

基于机器学习的电池容量变化预测报告

张昱

2024 年 10 月 22 日

摘要

本报告探讨了利用机器学习方法预测电池容量变化的研究。通过对现有数据集的分析，本文应用了一种特定的机器学习模型进行电池容量变化的预测，并对模型的性能和改进空间进行了讨论。基于 Pytorch 框架，我们构建并训练了一个长短时记忆网络 (LSTM)，实验结果表明该模型能够有效预测短期内的电池容量变化，并表现出一定的泛化能力。

1 引言

随着电池技术的不断进步，电池容量变化的精确预测变得愈发关键。准确的预测不仅能够优化电池的使用寿命，还能提升其性能表现。本研究旨在通过机器学习方法建立电池容量变化的模型，并对其进行预测，从而为电池管理提供有效的技术支持。

2 数据

2.1 数据来源

数据集来源于 [train] 文件，包含多个电池的容量随循环号的变化数据。我们的模型训练和测试主要使用了文件 M001、M012（训练）和 M013（测试）。同时，我们的代码中可以引入更多的训练数据来进一步提高模型的泛化能力。

2.2 数据特征

这一部分简单介绍数据集中包含的特征及对应的处理方式：

- 放电容量/Ah：待预测标签。
- 循环号：时间序列号。

- 工步号/工步状态：1-5 的整数，二者存在一一对应关系。
- 工步时间：特定工步号/工步状态下的运行时间。我们主要提取了在一个工步号下的总时间参与模型训练。

在具体模型训练中，我们用一个 1×6 的向量来表示特定循环号上的输入特征。前五个分量分别代表对应工步号/工步状态上的时间 (h)，第六个为经过预处理的待预测标签。针对外部参数，如温度等变量建模，可以在此处进行改动，即扩大向量维数以实现。

2.3 数据预处理

在模型训练之前，我们对数据进行了如下预处理：

- 数据清洗：去除缺失值和异常值，并通过插值进行修复。我们采用了四分位距 (IQR) 方法来识别异常值。该方法计算数据的第 25 百分位数 ($Q1$) 和第 75 百分位数 ($Q3$)，并将上下界设定为 $Q1 - 1.5 \times IQR$ 和 $Q3 + 1.5 \times IQR$ ，超出范围的数据被视为异常值。
- 数据归一化：使用 MinMaxScaler 方法将待预测标签缩放到 0 到 1 的区间。此过程能够减小不同特征值之间的尺度差异，提升模型的收敛速度和泛化能力。

3 方法

本研究选择了 [长短时记忆网络 (LSTM)] 作为主要的预测模型。选择该模型的原因包括其适用于时间序列预测的能力以及在类似任务中的良好表现。LSTM 是一种特殊的递归神经网络 (RNN)，通过引入门控机制来有效捕捉序列数据中的长期依赖关系，克服了 RNN 在处理长序列时梯度消失的问题。LSTM 的门控机制主要包括遗忘门（控制单元从细胞中丢弃信息）、输入门（将新的信息写入细胞状态）和输出门（输出当前单元的隐藏状态），这使得 LSTM 能够有效捕捉长序列中的依赖关系。

3.1 模型定义

我们的模型由两层组成。第一层为 LSTM 层，输入序列的时间步长为 6，隐藏层单元数为 50。第二层为全连接层，从隐藏状态中提取信息并生成最终的预测结果。

3.2 数据处理

我们采用滑动窗口方法。通过给定的窗口尺寸，将时间序列数据转化为适合模型输入的格式。每个输入样本包含 5 个时间步的特征，而对应的输出 y 则是下一个时间步的目标值。

3.3 损失函数和优化器

我们采用的损失函数为均方误差 (MSE)，优化器为 Adam，学习速率设定为 0.1。由于在训练过程中未出现过拟合现象，正则化参数 weight decay 设定为 0。

3.4 目标函数和损失计算

通过交替训练两个时间序列数据集，增强模型对不同数据集的适应性。这种方法能够使模型更好地捕捉时间序列的特征，提高泛化能力。我们分别计算每个数据集的损失并求和，得到总损失，从而避免因单一数据集特征差异导致的训练不均衡。我们共进行了 200 次循环的优化。

