基于机器学习的电池快速充电协议闭环优化（Closed-loop optimization of fast-charging protocols for batteries with machine learning）

典型的目标是最大化电池寿命，然而花费时间观察一个实验电池寿命需要几年时间，且大的参数空间和高的抽样可变性都需要大量的实验所以挑战点在于减少实验所需的数量和持续时间

本文开发并演示了一种机器学习方法，以有效优化参数空间，指定六步、十分钟快速充电协议的电流和电压分布，以最大化电池循环寿命，这可以缓解电动汽车用户的续航焦虑。

为了降低优化成本，我们结合了两个关键元素:一个是早期预测模型，它通过使用前几个周期的数据预测最终循环寿命，减少了每次实验的时间;另一个是贝叶斯优化算法，它通过平衡探索和开发，有效地探测充电协议的参数空间，减少了实验次数。

使用该方法，我们在16天内从224个候选充电协议中快速识别出高循环寿命充电协议(相比之下，在没有早期预测的情况下使用穷尽搜索超过500天)，随后验证了我们优化方法的准确性和效率

我们的闭环方法自动整合了来自过去实验的反馈，为未来的决策提供信息，并可以推广到电池设计中的其他应用，更广泛地说，可以推广到其他涉及时间密集实验和多维设计空间的科学领域

优化实验设计(OED)方法被广泛应用于降低实验优化成本。这些方法通常涉及一个闭环管道，完成的实验的反馈通知后续的实验决策，平衡探索(即测试高不确定性的实验参数空间的区域)和开发(即基于完成的实验结果测试有希望的区域)的竞争需求。

虽然闭环方法被设计为最小化多维参数空间优化所需的实验数量，但每次实验的时间(和成本)可能仍然很高，就像锂离子电池一样（所以OED方法应该考虑实验数量和每个实验的成本）

锂离子电池的低保真信号的设计和使用是具有挑战性的和未探索的。

开发了一个具有早期结果预测的闭环优化(CLO)系统，可在大参数空间中进行高效优化，且具有昂贵的实验费用和高抽样变异性

我们优化了一个由224个独特的6步、10分钟快速充电协议(即充电过程中如何控制电流和电压)组成的参数空间，以找到高循环寿命(定义为电池容量下降到其标称值的80%)的充电协议。

我们使用机器学习来根据电池到达寿命结束之前的早期周期数据预测实验结果，从而减少每次实验的时间5。其次，在选择下一轮实验时，我们使用贝叶斯优化(BO)算法来平衡勘探和开发的权衡，从而减少实验数量

系统由三个部分组成:并行电池循环，循环寿命早期预测器和BO算法。

我们的目标是找到一种充电方案，在固定充电时间(10分钟)和充电状态(SOC)范围(0 - 80%)下最大化预期电池循环寿命。我们的224个六步极限快速充电协议的设计空间如图2a所示。

多步充电协议，即在一次充电中应用一系列不同的恒流步骤，被认为比单步充电更有利于在快速充电过程中最大化循环寿命4,8，尽管最佳组合尚不清楚。

这些电池被设计成在17分钟内快速充电(速率测试数据见扩展数据图2)。随着充电时间的加快，循环寿命急剧下降4,5，这促使了这种优化。由于LFP正极通常被认为是稳定的4,5，我们选择这种电池化学方法来隔离极快充电对石墨的影响，这在锂离子电池中是普遍使用的。

最初，所有协议的估计周期寿命是相等的。经过两轮后，出现了参数空间的整体结构(即周期寿命对充电协议参数CC1、CC2和CC3的依赖关系)，并识别出一个高周期寿命协议的突出区域。从第2轮到第4轮，CLO在这一高性能区域的置信度进一步提高，但总体周期寿命估计并没有发生实质性变化(扩展数据图3)。

CLO反复测试几种具有高估计周期寿命的协议，以减少由于制造可变性和早期结果预测引入的误差而产生的不确定性。

电池文献中提出的大多数快速充电协议都表明，电流步长随SOC的单调减小是避免快速充电过程中普遍接受的降解模式——石墨上的锂镀层的最佳方法

但这一意想不到的结果强调了优化快速充电的数据驱动方法的需求。

我们通过这项验证研究来(1)确认CLO能够基于循环寿命对协议进行正确的排名，(2)比较CLO推荐的协议与受电池文献启发的协议的循环寿命，(3)比较CLO与实验设计的基线消融方法的性能。

