基于总体最小二乘回归的鲁棒电池寿命预测

IEEE会员、翟晓昂、陈思慧、、万家钰、、

*IEEE研究员*

***摘要-机器学习技术最近在可再生能源研究中获得了极大的关注，并被广泛应用于提供数据驱动的解决方案，以应对各种重大挑战。本文主要研究电池寿命预测问题。我们的目标是根据物理测试收集的测量数据，准确估计电池的剩余寿命。重要的是，我们建议采用总体最小二乘(TLS)回归，通过最小化与输入特征和预测结果相关的总误差来适应有噪声的测量。此外，一个新的数值求解器被开发，以确定所有未知的模型系数既有效又鲁棒。基于三个商用锂离子电池公共数据集的数值实验表明，与普通最小二乘(OLS)回归相比，TLS可以有效地将建模误差降低8.8%。***

***索引术语—电池寿命预测、总体最小二乘回归、测量误差、能量存储***

1. 介绍

稀有

随着可持续性的发展，可充电电池技术在现代社会正获得越来越多的关注

正在成为全球利益的关键问题。作为一种流行的可充电电池，锂离子电池由于其高能量密度和低成本的优点而得到广泛应用，这些优点使得在包括便携式设备、电动车辆和电网在内的多个应用领域中能够实现可靠和经济的能量存储[[1]](#_bookmark36)-[[2].](#_bookmark37) 锂离子电池的质量可以通过其寿命来定量评估，因此，电池寿命预测已在文献中得到广泛研究，为电池的制造、维护和使用提供了战略指导[[1],](#_bookmark36) [[3]](#_bookmark38)-[[5].](#_bookmark39)

最近已经采用了多种机器学习技术来以高精度和低成本预测电池寿命[[1],](#_bookmark36) [[6].](#_bookmark40) 这些方法旨在训练回归模型

T.中国上海东华大学的陆、翟和刘(电邮:[luting@dhu.edu.cn，](mailto:luting@dhu.edu.cn) [xiaoang\_zhai@mail.dhu.edu.cn，](mailto:xiaoang_zhai@mail.dhu.edu.cn)ghliu@dhu.edu.cn)。

南中国广东深圳南方科技大学的陈和J. Wan(电子邮件:[12232304@mail.sustech.edu.cn，](mailto:12232304@mail.sustech.edu.cn)wanjy@sustech.edu.cn)。

Y.刘和x .李就职于中国江苏昆山杜克大学(电子邮件:{yang.liu，xinli.ece}@dukekunshan.edu.cn)。

Y.本文通讯作者为刘、万俊杰和李。

它将电池测量值作为其输入来预测相应的寿命值。例如，在以下方面取得了重大进展[[1]](#_bookmark36) 通过从早期循环电池放电曲线中识别一组重要特征，这些特征与电池寿命密切相关。这项开创性的工作能够在早期阶段预测电池寿命，并在现实世界的应用中具有实际效用。此外，各种回归建模算法，如弹性网[[1],](#_bookmark36) 高斯过程回归[[7],](#_bookmark41) 支持向量回归[[6],](#_bookmark40)[[8],](#_bookmark42)XGBoost[[8]](#_bookmark42)和深度神经网络[[9]](#_bookmark43)-[[10],](#_bookmark44)已经被建议在实践中预测电池寿命。

作为回归建模中的常见实践，当训练回归模型时，大多数传统方法试图最小化电池寿命的预测误差。然而，输入特征(例如，电压、电流、电阻、温度等。)通常有噪声，因为它们是通过具有测量误差的探针技术测量的[[5].](#_bookmark39)此外，当所测量的物理度量被用来导出用于电池寿命预测的新特征时，这些物理度量上的测量误差可能被进一步传播到所导出的特征[[11]](#_bookmark45)-[[12].](#_bookmark46) 以放电容量曲线为例。虽然它是通过在测试周期内对电流进行积分来计算的，但它会受到时域采样和电流测量带来的累积误差的影响。这些与输入特征相关的测量误差没有被传统的回归建模方法适当地考虑。