3.5 交叉验证

我们将时间序列分为训练集与验证集，比例为 80:20。通过计算各自的损失函数变化，判断训练过程中的过拟合情况。

4 结果

这一部分展示了我们训练的模型在验证集上的预测效果，主要分为两块：第一块为训练集中未参与训练的部分，第二块为验证集整个序列的预测。

4.1 结果展示

我们首先展示机器学习过程中的训练集和验证集的损失曲线（图1）。

接下来展示验证集的预测效果，结果显示在图2、图3和图4中，其中红线表示预测值，蓝线表示实际值。

4.2 评估指标

我们采用均方误差 (Mean Squared Error, MSE) 来评估模型的性能。均方误差是评估预测模型性能的常用损失函数或评估指标，衡量的是预测值与真实值之间差值的平方的平均值，用于反映模型预测结果的误差。它的定义为：

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

其中， y_i 是真实值， \hat{y}_i 是模型预测值， n 是样本数量。经过计算，对于 (M013) 我们得到的结果为 0.34%。

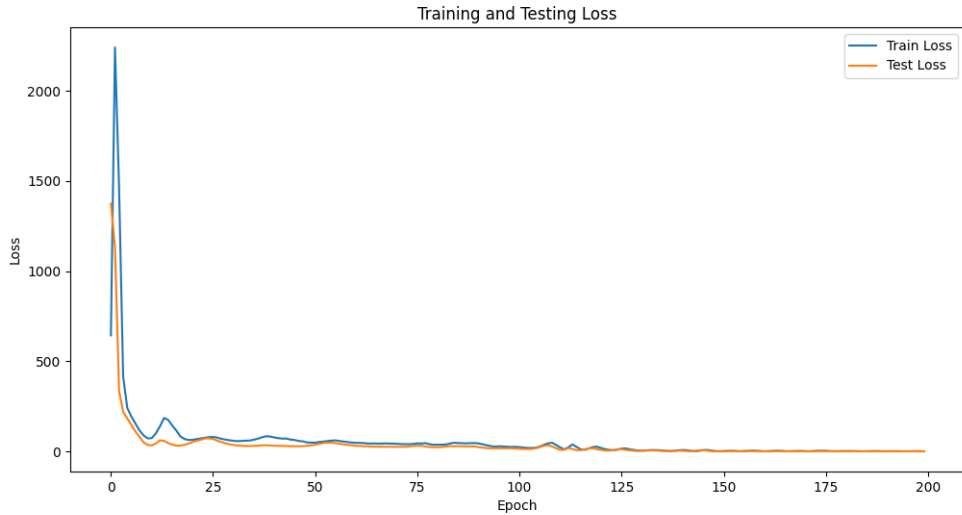


图 1: 损失曲线

5 讨论

5.1 结果分析

模型在电池容量变化的预测中表现良好，能够有效捕捉变化的短期趋势。然而，通过短时间序列预测之后的长期变化仍需进一步探索，主要的问题是误差的累计会导致后续逐步预测的失效。

5.2 局限性与改进方向

由于数据量有限，模型的泛化能力可能受到影响，建议未来增加数据集规模以提升预测准确性。未来的研究可以考虑引入更复杂的模型或增加更多特征以提高预测能力。同时，通过引入更广泛的超参数调节方法，可以进一步优化模型性能。对于长时间的预测，可能需要引入新的统计和机器学习方法。

