

工业化的兴起导致了对能源的大量需求,化石能源的枯竭以及环境污染促进了新能源的崛起[2,3]。然而,大多数可再生能源,如太阳能和风能,本质上是间歇性的,依靠自然现象来发电,此时能源存储便成为新的需求 [4-10]。可充电锂离子电池作为一种储能技术,由于拥有高密度、大容量、长寿命等优点,已经得到了广泛应用[11,12]。但广泛使用的同时也带来了续航、动力不足、电池爆炸等问题,如图 1 所示。如果能在电池老化之前对电池寿命进行预测的话,在避免上述问题的同时还可以加快电池开发周期、验证新的工艺、对电池进行二次回收[16-18],将为电池的制造、使用和优化带来重大机遇[19-21]。

目前的电池寿命估计所采用的模型主要可以分为等效电路模型(ecm)[22-23]、电化学模型[24-26]和数据驱动模型[27-32]三种。电化学模型和等效电路模型的准确性和鲁棒性有限,因此这两种模型并不是一个很好的可行解决方案。相反,数据驱动模型不需要了解电池内部的复杂化学反应,没有复杂的建立电路的过程等优势被研究者广泛应用。同时随着近几年的研究展开,发现了电池数据集中带有噪声是不可避免的,这主要源于充放电过程中的环境干扰,如温度变化、湿度波动的影响。此外实验条件并不能完全模拟现实,因此,越来越多的研究开始关注带有噪声的电池寿命预测。

线性参数估计问题出现在信号处理等广泛的科学学科中[33-34]。如[35]和[36]所示,在所有感兴趣的变量都具有参数线性关系且所有测量值都受到噪声污染的情况下,总最小二乘法是参数估计的最佳选择。

但实际情况中电池厂商提供的电池信息数据集来源不同,故由于温度、人为干扰以及传感器等造成的误差大不相同,此时便不能简单的假设数据集的噪声服从同一个分布,直接使用 TLS/OLS 并不能很好的建立电池寿命预测模型,故本文在建立线性模型计算电池的寿命时进行改进,对带有不同噪声分布的电池样本进行加权之后,使用 TLS/OLS 进行预测,经循环迭代能够准确的计算出噪声分布的标准差的同时建立适应不同噪声分布的预测模型对电池寿命进行预测,预测结果显示我们的方法比传统的 TLS/OLS 方法更好。

本文使用三个基于商用锂离子电池的公共数据集,三个数据集分别称为“数据集 1”、“数据集 2”和“数据集 3”,分别由 41 个、43 个和 40 个样本组成,虽然这些数据集总共提取了 20 个特征,但我们进一步根据领域专业知识手动选择 3 个重要特征的子集,具体含义见 table1,我们将三个数据集按照 9:1 划分之后合并作为训练集和测试集,每次实验随机打乱样本顺序,每次运行均独立随机生成训练和测试数据集。为每种方法报告 1000 个 RMSE 值的中位数,以便误差度量不会因随机波动而产生强烈偏差

Data-driven prediction of battery cycle life before capacity degradation

如图 3 所示,我们设置了四种不同的噪声比例。图 a,b,c,d 是噪声水平增大的实验结果:(1)随着噪声增大,TLS 和 OLS 效果明显变差,而改进的算法受噪声水平的影响不明显,具有较强的稳定性。(2)结合 EM 思想改进的算法(TLS_EM、OLS_EM)比传统算法(TLS、OLS)效果更佳,说明了改进的算法更能适应带有噪声的电池数据集。(3)TLS_EM 效果优于 OLS_EM,(TLS 效果也优于 OLS),在所有测量值都收到噪声污染的情况下,TLS 比 LS 有更大的优势。

图 e,f,g,h 是训练集比例增大的实验结果:(1)随着训练集占比增大,四种方

法效果更好，有了更多的训练数据，模型预测能力提升。(2) 不论训练集比例大小，改进的算法优于传统算法，说明了融入 EM 思想的算法有效性。(3) 在绝大部分情况下 (训练集占比大于 25%) TLS_EM 效果优于 OLS_EM, 说明了 TLS_EM 比 OLS_EM 适用性更强。

图 4 展示了算法经过循环迭代噪声收敛过程。(算法通过 TLS/OLS 拟合样本数据得到模型系数 w 和 b ，根据模型系数 w 和 b 对样本数据进行预测得到新一轮的电池寿命预测值，将预测值和真实值对比得到新一轮的误差，由此求出标准差对样本数据加权后通过 TLS/OLS 求出下一轮的模型系数 w 和 b)

本文在建立线性模型计算电池的寿命时进行改进，对带有不同噪声分布的电池样本进行加权之后，使用 TLS/OLS 进行预测，经循环迭代能够在准确的计算出噪声分布的标准差的同时建立适应不同噪声分布的预测模型，进而对电池寿命进行预测。预测结果显示我们的方法有着更好的效果。