关键的挑战是减少所需实验的数量和持续时间。在这里，我们开发并演示了一种机器学习方法，以有效优化参数空间，指定六步、十分钟快速充电协议的电流和电压分布，以最大化电池循环寿命，这可以缓解电动汽车用户的续航焦虑。为了降低优化成本，我们结合了两个关键元素:一个是早期预测模型，它通过使用前几个周期的数据预测最终循环寿命，减少了每次实验的时间;另一个是贝叶斯优化算法，它通过平衡探索和开发，有效地探测充电协议的参数空间，减少了实验次数。使用该方法，我们在16天内从224个候选充电协议中快速识别出高循环寿命充电协议(相比之下，在没有早期预测的情况下使用穷尽搜索超过500天)，随后验证了我们优化方法的准确性和效率。我们的闭环方法自动整合了来自过去实验的反馈，为未来的决策提供信息，并可以推广到电池设计中的其他应用，更广泛地说，可以推广到其他涉及时间密集实验和多维设计空间的科学领域。

优化实验设计(OED)方法被广泛应用于降低实验优化成本。这些方法通常涉及一个闭环管道，完成的实验的反馈通知后续的实验决策，平衡探索(即测试高不确定性的实验参数空间的区域)和开发(即基于完成的实验结果测试有希望的区域)的竞争需求。然而，虽然闭环方法被设计为最小化多维参数空间优化所需的实验数量，但每次实验的时间(和成本)可能仍然很高，就像锂离子电池的情况一样。因此，OED方法应该同时考虑实验数量和每个实验的成本。多保真度优化方法已经被开发用来从廉价的噪声信号和昂贵的精确信号中学习。例如，在机器学习算法的超参数优化中，几个用于预测算法配置的最终性能的低保真信号(例如，外推学习曲线，在完整训练数据集的子集上的快速测试)与更完整的配置评估一起使用。对于锂电池，经典方法已经被应用，但低保真信号的设计和使用是具有挑战性的和未探索的。这些先前考虑过的方法并没有发现和利用参数空间中存在的模式进行有效优化，也没有解决每次实验的时间问题。

在这项工作中，我们开发了一个具有早期结果预测的闭环优化(CLO)系统，可在大参数空间中进行高效优化，且具有昂贵的实验费用和高抽样变异性。我们利用该系统对锂离子电池的快速充电协议进行了实验优化;将充电时间缩短到接近汽油加油时间是减少电动汽车续航里程焦虑的关键，但这往往以牺牲电池寿命为代价。具体来说，我们优化了一个由224个独特的6步10分钟快速充电协议(即充电过程中如何控制电流和电压)组成的参数空间，以找到循环寿命高的充电协议(定义为电池容量下降到其标称值的80%)。系统使用两个关键元素来减少优化成本：首先，我们使用机器学习来根据电池到达寿命结束之前的早期周期数据预测实验结果，从而减少每次实验的时间。其次，在选择下一轮实验时，我们使用贝叶斯优化(BO)算法来平衡勘探和开发的权衡，从而减少实验数量。发现CLO根据生命周期对这些协议进行了准确的排名(肯德尔排名相关系数，0.83)和有效的排名(比使用随机搜索而不进行早期预测的基线“野蛮力”方法的时间少15倍)。由CLO识别的具有早期预测的最佳充电协议优于现有的快速充电协议，该协议旨在避免电镀锂(一种常见的快速充电退化模式)，这是传统电池智慧建议的方法。这项工作强调了将CLO与廉价的早期结果预测器结合起来以加速科学发现的效用。

该系统由三个部分组成:并行电池循环，循环寿命早期预测器和BO算法。在每个连续的回合中，我们迭代这三个组件。第一个组件是一个多通道电池循环器;本工作使用的循环器同时测试48节电池。在开始CLO之前，从完整的224个独特的多步骤协议(方法)中随机选择第一轮48节电池的充电协议(不更换)。每个电池重复充放电100次(约4天;平均预测周期寿命905个周期)，超过此周期实验终止。

然后，这些循环数据被输入到早期结果预测器中，根据前100次循环的数据，预测电池的最终循环寿命。早期预测器是一个线性模型，根据前100个周期的充电数据提取的特征进行弹性净回归训练27(补充表1)，类似于Severson等人提出的5。预测特征包括电压曲线差异和放电容量衰减趋势的转换。为了训练早期预测器，我们需要一个电池循环到故障的训练数据集。在这里，我们使用了一个已有的数据集，其中有41个电池循环到故障(交叉验证均方根误差，80.4循环;参见方法和补充讨论1)。尽管获取该数据集本身需要为一个小的电池训练集运行全循环实验(我们试图抵消的成本)，但如果可以使用预先训练的预测器或先前收集的数据集，则可以避免这种一次性成本。如果不可用，我们在收集这个数据集时支付前期费用;这个数据集也可以用于BO算法的热启动。收集到的数据集的大小应该最好地权衡获取数据集的前期成本，以训练一个精确的模型，并预期减少CLO的实验需求。

