1.研究背景及意义

锂离子电池具备高能量密度、循环寿命长的优点，是目前储能技术研发重点之一。为遏制环境污染、土地沙漠化、能源紧缺等问题，国家大力发展新型能源，大力挖掘绿色能源的开发，大力支持能源汽车的生产，以尽快实现“碳达峰、碳中和”的目标，促进经济转型，促进社会可持续发展。“十三五”指出为迎合国际上环保理念的趋势，坚持可持续发展理念，需要进一步创新发展电动汽车，推进我国汽车产业转型升级。电池的SOH和RUL是锂离子电池的重要参数，也是锂离子电池的重要研究方向，能准确辨识电池的SOH和RUL对新能源汽车的安全性和操作性具有重要意义[6-8]。由于锂离子电池的组成结构，目前无法利用传感技术直接得到电池实时的SOH，因此开发电池的SOH估计和RUL预测的模型有助于推动新能源产业的发展。锂离子电池具有使用寿命长、能量密度高、低自放电率、无记忆效应以及宽温度范围等优点，被广泛用做各类机器的储能部件，如：新能源汽车、移动电话、笔记本电脑等设备[18-31]。

与镍氢等其它储能设备相比，锂离子电池具有容量大、循环充放电寿命长、无记忆效应、内阻小等优点，被广泛应用于航天、新能源汽车、海上轮船等大型设备[32-33]。

锂离子电池是一个时变的电化学系统[34]，工作过程中内部反应复杂多变，在动态系统下处理各种物质和能量，会产生固体电解质界面膜增长、锂离子沉降、集流体腐蚀、隔膜损伤和电解液氧化等副反应[35]。这些副反应阻碍了电池正负极的嵌锂和脱锂过程，导致电池的性能衰减，宏观上表现为容量减少和内阻增加，最终降低了电池的使用寿命。在长时间使用过程中，因外部环境和内部电极材料腐蚀、隔膜老化等原因，锂离子电池性能会逐渐退化和失效。电池的突然失效容易导致电子设备失去能量来源而停止工作，进而造成事故。锂离子电池的可靠性、安全性和寿命已成为制约其成功应用的至关重要问题，因此如何准确地预测锂离子电池的SOH及电池的剩余使用寿命是目前锂电池研究亟待解决的问题和难题[36-38]。

锂离子电池的老化失效会导致设备的停电和断电，进而导致事故的发生。因此及时、准确地预测电池的SOH及RUL有利于及时规划和管理电池，有利于及时更换失效电池，提前预防事故的发生。由于锂电池内部的化学反应以及外部环境的影响，电池的使用寿命随着时间的推移逐渐衰退老化甚至失效。锂电池的衰弱退化无疑增加了许多电子设备的维修成本，而且电池的突然失效易导致新能源汽车等大型设备失去工作，甚至导致重大事故的发生[39-42]。

综上所述，准确的电池寿命预测对于提高能源利用率、保护环境、降低成本、增强可靠性以及推动电池技术的进步都具有重要意义。它对于改善各个领域的设备和系统性能，以及推动可持续发展和资源管理都具有深远的社会和经济价值。因此，电池寿命预测研究的重要性在现代社会愈发显著。

2.研究内容

研究带有噪声的电池寿命预测相对于没有噪声的方法具有显著优势：这种研究更接近实际应用场景，因为在现实生活中，电池性能受到各种噪声源的干扰，如温度变化、充电和放电速率的波动等。因此，噪声预测更准确地反映了电池在复杂环境下的真实运行情况。带有噪声的电池寿命预测能够提前发现电池可能出现问题的征兆，因为噪声通常是由电池性能的微小变化引起的。这有助于及早采取维护措施，避免电池失效或损坏，从而延长其寿命。另外，通过噪声分析，可以更好地了解电池性能的波动性，为资源的更有效利用提供了机会。这有助于降低不必要的电池更换和维护成本，提高能源系统的可维护性和性能。

我们使用的电池数据集大多是在实验室条件下测量的，包括温度、湿度、充放电电压电流、充放电使用的协议等。但在现实情况下的电池循环信息往往掺杂一些因环境因素、人为因素造成的有噪声的电池数据集。因此，本文的研究将聚焦于当电池数据集带有噪声时，如何使用早起电池循环数据对电池寿命进行预测。