为了应对上述挑战，本文提出采用总体最小二乘法(TLS)回归来预测电池寿命。与传统方法不同，TLS最小化了与输入特征和预测结果相关的总误差[[13]](#_bookmark47)-[[14].](#_bookmark48) 换句话说，它能够学习考虑噪声特征的精确回归模型。由于这一优点，TLS在许多具有EIV性质的实际应用中得到了广泛的应用，例如图像重建和金融数据预测[[14].](#_bookmark48) 我们的数值实验基于三个公共数据集

[[1]](#_bookmark36) 对于商用锂离子电池，与普通最小二乘(OLS)回归相比，TLS可以将建模误差减少8.8%[[15].](#_bookmark49)

本文的其余部分组织如下。在第二节中，我们推导了TLS的数学公式

电池寿命预测。TLS的数值求解器将在第三节中讨论。在第四节中，给出了实验结果来证明TLS的有效性。最后，我们在第五节得出结论。

其中εy，k和εg，m，k分别表示εy和εg，m的第k个样本。此外，通过假设数据集中的所有样本{(xk，yk)；k =1，，K}在统计上是独立的，观察这k个样本的可能性等于:

*  2*

* 2 *

1. 问题定式化

纽约州·x

  exp 

*y·k·*

**

考虑一组特征x = [x1 x2  xN]T，其中n

*k1*

*k k*

*k1*

 2米1 2 

。（[8](#_bookmark0))

表示特征的总数，上标T

·2

* 2 *

表示一个向量或矩阵的转置，我们的目标是

# 出口

*y·k·*

**

学习从x到电池寿命y的映射[[16]](#_bookmark50)-[[17]:](#_bookmark51)

十二年级

*k1m 1*

# 2 

*M*

*yWMGMxy，(*[1](#_bookmark1))

因此，最大似然解w可以通过求解以下优化问题来找到:

*m1*

最小ε2ε2

其中gm(x)是第M个基函数，wm表示第M个模型系数，εi表示遵循零均值高斯分布N(0，σy2)的建模误差，M表示

**ε，E，w 2 F**

·ge·w

, ([9](#_bookmark2))

基函数的总数。

其中| | | | | F表示矩阵的Frobenius范数，并且

给定一组样本{(xk，yk)；k =1，，K}，模型

 y，1

 y，2 L

·t

（[10](#_bookmark3))

系数{ wmm =1，，M}可以通过最小化总平方误差来确定[[15]](#_bookmark49)：

 g，1.1

*g，2.1*

L g，m，1 

部





y  G  w 2，([2](#_bookmark5))

g，1，2

g，2.2升

g，m，2 。（[11](#_bookmark4))

w2 



其中| | | | | 2表示向量的L2范数

   

·x *x·x*

 g，1，K g，2，K g，m，K 

1 1 2 1





 g1 x2  *g2 x2  L*

*M 1*

*通用汽车x2  公司*

（[3](#_bookmark6))

注意，最小化([9](#_bookmark2))相当于最大化([8](#_bookmark0)).这种方法被称为

 

文献中的TLS回归[[13].](#_bookmark47)

·x·x·x

 1 K 2 K M K 

w1 *西部和西部*

*M*

*T*

（[4](#_bookmark7))

1. 数值求解器

中的最优化问题([9](#_bookmark2))是非凸的，因为它的

yy1 *。（*[5](#_bookmark8))上述方法在文献中被称为OLS回归，其目的是找到未知模型系数的最大似然解w[[15].](#_bookmark49)

*K*

*T*

实际上，每个基函数gm(x)可能是有噪声的，因为特征向量x是用测量误差测量的。在这种情况下，Eq。（[1](#_bookmark1))应改写为:

*M*

二次等式约束[[18].](#_bookmark52) 因此，很难找到既高效(即，具有低计算成本)又鲁棒(即，具有保证的全局收敛性)的最优解。在这一节中，我们将原始的非凸问题转化为一个易于求解的等价问题。

