提纲1提纲2主要区别是提纲1使用了一张电池背景图分析，提纲2 则是将电池的应用和电池寿命预测分为两张图

**提纲1：**

1. 引入带噪声的电池寿命预测，图1分析（新能源崛起，锂电池应用广泛，电池寿命预测的意义）
2. LS、TLS在降噪方面的广泛运用，提出存在不同数据集的噪声分布不同这个问题
3. 总结本文贡献，算法大致结构，图2分析
4. 所使用的特征分析（三个特征数据图，分别分析它们的趋势等）
5. 噪声水平增大，图5分析（详细讲述算法改进，算法优势）
6. 训练集比例增大，图6分析（算法在训练集较少情况下仍然有效）
7. 总结全文

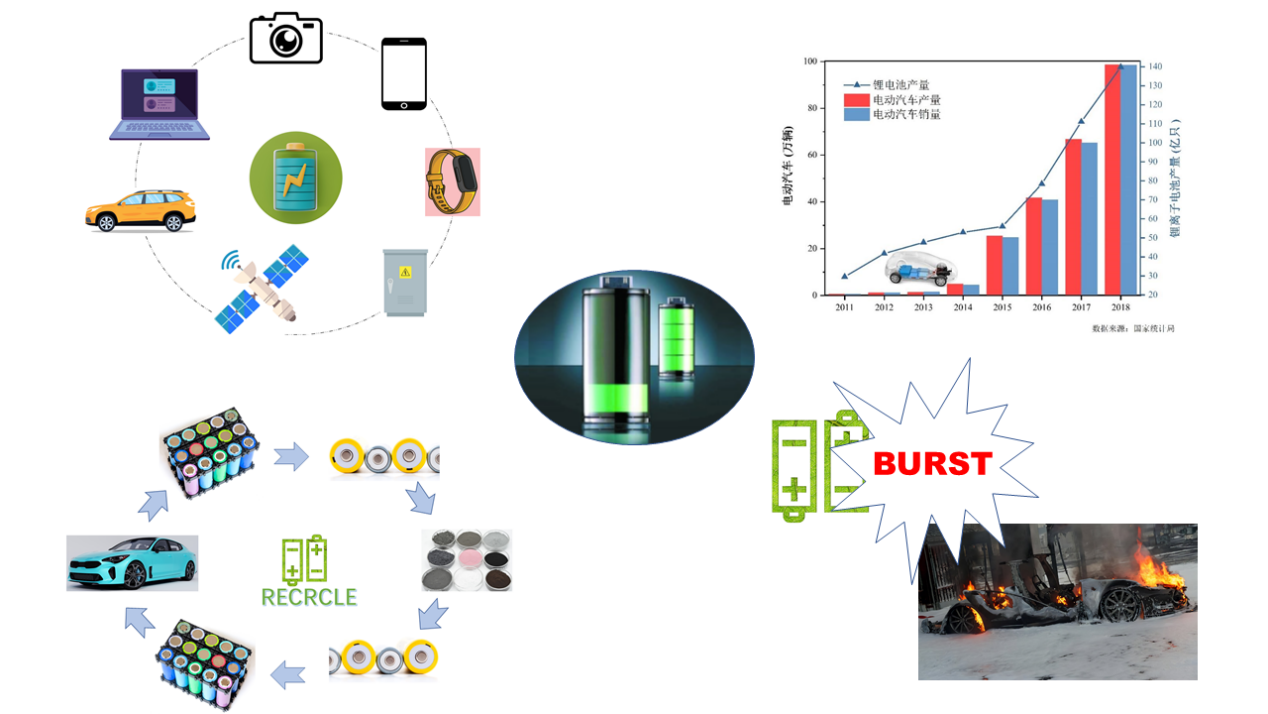


图1

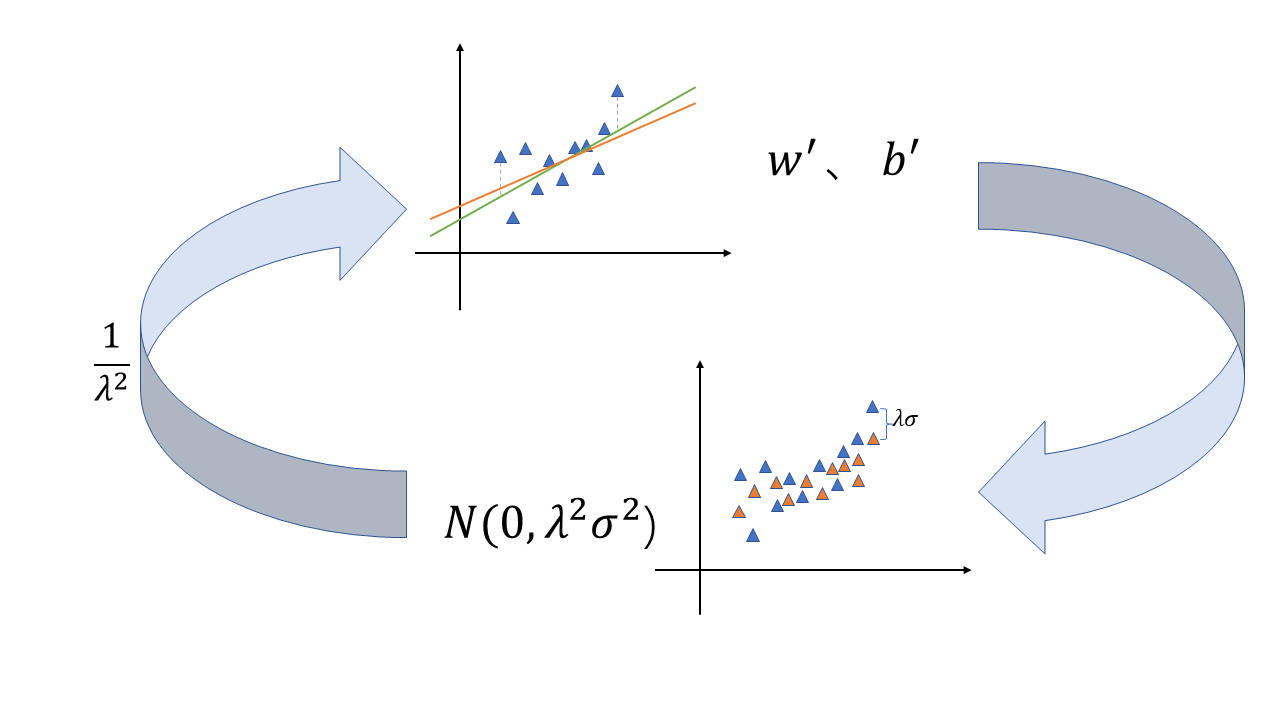


图2

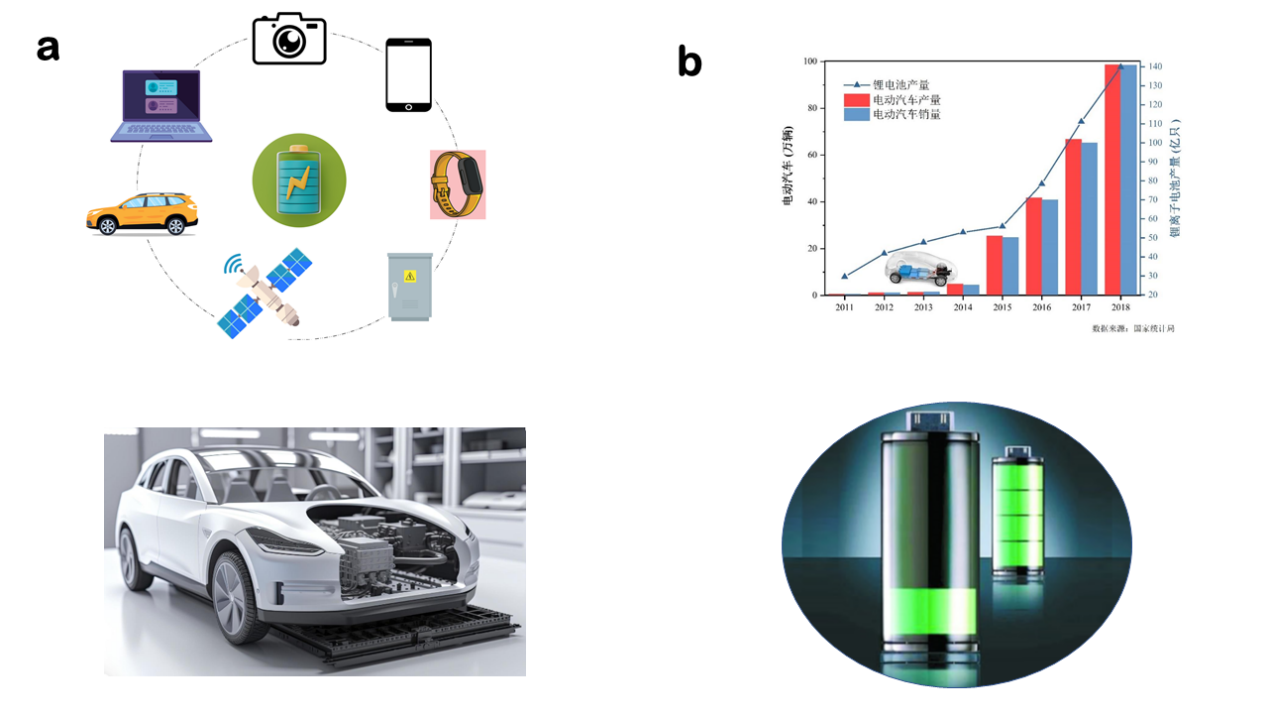


图3

图4

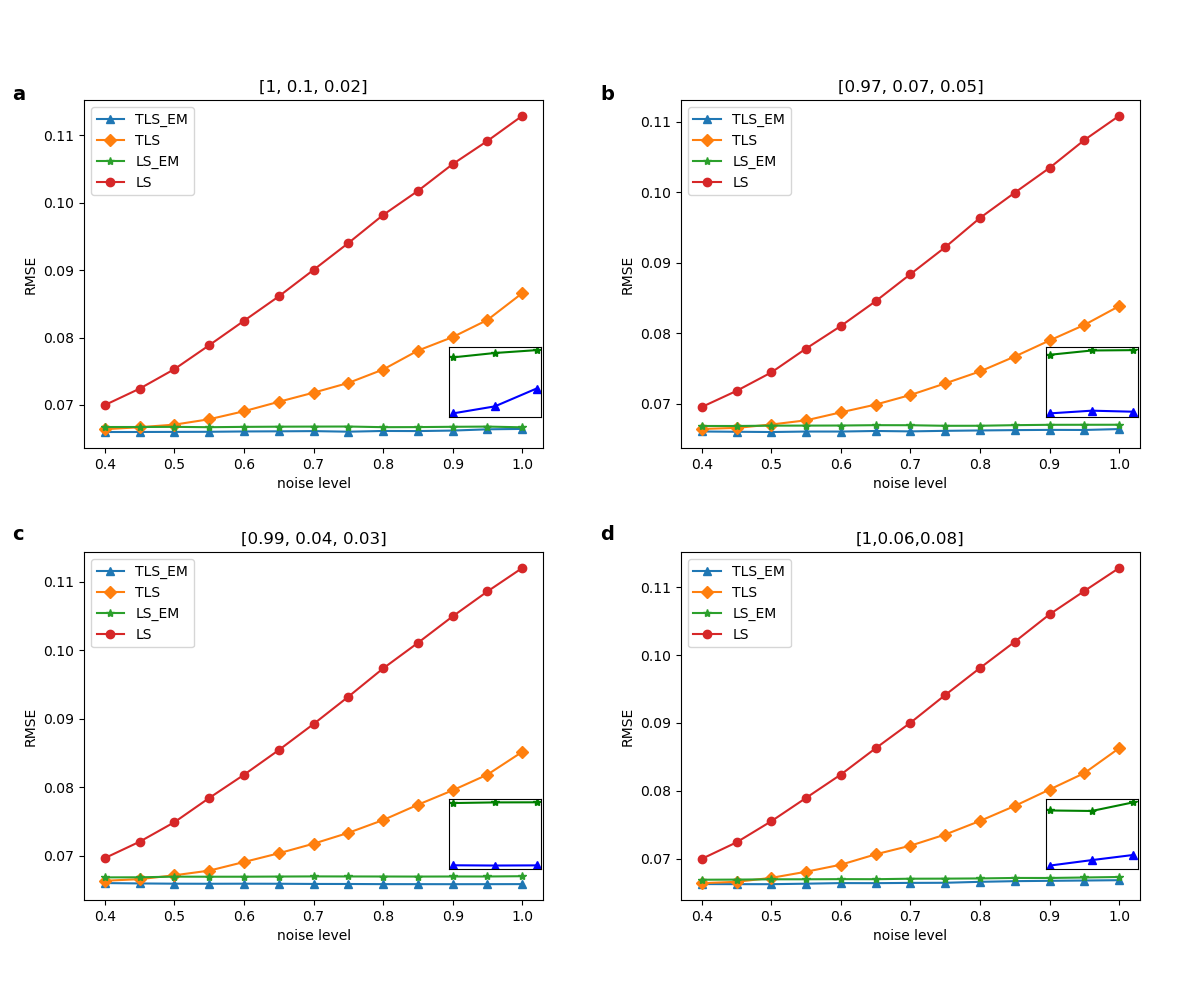


图5

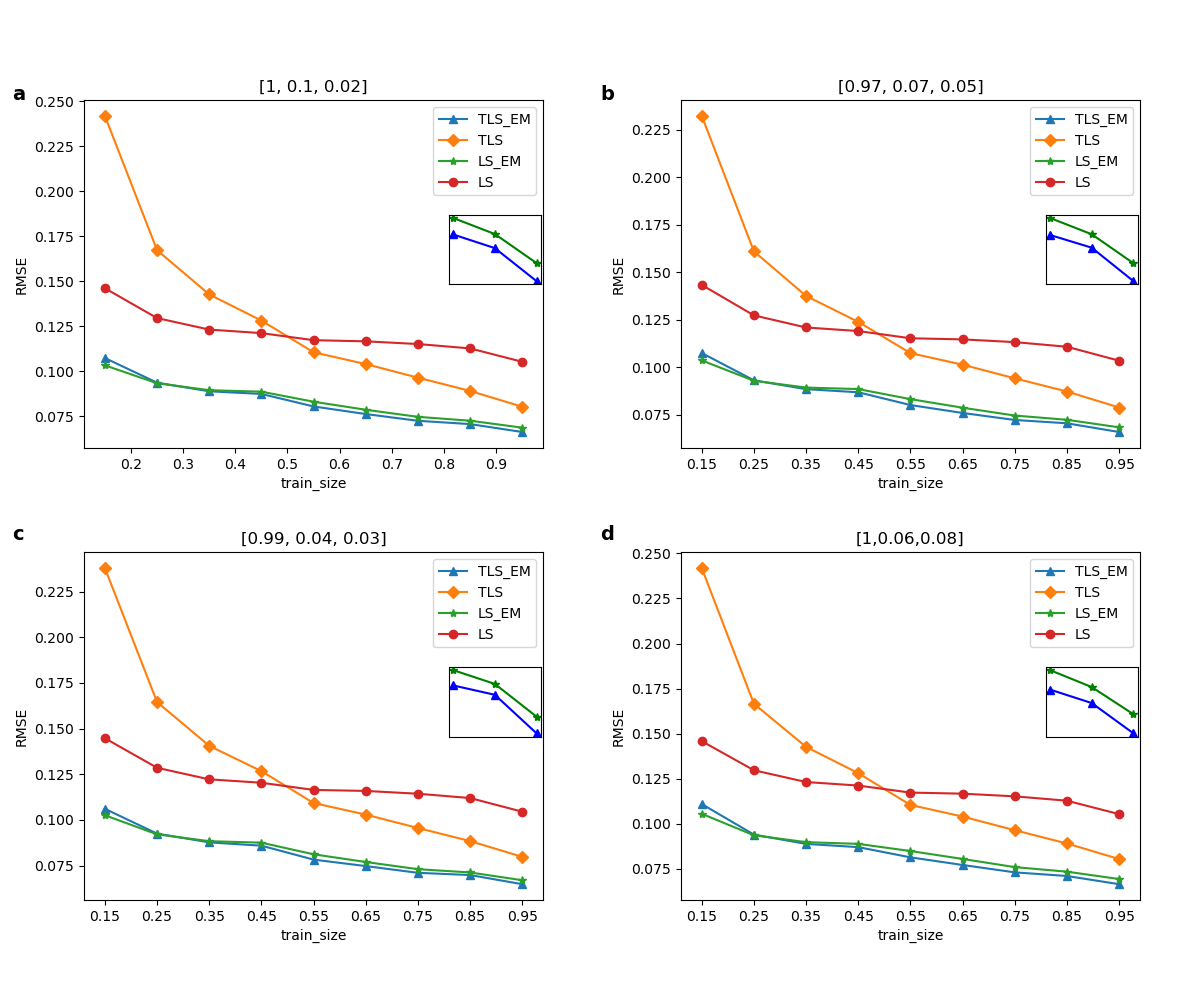


图6

锂离子电池由于具有高额定功率、高能量密度和长循环寿命等优点，在储能系统和电动汽车中得到了广泛的应用。为了提高电池系统的安全性、可靠性、效率和寿命，需要设计好电池管理系统(BMS)，对电池系统的基本状态进行监控。准确、可靠地监测电池荷电状态(SOC)是BMS的主要要求之一。目前，SOC的估计方法主要有四大类，分别是基于查找表的方法、安培小时积分法、数据驱动的估计方法和基于模型的估计方法[1]。基于查找表的方法从SOC与一个或多个参数(如开路电压(OCV)或阻抗)之间的关系推断SOC[2-4]。精确测量OCV需要电池在结束电流中断后有足够的休息时间，因此查找SOC和OCV之间的表无法实时获取SOC。电池阻抗的测量需要额外的测量因此，查找SOC和阻抗之间的表不适合运行电动汽车。安培-小时积分法易于实现，但存在初始状态不确定、电池电流测量误差、误差随时间累积等问题[1]。数据驱动的估计方法不需要精确的电池模型，但对训练数据的依赖程度高，计算复杂度高[5-10]

[1] R. Xiong, J. Cao, Q. Yu, H. He, F. Sun, Critical review on the battery state of charge

estimation methods for electric vehicles, IEEE Access 6 (2018) 1832–1843, https://

doi.org/10.1109/ACCESS.2017.2780258.

