 模型。反过来，随机微分方程-双支 ResNet 模型以其丰富的理论支持及其时间序列特征捕获能力，补充了 LSTM-Attention 的能力。这种协同互动增强了学生网络的整体预测性能，使其能够对输入数据进行更细致和全面的表示。

采用跨模态 K-D 框架的主要优点之一在于其减轻模型冗余的能力。传统的 K-D 框架通过使用相同的教师-学生模型，可能会强化对冗余特征的学习，从而限制模型提升的潜力。相比之下，本文的方法鼓励从不同角度探索数据，减少模型冗余和复杂性，而不会牺牲核心预测能力。此外，教师和学生网络的异构性引入了对数据噪声和异常值的不同敏感性。通过知识蒸馏过程，学生模型继承了教师模型的韧性和稳健性，从而在不利的数据条件下增强了其稳定性和可靠性。这一方面在数据质量可能变化或容易受到干扰的应用中尤为有利。本文框架的核心在于促进从随机微分方程-双支 ResNet 模型到 LSTM-Attention 的知识转移，前者在数据中捕捉了复杂的模式和依赖关系。这种知识的转移使学生模型能够利用教师模型的分析能力，显著减少预测误差，提高 K-D 过程的整体效能。

为了更好地理解模型的内部计算过程，下面将以数学方式进行解释。给定一个 SOH 退化数据集 $D = \{y^{(k)} \mid k = 1, \dots, N\}$ ，其中 $y^{(k)}$ 表示第 k 个循环的 SOH 值，教师网络的预训练数据集 $D_T = \{X^{(k)}, y^{(k)} \mid k = 1, \dots, N-1\}$ ，其中每个样本 $X^{(k)} = y^{(k-1)}$ 有一个通道和一个时间步，并且与真实值 $y^{(k)}$ 相关联。然后，建立一个简单的随机微分方程-双支 ResNet 模型作为教师网络，以学习非线性映射 $F_T: R^1 \rightarrow R^1$ ，该映射反映了 SOH 的退化。给定上述教师网络 F_T ，可以得到一系列来自 F_T 的估计 SOH 值 $\tilde{y}_T^{(k)}$ ，其计算公式如方程 4-1 所示：

$$\tilde{y}_T^{(k)} = F_T(X^{(k)}; \theta_T) \quad (4-1)$$

F_T 通过参数 θ_T 输入序列进行一系列变换，以导出真实值的表示。在这里， $X^{(k)}$ 是第 $(k-1)$ 次循环的真实值 SOH，而 $\tilde{y}_T^{(k)}$ 代表教师网络在第 k 个时间步的输出，也就是第 k 次循环的预测值。通过这种设计，可以从输入序列中提取重要的时间信息并将其转化为相应的真实值。

以类似的方式，当处理包含直接可测特征（电压、电流等）的数据集 $D_{\text{origin}} = \{x_{\text{origin}}^{(k)} \mid k=1,2,\dots,N\}$ 时，其中 $x_{\text{origin}}^{(k)}$ 代表第 k 次循环的原始特征数据。需要注意的是，由于不同循环期间数据点长度的变化，这些 $x_{\text{origin}}^{(k)}$ 可能并不嵌入在同一个欧几里得空间中。因此，本研究选取能接受不等长输入的模式用作学生模型。此外，在实际操作中，值范围较大的特征可能会不成比例地影响模型的决策过程，潜在地引入偏差。为了解决这个问题并促进所有特征在模型中的更公平贡献，数据标准化被执行，从而消除了尺度差异，同时提升了模型的收敛速度。基于这些考虑，使用标准化技术为学生网络构建一个训练数据集，记作 $D_S = \{x^{(k)}, y^{(k)}\}$ ，其中 $y^{(k)}$ 代表第 k 次循环的真实值。随后，建立一个简单的 LSTM-Attention 作为学生网络，旨在学习一个非线性映射 $F_S: R^{d \times N} \rightarrow R^1$ (d 代表特征数)，有效地捕捉测量特征与 SOH 之间的关系。通过利用上述学生网络 F_S ，可以实现 SOH 的估计。学生网络的输出，记作 \tilde{y}_S ，可以使用方程 4-2 进行计算。

$$\tilde{y}_S^{(k)} = F_S(x^{(k)}; \theta_S) \quad (4-2)$$

其中， θ_S 表示学生网络的参数， $\tilde{y}_S^{(k)}$ 是第 k 次循环的估计 SOH。

4.2 模型结构

4.2.1 教师网络

在跨模态知识蒸馏框架中，教师模型旨在学习 SOH 的退化规律，因此在选择合适的模型作为教师网络时，什么模型能更好地捕捉 SOH 退化规律是一个关键考虑因素。本文选择了随机微分方程-双支 ResNet 网络作为教师网络的基础，与基于 RNN 的架构（如 LSTM 或 GRU）相比，由于 Skip Connection 的原因，随机微分方程-双支 ResNet 不会遇到梯度消失或爆炸的问题，因此它能更有效地捕捉数据中的长期依赖关系。这一属性在 SOH 估计任务中至关重要，因为电池健康的退化是由广泛的操作条件和使用模式的历史驱动的。此外，随机微分方程-双支 ResNet 不对输入数据的顺序施加任何假设，从而赋予它一定程度的尺度不变性。在 SOH 预测的背景下，由于不同的操作条件导致退化速度可能显著不同，这一特性使得模型能够专注于健康状态的相对变化，而不是过度受数据尺度的影响。除此之外，本研究对随机微分方程-双支 ResNet 模型的误差进行了理论分析，使其相

对于其他模型有更强的理论完备性，增加了其的模型可解释性。

4.2.2 学生网络

本研究选择 LSTM-Attention 模型作为跨模态知识蒸馏框架中的学生网络。LSTM 在处理和预测时间序列数据方面表现尤为出色，因为其特殊的门控机制能够有效保留和利用历史信息，从而捕捉数据中的长时间依赖关系。在进行 SOH 预测的过程中，电池的历史使用数据在预测未来的健康状况方面具有关键性作用，这使得 LSTM 显著提升了预估准确度。另外，Attention 的机制还增强了 LSTM 的工作表现。Attention 机制能实时监控输入序列的各个组成部分，这让模型能更精确地检测和运用最有价值的信息，同时能忽视无关数据。这个预测机制尤其适用于 SOH，鉴于各个时间点数据对电池健康状况的不同影响，Attention 机制能够自动筛选出对预测最具价值的时间点，从而进一步增强预测性能。LSTM-Attention 模型在研究数据中的复杂模式与非线性关联方面表现出色，对于识别电池生命周期中的复杂 SOH 变化模式来说，这是一个极为关键的预测任务。Attention 机制的加入不仅提升了模型在捕捉复杂模式方面的性能，而且在预测精度和稳定性方面表现更为卓越。Attention 机制在某种程度上本研究提供了解读能力，助于本研究明确预测模型在处理特定数据部分时，特别关注哪些内容。电池 SOH 预测等核心应用场合对此尤为关注，因为这可以为用户和工程师带来宝贵的建议和参考，辅助他们深入了解电池当前的运行和健康状况。

4.3 损失函数

在之前的分析中，本研究深入研究了教师损失（Wasserstein-2 距离），这对教师模型的离线训练阶段至关重要。然而，在在线实时训练领域，需要整合各种损失函数，因此必须同时考虑预测准确性和知识转移。这就需要同时使用硬损失和软损失，前者表明从地面实况标签中直接学习，后者则反映从教师模型中吸收知识。因此，本文提出了一个综合损失函数，通过对这些元素进行加权合并来计算。在知识蒸馏范式中，这些不同的损失函数发挥着重要作用。

接下来的本文将阐明学生网络在线训练中使用的损失函数--软损失、硬损失和总损失。这些函数的主要目的是简化从教师到学生模型的知识传递，促进后者的自主模式识别和精确预测。通过对这些不同的损失函数进行明智的整合和加权，本文制定了一个整体评估指标，其中既包括从真实标签中直接获取知识（硬损失），也包括从教师模型中传递知识（软损失）。这种综合损失指标使学生模型不仅能模仿教师模型，还能准确预测真实标签，从而增强其在新数据上的泛化能力。

4.3.1 软损失

在跨模态知识蒸馏框架中，基于随机微分方程-双支 ResNet 的教师网络和基于 LSTM-Attention 的学生网络的输出计算出的软损失是指导模型优化方向的关键。该损失由教师输出（称为“软标签”）与学生输出之间的均方误差确定，衡量它们对 SOH 模式理解的差异。软损失量化了教师网络传授的知识与学生网络当前理解之间的差距，模型通过最小化这一差距来促进学生网络从教师网络那里学习复杂的 SOH 模式。这一过程不仅增强了学生网络在模式检测与预测方面的准确率，同时也保证了在知识提炼阶段知识能够被有效地传播出去。软损失作为损失函数中的一环，旨在辅助学生网络进一步的完善和调整。其表达式如公式 4-4 所示，其中 $\hat{y}_T^{(k)}$ 表示通过 F_T 获得的教师网络的第 k 个输出， $\tilde{y}_S^{(k)}$ 表示学生网络的相应输出。

$$\hat{y}_T^{(k)} = F_T(\tilde{y}_S^{(k-1)}; \theta_T) \quad (4-3)$$

$$L_{\text{soft}}(\hat{y}_T, \tilde{y}_S) = \frac{1}{N-1} \sum_{k=1}^{N-1} (\hat{y}_T^{(k)} - \tilde{y}_S^{(k)})^2 \quad (4-4)$$

4.3.2 硬损失

在跨模态知识蒸馏体系内，依据 LSTM-Attention 的学生网络进行硬损失计算的重要性不容忽视。该方法是为了对比 LSTM-Attention 预测中的实际数值与所使用的“硬标签”之间的偏差。该差异引导着学生对网络进行优化，引导模型向最小化预估数据与真实的 SOH 值之间差距的目标发展。模型的反传播以及梯度减少都基于硬损失的原理，这确保了学生网络在数据模式上的自我认知并能对 SOH 做出准确的预测。尽管教师网络的知识助力学生的学习进展，学生具备独立进行准确预测的能力也同样显得非常关键。硬损失可以如方程 4-5 计算得到，其中 $y^{(k)}$ 表示第 k 个真实值，它是评估学生网络独立学习和预测准确性的重要指标。

$$L_{\text{hard}}(\tilde{y}_S, y) = \frac{1}{N} \sum_{k=1}^N (\tilde{y}_S^{(k)} - y^{(k)})^2 \quad (4-5)$$