首先，我们注意到等式。（[9](#_bookmark2))被简化为下面的凸

如果向量w已知，则优化问题:

*伊··伊·wmgmxg·，* （[6](#_bookmark9))

最小ε2ε2

*m1*

ε，E 2 F

, ([12](#_bookmark10))

其中εg，m表示与第m个基函数gm(x)相关的测量误差。本文假设εg，m服从零均值高斯分布N(0，σg，m2)。

为了导出未知模型系数的最大似然解w([6](#_bookmark9))，我们进一步做两个假设。第一，既εy又{εg，m；m =1，，M}可以通过适当缩放y和{gm(x)归一化为标准高斯分布N(0，1)；m =1，，M}分别为。第二，εy和{εg，m；m =1，，M}在统计上是独立的。

有了这两个假设，很容易证明观察样本的可能性(xk，yk)等于:

·ge·w

其中ε和E是问题未知数。情商。（[12](#_bookmark10))最小化服从一组线性等式约束的凸二次成本函数。利用KKT条件可以解析地导出它的最优解[[18].](#_bookmark52) 为此，我们计算了([12](#_bookmark10)):

*lεeλε2e 2λt gewyε，(*[13](#_bookmark11))其中λ是包含K个拉格朗日乘数的向量。通过取L(ε，E，λ)对ε和的偏导数

2楼 

e并将它们设置为零，我们得到:

1   2 *M* 一

  2 

**2  ε  λ  0 (**[14](#_bookmark13)**)**

*·xk*

# exp 

*y·k·*

# exp 

** *T*

#  2 



2



*m12*  2 

2eλw0。（[15](#_bookmark14))

1   2

* 2 *

, ([7](#_bookmark12))

ε和E的最优解必须满足([14](#_bookmark13))-([15](#_bookmark14))和

#  出口

*y·k·*

**

线性等式约束:

 2 M 1

#  2

*m1*

# 2 

yεgew[16](#_bookmark15))

求解线性系统([14](#_bookmark13))-([16](#_bookmark15))产量:

··y

w 2 1

2

g和 wT

w 2 1

2

（[17](#_bookmark16))

样本集{(xk，yk)；k =1，，K}。

1. 计算矩阵G在([3](#_bookmark6))和向量y在([5](#_bookmark8)).
2. 基于([22](#_bookmark21)).
3. 计算B的最小特征值对应的特征向量θ\*。

东部

。（[18](#_bookmark17))

1. 计算矢量θ in([25](#_bookmark26))通过归一化θ\*。
2. 计算矢量w在([26](#_bookmark27))，包含最优模型

替换([17](#_bookmark16))-([18](#_bookmark17))变成([9](#_bookmark2))，TLS回归的代价函数可以表示为:

系数。

ε 2 第二季

 w  y 2

 2 .（[19](#_bookmark18))

1. 实验结果

2 *F* w 2 1

2

注意，成本函数在([19](#_bookmark18))已经在ε上最小化

和E，它现在是w的非凸函数。

为了有效地最小化([19](#_bookmark18))，我们构造以下优化问题[[18]:](#_bookmark52)

y 2

部 2

2

在本节中，三个基于商用锂离子电池的公共数据集[[1]](#_bookmark36) 用于证明我们提出的用于电池寿命预测的TLS方法的有效性。这三个数据集被称为“数据集1”、“数据集2”和“数据集3”，它们分别由41、43和40个样本组成。

虽然已经为这些数据集提取了总共20个特征[[1],](#_bookmark36) 我们进一步手动选择5个重要的子集

w，t

w 2  t 2

。（[20](#_bookmark19))

特征，表示为x = [x1

*x2  x5]T，基于领域专业知识。*

它相当于:

标准时间t1

 B 

在本例中，由于每个数据集中只有少量样本，因此我们仅使用5个特征进行回归

建模以避免过度拟合。这些选定特性的物理意义总结在[TABLE I.](#_bookmark23)到

在哪里

部

**θ**

# 科学技术。

**θ 2**

2

1先生1

, ([21](#_bookmark20))