[2] R. Xiong, J. Tian, H. Mu, C. Wang, A systematic model-based degradation behavior

recognition and health monitoring method for lithium-ion batteries, Appl. Energy

207 (2017) 372–383, https://doi.org/10.1016/j.apenergy.2017.05.124.

[3] X. Hu, F. Sun, Y. Zou, Comparison between two model-based algorithms for Li-ion

battery SOC estimation in electric vehicles, Simulat. Model. Pract. Theor. 34 (2013)

1–11, https://doi.org/10.1016/j.simpat.2013.01.001.

[4] S. Rodrigues, N. Munichandraiah, A.K. Shukla, A review of state-of-charge

indication of batteries by means of a.c. impedance measurements, J. Power Sources

87 (2000) 12–20, https://doi.org/10.1016/S0378-7753(99)00351-1.

[5] H. Sheng, J. Xiao, Electric vehicle state of charge estimation: nonlinear correlation

and fuzzy support vector machine, J. Power Sources 281 (2015) 131–137, https://

doi.org/10.1016/j.jpowsour.2015.01.145.

[6] E. Walker, S. Rayman, R.E. White, Comparison of a particle filter and other state

estimation methods for prognostics of lithium-ion batteries, J. Power Sources 287

(2015) 1–12, https://doi.org/10.1016/j.jpowsour.2015.04.020.

[7] L. Kang, X. Zhao, J. Ma, A new neural network model for the state-of-charge

estimation in the battery degradation process, Appl. Energy 121 (2014) 20–27,

https://doi.org/10.1016/j.apenergy.2014.01.066.

[8] T. Weigert, Q. Tian, K. Lian, State-of-charge prediction of batteries and batterysupercapacitor hybrids using artificial neural networks, J. Power Sources 196

(2011) 4061–4066, https://doi.org/10.1016/j.jpowsour.2010.10.075.

[9] W. He, N. Williard, M. Osterman, M. Pecht, Prognostics of lithium-ion batteries

based on Dempster–Shafer theory and the Bayesian Monte Carlo method, J. Power

Sources 196 (2011) 10314–10321, https://doi.org/10.1016/j.

jpowsour.2011.08.040.

[10] Y. Shen, Adaptive online state-of-charge determination based on neuro-controller

and neural network, Energy Convers. Manag. 51 (2010) 1093–1098, https://doi.

org/10.1016/j.enconman.2009.12.015.

化石能源的问题-》新能源崛起-》锂电池能源的广泛应用-》锂电池的问题-》寿命预测的必要性、

能源是所有科学和工程技术中的基础，没有能源人类世界将难以运转。自18世纪末和19世纪初以来，工业化的兴起导致了对大量能源的需求，以驱动机械设备、生产线和交通工具等。因此化石能源开始被大量使用。早期使用的的煤炭、后来兴起的油和天然气等其他化石燃料使用的不断扩大，成为支撑现代社会和经济体系的关键因素。然而这些化石能源的枯竭以及随之产生的一系列环境问题，加上传统能源需求的加速迫使规划人员和决策者寻找替代能源[2,3]

Renewable energy technologies in Pakistan: Prospects and challenges

[2] Mirza UK, Ahmad N, Majeed T, Harijan K. Wind energy development in Pakistan. Renewable and Sustainable Energy Reviews 2007;11(9): 2179–90.

[3] Tiwari, GN, Ghosal, MK. Renewable Energy Resources: Basic Principles and Applications. Alpha Science Int’l Ltd., 2005. ISBN 1-84265-125-0

可充电电池技术在现代世界得到了越来越多的关注，因为可持续性正成为全球利益的关键问题。锂离子电池作为一种受欢迎的可充电电池，因其能量密度高、成本低等优点，在便携式设备、电动汽车、电网等多个应用领域实现了可靠、经济的储能

[2] B. Scrosati and J. Garche, “Lithium batteries: Status, prospects and future,” Journal of Power Sources, vol. 195, no. 9, pp. 2419-2430, 2010

许多可再生能源技术，包括太阳能、风能、潮汐能、生物质能和水能，已经得到了广泛的发展，以减轻对化石燃料的依赖。[1]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，必须储存和按需使用。[1] B. Obama, Science 2017, DOI: 10.1126/science.aam6284.