4.3.3 总损失

在计算出软损失和硬损失后，将其进行加权从而得到总损失即最终的优化目标，其表达式如下：

$$L(y_i, \tilde{y}_i) = \alpha L_{\text{soft}}(\hat{y}_T, \tilde{y}_S) + (1-\alpha) L_{\text{hard}}(\tilde{y}_S, y) \quad (4-6)$$

在 α 值的选取上，由于计算资源的限制，本研究采用 Xie 和 Zeng（2024）所得到的结论即 $\alpha = 0.5$ ^[33]。

4.4 实验验证

4.4.1 训练策略

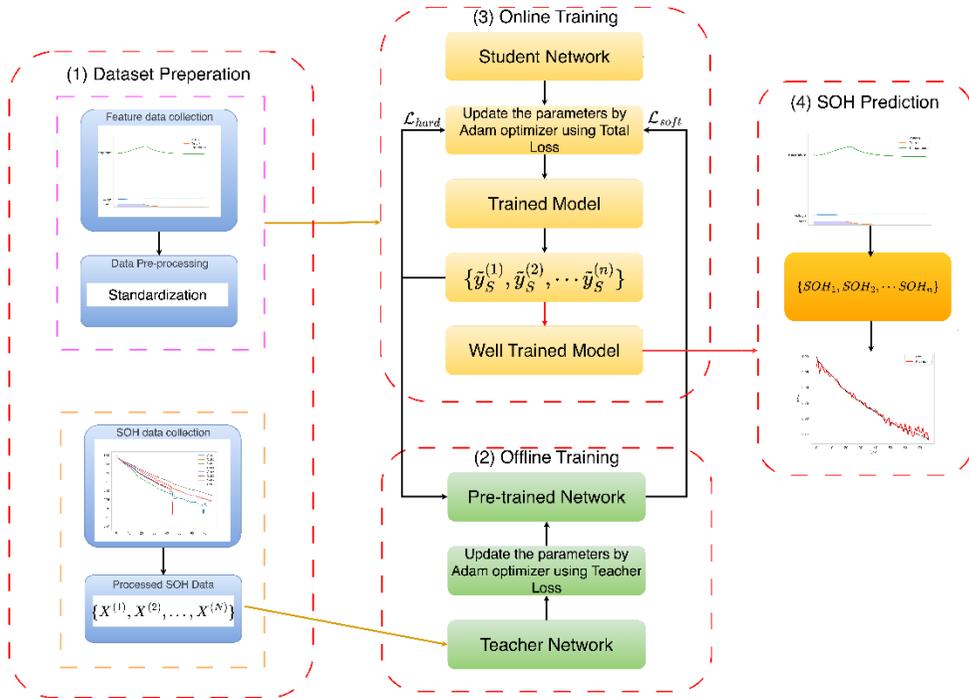


图 4-1: 跨模态知识蒸馏框架模型工作示意图

在本文提出的基于跨模态知识蒸馏框架的模型中，SOH 估算过程大致可分为四个部分：数据准备、教师网络离线训练、学生网络在线训练和 SOH 预测，如图 4-1 所示，接下来本文将对各个部分进行阐述：

(1) 数据准备

在数据准备过程中，实验数据集来自于采用不同充电和放电协议的电池，在训练前本研究对其进行了简单的预处理。具体而言，对于用于教师网络离线训练的 SOH 退化数据，将数据集通过标准化后，以 $y_{ii} = x_{ii+1}$ 的形式构建数据集，随后送入模型进行训练。其次，用于学生模型在线训练的原始数据经过标准化后，直接送入模型进行训练。预处理过程的详细细节在第 2.5.3 和第 3.4.1 节中进行了详细说明，提供了为有效模型训练准备数据所采用的方法的全面概述。

(2) 教师网络离线训练

数据准备完成后，通过 SOH 退化数据对教师网络进行离线训练。使用 Adam 优化器并以教师损失为优化目标，对教师网络的所有参数进行迭代更新。离线训练算法在算法 1 进行了说明。

(3) 学生网络在线训练

算法 2：学生网络在线训练策略
算法 2：学生网络在线训练策略

输入：初始可测量数据 $D_{\text{origin}} = \{X_{\text{origin}}^{(k)} \mid k=1,2,\dots,N\}$

初始化： θ (随机选择)

1: 标准化 $D_{\text{origin}} = \{X_{\text{origin}}^{(k)} \mid k=1,2,\dots,N\}$ 得到预处理数据集

$D_S = \{X^{(k)}, y^{(k)} \mid k=1,2,\dots,N\}$

2: 冻结预训练教师网络的参数 θ_T

3: **while** Not converged **do**

4: **for each batch do**

5: 将 D_S 输入 F_S 得到 $\tilde{y}_S^{(k)}$

6: 将 $\tilde{y}_S^{(k)}$ 输入 F_T 得到 $\hat{y}_T^{(k)}$

7: $L_{\text{soft}} \leftarrow \frac{1}{N-1} \sum_{k=1}^{N-1} (\hat{y}_T^{(k)} - \tilde{y}_S^{(k)})^2$

8: $L_{\text{hard}} \leftarrow \frac{1}{N} \sum_{k=1}^N (\tilde{y}_S^{(k)} - y^{(k)})$

9: $\ell \leftarrow \frac{1}{2} L_{\text{hard}} + \frac{1}{2} L_{\text{soft}}$

10: $\theta = \theta - \gamma \nabla_{\theta} \ell$

11: **end while**

输出： \hat{y}

离线训练完成后，预训练教师网络的参数被冻结，然后通过形成的原始数据对学生网络进行训练。在线训练过程如图 4-2 所示。对于每个 batch，当前 batch 的学生网络的输出序列作为输入传入到教师网络。然后，通过软损失 L_{soft} （通过学生网络的输出 \tilde{y}_S 和教师网络的输出 \hat{y}_T 计算得出）和硬损失 L_{hard} （通过学生网络的输出 \tilde{y}_S 和真实值 y 计算）来计算总损失 L_{total} 。之后，使用 Adam 优化器迭代更新学生网络的参数以最小化总损失 L_{total} 。在线训练算法总结如算法 2 所示。

(4) SOH 预测

在对学生网络进行在线训练后，利用训练好的学生网络，将处理好的原始数据输入模型预测其他电池的 SOH。

4.4.2 评价指标

除了在第 2.5.4 节中提到的 RMSE 和 MAE 外，为了进一步证明所提出的跨模态知识蒸馏框架的有效性，本研究引入了两个额外的评估指标，即

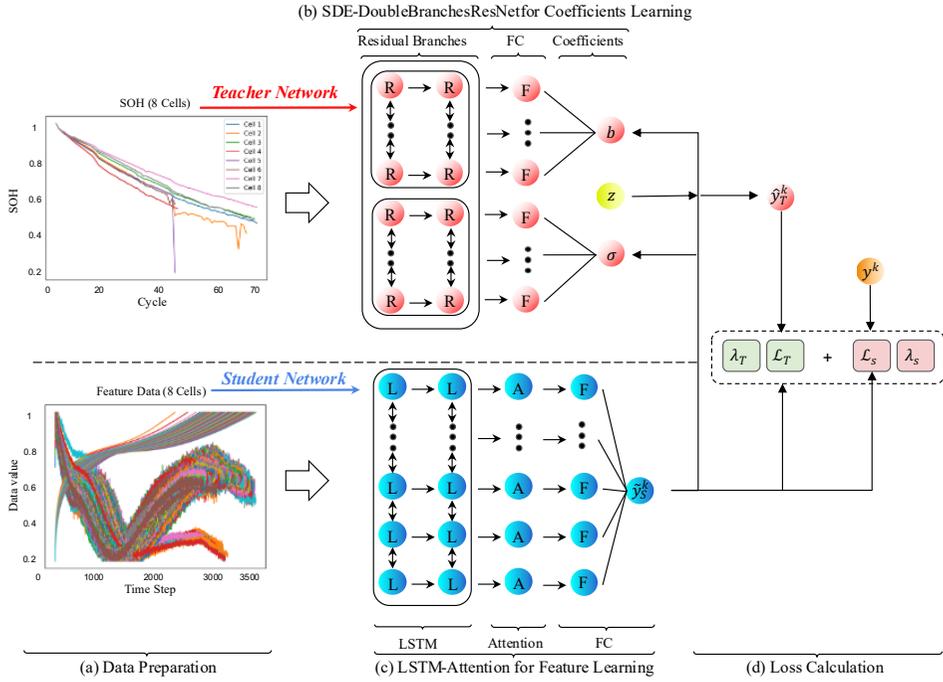


图 4-2: 学生网络在线训练示意图

知识嵌入的直接增益和有效增益。这些指标作为跨模态 K-D 策略相对于基线模型 LSTM-Attention 所取得的性能提升的有力证明。直接增益和有效增益的计算基于从基线模型 (ζ_1)、所提出的框架 (ζ_2) 以及所提出的教师网络 (ζ_3) 获得的指标结果，并根据以下方程进行计算：

直接增益 (Direct Gain) 计算方式为基线模型的指标结果与所提出的框架的指标结果之间的差值，它描述使用该框架所取得的直接改进。其计算公式如公式 4-7 所示：

$$\text{Direct Gain} = \zeta_1 - \zeta_2 \quad (4-7)$$

有效增益 (Effective Gain) 计算方式的核心是将直接的增益和由教师网络得到的数据进行统一整合。借助这种标准化手段，使得研究者能评定知识蒸馏架构在提升性能方面的有效性。其计算方法如公式 4-8 所示：

$$\text{Effective Gain} = \frac{\zeta_1 - \zeta_2}{\zeta_3} \quad (4-8)$$

通过直接和有效增益，能够深化本文对于跨模态知识蒸馏框架应用带来的性能增益的理解。

4.4.3 实验结果与分析

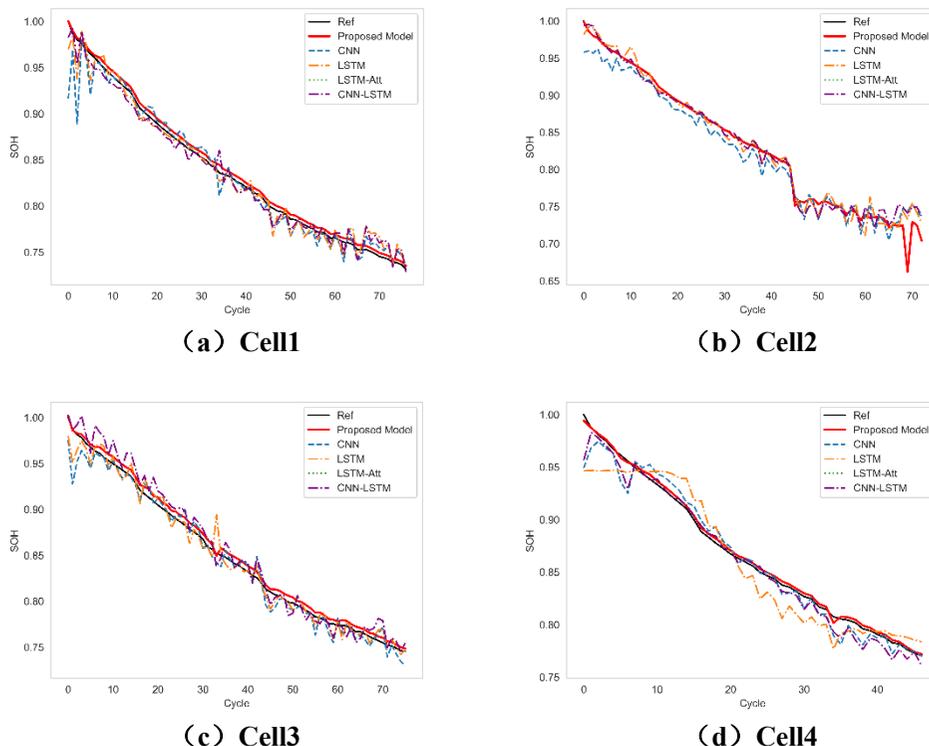


图 4-3 K-D 框架在 Oxford 数据集上测试结果 (Cell1-Cell4)