最后，从早期周期数据中预测的周期寿命被输入BO算法(方法和补充讨论2)，该算法推荐下一轮的48个收费协议，以最佳地平衡勘探和开发的权衡。该算法(方法和补充讨论2)建立在Hoffman等人10和Grover等人11之前的工作之上。该算法对每个协议的平均周期寿命和不确定性边界保持估计;对于所有协议，这些估计值最初是相等的，并随着收集到的额外数据而进行细化。算法使用更新使用的是早期预测寿命而非实际寿命。周期寿命的平均值和不确定估计是通过高斯过程(方法)获得的，该方法具有平滑效果，并允许用相关协议的预测更新未测试协议的周期寿命估计。闭环过程不断重复，直到优化预算耗尽，在我们的例子中，测试了192个电池(每个电池100循环)。

我们的目标是找到一种充电方案，在固定充电时间(10分钟)和充电状态(SOC)范围(0 - 80%)下最大化预期电池循环寿命。我们的224个六步极限快速充电协议的设计空间如图2a所示。多步充电协议，即在一次充电中应用一系列不同的恒流步骤，被认为比单步充电更有利于在快速充电过程中最大化循环寿命4,8，尽管最佳组合尚不清楚。如图2b所示，每个协议由三个独立的参数(CC1、CC2和CC3)指定;每个参数都是在固定的SOC范围内(分别为0-20%，20-40%和40-60%)应用的电流，第四个参数CC4取决于CC1、CC2、CC3和充电时间。在给定当前值(Methods)的约束下，总共允许224个计费协议。我们在对流环境室(30°C环境温度)中测试商用磷酸铁锂(LFP)/石墨圆柱形电池(A123系统)。施加的最大电压为3.6 V。这些电池被设计成在17分钟内快速充电(速率测试数据见扩展数据图2)。随着充电时间的加快，循环寿命急剧下降4,5，这促使了这种优化。由于LFP正极通常被认为是稳定的4,5，我们选择这种电池化学方法来隔离极快充电对石墨的影响，这在锂离子电池中是普遍使用的。

我们总共连续运行了4个CLO回合，共包含185个电池(不包括7个电池;见的方法)。使用早期预测，每一轮CLO需要4天来完成100个周期，导致总测试时间为16天——从每个充电协议测试到故障3次所需的560天大幅减少。图3展示了预测和选择的协议(图3a)，以及随着优化的进展，在参数空间中周期寿命估计的演化(图3a)。最初，所有协议的估计周期寿命是相等的。经过两轮后，出现了参数空间的整体结构(即周期寿命对充电协议参数CC1、CC2和CC3的依赖关系)，并识别出一个高周期寿命协议的突出区域。从第2轮到第4轮，CLO在这一高性能区域的置信度进一步提高，但总体周期寿命估计并没有发生实质性变化(扩展数据图3)。通过学习和利用参数空间的结构，我们避免评估估计周期寿命较低的充电协议，将更多的资源集中在高性能区域(扩展数据图3 - 5)。具体来说，224个协议中有117个从未经过测试(图3c);我们用67%的电池测试21%的协议(平均每个协议0.83块电池)。CLO反复测试几种具有高估计周期寿命的协议，以减少由于制造可变性和早期结果预测引入的误差而产生的不确定性。不确定性表示为后验预测分布在周期寿命中的预测区间(扩展数据图3g, 4,5)。据我们所知，这项工作展示了作为充电条件函数的循环寿命的最大已知地图(扩展数据图5)。该数据集可用于验证基于物理的电池退化模型。电池文献中提出的大多数快速充电协议都表明，电流步长随SOC的单调减小是避免快速充电过程中普遍接受的降解模式——石墨上的锂镀层的最佳方法。相比之下，CLO确定为最佳的协议(例如，图3d)一般类似于单步恒流充电(即CC1≈CC2≈CC3≈CC4)。具体来说，在估计周期寿命最高的75个协议中，只有10个协议是单调递减的(即CCi≥CCi+1对于所有i)， 2个协议是严格递减的(即CCi > CCi+1)。我们推测，与最小化镀锂倾向相反，最小化发热引起的寄生反应可能是这些电池的操作优化策略(补充讨论3)。虽然新场景的最佳方案将取决于所选的充电时间、SOC窗口、温度控制条件和电池化学成分，但这一意想不到的结果强调了优化快速充电的数据驱动方法的需求。