Porous Carbon Composites for Next Generation Rechargeable Lithium Batteries

可充电电池作为一种储能技术，已经被人们广泛应用于航空航天、便携式电子设备、电动汽车等领域【4,5】。其中锂电池由于拥有更高的能量密度，更小的体积、更长的寿命、更大的容量等优点，被认为是最佳的储能方式。

[4] P. Poizot, S. Laruelle, S. Grugeon, L. Dupont, J. M. Tarascon, Nature 2000, 407, 496.

[5] B. Dunn, H. Kamath, J. M. Tarascon, Science 2011, 334, 928.

Recent progress of magnetic field application in lithium-based batteries

可充电电池被广泛认为是下一代大规模储能系统或电动汽车(ev)取代不可再生能源的最佳电源。[2]可充电锂电池通常被认为是电动汽车的最佳电池技术。[3]自20世纪90年代首次商业化以来，可充电锂电池的研究与开发(R&D)进展迅速。可充电锂电池已经彻底改变了便携式电子设备，并已成为移动电话、笔记本电脑、数码相机和视频的主要电源，因为它们具有优越的能量密度，与传统的可充电电池相比，每单位重量和体积能够存储更高的能量。[4,5]然而，可充电锂离子电池在高电流速率下的充放电过程会导致大块材料的高度极化，降低电池的电化学性能。电动汽车或混合动力汽车(hev)的发展需要能够在大电流条件下工作的高功率电池

[2] J. M. Tarascon, M. Armand, Nature 2001, 414, 359.

[3] J. B. Goodenough, Y. Kim, Chem. Mater. 2010, 22, 587.

同时城市可持续发展需求和新一代信息技术应用，智慧城市将成为人们未来的模式[1,2]。此时，电动汽车能够很好的解决智慧城市的节能型发展和环境污染问题【3】，发展新能源电动汽车已成为全球共识[7]，

An energy matching method for battery electric vehicle and hydrogen fuel cell vehicle based on source energy consumption rate

[1] Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev 2015;50:1352e72.

[2] Xu J, Zhang R. CoMP meets smart grid: a new communication and energy cooperation paradigm. IEEE Trans Veh Technol 2013;64(6):2476e88.

[3] Wang G, Xu Z, Wen F, et al Traffic-constrained multiobjective planning of electric-vehicle charging stations.

IEEE Trans Power Deliv 2013;28(4):2363e72.

[7] Hu J, Zheng L, Jia M, et al Optimization and model validation of operation control strategies for a novel dual-motor coupling-propulsion pure electric vehicle. Energies 2018;11.

而可充电锂离子电池作为新能源电动汽车的最佳选择，在满足促进电动汽车发展的同时也比可避免的产生了一系列的问题。无论何种应用锂电池都会随着时间的推移而退化，具体表现为电池容量的丧失和阻抗的增加。随之产生的便是电动汽车的续航减少、动力不足等问题，且随时间推移，锂离子电池的老化可能会造成安全事故。电池的降解速率受动态运行条件的影响，包括不同的充放电速率、不同的电压运行极限和温度波动。

Predicting the State of Charge and Health of Batteries using Data-Driven Machine Learning

MACHINE LEARNING PIPELINE FOR BATTERY STATE OF HEALTH ESTIMATION

如果我们能在电池老化之前对其寿命进行预测将为电池生产、使用和优化带来新的机遇。例如，制造商可以加快细胞开发周期，对新的制造工艺进行快速验证，并根据预期寿命对新细胞进行分类/分级。同样，终端用户可以估计他们的电池寿命4 - 6。

Data-driven prediction of battery cycle life before capacity degradation

4. Peterson, S. B., Apt, J. & Whitacre, J. F. Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization. J. Power Sources 195, 2385–2392 (2010).

5. Ramadesigan, V. et al Modeling and simulation of lithium-ion batteries from a systems engineering perspective. J. Electrochem. Soc. 159, R31–R45 (2012).

6. Waag, W., Fleischer, C. & Sauer, D. U. Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. J. Power Sources 258, 321–339 (2014)

此外，电池预测对扩大回收部门至关重要，使设施能够决定电池是作为废金属回收还是用于要求较低的“第二生命”应用。Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning

总之，对电池当前和未来状态的准确预测将为电池的制造、使用和优化带来巨大的机会【3、4】。

3. Severson, K. A. et al Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy 4, 383–391 (2019).

4. Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M. & Dietmayer K. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using datadriven methods. J. Power Sources 239, 680–688 (2013).