为了降低建模任务的非线性，我们对电池寿命y和第一特征x1取对数[[1].](#_bookmark36)通过这些非线性变换，我们采用以下线性模型模板来预测log(y):

** GT G GT y**

log和w1logx1w2 x2w3 x3w4 x4w5 X5w6。（[27](#_bookmark22))

**yt·居伊特·**

 

（[22](#_bookmark21))

表一

θw，([23](#_bookmark24))

·

 

并且γM+1是γ的第(M+1)个元素。学习([21](#_bookmark20))，我们注意到它的成本函数与θ的大小无关。因此，在不使用

所选功能的描述

约束θ

*M+1*

= 1产生相同的最佳成本函数值:

 B 

|  |  |
| --- | --- |
| 功能名称 | 描述 |
| *x1* | 第10次和第100次循环之间作为电压函数的放电容量曲线差异的方差 |
| *x2* | 由线性函数拟合的容量衰减曲线的斜率 |
| *x3* | 第二次循环的放电容量 |
| *x4* | 从第2个周期到第6个周期的平均充电时间 |
| *x5* | 第2次和第100次循环之间的内阻差异 |

# 部

**θ 2**

2

**θ**

# 科学技术。

θ  0

。（[24](#_bookmark25))

出于测试和比较的目的，已经实施了两种回归建模方法:(i) TLS和(ii) OLS。为了评估建模的准确性，我们对每个模型进行了划分

情商。（[24](#_bookmark25))代表了众所周知的瑞利商最小化问题[[18].](#_bookmark52)它的最优解θ\*是b的最小特征值对应的特征向量.一旦θ\*

数据集转换成(I)用于确定未知模型系数的训练数据集，以及(ii)用于基于均方根误差(RMSE)评估建模精度的测试数据集

已知，我们可以用第(M+1)个元素θ\*来缩放θ\*

确定()的最优解θ[21](#_bookmark20)):

**θ\***

到[[19].](#_bookmark53)我们随机选择30%的样本进行测试

数据集和剩余的70%样本形成训练

*M+1*

数据集。此外，我们改变了70%的训练数据集大小

θ  



*1先生*

。（[25](#_bookmark26))

样本为20%的样本，以便研究TLS和OLS的鲁棒性时，只有极少数的训练

因此，未知模型系数的最优解w可以通过取θ的前M个元素来获得:

1 2l。 （[26](#_bookmark27))中总结了TLS回归的主要步骤[Algorithm 1.](#_bookmark28)

