

数据驱动的动力电池能量特性预测研究*

王燕^{1,2,3} 闵海涛¹ 霍云龙^{2,3} 杨钊^{2,3}

(1. 吉林大学, 长春 130000; 2. 中国第一汽车股份有限公司研发总院, 长春 130013; 3. 高端汽车集成与控制
全国重点实验室, 长春 130013)

【摘要】为实现纯电动汽车电池能量信息的准确预测,提出了一种基于充电型纯电动汽车大数据的电池能量分析和预测方法。首先,通过大数据平台获取搭载相同型号电池车型的不区分地域大数据,然后使用区间平均法和支持向量回归(SVR)方法对总数据和典型地域数据进行里程-总能量关系的拟合,完成电池总能量衰减的预测,最后,将预测结果与长短时记忆(LSTM)神经网络的预测结果进行对比,并利用实车试验验证所提出方法的准确性。验证对比结果表明:基于SVR的模型能够对分散电池容量进行量化拟合,具有较高的预测精度。

关键词: 新能源汽车大数据 电池能量衰减 支持向量回归

中图分类号: U469.72 **文献标志码:** A **DOI:** 10.19620/j.cnki.1000-3703.20230754

Research on Power Battery Energy Characteristic Prediction Based on Data-Driven

Wang Yan^{1,2,3}, Min Haitao¹, Huo Yunlong^{2,3}, Yang Fang^{2,3}

(1. Jilin University, Changchun 130000; 2. Global R&D Institute, China FAW Corporation Limited, Changchun 130013;
3. National Key Laboratory of Advanced Vehicle Integration and Control, Changchun 130013)

【Abstract】To achieve accurate prediction of EV battery energy information, this paper proposed a method for battery energy analysis and prediction based on big data of chargeable pure electric vehicles. Firstly, the big data of vehicles with the same battery model from different regions were obtained through a big data platform, and then the interval average method and Support Vector Regression (SVR) were used to fit the relationship between mileage and total energy for both the total data and typical regional data, to predict degradation of the battery total energy. Finally, the predicted results were compared with that obtained from Long Short-Term Memory (LSTM) neural network, and the accuracy of the proposed method was verified by vehicle test. The results show that: the SVR-based model can quantitatively fit the degraded battery capacity, which has high prediction accuracy.

Key words: New energy vehicle big data, Battery energy degradation, Support Vector Regression (SVR)

【引用格式】王燕, 闵海涛, 霍云龙, 等. 数据驱动的动力电池能量特性预测研究[J]. 汽车技术, 2024(8): 22-26.

WANG Y, MIN H T, HUO Y L, et al. Research on Power Battery Energy Characteristic Prediction Based on Data-Driven[J]. Automobile Technology, 2024(8): 22-26.

1 前言

纯电动汽车动力电池在使用过程中会受到环境温度、用户驾驶行为、充电和放电次数等多种因素的影响,导致电池能量衰减,从而缩短整车续航里程,降低电池的使用寿命^[1]。

锂离子电池内部反应机理极其复杂,对其进行准确

的能量衰减分析面临多重困难^[2]。目前,动力电池能量的估计和预测方法主要包括基于电池模型的方法和数据驱动方法。基于电池模型的方法有经验模型法、等效电路模型法和电化学模型法,可建立能够直接或间接表征电池衰减程度的精确模型,但需通过大量试验数据和其他算法对模型参数进行估计和验证,成本较高^[3]。数据驱动方法以试验数据或运行数据为基础,提取能够表

*基金项目:国家自然科学基金项目(52372384);吉林省重大科技专项(20210301023GX)。

通信作者:闵海涛(1970—),博士,教授,主要研究方向为新能源汽车能量管理等,minht@jlu.edu.cn。

征电池老化因素的特征,并通过不同的机器学习算法,如神经网络、高斯过程回归等,建立特征和目标总能量衰减的拟合模型,以实现总能量衰减的预测或估算^[4-6],但也需要大量的线下实车数据作为输入。

然而,传统电池试验只能针对数量有限的电池开展,测试周期漫长,且工况相对单一。本文采用基于大数据的机器学习和深度学习模型的方法,通过搭建电池支持向量回归(Support Vector Regression, SVR)模型,实现数据的高效处理和分析,并采用长短时记忆(Long Short-Term Memory, LSTM)神经网络算法验证算法的有效性。