关于使用TLS\_EM方法对有噪声的电池寿命进行预测：TLS和OLS数据集样本的噪声假设都有一个前提就是噪声是服从一个零均值，相同标准差的高斯白噪声，然而现实情况中的电池样本数据噪声可能来自不同噪声分布，即不同标准差。针对这一现象，我们将EM算法思想和TLS/OLS结合，将现实数据集样本中噪声分布的标准差作为隐变量，利用EM迭代不断更新TLS/OLS算法的模型系数和噪声的标准差，即隐变量，收敛得到噪声标准差和模型系数进而对电池寿命进行预测。

使用完全自适应模态分解算法对数据集降噪预测：考虑到在实验条件下无法完全模拟现实情况下的数据噪声，我们在已知数据集上加上噪声之后作为要模拟的现实情况。使用完全自适应模态分解算法对带噪声的数据集进行分解之后得到分量，使用随机森林等算法对得到的分量进行一系列处理得到与电池寿命相关性更高的特征，将其作为XGB预测模型的输入对电池寿命进行预测。除了完全自适应模态分解算法去噪之外，本实验还使用了小波分析、季节性分解算法等去噪算法，对比结果显示使用完全自适应模态分解算法能够更有效地降噪得到更准确的结果。

3.研究方法

3.1传统的TLS和OLS算法

TLS算法：

, Feature Matrix , Label Vector , Noise and , Model Coefficients

Total Least Square:

Two steps:

Transformation

,

KKT Condition: (1), (2), (3)

From (1) and (2): , , (4)

From (3) and (4): , , (5)

From (4) and (5): (6)

Put (5) and (6) into the objective function:

->

Solution

-> -> ,

Ignore the constraint , directly solve

* The scale of has no impact to the objective value, we set
* , →, ->
* KKT condition: , is an eigenvalue of
* , is the smallest eigenvalue of
* is the eigenvector of corresponding to the smallest eigenvalue

OLS算法：

最小二乘估计：

但数据都带有一定噪声，假设噪声满足零均值的高斯分布

因此，

则,

使用MLE求解W：

首先是最大似然函数：

再求W：

3.2EM算法思想

给定数据集，假设样本间相互独立，我们想要拟合模型到数据的参数。根据分布得到似然函数：

我们可以最大化似然函数以求解，但当存在隐变量时，似然函数表示为：



对于某个样本i，我们使用表示样本i隐含变量z的分布概率，则

引入，并利用Jensen不等式：



似然函数变换为：



我们得到了的形式，

不等式取等号：当为常数时，不等式可以取等号，对等式变换并两边求和

,,

E-Step：求条件期望，本质是求出来自哪种类别的概率，为简化将log中的去掉，

关于：

两边同时乘以且求和即可得到期望：

可以发现:

M-Step: 求，更新参数

3.3降噪算法

CEEMDAN算法：

假设为经过EMD分解之后得到的第个本证模态分量，CEEMDAN分解得到的第个本征模态分量为，为满足标准正态分布的高斯噪声信号，是加入噪声的次数，为白噪声的标准差，为待分解信号，CEEMDAN分解步骤如下：

1)将高斯白噪声加入到待分解信号得到新信号，其中…对新信号进行EMD分解，得到第一阶本征模态分量

2)对产生的N个模态分量进行总体平均就得到CEEMDAN分解的第1个本征模态分量:

3)计算去除第─个模态分量后的残差:

4)在中加入正负成对高斯白噪声得到新信号，以新信号为载体进行EMD分解，得到第一阶模态分量，由此可以得到CEEMDAN分解的第2个本征模态分量:

5)计算去除第二个模态分量后的残差:

6）重复上述步骤，直到获得的残差信号为单调函数，不能继续分解，算法结束。此时得到的本征模态分量数显为K，则原始信号被分解为:

3.4预测算法

BA算法：

蝙蝠是通过其发出超声波的频率、速度来确定物体的位置信息。所以，为了便于模拟蝙蝠的回声定位行为。Yang给出了两个理想化的规则：

搜索规则：蝙蝠随机飞行，同时以固定的频率、可变的波长和音量来搜索猎物；

参数变化规则：蝙蝠会根据自身与猎物的距离来自动调整脉冲波长和脉冲发射率，并限定音量在指定范围内依照给定方式由大到小而发生变化。

在以上的假设的基础上，BA算法的步骤如下所示：  
步骤一：初始化相关参数，包括蝙蝠的种群数量n nn、蝙蝠种群的位置和速度。  
步骤二：根据目标函数和初始位置计算适应度值，从而得出初始解；  
步骤三：对于每只蝙蝠k kk，根据下列公式更新脉冲频率、速度和位置信息，从而产生后代个体：