结果显示，基于知识蒸馏框架的模型与实际 SOH 值的匹配度更高，显著优于 CNN、CNN-LSTM 和 LSTM-Attention 模型等这些已被验证在特征提取和时间序列数据处理以及 SOH 预测任务上表现出色的模型。通过结合 LSTM-Attention 模型的强悍能力与随机微分方程-双支 ResNet 对 SOH 趋势变化的洞察力，本研究提出的模型在预测锂电池 SOH 方面超越了基线模型 (LSTM-Attention) 和之前占主导地位的 CNN-LSTM。

数值测试结果如表 4-1 中所示，同样采用 RMSE 和 MAE 作为评估指标。结果显示，所有电池单元的 RMSE 均小于 0.01，且除 Cell3 外，所有电池单元的表现均优于基线模型 (LSTM-Attention)，进一步说明本研究所提出的 K-D 策略的有效性和泛化性。特别是 Cell2 和 Cell5 的预测最为准确，RMSE 分别为 0.0006 和 0.0009。通过表 4-2 中的直接和有效增益计算，进一步验证了所提模型的有效性。与基线 LSTM-Attention 模型进行对比，所提模型在大多数情况下表现出正增益，证明了跨模态 K-D 策略对准确预测的帮助。虽然 Cell3 的直接和有效增益小于 0，本文推测这是因为 Cell3 的 SOH 退化存在奇异数据，导致教师模型不能很好地捕捉其中的退化规律。

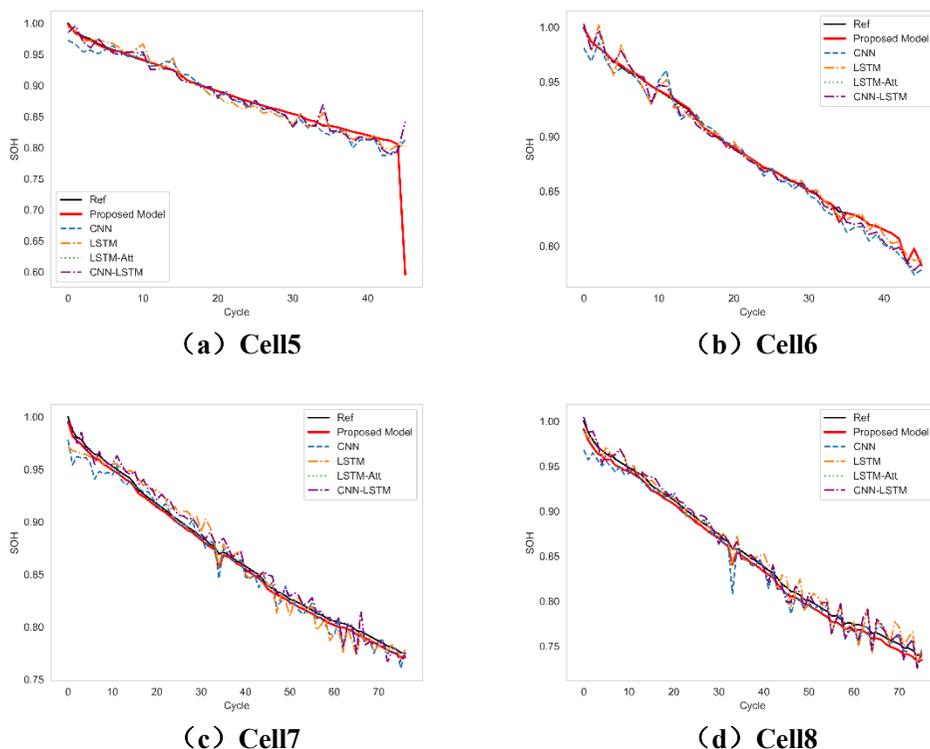


图 4-4 K-D 框架在 Oxford 数据集上测试结果 (Cell5-Cell8)

表 4-1 K-D 框架在 Oxford 数据集上数值测试结果

Criteria	Cell No.	CNN	LSTM	CNN-LSTM	LSTM-Att	Proposed Model
RMSE	Cell1	0.1676	0.0116	0.0106	0.0056	0.0048
	Cell2	0.0171	0.0128	0.0122	0.0021	0.0006
	Cell3	0.0112	0.0102	0.0119	0.0017	0.0041
	Cell4	0.1398	0.0104	0.0084	0.0033	0.0030
	Cell5	0.0348	0.3200	0.0378	0.0027	0.0009
	Cell6	0.0097	0.0073	0.0075	0.0056	0.0015
	Cell7	0.009	0.0103	0.0067	0.0053	0.0037
	Cell8	0.0107	0.0091	0.0091	0.0062	0.0061
MAE	Cell1	0.0111	0.0077	0.0078	0.0054	0.0047
	Cell2	0.0107	0.0085	0.0078	0.0017	0.0005
	Cell3	0.0086	0.0072	0.0064	0.0012	0.0038
	Cell4	0.0101	0.0080	0.0136	0.0028	0.0028
	Cell5	0.0134	0.0125	0.0136	0.0021	0.0007
	Cell6	0.0081	0.0055	0.0060	0.0032	0.0009
	Cell7	0.0069	0.0083	0.0053	0.0032	0.0035
	Cell8	0.0076	0.0067	0.0067	0.0061	0.0059

表 4-2 K-D 框架与基线模型间的直接和有效增益结果

Criteria	Cell1	Cell2	Cell3	Cell4	Cell5	Cell6	Cell7	Cell8
Direct Gain	0.0008	0.0015	-	0.0003	0.0018	0.0041	0.0016	9.9-5e
Effective Gain	0.2051	0.1271	-	0.0698	0.6000	0.8723	0.3902	0.0154
			0.5854					

4.5 本章小结

本章介绍了一种基于跨模态知识蒸馏框架的锂电池 SOH 预测手段。首先，本文对 K-D 框架进行了详尽的阐述并强调其优势，接着是基于这一框架来建构相应的模型。本研究采用随机微分方程和双支 ResNet 模型作为教师网络，同时利用 LSTM-Attention 模型作为学生网络，构建了一种既新颖又高效的跨模态知识传输机制。

在实验环节，本研究使用 Oxford 数据集作为验证工具，并把这些数据与众多的基线模型（如 CNN、CNN-LSTM、LSTM-Attention）作了对照和比较。研究结果揭示，本研究提出的 K-D 模型在预测准确度和稳健性上的表现显然超越了标准模型。尤为显著的是，Cell2 和 Cell5 的预测准确性极高，其 RMSE 的数值分别达到了 0.0006 和 0.0009，进一步证实了 K-D 框架的高效与优越特性。

结论

本文提出了三种锂电池健康状况评估方法，分别是基于 SOH 退化数据的预测方法（随机微分方程-双支 ResNet），基于片段数据的预测方法（LSTM-Attention），和基于知识蒸馏框架的评估方法，并对三种方法均进行了实验验证。基于知识蒸馏框架的 SOH 评估方法可以认为是前两种方法的结合，目的是提高锂电池状态（SOH）的预测准确性和稳定性。通过将随机微分方程-双支 ResNet 模型作为教师网络，LSTM-Attention 模型作为学生网络，实现了跨模态的知识转移机制，充分利用了两种模型的优势。实验结果表明，所提出的知识蒸馏框架在 Oxford 数据集上表现出色，所有测试单元的 RMSE 均小于 0.01，且大多数电池单元的预测性能显著优于基线模型（LSTM-Attention），特别是 Cell2 和 Cell5 的 RMSE 分别为 0.0006 和 0.0009，进一步验证了框架的有效性。本文的主要贡献主要包括以下方面：

- (1) **一定的理论基础：**本研究对教师网络（随机微分方程-双支 ResNet），学生网络（LSTM-Attention）均进行了详细的理论误差分析和泛化能力评估，为模型提供了理论结果，增强了模型的可解释性和可靠性。
- (2) **跨模态知识蒸馏框架：**通过融合随机微分方程-双支 ResNet 和 LSTM-Attention 模型，本研究设计了一个独特的跨模态知识蒸馏策略，这大幅增强了 SOH 的预测准确率和稳定性。传统的知识蒸馏方法通常依赖同构网络，而本研究则选择异构网络，这充分利用了多种模型的优势，从而进一步优化了预测的效能。
- (3) **一定的实验基础：**在 Oxford 数据集内，本研究设计了相关的试验，涉及到 8 个电池单元的实验。实验结果还与一系列已经在 SOH 预测任务中显示出优越表现的基线模型进行了深入比较，从而更好地展示了模型在性能和泛化能力上的卓越表现。此外，为了深入研究知识蒸馏框架与模型效果之间的关系，此项研究也采用了直接增益以及有效增益的方法来评估知识蒸馏框架中的教师模型为模型效果带来的增益。
- (4) **提升预测效果：**在 Oxford 数据集进行的测试显示，本研究提出的跨模态知识蒸馏框架在进行 SOH 的预测时表现卓越。在数值和趋势预测方面，所有的测试单位都展示出相当强的匹配度。RMSE 的数值都低于 0.01，而且对于绝大多数电池单元的预测能力明显优越于基线模型（LSTM-Attention）。尤其是在 Cell2 与 Cell5 方面，RMSE 的数值表现分别为 0.0006 和 0.0009，这