*M*

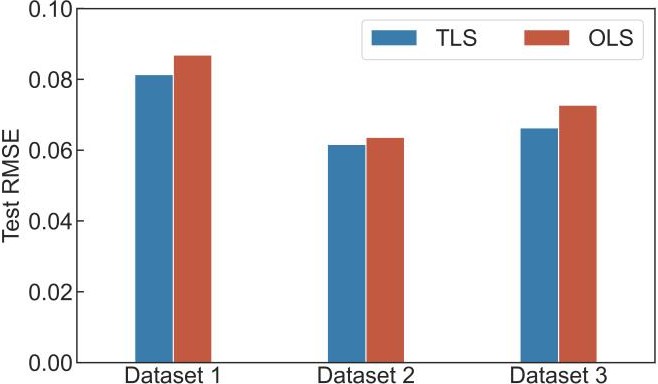
*T*

**算法1:总体最小二乘回归**

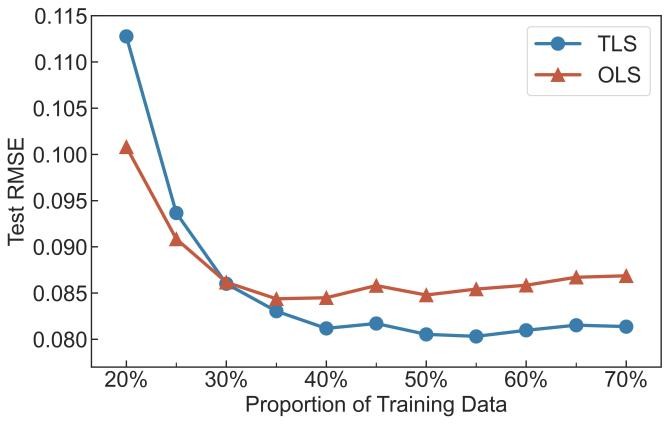
1.从一组基函数{gm(x)开始；m =1，，M}和a

提供样品。为了提高数值稳定性，我们对预测结果log(y)和所有特征{log(x1)，x2，x3，x4，x5}进行了归一化，以便它们在训练数据集上具有零均值和单位方差。前述实验被重复200次，其中训练和测试数据集在每次运行时被独立地和随机地生成。报告了每种方法的200个RMSE值的中值，以使误差度量不会因随机波动而有很大偏差。

[Fig. 1](#_bookmark29) 显示了三个数据集的TLS和OLS的RMSEs，其中在每种情况下，70%的样本用于形成训练数据集。通过考虑x的测量误差，在所有三个数据集上，TLS始终比OLS获得更高的建模精度。最值得注意的是，“数据集3”的建模误差降低了8.8%。这反过来表明，当输入特征有噪声时，TLS可以通过最小化x和y的总平方误差来有效地提高建模精度。



图一。显示了三个数据集的TLS和OLS的RMSEs，其中在每种情况下使用70%的样本来形成训练数据集。



图二。随着“数据集1”的训练数据集大小从20%样本变化到70%样本，TLS和OLS的RMSEs降低。

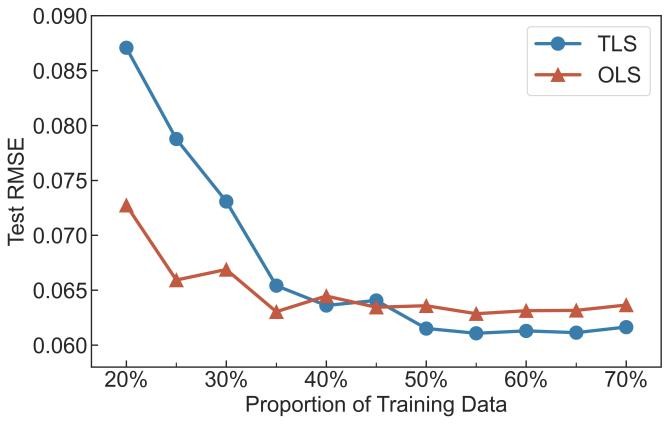


图3。随着“数据集2”的训练数据集大小从20%样本变化到70%样本，TLS和OLS的RMSEs降低。

[Fig. 2,](#_bookmark30) [Fig. 3](#_bookmark31) 和[Fig. 4](#_bookmark32) 进一步示出了分别作为三种情况的训练数据集大小的函数的TLS和OLS的RMSEs。研究这些数字揭示了三个重要的观察结果。首先，TLS和OLS的RMSEs是

减少，因为有更多的训练样本可用，因此为模型训练纳入了更多的信息。第二，当只使用少量训练样本时，OLS优于TLS。注意TLS需要在([9](#_bookmark2))，而OLS只需要在([2](#_bookmark5)).由于TLS涉及更多的问题未知，它对过拟合的鲁棒性较差，并且需要更多的训练样本来可靠地确定所有未知。第三，当训练数据集足够大时，TLS比OLS更准确，这与我们在[Fig. 1.](#_bookmark29)