我们在9个极端快速充电协议的子集上通过早期预测验证了CLO的性能。对于每一种方案，我们都将5个电池循环到失效，并使用最终循环寿命的样本平均值作为真实寿命的估计。我们通过这项验证研究来(1)确认CLO能够基于循环寿命对协议进行正确的排名，(2)比较CLO推荐的协议与受电池文献启发的协议的循环寿命，(3)比较CLO与实验设计的基线消融方法的性能。验证中使用的充电协议，其中一些是受到现有电池快速充电文献的启发(参见方法)，跨越估计循环寿命的范围(扩展数据图6和扩展数据表1)。我们调整这些文献协议的电压限制和充电时间，以匹配我们的协议，同时保持类似的电流比作为SOC的函数。在这些验证实验中使用的文献协议通常是针对具有高压正极化学结构的电池设计的，而快速充电优化策略通常侧重于石墨负极，我们根据这9个协议根据5个最终循环寿命验证CLO估计循环寿命。

验证结果如图4所示。放电容量衰减曲线(图4a)显示出快速充电时典型的非线性衰减5,7。如果我们将我们的早期预测模型应用于验证实验中的电池，这些早期预测(每个协议的平均值)与clo估计的平均循环寿命匹配得很好(皮尔逊相关系数r = 0.93;图4 b)。这一结果尤其验证了CLO的BO组件的性能，因为clo估计的周期寿命是从早期预测中推断出来的。然而，我们的早期预测模型表现出一些偏差(图4c)，可能是由于不同电池存储时间对日历老化的影响28(补充表2和补充讨论4)。尽管我们的预测模型存在这种偏差，但我们通常很好地捕获了排名(肯德尔排名相关系数，0.83;图4d和扩展数据图7)。同时，我们注意到排名靠前的协议的最终周期寿命是相似的。此外，由CLO确定的最佳协议优于先前发布的快速充电协议(平均895次vs . 728次;图6和表1)。这一结果表明，我们方法的效率并不以牺牲准确性为代价。

与基线优化方法相比，我们的方法大大减少了所需的优化时间(图4e)。例如，一个不使用早期结果预测，而只是随机选择协议进行测试的程序，在大约7700电池小时的测试后，开始在具有竞争力的性能水平上饱和。为了达到类似的性能水平，同时具有早期结果预测和BO算法的CLO只需要500电池小时的测试。对于这个小规模的验证实验，我们观察到CLO的早期预测成分大大减少了每次实验的时间。在这里，随机选择等同于纯探索策略，可以在较小的实验预算下实现与基于bo的方法类似的性能。在后期阶段，基于bo的方法最终优于随机选择，这种方法利用了跨协议的结构，并专注于减少参数空间有希望区域的不确定性。尽管这些结果是针对本验证研究的，但我们观察到，当可用的电池较少或并行实验(相对于参数空间的大小)较少时，模拟得到了类似或更大的增益(扩展数据图8)。在资源最少的情况下，BO相对于随机选择的相对增益最大。

最后，我们将我们的早期预测器与最先进的多保真优化算法中提出的其他低保真预测器进行比较，并发现我们的方法优于这些算法(补充讨论2和补充表4)。这些先前工作中的通用早期预测模型将参数函数的组合与容量衰减曲线相匹配，而我们的模型使用了在每个周期记录的附加特征(例如电压)。这一结果强调了为目标应用设计预测模型在多保真度优化中的价值。

综上所述，我们已经成功地加速了锂离子电池极快充电的优化和早期结果预测。这种方法可以扩展到其他快速充电设计领域，如脉冲充电、28充电和恒功率充电，以及其他目标，如慢充电和日历老化。

此外，这项工作为电池优化开辟了新的应用领域，如电池管理系统模型的编队、自适应循环和参数估计。此外，如果存在一个合适的早期结果预测器，该方法也可以用于优化电池开发的其他方面，如电极材料和电解质化学。除了电池，我们的CLO方法结合了黑箱优化和早期结果预测，可以扩展到有效优化其他物理1,2,18和计算22,32多维参数空间，涉及时间密集的实验，说明了数据驱动方法在加快科学发现速度方面的力量。