进一步证实了该模型在准确性和性能方面的优越性。

致谢

行文至此，我的大学生涯也即将画上句号。回顾这段难忘的时光，我心中充满了无限的感激与不舍。在这里，我不仅学到了丰富的专业知识，也在探索自我兴趣和追求中取得了长足的进步，面对了许多挑战。从大一的迷茫，到逐步明确自己的方向，转专业进入全新领域的学习，每一步都是宝贵的经历。特别是在大二期间，突如其来的疫情使我们不得不在宿舍上网课，这对我们来说无疑是一次巨大的挑战。然而，这些经历不仅让我有所收获，也在很大程度上改变了我。

首先，我要由衷地感谢我的指导老师谢巍老师。从论文选题、开题到定稿，谢老师无数次地给予我悉心的指导和帮助，不断完善我的知识体系，以严谨细致的态度要求我，使得我的论文得以顺利完成。我同样要感谢大学期间所有的老师，感谢你们在我学习和生活中的悉心教诲和无私帮助。

其次，我要感谢我的家人。你们无条件的支持和关爱，见证了我每一步的成长。你们教会我真诚待人，正直勇敢，是我最坚强的后盾。

此外，我要感谢我的朋友们。山水一程，三生有幸。我们一起学习、相互支持、共同进步。你们的陪伴和帮助，使我的学术之旅充满了乐趣和动力。

最后，我要感谢参与我论文评审和答辩的各位老师。您的宝贵意见和建议，对我的论文质量的提高有着极大的帮助。您的严谨和认真，使我更加深刻地理解了学术研究的重要性和责任感。

感谢所有曾经帮助过我的人，你们的支持和鼓励将是我前行的不竭动力。

参考文献

- [1] Ouyang M, et al. A Dynamic Capacity Degradation Model and Its Applications Considering Varying Load for A Large Format Li-Ion Battery [J]. *Applied Energy*, 2016, 165: 48-59.
- [2] Ge M-F, et al. A Review on State of Health Estimations and Remaining Useful Life Prognostics of Lithium-Ion Batteries [J]. *Measurement*, 2021, 174: 109057.
- [3] Salucci C B, et al. Multivariable Fractional Polynomials for Lithium-Ion Batteries Degradation Models Under Dynamic Conditions [J]. *Journal of Energy Storage*, 2022, 52: 104903.
- [4] Su X, et al. Interacting Multiple Model Particle Filter for Prognostics of Lithium-Ion Batteries [J]. *Microelectronics Reliability*, 2017, 70: 59-69.
- [5] 韩槿萱. 基于分数阶模型的锂电池荷电状态和健康状态协同估计 [D]. 西安: 西安理工大学, 2024.
- [6] 朱浩然, 陈自强, 杨德庆. 基于差分热伏安法和高斯过程回归的锂电池健康状态估计[J]. *上海交通大学学报*, 2024: 1-16.
- [7] 李煜, 蔡玉梅, 曾凯, 等. 基于 EKF 算法的纯电动汽车锂电池 SOC 与 SOH 联合估算[J]. *邵阳学院学报(自然科学版)*, 2024, 21(02): 45-55.
- [8] Wang Y, et al. A Comprehensive Review of Battery Modeling and State Estimation Approaches for Advanced Battery Management Systems [J]. *Renewable and Sustainable Energy Reviews*, 2020, 131: 110015.
- [9] Zhang J, et al. SOH Estimation of Lithium-Ion Batteries Based on Least Squares Support Vector Machine Error Compensation Model [J]. *Journal of Power Electronics*, 2021, 21: 1712-1723.
- [10] Li Y, et al. Random Forest Regression for Online Capacity Estimation of Lithium-Ion Batteries [J]. *Applied Energy*, 2018, 232: 197-210.
- [11] Feng H, Shi J. Research on Battery Life Prediction Method Based on Gaussian Process Regression [C]. *2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, IEEE, 2022.
- [12] Chemali E, et al. State-of-Charge Estimation of Li-Ion Batteries Using Deep Neural Networks: A Machine Learning Approach [J]. *Journal of Power Sources*, 2018, 400: 242-255.
- [13] Chemali E, et al. A Convolutional Neural Network Approach for Estimation of Li-Ion Battery State of Health from Charge Profiles [J]. *Energies*, 2022, 15(3): 1185.
- [14] Park K, et al. LSTM-Based Battery Remaining Useful Life Prediction with Multi-Channel Charging Profiles [J]. *IEEE Access*, 2020, 8: 20786-20798.
- [15] Chen D, Hong W, Zhou X. Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries [J]. *IEEE Access*, 2022, 10: 19621-19628.

- [16] 方斯顿, 刘龙真, 等. 基于双向长短期记忆网络含间接健康指标的锂电池 SOH 估计[J]. *电力系统自动化*, 2024, 48(04): 160-168.
- [17] Gong L-H, et al. Quantum Convolutional Neural Network Based on Variational Quantum Circuits [J]. *Optics Communications*, 2024, 550: 129993.
- [18] Tian J, et al. Deep Learning Framework for Lithium-Ion Battery State of Charge Estimation: Recent Advances and Future Perspectives [J]. *Energy Storage Materials*, 2023: 102883.
- [19] Jovanovic L, et al. Multi-Step Crude Oil Price Prediction Based on LSTM Approach Tuned by Salp Swarm Algorithm with Disputation Operator [J]. *Sustainability*, 2022, 14(21): 14616.
- [20] Song X, et al. Combined CNN-LSTM Network for State-of-Charge Estimation of Lithium-Ion Batteries [J]. *IEEE Access*, 2019, 7: 88894-88902.
- [21] Mousapour M M, et al. Hybrid Neural Network-Based Metaheuristics for Prediction of Financial Markets: A Case Study on Global Gold Market [J]. *Journal of Computational Design and Engineering*, 2023, 10(3): 1110-1125.
- [22] Guo Y, et al. State of Health Estimation for Lithium-Ion Battery Based on Bi-Directional Long Short-Term Memory Neural Network and Attention Mechanism [J]. *Energy Reports*, 2022, 8: 208-215.
- [23] 潘子良, 朱成杰, 余梦书. 基于贝叶斯优化 LSTM 的锂电池健康状态评估方法[J]. *仪表技术*, 2024, (03): 66-69.
- [24] 辛付宇, 邢丽坤, 刘笑. 基于 CNN-GRU 神经网络的锂电池 SOH 估计与 RUL 预测[J]. *上海节能*, 2024, (05): 819-826.
- [25] 莫易敏, 余自豪, 等. 基于迁移学习与 GRU 神经网络结合的锂电池 SOH 估计[J]. *太阳能学报*, 2024, 45(03): 233-239.
- [26] 雷孟飞, 梁泉, 等. 基于自注意力和 GRU 的锂电池
- [27] Remlinger C, Mikael J, Elie R. Conditional Loss and Deep Euler Scheme for Time Series Generation [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(7).
- [28] He K, et al. Deep Residual Learning for Image Recognition [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [29] Xiao Y, Lu Z, Yang F. A ResNet-SVR Model for Lithium-Ion Battery State of Health Estimation [C]. *2023 5th International Conference on System Reliability and Safety Engineering (SRSE)*, IEEE, 2023.
- [30] Yoon J, Jarrett D, Van der Schaar M. Time-Series Generative Adversarial Networks [C]. *Advances in Neural Information Processing Systems*, 2019, 32.
- [31] Gelbrich M. On a Formula for the L2 Wasserstein Metric Between Measures on Euclidean and Hilbert Spaces [J]. *Mathematische Nachrichten*, 1990, 147(1): 185-203.
- [32] Bhatia R, Jain T, Lim Y. On the Bures–Wasserstein Distance Between Positive Definite Matrices [J]. *Expositiones Mathematicae*, 2019, 37(2): 165-191.

- [33]Xie W, Zeng Y. A Knowledge Distillation Based Cross-Modal Learning Framework for the Lithium-Ion Battery State of Health Estimation [J]. *Complex & Intelligent Systems*, 2024: 1-23.

附录 A



ELSEVIER



Available online at www.sciencedirect.com

ScienceDirect

Energy Reports 8 (2022) 208–215



www.elsevier.com/locate/egy

International Conference on the Energy Internet and Energy Interactive Technology, South China University of Technology, 2022, 25–27 March 2022, Wuhan, China

State of health estimation for lithium-ion battery based on Bi-directional long short-term memory neural network and attention mechanism

Yu Guo^a, Dongfang Yang^b, Kun Zhao^c, Kai Wang^{a,*}

^a School of Electrical Engineering, Weihai Innovation Research Institute, Qingdao University, Qingdao, 266000, China

^b Xi'an Traffic Engineering Institute, Xi'an 710300, China

^c Shandong Wide Area Technology Co., Ltd, Dongying 257081, China

Received 4 October 2022; accepted 7 October 2022

Available online 25 October 2022

Abstract

At present, lithium-ion batteries (LIBs) play an irreplaceable role in various fields of production and life as an efficient energy storage element. The state of health (SOH) for LIB is critical to the safe operation of energy storage system. In fact, it is currently difficult to estimate SOH of LIB quickly and accurately. This paper proposes a method for SOH estimation that combines bidirectional long short-term memory (BiLSTM) neural network and attention mechanism. We extract three features from the incremental capacity (IC) curve as inputs to the model. The correlation rates between the proposed features and battery capacity are more than 0.98. Finally, the NASA dataset is introduced for experimental verification. The verification results demonstrate that the proposed method achieves accurate estimation of the SOH for LIBs. In the experimental results, the root mean square error (RMSE) and mean absolute percentage error (MAPE) of the proposed method can be as low as 0.0051 and 0.34%, respectively.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the International Conference on the Energy Internet and Energy Interactive Technology, South China University of Technology, 2022.

Keywords: Lithium-ion battery; State of health; BiLSTM; Attention mechanism

1. Introduction

Today LIBs are widely used in various fields such as mobile communications, power storage, and aerospace. As the cycles that LIBs are used increase, the LIBs will slowly age, resulting in performance degradation. SOH is an important index to measure battery aging. Accurate estimation of SOH is critical for equipment operation, because accurate estimation of SOH reduces the risk of battery aging [1–3]. However, the SOH of LIB cannot be obtained directly. We can only estimate the SOH using parameters such as the voltage, current, and internal resistance of

* Corresponding author.

E-mail address: wangkai@qdu.edu.cn (K. Wang).

<https://doi.org/10.1016/j.egy.2022.10.128>

2352-4847/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the International Conference on the Energy Internet and Energy Interactive Technology, South China University of Technology, 2022.

the battery. This paper proposes a SOH estimation method that combines BiLSTM neural network and attention mechanism. The SOH estimation of LIB can be realized by using IC curve.