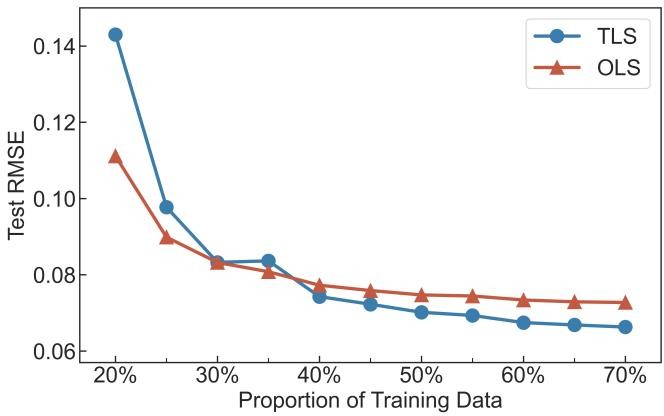


图4。随着“数据集3”的训练数据集大小从20%样本变化到70%样本，TLS和OLS的RMSEs降低。

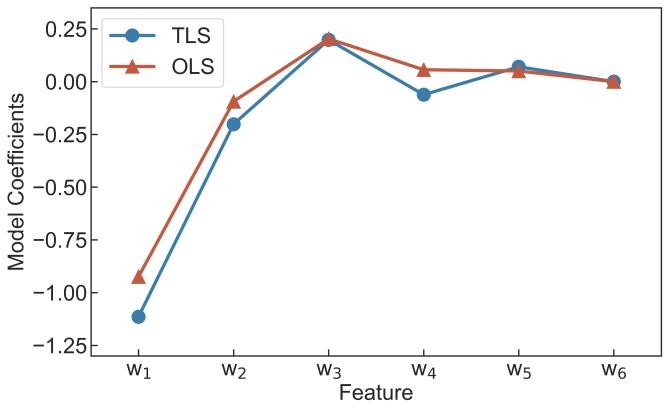


图5。针对“数据集1”示出了TLS和OLS的模型系数，其中70%的样本用于形成训练数据集。

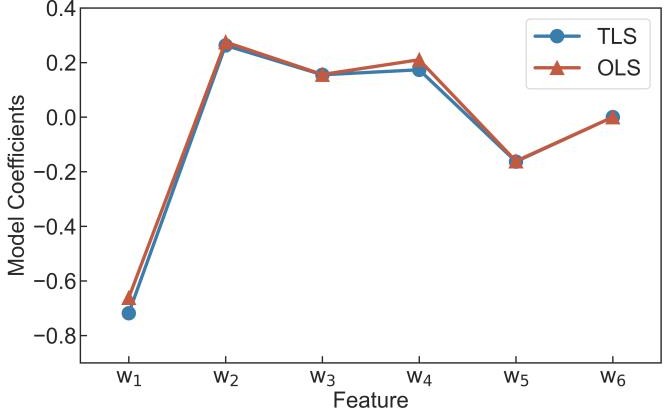


图6。针对“数据集2”示出了TLS和OLS的模型系数，其中70%的样本用于形成训练数据集。

最后，我们可视化三个数据集的TLS和OLS模型系数，如所示[Fig. 5,](#_bookmark33)[Fig. 6](#_bookmark34)和[Fig. 7](#_bookmark35)其中70%的样本用于形成训练

数据集。注意，模型系数w6等于0，因为预测结果log(y)已经被标准化为具有零均值。由于与输入特征相关的测量误差，由TLS和OLS求解的模型系数相似，但不相同。在这个例子中，由于TLS误差小于OLS误差，由TLS求解的模型系数预计会更准确。