电池健康状态可用于预测电池的预期寿命，但通过直接测量电池内部化学反应参数在线估计电池健康状态的可行性有限[4]。

[4] Anthony Barré, Benjamin Deguilhem, Sébastien Grolleau, Mathias Gérard, Frédéric Suard, and Delphine Riu. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. Journal of Power Sources, 241:680–689, 2013.

直接预测电池寿命的条件难以满足》基于数据驱动的方法优势体现》寿命预测的现状》有噪声寿命预测的必要性

目前的电池寿命估计所采用的模型主要可以分为以下三种，等效电路模型(ecm)[21-23]、电化学模型[24-26]或数据驱动模型[27-32]。电化学模型近似于电池运行过程中在电池内部发生的化学过程，需要电池详细规格信息和复杂的电化学知识。等效电路模型采用具有经验非线性参数的电路元件[9]，然而简单的等效电路无法完全模拟电池内部的化学反应，复杂的模型计算量又太大，且对电池行为的假设，上述的两种模型的准确性和鲁棒性有限[8]。因此这两种模型并不是一个很好的可行解决方案。相反，数据驱动的方法有着一系列的优势，比如不需要了解电池内部的复杂化学反应，分析各种电池降解原理，没有复杂的建立电路的过程等迄今为止，许多研究都使用机器学习工具来分析电池寿命预测估计。

[21] Xiaosong Hu, Shengbo Li, and Huei Peng. A comparative study of equivalent circuit models for li-ion batteries.

Journal of Power Sources, 198:359–367, 2012.

[22] Tianheng Feng, Lin Yang, Xiaowei Zhao, Huidong Zhang, and Jiaxi Qiang. Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction. Journal of Power Sources, 281:192–203, 2015.

[23] D Andre, M Meiler, K Steiner, H Walz, T Soczka-Guth, and DU Sauer. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. ii: Modelling. Journal of Power Sources, 196(12):5349–5356, 2011.

[24] Matthew J Daigle and Chetan Shrikant Kulkarni. Electrochemistry-based battery modeling for prognostics. 2013.

[25] Brian Bole, Chetan S Kulkarni, and Matthew Daigle. Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use. Technical report, SGT, Inc. Moffett Field United States, 2014.

[26] Githin K Prasad and Christopher D Rahn. Model based identification of aging parameters in lithium ion batteries.

Journal of power sources, 232:79–85, 2013.

[27] Kristen A Severson, Peter M Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H Chen, Muratahan Aykol, Patrick K Herring, Dimitrios Fraggedakis, et al Data-driven prediction of battery cycle life before capacity degradation. Nature Energy, 4(5):383, 2019.

[8] Bhaskar Saha, Kai Goebel, Scott Poll, and Jon Christophersen. Prognostics methods for battery health monitoring using a bayesian framework. IEEE Transactions on instrumentation and measurement, 58(2):291–296, 2008.

[29] Kai Goebel, Bhaskar Saha, Abhinav Saxena, Jose R Celaya, and Jon P Christophersen. Prognostics in battery health management. IEEE instrumentation & measurement magazine, 11(4):33–40, 2008.

[30] Xiaosong Hu, Jiuchun Jiang, Dongpu Cao, and Bo Egardt. Battery health prognosis for electric vehicles using sample entropy and sparse bayesian predictive modeling. IEEE Transactions on Industrial Electronics, 63(4):2645– 2656, 2015.

[31] Verena Klass, Mårten Behm, and Göran Lindbergh. A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. Journal of Power Sources, 270:262–272, 2014

[32] Peter M Attia, Aditya Grover, Norman Jin, Kristen A Severson, Todor M Markov, Yang-Hung Liao, Michael H Chen, Bryan Cheong, Nicholas Perkins, Zi Yang, et al Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature, 578(7795):397–402, 2020.

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，大多公用数据集都是在实验条件下完全充放电测量得到，但实际情况下的电池充放电是不完全的，因此，研究带有噪声的电池数据集才更贴近实际生活，将模型推广到实际应用时才能保证其健壮性。

线性参数估计问题出现在信号处理等广泛的科学学科中[1]。它从线性(参数内)模型开始，其中表示可以测量或可以从其他测量推断或可以通过非线性变换计算的过程变量;所有变量都受到测量噪声的影响;包含表征流程变量的基础关系的参数。如[5]和[16]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。

Detection of Abrupt Changes of Total Least Squares Models and Application in Fault Detection

[1] S. Van Huffel, “Tls applications in biomedical signal processing,” in Recent Advances in Total Least Squares Techniques and Error-in-Variables Modeling, S. Van Huffel, Ed. Philadelphia, PA: SIAM, 1997

[5] S. Van Huffel and J. Vandewalle, Frontiers in Applied Mathematics: The Total Least Squares Problem—Computational Aspects and Analysis. Philadelphia, PA: SIAM, 1991.