方法：

Experimental：

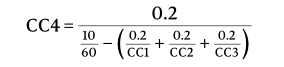
电池按照不同的充电方案循环，但放电相同。电池使用224个候选六步10分钟充电协议中的一种进行充电，充电范围从0%到80% SOC，详情如下。5秒钟休息后，所有细胞从80%充电至100% SOC, 1C CC-CV充电步骤至3.6 V，电流截止为C/20。再休息5秒后，所有细胞随后放电，在4C到2.0 V的CC-CV放电，电流截止为C/20。电池又休息了5秒钟，然后才开始后续的充电步骤。下限和上限截止电压分别为2.0 V和3.6 V，符合制造商的建议。在这项工作中，循环寿命被定义为在放电容量低于标称容量80%之前的循环次数。所有电池都在圆柱形固定装置中进行测试，固定装置上有48通道Arbin实验室电池测试电池循环器，放置在环境室(Amerex Instruments)中，温度为30°C。循环器校准在实验状态之前进行了验证。在闭环实验中，有4个实验都没有达到100个循环，原因有在实验开始时接触问题，也有部分实验结束时接触问题。这些实验分别在第一轮(oed\_0)中的通道17和27和第二轮(oed\_1)中的通道4和5上运行。此外，在每一轮中，由于处理错误(扩展数据图4)，每一轮中应该被选择的一个协议(即上限为48的协议)没有被选择，而被上限为第49的协议所取代，但该错误预计不会产生很大的影响。在数据存储库的注释中记录了其他实验问题。

Charging protocol and parameter space design：

电池按照224种不同的四步充电方案中的一种进行充电。前四个步骤中的每一个都是应用于20% SOC范围的单个恒流步骤;因此，224种充电协议代表0%至80% SOC范围内的不同电流步骤组合。

定义充电时间（0-80%的SOC）为：

在这里考虑的所有协议中，我们将t0-80%限制为10分钟。

现在我们将CC4写成前三个充电步骤的函数，如下:

因此，每个协议都可以由CC1、CC2和CC3唯一定义。

每个独立的参数可以采用以下离散值之一:3.6C, 4.0C, 4.4C, 4.8C, 5.2C和5.6C。此外，CC1可以接受6.0C、7.0C和8.0C的值，CC2可以接受6.0C和7.0C的值。CC4不允许超过4.81 1c。各参数的最大允许电流随SOC的增加而减小，以避免达到3.6 V的上截止电压。在这些约束条件下，总共允许有224个充电协议，为了一致的协议命名，我们将每个快速充电协议定义为CC1-CC2-CC3-CC4。例如，co -估计平均循环寿命最高的充电协议写为4.8C-5.2C5.2C-4.160C。

Early outcome predictor：

周期寿命的早期预后预测与Severson等人提出的相似5。该线性模型使用前100个循环的特征预测最终log10循环寿命(达到标称容量80%的循环次数，或0.88 A h)。这套训练装置与Severson等人使用的训练装置相同，由41节电池组成。线性模型的形式为:

其中，yi是电池i的预测循环寿命，xi是电池i的p维特征向量，w?为p维模型系数向量。在对模型进行评估之前，对训练集的特征进行z评分(均值减去标准差并归一化)。

正则化，或者同时特征选择和模型拟合，使用弹性网进行。正则化惩罚过于复杂的拟合，以提高可泛化性和可解释性。

具体系数向量w定义如下：

这里λ和α是超参数;λ是一个非负标量，α是一个介于0到1之间的标量。第一项使平方损失最小化，第二项同时执行连续收缩和自动特征选择。在模型开发过程中，我们采用四次交叉验证和蒙特卡罗抽样的训练集来优化超参数λ和α的值。

正如Severson et al5所述，可用的特征是基于循环100和10的放电电压曲线之间的差异，或放电容量的趋势。所选的五个特征及其对应的权重和z评分值见补充表1。训练(交叉验证)误差为80.4个周期(10.2%);Severson等5的测试集的测试误差为122个周期(12.6%)。

如果95%的预测间隔超过2000个周期，早期预测器自动将预测标记为异常。双尾95%预测区间计算公式为:

，

其中，t为Student的t值，α为显著性水平(95%置信度为0.05)，n为样本数量，p为特征数量，RMSE为训练集的均方根误差(以周期为单位)，xi为电池i的选定特征向量，X为训练集中所有观测结果的选定特征矩阵。

在闭环实验中，有三个测试返回的预测区间在阈值之外;这些反常的预测被排除在外。这些测试在第1轮(oed\_0)的频道27上运行，在第3轮(oed\_2)的频道12上运行，在第4轮(oed\_3)的频道6上运行。此外，在验证实验中，一个测试返回的预测区间超出阈值(通道12;3.6C-6.0C-5.6C-4.755C)，虽然最终循环寿命是合理的。

我们注意到，该模型的预测对验证实验中的细胞显示出系统性偏差，我们将其归因于这些细胞相对于训练集的日历老化增加(补充表2和补充讨论4)。

Bayesian optimization algorithm：