SOH can be defined in terms of internal resistance or capacity. This paper defines SOH from the perspective of capacity, as shown in Eq. (1).

$$SOH = \frac{C_A}{C_N} \quad (1)$$

Where, C_A represents current available capacity, C_N represents nominal capacity.

SOH estimation can be achieved by different methods, such as electrochemical analysis methods, Kalman filter methods, particle filter methods, and data-driven methods. Each method has different strengths, weaknesses and levels of complexity, as well as the achievable estimation accuracy.

The method based on electrochemical analysis is to study the aging of LIB from the perspective of chemical reaction inside the battery. It realizes the estimation of SOH through lithium-ion diffusion, discharge current density, average radius of solid particles and changes of negative electrode materials. The accuracy of this method is very high, but complex partial differential equations need to be established. The parameters of equations are extremely difficult to calibrate and calculation cost is high. The most important is that precision measurement needs to be realized in the laboratory environment, so it cannot be applied in practice [4–6].

The methods based on Kalman filter or particle filter usually realize the estimation of SOH using the equivalent circuit model [7]. The advantage is that same model can be used to batteries with different capacities and even different internal chemical properties. The disadvantage is that the factors such as battery discharge power, charging rate and temperature, have nonlinear effects on the battery. It is difficult to take these factors account into the parameters of battery model [8–10].

LIB is a highly complex and nonlinear system. It is often difficult to establish a reliable and accurate equivalent model for the aging process of batteries [11,12]. Estimating SOH by a data-driven approach does not require a deep understanding of the internal aging principles for LIBs like electrochemical analysis methods, but extracts features closely related to battery aging from parameters such as terminal voltage, terminal current, and surface temperature of the battery [13]. These parameters can be obtained directly through the sensor. Finally, the nonlinear mapping relationship between these features and battery capacity can be established by data-driven method, which avoids the problems of chemical analysis and parameter identification [14]. Among the data-driven methods, neural networks, especially deep neural networks, are especially suitable for classification and estimation with complex causality. In this paper, the BiLSTM neural network is used to estimate SOH, which has the best prediction performance currently.

The rest of the paper is as follows. Section 2 presents the data sources and extracted features. Section 3 details the BiLSTM neural network and estimation method. Section 4 shows the experimental results based on the proposed method. Finally, Section 5 presents the conclusion of this paper.

2. Data analysis

2.1. LIB data set

This experiment used four batteries in NASA data set, B0005, B0006, B0007 and B0018. Fig. 1 shows the capacity decline curves of four batteries with the increase of the number of cycles. In Fig. 1, the capacity of the battery has local recovery phenomena during the decline process [15]. Accurate estimation of SOH should be able to accurately estimate the capacity of local recovery, rather than ignore the fluctuation of these capacity.

The experimental process of LIBs charged and discharged is as follows. First, the LIB is charged with a constant current of 1.5 A while the terminal voltage is gradually increased. When the terminal voltage reaches 4.2 V, the LIB is converted to 4.2 V constant voltage charging and the charging current starts to decrease gradually. The charging process ends when the charging current drops to 20 mA. During the discharge process, the LIB was discharged with a constant current of 2 A and the terminal voltage gradually decreased. This process continues until the terminal voltages of B0005, B0006, B0007, and B0018 drop to 2.7 V, 2.5 V, 2.2 V, and 2.5 V, respectively.

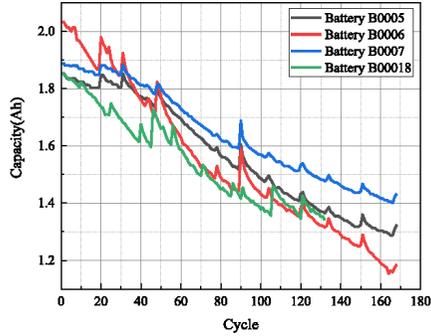


Fig. 1. Capacity decline curves of B0005, B0006, B0007 and B0018.

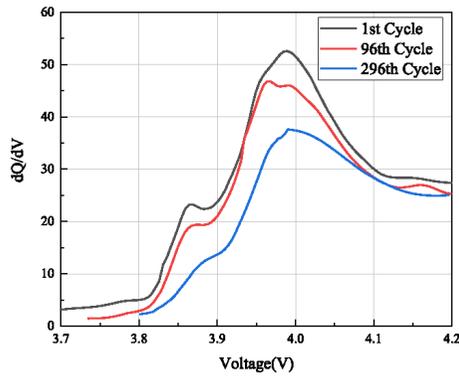


Fig. 2. IC curve of B0005 battery in different aging states.

2.2. Extract features

Increment capacity (IC) analysis is an important method to analyze the mechanism of battery decline. Its core is IC curve of LIB. Fig. 2 shows the IC curve of the B0005 battery in different aging states. The greater the number of cycles, the more serious the battery aging [16].

This paper uses the maximum peak of IC curve as the first feature. The second feature is the area included in IC curve when voltage is from 3.9v to 4.0 V. The third feature is the area included in IC curve when voltage is from 3.9v to 4.1 V. The three features are used as inputs to BiLSTM neural network.

Pearson correlation coefficient is widely used as an important index to measure correlation rate between two vectors. The calculation method of the Pearson correlation rate is shown in Eq. (2).

$$P_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

X and Y represent two vectors, $Cov(X,Y)$ represents the covariance of two vectors, σ_X and σ_Y represent the covariance of two vectors respectively.

Table 1. Correlation rate between three features and four battery capacities.

	B0005	B0006	B0007	B0018
Feature1	0.9842	0.9845	0.9889	0.9848
Feature2	0.9856	0.9889	0.9878	0.9881
Feature3	0.9875	0.9912	0.9921	0.9885

Table 1 shows the correlation rates between the three selected features and the capacities of the four batteries. It can be seen that the correlation rate is higher than 0.98, so a good estimation can be achieved.

3. Method proposed

3.1. BiLSTM neural network

Considering the temporality of the data of LIB, this paper selects BiLSTM neural network. BiLSTM neural network consists of two LSTM neural networks in opposite directions, so it has better performance in sequence related problems. BiLSTM neural network can forward and reverse process data through two LSTMs in opposite directions, so it can utilize both past and future information. Because the BiLSTM neural network can be trained with both past and future data, its estimation results are more accurate. Its network structure is shown in Fig. 3.

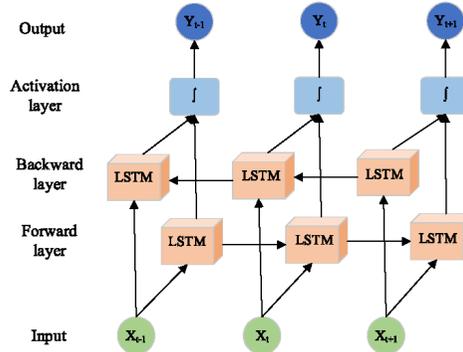


Fig. 3. The structure of BiLSTM neural network.

The structure of LSTM neural network is shown in Fig. 4. LSTM neural network consists of input x_t , cell state c_t , temporary cell state \tilde{c}_t , hidden state h_t , forget gate f_t , memory gate i_t , and output gate o_t , at time t . The three operations of forgetting, memory and output are controlled by forget gate, memory gate and output gate. The specific states of the three gates are calculated from the hidden state at the previous moment and the current input 17.

The formulas for forgetting gate f_t , memory gate i_t , temporary cell state \tilde{c}_t , cell state c_t , output gate o_t and hidden state h_t are as follows:

$$f_t = \sigma(\omega_f[h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(\omega_i[h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{c}_t = \tanh(\omega_c[h_{t-1}, x_t] + b_c) \tag{5}$$

$$o_t = \sigma(\omega_o[h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t * \tanh(c_t) \tag{7}$$

Where, h_{t-1} represents the hidden state layer at the last moment, σ and \tanh represents the nonlinearity, b represents the threshold and ω represents the weight.

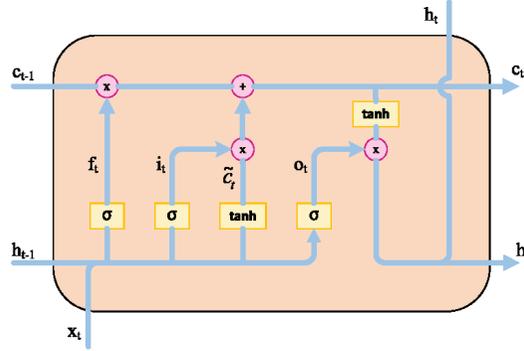


Fig. 4. Schematic of LSTM neural network.

3.2. Attention mechanism

For different features, their impact to the final output is different. Therefore, this paper adds an attention mechanism to improve the estimation accuracy, which can assign different weights to different features [17]. Its principle is as follows:

$$s_i = \Phi(\omega^T h_i + b) \quad (8)$$

h_i is the hidden state of the BiLSTM, T represents to transpose the matrix. ω and b represent the weight matrix and threshold vector respectively, $\Phi(\cdot)$ represents the nonlinear activation function, and s_i represent the importance of h_i .

After obtaining the importance for the i th feature vector, it can be normalized using a function as follows:

$$\alpha_i = \text{soft max}(s_i) = \frac{\exp(s_i)}{\sum_{i=1}^n \exp(s_i)} \quad (9)$$

After getting α_i , we can assign different weights to the features as follows:

$$O = H \otimes A \quad (10)$$

$H = \{h_1, h_2, \dots, h_n\}^T$, $A = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, O represents final output features of the attention mechanism, \otimes represents the dot product of corresponding elements between two vectors.

3.3. The method proposed

Fig. 5 shows the overall framework of the proposed estimation method. As shown in Fig. 5, the method proposed is divided into training module and estimation module. Firstly, we divide each battery data set into training set and test set. The training set and test set account for 70% and 30% of each battery dataset, respectively. In the training module, firstly, the features proposed are extracted from the training set, and then the features extracted are used to train the BiLSTM neural network with attention mechanism, which is called ABiLSTM neural network. In the estimation module, firstly, the features proposed are extracted from the test set, and then the ABiLSTM neural network trained in the training module is used to estimate the SOH of the data in the test set.

4. Experimental results

In this section, the proposed SOH estimation method is tested on four batteries (B0005, B0006, B0007, and B0018) in the NASA dataset. The experimental of B0005 results are shown in Fig. 6. Comparing with BiLSTM,

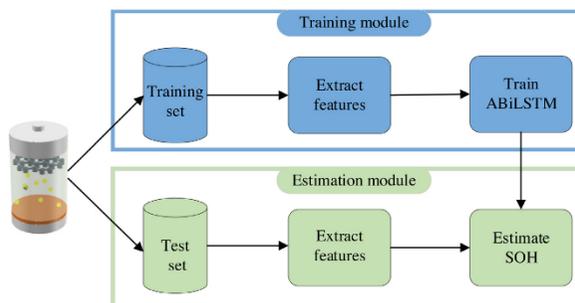


Fig. 5. The overall framework of the proposed SOH estimation method.