[6] S. Van Huffel, Ed., Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modeling. Philadelphia, PA: SIAM, 1997.

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。此时直接使用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。

Considering a set of features **x** = [*x*1 *x*2 ⋅⋅⋅ *xN*]*T*, where *N* denotes the total number of features and the superscript *T* represents the transpose of a vector or matrix, our objective is to learn a mapping from **x** to the battery lifetime *y* [23],[29]:

1 , (1)

where *gm*(**x**) is the *m*-th basis function, *wm* denotes the *m*-th model coefficient, *εi* stands for the modeling error following a zero-mean Gaussian distribution *N*(0, *σ*y2), and *M* represents the total number of basis functions.

Given a set of data samples {(**x***k*, *y*k); *k* =1, ⋅⋅⋅, *K*}, the model coefficients {*wm*; *m* =1, ⋅⋅⋅, *M*} may be determined by minimizing the total squared error [14]:

2 , (2)

where ||•||2 denotes the L2 norm of a vector and

3 (3)

4 (4)

5 . (5)

The aforementioned approach is referred to as OLS regression in the literature and it aims to find the maximum-likelihood solution **w** for the unknown model coefficients [14].

In practice, each basis function *gm*(**x**) may be noisy, as the elements in feature vector **x** is obtained through the physical measurements of batteries (e.g., voltage, current, temperature, etc.), which are usually associated with measurement errors. In this case, Eq. (1) should be re-written as:

6 , (6)

where *εg*,*m* represents the measurement error associated with the *m*-th basis function *gm*(**x**). In this paper, we assume that *εg*,*m* follows a zero-mean Gaussian distribution *N*(0, *σg*,*m*2).

In order to compute the solution of the unknown model coefficientsin (6), we formulate an optimization problem of **w** based on maximum-likelihood estimation. To achieve this goal, we make two further assumptions. First, both *εy* and {*εg*,*m*; *m* =1, ⋅⋅⋅, *M*} can be normalized to standard Gaussian distribution *N*(0, 1) by appropriately scaling *y* and {*gm*(**x**); *m* =1, ⋅⋅⋅, *M*} respectively. Second, *εy* and {*εg*,*m*; *m* =1, ⋅⋅⋅, *M*} are statistically independent.

With these two assumptions, it is straightforward to show that the likelihood of observing a sample (**x***k*, *yk*) is equal to:

7, (7)

where *εy*,*k* and *εg*,*m*,*k* denote the *k*-th samples for *εy* and *εg*,*m* respectively. Furthermore, by assuming that all samples in the dataset {(**x***k*, *yk*); *k* =1, ⋅⋅⋅, *K*} are statistically independent, the likelihood for observing these *K* samples is equal to:

8 . (8)

Hence, the maximum-likelihood solution **w** can be found by solving the following optimization problem:

9 , (9)

where ||•||F denotes the Frobenius norm of a matrix and

10 (10)

11 . (11)

Note that minimizing the cost function in (9) is equivalent to maximizing the likelihood in (8). Such an approach is referred to as the TLS regression in the literature [12].

According to the sample relationship of data set, we can know that our goal is to minimize the error when the samples obey different noises：

6 , (6)

To solve this problem, we give each sample a different weight:

6 , (6)

Assuming that the noise obeys a Gaussian distribution with zero mean and different variances: , then obeys a Gaussian distribution . The likelihood function is:

6 , (6)

Maximizing the likelihood function is equivalent to minimizing the objective function , and we can get by comparison.

EM algorithm can solve the problem with hidden variables well. In this paper, we assume that the standard deviation of noise obeys the distribution as hidden variables, and improve the traditional OLS/TLS algorithm. The specific steps are as follows:

1. Initialize model coefficients **w,b**

2. Predict the battery life according to the model coefficient, and calculate the error between the predicted value and the real value to update the noise standard deviation of the three data sets.

6 , (6)

3. The samples are weighted according to the formula, and a new round of model coefficients are obtained by using TLS/OLS.

4. Repeat steps 2 and 3 until convergence.

数据集由三个不同来源的三个小数据集组成，我们对每个小数据集按照9:1比例划分训练集和测试集，再将训练集和测试集分别合并组成最终的训练集和测试集。实验重复1000次，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。

Figure3