其中，表示第k只蝙蝠的脉冲频率信息，​he和分别表示脉冲频率的最大值和最小值，是均匀分布的随机数，,，表示第t个迭代过程中蝙蝠k所在的位置和速度，表示当前全局的最优解。  
步骤四：生成随机数rand1，如若rand1 > ​，则在当前最优解附近进行邻域位置的搜索，搜索方式如下公式所确定：

其中， ​表示第t代蝙蝠k的脉冲发射率，​表示随机扰动得到的新解，​即，表示第t迭代中所有蝙蝠群体音量的平均值，表示均匀分布的随机数。  
步骤五：生成随机数rand2，如若rand2 < ​并且​的适应度优于​，则接受​并且更新 ​​和​，更新的公式如下所示：

其中，, ,α , γ均为常量；  
步骤六：计算出当前所有蝙蝠的适应度并找出当前最优解；  
步骤七：重复步骤三至步骤六，直到达到循环条件，退出算法。

XGB算法：

XGBoost方法是一种可伸缩的端到端树提升系统，旨在提高效率、灵活性和可移植性。实现了梯度增强框架下的机器学习算法。与多元线性回归相比，XGBoost具有处理非线性关系的优势。Treef(X)被定义为：

其中t表示一棵树，q表示每棵树的结构，该结构将一个示例映射到相应的叶索引。每一棵树对应着一个独立的树结构Q和叶重o(一棵树的输出)。

目标函数的定义是：

其中L是一个可微凸损失函数，它度量

预测$和目标y。第二项惩罚模型的复杂性，这有助于平滑最终学习的权重，以避免过度拟合。

其中OJ是JT叶节点的重量。Y和1是惩罚tc的系数。

利用二阶Taylor公式，可以给出目标函数如下：

其中；是样本的输入和，

除去常数后，步骤t处的目标函数变为

其中，,是叶j的实例集。

对于一个固定结构，叶j的最优重量o；可以通过以下方法计算：,

,

是标记树结构和测量树结构质量的功能 的值越小 ，越好

4.预期目标

1.使用TLS\_EM和OLS\_EM这两种方法相较于传统的TLS和LS有更高的预测精度，且在噪声增强的情况下比传统算法优势明显。

2.使用完全自适应模态分解算法对电池数据集进行去噪效果明显，比其他降噪方法有更高的预测精度，且使用了模态分解算法之后比原始的算法有优势。

3.使用特征工程的预测结果比未使用结果更好。

4.难点及存在问题

1．给带不同噪声的样本加权的权值确定。不同样本所带噪声服从不同的噪声标准差，如何确定加权时的权值以达到更好地消除噪声的目的，是本研究的难点

2.降噪方法的选择。降噪算法种类数不胜数，如何选择合适算法是本研究的难点。电池预测中选择合适的降噪方法是一个难点，主要因为电池系统的复杂性和不确定性。首先，电池的性能受多种因素影响，包但不仅限于充放电速率、温度、老化程度等。这导致电池系统产生的数据具有高度非线性和时变性质，难以简单建模。其次，电池预测受到外部环境和使用条件的影响，如温度变化、负载波动等，这些因素引入了随机性和不确定性。传统的降噪方法往往难以捕捉到这些复杂的非确定性变化，从而影响预测的准确性。此外，电池系统中可能存在噪声干扰，如传感器误差、测量噪声等，这些噪声源对预测模型的训练和性能评估都构成挑战。选择适当的降噪方法需要兼顾对系统复杂性、外部环境变化和噪声的适应性，以确保预测模型在真实工作条件下具有稳健性和准确性。因此，电池预测中的降噪问题需要综合考虑多方面因素，这增加了该任务的难度。

3.模型方法的健壮性。本研究中出现的电池预测健康因子是电池地100次循环和第150次循环的电压-放电曲线面积差，目前公开的数据集中只有一个数据集包含本研究所需要的健康因子，其他公开的电池数据集中，马里兰大学的数据集和美国航天局的数据集并没有包含所需信息，故模型的健壮性有待扩展。

5.工作的初步计划

|  |  |
| --- | --- |
| **时间阶段** | **具体工作** |
| 2024.01~2023.02 | 查阅资料，阅读文献，复现代码。 |
| 2024.03~2023.05 | 实现所提出的方法，并选取部分数据集进行实验，验证方法的性能。 |
| 2023.06~2023.10 | 根据实验结果调整方法的细节，并进行大规模数据集的实验。 |
| 2024.10~2024.11 | 与现有方法进行对比，进一步优化算法。 |
| 2024.12~2025 | 撰写毕业论文，准备毕业答辩。 |