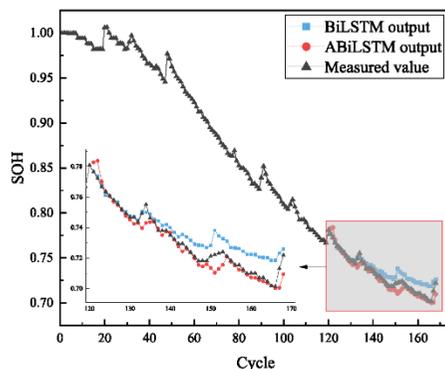


Fig. 6. B0005 battery estimation results.

ABiLSTM is able to understand the downward trend of the capacity curve. Due to the chemical side reactions inside LIB, local recovery and fluctuation will occur during the process of capacity reduction, which affects the capacity estimation. But the ABiLSTM model understands this volatility well and predicts an overall trend that matches the actual curve. Due to the role of attention mechanism, the mapping relationship between the features proposed and the capacity is refined, so that the estimation of the capacity is More accurate. In particular, the BiLSTM model cannot predict local trends well because it cannot accurately learn local fluctuations in capacity. This phenomenon proves that the attention mechanism is very effective in capturing local fluctuations and significantly improves the estimation performance.

Next, in Fig. 7 we present the errors of the estimation result. The error used is the difference between the estimated value and the measured value. As shown in Fig. 7, comparing with BiLSTM, the overall error of ABiLSTM is smaller, although the error may change significantly in some local areas. This further illustrates the accuracy and effectiveness of the proposed method.

To further illustrate the advantages of the ABiLSTM model, we present the RMSE and MPAE of the estimation results for four batteries, as shown in Table 2. In different batteries, the ABiLSTM model performs better because RMSE and MPAE are relatively small. Comparing with other methods [18,19], the proposed method requires less training data to achieve the same estimation accuracy. This proves that the proposed ABiLSTM model is more

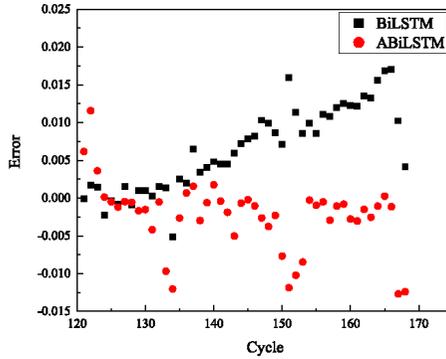


Fig. 7. B0005 battery estimation errors.

Table 2. RMSE and MAPE of four battery estimation results.

	B0005		B0006		B0007		B0018	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
BiLSTM	0.0085	0.69%	0.0156	1.34%	0.0089	0.72%	0.0105	0.99%
ABiLSTM	0.0051	0.34%	0.0093	1.02%	0.0055	0.36%	0.0078	0.68%

accurate and reliable in SOH estimation for LIBs. This advantage increases the possibility of applying the method proposed to practice [20]. In general, the method proposed in this paper has great application prospects compared to the existing methods.

5. Conclusion

Estimating SOH is an important function of the battery management system. This paper proposes a novel SOH estimation method for LIB, which combines BiLSTM neural network and attention mechanism. Three features were first extracted from the IC curve. Then, the output of the BiLSTM neural network is adjusted according to the importance of the feature to the output using the attention mechanism. Through the specific application of the above method, the numerical experimental results show that the method has less error in the SOH estimation of LIBs. In literature [21], it is proposed that BiLSTM performs better than CNN, LSTM and other neural networks in estimating SOH, so the method we proposed is also better than other neural networks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the Youth Fund of Shandong Natural Science Foundation, China (No. ZR2020QE212), Key Projects of Shandong Natural Science Foundation, China (No. ZR2020KF020).

References

- [1] M. Zhang, K. Wang, Y. Zhou, Online State of Charge Estimation of Lithium-Ion Cells Using Particle Filter-Based Hybrid Filtering Approach, *Complexity* 2020 (2022) 8231243, <http://dx.doi.org/10.1155/2020/8231243>.
- [2] Y. Guo, D. Yang, Y. Zhang, L. Wang, K. Wang, Online estimation of SOH for lithium-ion battery based on SSA-Elman neural network, *Prot. Control Mod. Power Syst.* 7 (40) (2022) <http://dx.doi.org/10.1186/s41601-022-00261-y>.
- [3] H. Sun, D. Yang, L. Wang, K. Wang, A method for estimating the aging state of lithium-ion batteries based on a multi-linear integrated model, *Int J Energy Res* (2022) <http://dx.doi.org/10.1002/ER.8709>.
- [4] Y. Guo, P. Yu, C. Zhu, K. Zhao, L. Wang, K. Wang, A state-of-health estimation method considering capacity recovery of lithium batteries, *Int J Energy Res* (2022) <http://dx.doi.org/10.1002/ER.8671>.
- [5] D. Li, D. Yang, L. Li, L. Wang, K. Wang, Electrochemical impedance spectroscopy based on the state of health estimation for lithium-ion batteries, *Energies* 15 (2022) 6665, <http://dx.doi.org/10.3390/en15186665>.
- [6] Z. Cui, L. Kang, L. Li, L. Wang, K. Wang, A hybrid neural network model with improved input for state of charge estimation of lithium-ion battery at low temperatures, *Renew Energy* 98 (2022) 1328–1340, <http://dx.doi.org/10.1016/j.renene.2022.08.123>.
- [7] D. Li, S. Li, S. Zhang, J. Sun, L. Wang, K. Wang, Aging state prediction for supercapacitors based on heuristic kalman filter optimization extreme learning machine, *Energy* 250 (2022) (2022) 123773.
- [8] Z. Cui, L. Kang, L. Li, L. Wang, K. Wang, A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF, *Energy* 259 (2022) 124933, <http://dx.doi.org/10.1016/j.energy.2022.124933>.
- [9] Y. Hua, N. Wang, K. Zhao, Simultaneous unknown input and state estimation for the linear system with a rank-deficient distribution matrix, *Math Probl Eng* 21 (2021) (2021) 6693690.
- [10] Z. Cui, L. Wang, Q. Li, K. Wang, A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network, *Int J Energy Res* 46 (2021) (2021) 5423–5440.
- [11] D. Feng, H. Du, H. Ran, T. Lu, S. Xia, L. Xu, et al., Antiferroelectric stability and energy storage properties of Co-doped AgNbO₃ ceramics, *J Solid State Chem* 310 (2022) (2022) 123081.
- [12] J. Xu, X. Mei, X. Wang, M. Fu, Y. Zhao, J. Wang, A relative state of health estimation method based on wavelet analysis for lithium-ion battery cells, *IEEE Trans Ind Electron* 68 (2021) (2021) 6973–6981.
- [13] C. Liu, D. Li, L. Wang, L. Li, K. Wang, Strong robustness and high accuracy remaining useful life prediction on supercapacitors, *APL Mater* 10 (6) (2022) 061106, <http://dx.doi.org/10.1063/5.0092074>.
- [14] C. Liu, Y. Zhang, J. Sun, Z. Cui, K. Wang, Stacked bidirectional LSTM RNN to evaluate the remaining useful life of supercapacitor, *Int J Energy Res* 46 (2021) (2021) 3034–3043.
- [15] Z. Xia, A. Jaber, Lithium-ion battery ageing behavior pattern characterization and state-of-health estimation using data-driven method, *IEEE Access* 9 (2021) (2021) 98287–98304.
- [16] E. Mahidin, M. Nasaruddin, Muhibbuddin, A critical review of the integration of renewable energy sources with various technologies, *Prot Control Mod Power Syst* 6 (2021) 1.
- [17] W. Wang, D. Yang, Z. Huang, H. Hu, L. Wang, K. Wang, et al., Electrodeless Nanogenerator for Dust Recover, *Energy Technol.* 2022 (2022) 202200699, <http://dx.doi.org/10.1002/ente.202200699>.
- [18] Z. Chen, Q. Xue, R. Xiao, Y. Liu, J. Shen, State of health estimation for lithium-ion batteries based on fusion of autoregressive moving average model and elman neural network, *IEEE Access* 7 (2019) (2019) 102662–102678.
- [19] W. Wang, D. Yang, X. Yan, L. Wang, H. Hu, K. Wang, Triboelectric nanogenerators: The beginning of blue dream, *Front. Chem. Sci. Eng.* (2022) <http://dx.doi.org/10.1007/s11705-022-2271-y>.
- [20] N. Khan, U. Fath, U. Amin, S. Baik, Batteries state of health estimation via efficient neural networks with multiple channel charging profiles, *IEEE Access* 9 (2021) (2021) 7797–7813.
- [21] Z. Yi, K. Zhao, J. Sun, L. Wang, K. Wang, Y. Ma, Prediction of the remaining useful life of supercapacitors, *Math Probl Eng* 2022 (2022) (2022) 7620382.

Further reading

- [1] H. Sun, J. Sun, K. Zhao, L. Wang, K. Wang, Data-driven ICA-Bi-LSTM-combined lithium battery SOH estimation, *Math. Probl. Eng.* 2022 (2022) (2022) 9645892.