6 结论

本研究展示了使用机器学习方法对电池容量变化进行预测的有效性。通过合理的数据预处理和模型选择，能够为电池管理提供有价值的参考。

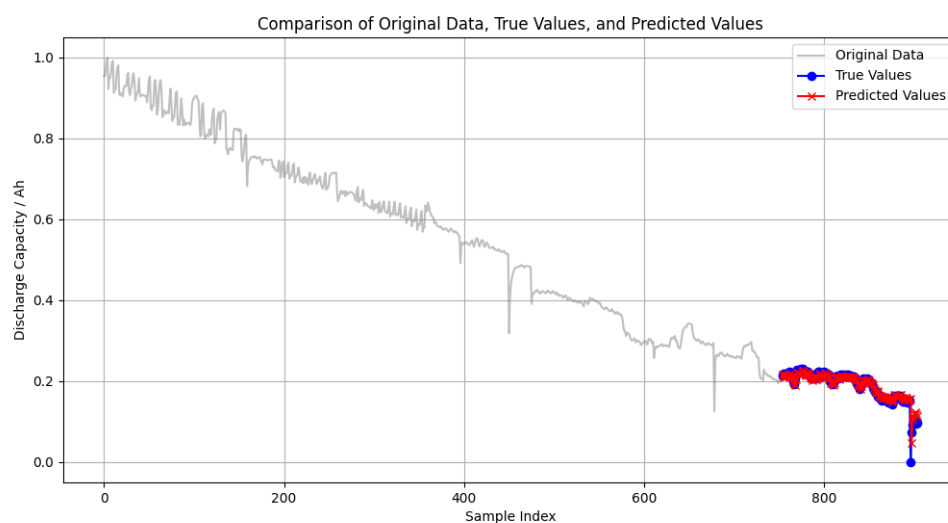


图 2: 数据 M001 验证集预测效果

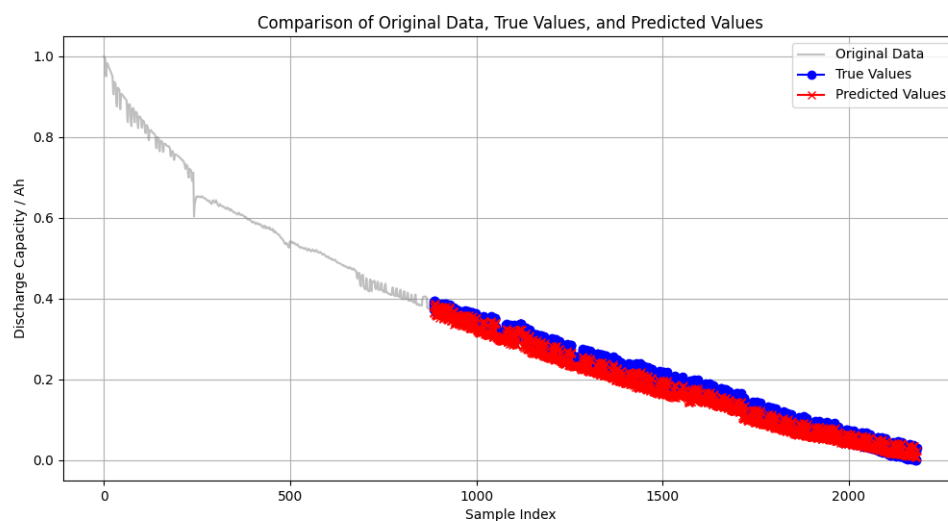


图 3: 数据 M012 验证集预测效果

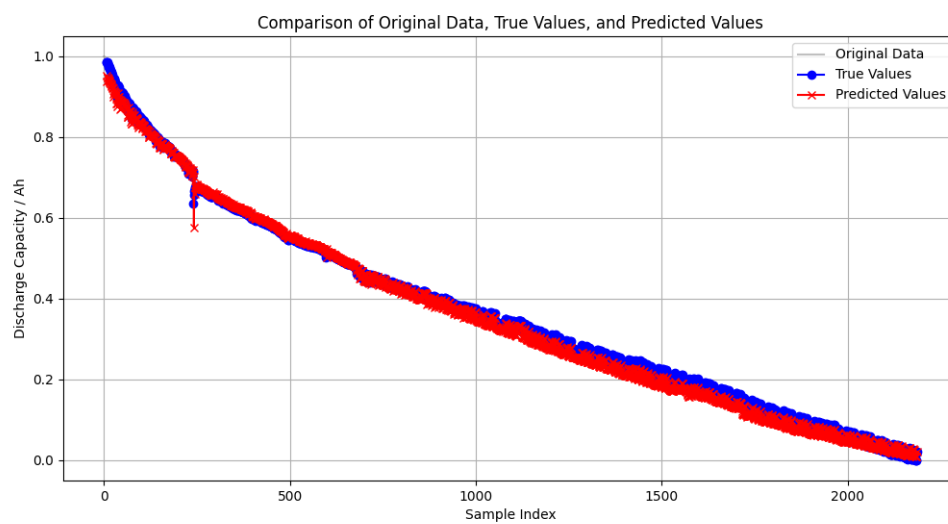


图 4: 数据 M013 预测效果