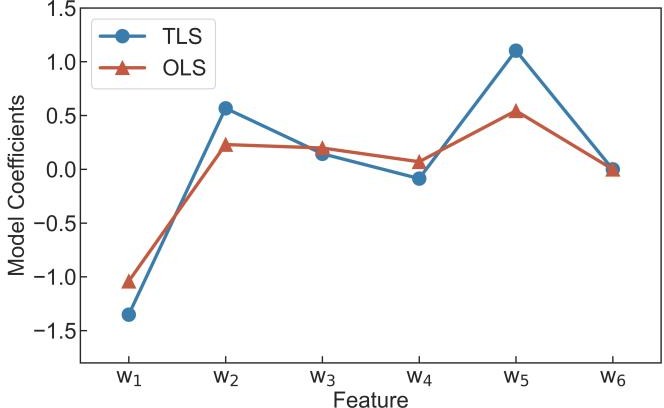


图7。针对“数据集3”示出了TLS和OLS的模型系数，其中70%的样本用于形成训练数据集。

1. 结论

本文提出了一种基于回归建模的预测电池寿命的TLS方法。TLS可以有效地考虑与输入特性和电池寿命相关的测量噪声。虽然TLS问题被视为非凸优化，但是开发了数值求解器来高效地(即，具有低计算成本)和鲁棒地(即，具有保证的全局收敛性)确定未知模型系数。基于三个商用锂离子电池公共数据集的数值实验表明，与传统的OLS方法相比，TLS方法可以有效地将建模误差降低8.8%。这反过来证明了TLS作为一种有前途的电池寿命预测方法在实践中的功效。

参考

1. K.A. Severson等，“容量退化前电池循环寿命的数据驱动预测”，《自然能源》第4卷第5期，第383–391页，2019年。
2. B.Scrosati和J. Garche，“锂电池:现状、前景和未来”

*电源杂志，第195卷，第9期，第2419-2430页，2010年。*

1. 南Chetoui和S. Reda，“面向移动SOC的工作负载和用户感知电池寿命管理”，《欧洲设计、自动化和测试会议与展览论文集》，第1679-1684页，2021年。
2. 米（meter的缩写））Rossi，A. Toppano和D. Brunelli，“混合住宅电气系统中电池组寿命的实时优化”，《欧洲设计、自动化和测试会议与展览论文集》，第1-6页，2014年。
3. J.魏，董国光，陈，"基于粒子滤波和支持向量回归的锂离子电池剩余寿命预测和健康状态诊断"，电气工程学报。《论工业电子学》，第65卷第7期，第5634-5643页，2018年。
4. 米（meter的缩写））A. Patil等人，“一种新的基于多级支持向量机的锂离子电池剩余使用寿命估计方法”，《应用能源》，第159卷，第285-297页，2015年。
5. Y.张等，“利用机器学习从阻抗谱识别锂离子电池退化模式”，自然通讯，第11卷第1期，第1706页，2020。
6. J.朱等，“基于电压弛豫的商用锂离子电池数据驱动容量估算”，自然通讯，第13卷第1期，2261页，2022。
7. D.陈等，“一种新颖的基于深度学习的多循环剖面下泛化能力强的锂离子电池寿命预测方法”，应用能源，第327卷，第120114页，2022。
8. C.-W. Hsu等，“仅使用一个周期的数据进行深度神经网络电池寿命和电压预测”，《应用能源》，第306卷，第118134页，2022年。
9. J.泰勒等人，“洞察锂离子电池表征实验的误差和不确定性”，《能量存储杂志》，第24卷，第100761页，2019年。
10. A.Moradpour等人，“电池电化学阻抗谱的测量不确定性”，IEEE Trans。《论仪器与测量》，第71卷，第1-9页，2022年。
11. A.贝克、本塔尔和特布尔先生。“寻找二次约束分数二次问题的全局最优解及其在正则化总体最小二乘法中的应用”，SIAM矩阵分析与应用杂志，第28卷，第2期，第425-445页，2006年。
12. H.朱，G. Leus和G. B. Giannakis，“扰动压缩采样的稀疏认知总体最小二乘”，IEEE Trans .《论信号处理》，第59卷第5期，2002-2016页，2011年。
13. C.模式识别和机器学习。斯普林格，2006年。
14. X.李，“从欠定方程中寻找确定性解:模拟/射频电路大规模性能可变性建模”，IEEE Trans .集成电路和系统的计算机辅助设计，第29卷，第11期，第1661-1668页，2010年。
15. F.王等，“贝叶斯模型融合:通过重用早期数据对模拟和混合信号电路进行大规模性能建模”，IEEE Trans .论集成电路与系统的计算机辅助设计，第35卷第8期，第1255-1268页，2016。
16. A.非线性最优化导论:理论、算法和MATLAB应用。工业和应用数学学会，2014年。
17. W.张等，“晶片空间特征的自动聚类”，

*设计自动化会议论文集，第1-6页，2013。*