附录 B

基于双向长短期记忆神经网络和注意力机制的锂电池健康状况评估 摘要

目前，锂电池（LIB）作为一种高效的储能元件，在生产和生活的各个领域发挥着不可替代的作用。锂电池的健康状态（SOH）对储能系统的安全运行至关重要。事实上，目前很难快速准确地估算锂电的健康状态。本文提出了一种结合双向长短期记忆（BiLSTM）神经网络和注意力机制的 SOH 估算方法。我们从增量容量（IC）曲线中提取了三个特征作为模型的输入。所提出的特征与电池容量之间的相关率超过了 0.98。最后，引入 NASA 数据集进行实验验证。验证结果表明，所提出的方法能够准确估计 LIB 的 SOH。在实验结果中，所提方法的均方根误差（RMSE）和平均绝对百分比误差（MAPE）分别低至 0.0051 和 0.34%。

关键词 锂离子电池；健康状况；BiLSTM；注意机制

1. 介绍

如今，锂电池广泛应用于移动通信、电力存储和航空航天等各个领域。随着锂电池使用周期的增加，锂电池会慢慢老化，导致性能下降。SOH 是衡量电池老化的一个重要指标。准确估算 SOH 对设备运行至关重要，因为准确估算 SOH 可以降低电池老化的风险 [1-3]。然而，LIB 的 SOH 无法直接获得。我们只能通过电池的电压、电流和内阻等参数来估算 SOH。本文提出了一种结合 BiLSTM 神经网络和注意力机制的 SOH 估算方法。利用 IC 曲线可以实现对 LIB 的 SOH 估算。

SOH 可以用内阻或容量来定义。本文从容量的角度定义 SOH，如公式 (1) 所示。

$$SOH = \frac{C_A}{C_N} \quad (1)$$

其中， C_A 代表当前可用容量、 C_N 代表额定容量。

SOH 估算可以通过不同的方法实现，如电化学分析方法、卡尔曼滤波方法、粒子滤波方法和数据驱动方法。每种方法都有不同的优缺点、复杂程度以及可达到的估计精度。

基于电化学分析的方法是从电池内部化学反应的角度来研究 LIB 的老化。它通过锂扩散、放电电流密度、固体颗粒平均半径和负极材料的变化来估算 SOH。这种方法的精度非常高，但需要建立复杂的偏微分方程。方程参数极难校准，计算成本高。最重要的是，精确测量需要在实验室环境中实现，因此无法在实际中应用。

基于卡尔曼滤波器或粒子滤波器的方法通常使用等效电路模型来实现 SOH 的估计。其优点是同一模型可用于不同容量甚至不同内部化学特性的电池。缺点是电池放电功率、充电速率和温度等因素会对电池产生非线性影响。很难将这些因素考虑到电池模型的参数中。

LIB 是一个高度复杂的非线性系统。通常很难为电池老化过程建立可靠、准确的等效模型。通过数据驱动法估算 SOH 不需要像电化学分析方法那样深入了解 LIB 的内部老化原理，而是从电池的端电压、端电流和表面温度等参数中提取与电池老化密切相关的特征。这些参数可直接通过传感器获得。最后，这些特征与电池容量之间的非线性映射关系可通过数据驱动方法建立，从而避免了化学分析和参数识别的问题。在数据驱动方法中，神经网络，尤其是深度神经网络，特别适用于具有复杂因果关系的分类和估计。本文采用目前预测性能最好的 BiLSTM 神经网络来估计 SOH。

本文接下来的内容如下。第 2 节介绍数据源和提取的特征。第 3 节详细介绍 BiLSTM 神经网络和估算方法。第 4 节展示了基于所提方法的实验结果。最后，第 5 节给出本文的结论。

2. 数据分析

2.1 LIB 数据集

本实验使用了 NASA 数据集中的四块电池：B0005、B0006、B0007 和 B0018。图 1 显示了四块电池的容量随循环次数增加而下降的曲线。在图 1 中，电池容量在下降过程中出现了局部恢复现象。准确估算 SOH 时应能准确估算局部恢复的容量，而不是忽略这些容量的波动。

LIB 充放电实验过程如下。首先，以 1.5 A 的恒定电流对 LIB 充电，同时逐渐增加终端电压。当端电压达到 4.2 V 时，LIB 转换为 4.2 V 恒压充电，充电电流开始逐渐减小。当充电电流降至 20 mA 时，充电过程结束。在放电过程中，LIB 以 2 A 的恒定电流放电，终端电压逐渐降低。这一过程一直持续到 B0005、B0006、B0007 和 B0018 的端电压分别降至 2.7 V、2.5 V、2.2 V 和 2.5 V。

增量容量 (IC) 分析是分析电池衰退机理的一种重要方法。其核心是 LIB 的 IC 曲线。图 2 显示了 B0005 电池在不同老化状态下的 IC 曲线。循环次数越多，电池老化越严重。

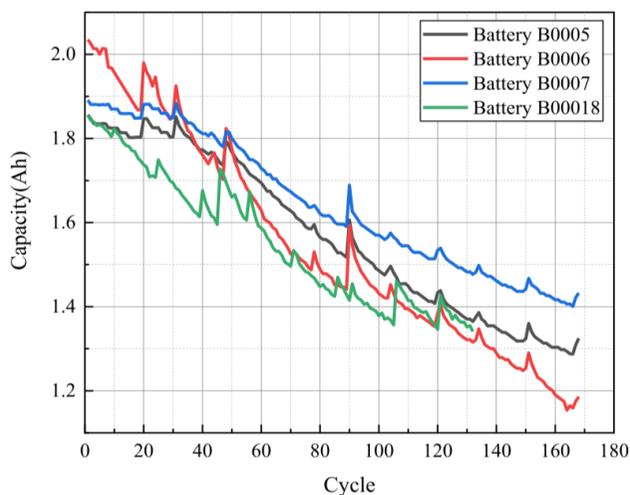


图 1. B0005、B0006、B0007 和 B0018 的容量下降曲线

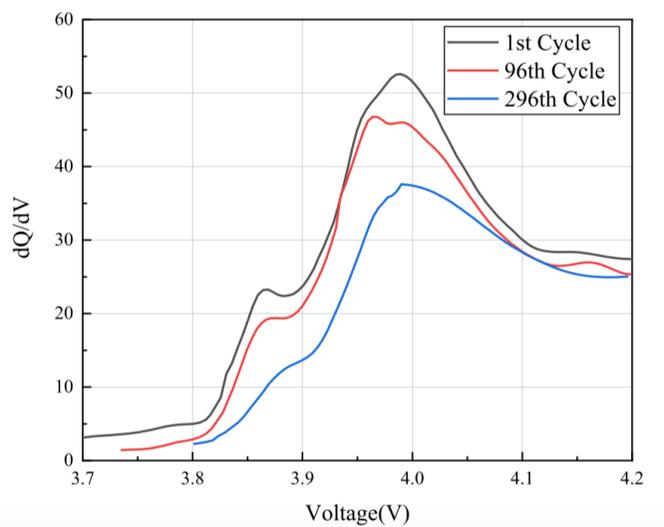


图 2. B0005 电池在不同老化状态下的 IC 曲线

2.2 特征提取

本文将 IC 曲线的最大峰值作为第一个特征。第二个特征是在电压为 3.9V 至 4.0V 时 IC 曲线所包含的面积。第三个特征是在电压为 3.9V 至 4.1V 时 IC 曲线所包含的面积。这三个特征被用作 BiLSTM 神经网络的输入。

皮尔逊相关系数被广泛用作衡量两个向量之间相关性的重要指标。皮尔逊相关率的计算方法如公式 (2) 所示。

$$P_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} \quad (2)$$

X 和 Y 代表两个向量， $\text{Cov}(X,Y)$ 代表两个向量的协方差， σ_X 和 σ_Y 分别代表两个向量的协方差。

表 1. 三种特征与四种电池容量之间的相关性

	B0005	B0006	B0007	B0018
Feature1	0.9842	0.9845	0.9889	0.9848
Feature2	0.9856	0.9889	0.9878	0.9881
Feature3	0.9875	0.9912	0.9921	0.9885

表 1 显示了所选三个特征与四种电池容量之间的相关率。可以看出，相关率高于 0.98，因此可以实现较好的估算。

3. 方法提出

3.1 BiLSTM 神经网络

考虑到 LIB 数据的时间性，本文选择 BiLSTM 神经网络。BiLSTM 神经网络由两个方向相反的 LSTM 神经网络组成，因此在处理序列相关问题时具有更好的性能。BiLSTM 神经网络可以通过两个方向相反的 LSTM 对数据进行正向和反向处理，因此可以同时利用过去和未来的信息。由于 BiLSTM 神经网络可以利用过去和未来的数据进行训练，因此其估计结果更为准确。其网络结构如图 3 所示。

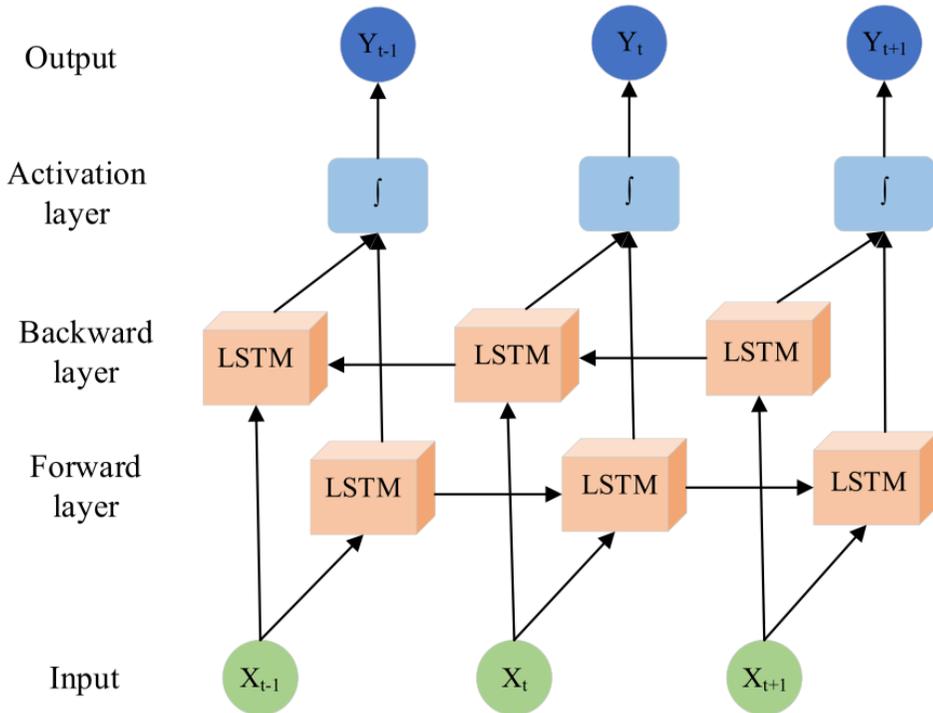


图 3. BiLSTM 结构图

LSTM 神经网络的结构如图 4 所示。LSTM 神经网络由输入 x_t 、单元状态 c_t 、临时单元状态 \tilde{c}_t 、隐藏状态 h_t 、遗忘门 f_t 、记忆门 i_t 和时间 t 的输出门 o_t 组成。遗忘、记忆和输出三个操作分别由遗忘门、记忆门和输出门控制。三个门的具体状态由上一时刻的隐藏状态和当前输入计算得出。

遗忘门 f_t 、记忆门 i_t 、临时细胞状态 \tilde{c}_t 、细胞状态 c_t 、输出门 o_t 和隐藏状态 h_t 的计算公式如下：

$$f_t = \sigma(\omega_f [h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(\omega_i [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\tilde{c}_t = \tanh(\omega_c [h_{t-1}, x_t] + b_c) \quad (5)$$

$$o_t = \sigma(\omega_o [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(c_t) \quad (7)$$

其中， h_{t-1} 代表最后时刻的隐藏状态层， σ 和 \tanh 代表非线性， b 代表阈值， ω 代表权重。

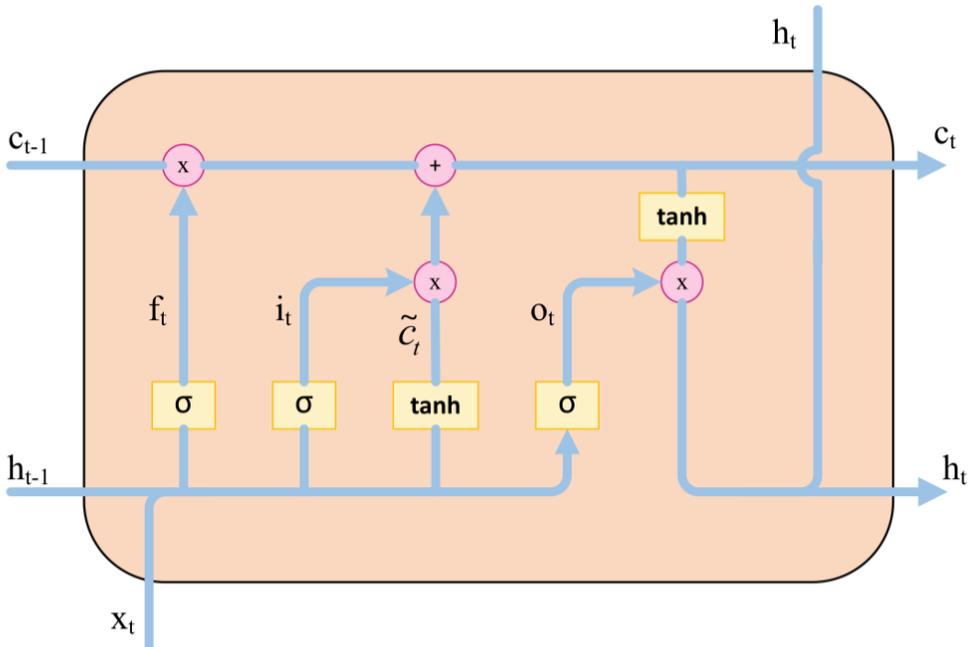


图 4. LSTM 神经网络示意图

3.2 Attention 机制

对于不同的特征，它们对最终输出的影响是不同的。因此，本文增

加了一种注意力机制来提高估计精度，它可以为不同的特征分配不同的权重。其原理如下：

$$s_i = \Phi(\omega^T h_i + b)$$

h_i 是 BiLSTM 的隐藏状态， T 代表矩阵的转置。 ω 和 b 分别代表权重矩阵和阈值向量， $\Phi(\cdot)$ 代表非线性激活函数， s_i 代表 h_i 的重要性。

在获得第 i 个特征向量的重要性后，可以使用如下函数对其进行归一化处理：

$$\alpha_i = \text{softmax}(s_i) = \frac{\exp(s_i)}{\sum_{i=1}^n \exp(s_i)} \quad (8)$$

在得到 α_i 我们可以给特征分配如下不同的权重：

$$O = H \otimes A$$

$H = \{h_1, h_2, \dots, h_n\}^T$, $A = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, O 表示注意力机制的最终输出特征， \otimes 表示两个向量之间对应元素的点积。

3.3 提出的方法

图 5 显示了拟议估算方法的整体框架。如图 5 所示，所提出的方法分为训练模块和估计模块。首先，我们将每个电池数据集分为训练集和测试集。训练集和测试集分别占每个电池数据集的 70% 和 30%。在训练模块中，首先从训练集中提取所提出的特征，然后利用所提取的特征训练具有注意力机制的 BiLSTM 神经网络，即 ABiLSTM 神经网络。在估计模块中，首先从测试集中提取建议的特征，然后使用在训练模块中训练的 ABiLSTM 神经网络来估计测试集中数据的 SOH。

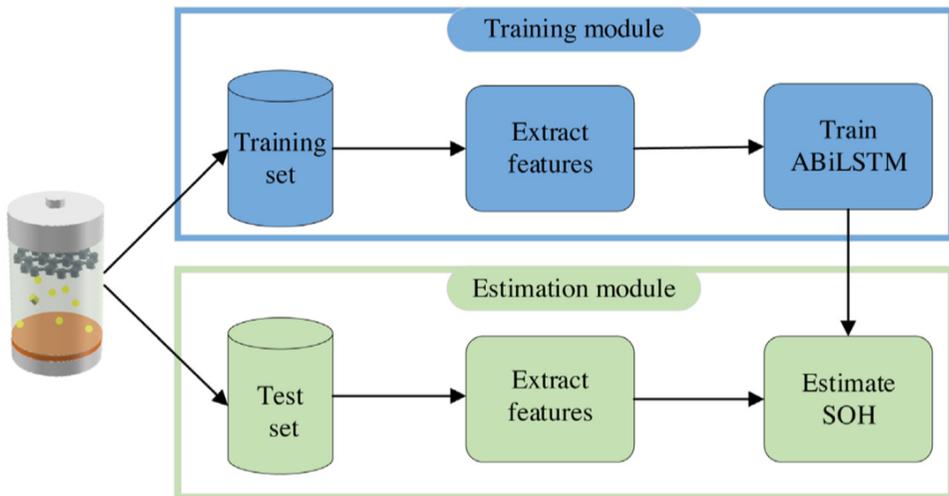


图 5. LSTM 神经网络示意图

4. 实验结果

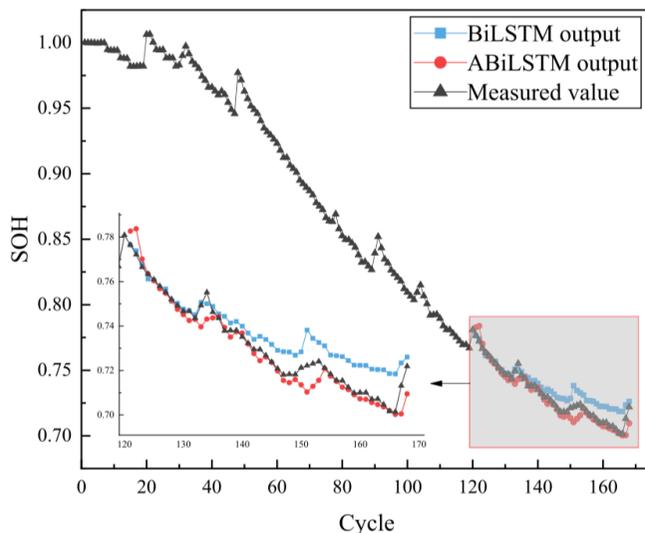


图 6. B0005 电池估计结果

本节将在 NASA 数据集中的四种电池（B0005、B0006、B0007 和 B0018）上测试所提出的 SOH 估算方法。B0005 的实验结果如图 6 所示。与 BiLSTM 相比，ABiLSTM 能够理解容量曲线的下降趋势。由于 LIB 内部存在化学副反应，在容量下降过程中会出现局部恢复和波动，从而影响容量估计。但 ABiLSTM 模型能很好地理解这种波动，并预测出与实际曲线相符的整体趋势。由于注意力机制的作用，所提出的特征与容量之间的映射关系得到了完善，从而使容量估计更加准确。特别是，BiLSTM 模型不能很好地预测局部趋势，因为它不能准确地学习容量的局部波动。这一现象证明，注意力机制能非常有效地捕捉局部波动，并显著提高估计性能。

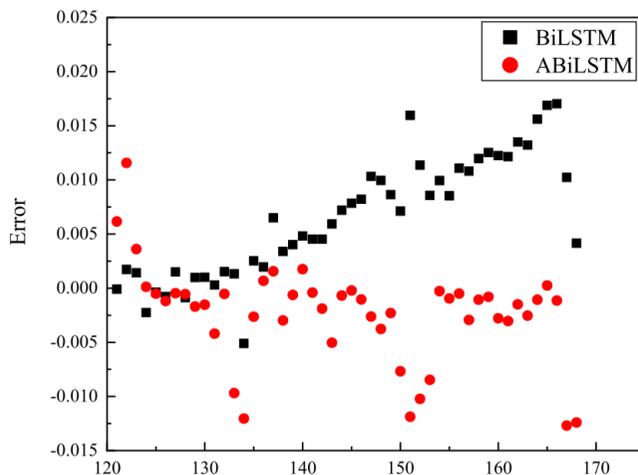


图 7. B0005 电池估计误差

接下来，我们在图 7 中给出了估计结果的误差。所使用的误差是估计值与测量值之间的差值。如图 7 所示，与 BiLSTM 相比，ABiLSTM 的整体误差较小，但在某些局部区域误差可能会有显著变化。这进一步说明了所提方法的准确性和有效性。

表 2. 四组电池的 RMSE 和 MAPE 结果

	B0005 (RMS E)	B0005 (MAP E)	B0006 (RMS E)	B0006 (MAP E)	B0007 (RMS E)	B0007 (MAP E)	B0018 (RMS E)	B0018 (MAP E)
BiLSTM	0.0085	0.69%	0.0156	1.34%	0.0089	0.72%	0.0105	0.99%
ABiLSTM	0.0051	0.34%	0.0093	1.02%	0.0055	0.36%	0.0078	0.68%

为了进一步说明 ABiLSTM 模型的优势，我们给出了四种电池的估计结果的 RMSE 和 MPAE，如表 2 所示。在不同的电池中，由于 ABiLSTM 模型的 RMSE 和 MPAE 相对较小，因此表现更好。与其他方法相比，本文提出的方法需要较少的训练数据就能达到相同的估计精度。这证明了所提出的 ABiLSTM 模型在 LIB 的 SOH 估算方面更加准确可靠。这一优势增加了将所提方法应用于实践的可能性。总体而言，与现有方法相比，本文提出的方法具有广阔的应用前景。

5. 结论

估算 SOH 是电池管理系统的一项重要功能。本文结合 BiLSTM 神经网络和注意力机制，提出了一种新型的 LIB SOH 估算方法。首先从 IC 曲线中提取三个特征。然后，根据特征对输出的重要性，利用注意力机制调整 BiLSTM 神经网络的输出。通过上述方法的具体应用，数值实验结果表明，该方法在 LIB 的 SOH 估算中误差较小。文献提出，BiLSTM 在估计 SOH 方面的表现优于 CNN、LSTM 和其他神经网络，因此我们提出的方法也优于其他神经网络。