2 大数据提取与样本分析

新能源汽车大数据系统通过车辆内部传感器、车辆通信网络以及外部环境数据等多渠道获取车辆相关数据,并结合云计算、人工智能和数据分析等技术进行处理和分析。

2.1 数据样本总体情况

本文的研究对象为某B级纯电动轿车的实际充电数据,该车型主要分布在吉林省、广东省、四川省、江苏省、北京市等地区,其中吉林省分布车辆最多,已超过20 000辆,其他地区车辆样本均超过5 000辆,以私人用户和移动出行用户为主。由于移动出行用户数量较多,单日行驶里程普遍超过300 km,且能获取的信号周期为500 ms,所以该车型行驶和充电数据十分丰富。

该车型电池荷电状态(State Of Charge, SOC)与能量基本呈线性关系,且本文主要统计电池能量衰减相对比例,故定义电池等效总能量 E_{total} 作为电池总能量近似计算的统计概念:

$$E_{total} = \frac{P_{bat} \times 0.5/3\ 600}{S_{start} - S_{end}} \times 100 \quad (1)$$

式中: P_{bat} 为电池功率, S_{start} 、 S_{end} 分别为充电开始、结束时的电池荷电状态。

从充电相关数据中提取和处理充电地区位置、环境温度、充电始末时间、车辆里程、充电始末SOC、充电量,如表1所示。

电池能量与温度、充放电循环工况等因素有关,本文采用地域差异在一定程度上代替温度差异。以往电池能量测试中的放电工况为几种特定工况,或者聚类获取的综合工况。而用户驾驶工况复杂多变,利用大数据可将各类用户行驶习惯和各类驾驶场景及交通环境、气候环境等包含在内。本文主要分析电池能量随行驶里程变化的规律并开展预测,为了体现方法的通用性,更贴近

用户实际使用特性,选取吉林、广东、四川、江苏4个地区的数据进行分析,充分保证数据的规模和多样性。

表1 总数据提取样本示例

序号	地区	温度/°C	里程/km	SOC变化/%	等效总能量/kW·h
1	山东	14.1	29 135	31.1	55.1
2	天津	9.1	655	32.5	46.4
3	上海	15.5	15 377	56.0	53.4
4	江苏	15.0	15 545	50.5	55.5
5	江苏	12.6	15 699	57.7	55.3
6	吉林	-9.9	8 408	32.1	54.3

2.2 数据样本分析

电池总能量的分布无规律性且较为分散,需要对原始数据进行预处理,包括数据清洗和数据划分等^[7]。本文对预处理后的数据分别采用滑动平均法和区间平均法实现对地区数据和总体数据的处理,如表2所示,同时增加样本车辆的总行驶里程,研究不同行驶里程与电池能量的关系。

表2 样本数据属性

里程/ $\times 10^4$ km	数量/辆				
	总计	吉林省	四川省	江苏省	广东省
>1~2	81 956	23 134	8 108	4 669	3 711
>2~3	61 657	11 265	5 884	4 316	3 040
>3~4	46 004	8 485	4 417	3 424	2 498
>4~5	27 571	4 336	2 073	2 620	2 275
>5~6	13 773	1 784	902	1 810	1 855
>6~7	6 296	577	214	701	1 369
>7~8	2 920	288	118	282	703
>8~9	1 659	153	112	62	351
>9~10	1 046	87	103	45	229
>10~11	530	54	66	23	41
>11~12	325	23	53	13	28
>12~13	199	13	25	1	25
>13~14	81	4	23	1	15

由表2可知,该车型里程数据主要集中在 $1 \times 10^4 \sim 5 \times 10^4$ km区间,本文结合区间平均和滑动平均两种算法的特点,分别对数据进行处理。区间平均是将整体范围划分为若干个小区间,统计每个小区间的数据量;滑动平均是利用一定长度的窗口对数据进行平均计算,然后移动窗口继续计算下一段数据的平均值,其优点是能消除随机波动。

2.2.1 整体数据处理

整体数据的数据量较大,可以采用区间平均算法进

行拟合,以200 km作为区间长度,以50 000 km为最后一个区间的终点。

最后得到数据整体分布情况和使用三次多项式拟合得到的里程-总能量关系,如图1所示。

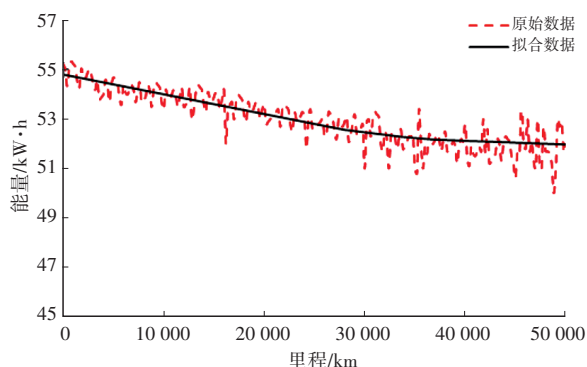


图1 区间平均拟合里程-总能量关系示意

2.2.2 地区数据处理

对于地区数据,采用滑动平均算法处理。针对总能量数据 E_1, E_2, \dots, E_n ,需计算窗口大小为 w 时的滑动平均值。则第 i 个滑动平均值 M_{Ai} 的计算公式为:

$$M_{Ai} = \frac{1}{w} \sum_{j=i-w+1}^i E_j \quad (2)$$

式中: j 为当前时刻。

因此,对能耗数据进行滑动平均处理后,得到的新数据序列为 $M_{A1}, M_{A2}, \dots, M_{An}$ 。最后得到数据量充足的4个地区数据分布以及里程与能量的拟合曲线,以江苏省为例,拟合曲线如图2所示。

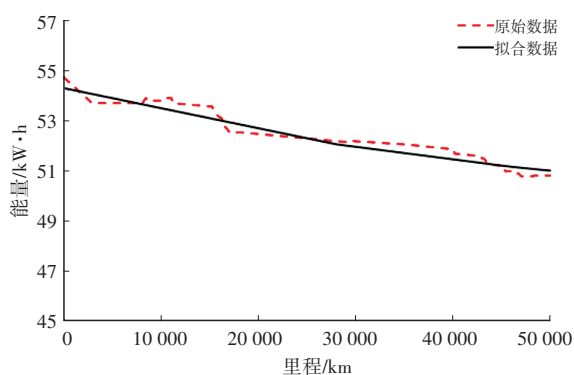


图2 滑动平均拟合里程-总能量关系示意

3 电池总能量衰减预测与分析

3.1 基于支持向量回归的数据拟合分析

SVR利用核函数对数据进行映射,并在高维空间中寻找最优拟合超平面,从而实现非线性回归。相对于多项式回归算法,SVR在电池数据处理过程中鲁棒性更好、可解释性强,可避免过拟合问题,且可以通过交叉验证来提高电池总能量衰减拟合和数据预测精度。本文

对样本数据的能量衰减拟合步骤如下:

a. 数据预处理。将数据集划分为训练集和测试集,其中训练集用于训练模型,测试集用于评估模型的预测性能。

b. 特征工程处理。本文问题只有1个特征,即车辆里程,因此不需要进行特征工程处理。

c. 模型训练。使用SVR模型拟合训练集数据。设训练集样本数量为 n ,第 i 个样本的特征为 x_i 、标签为 y_i ,则SVR模型的目标是最小化带有正则化项的损失函数:

$$\begin{aligned} \min_{w,b,\xi_i,\xi_i^*} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{subject to:} & \begin{cases} y_i - \mathbf{w}^T \varphi(x_i) - b \leq \delta + \xi_i \\ \mathbf{w}^T \varphi(x_i) + b - y_i \leq \delta + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (3)$$

式中: \mathbf{w} 为权重向量; b 为偏置项; ξ_i, ξ_i^* 为松弛变量; C 为正则化参数; δ 为间隔参数; $\varphi(x_i)$ 为核函数,将输入特征映射到高维空间,目标是最小化误差并限制支持向量到达边界,以提高泛化能力。

d. 网络搜索超参数选择。假设有 k 个超参数 $\theta_1, \theta_2, \dots, \theta_k$,每个超参数有 n_i 个可能的取值,即 $\theta_i \in \{\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,n_i}\}$,则共有 $n_1 \times n_2 \times \dots \times n_k$ 种超参数组合 $(\theta_{1,j_1}, \theta_{2,j_2}, \dots, \theta_{k,j_k})$,其中 $j_i \in 1, 2, \dots, n_i$ 。对于每个超参数组合 $(\theta_{1,j_1}, \theta_{2,j_2}, \dots, \theta_{k,j_k})$,假设使用 m 折交叉验证,则可得到 m 个平均误差 $E_{j_1, j_2, \dots, j_k}^{(1)}, E_{j_1, j_2, \dots, j_k}^{(2)}, \dots, E_{j_1, j_2, \dots, j_k}^{(m)}$,其中 $E_{j_1, j_2, \dots, j_k}^{(i)}$ 为第 i 折交叉验证的平均误差。最终,可以计算所有平均误差的平均值 E_{j_1, j_2, \dots, j_k} 作为超参数组合 $(\theta_{1,j_1}, \theta_{2,j_2}, \dots, \theta_{k,j_k})$ 的性能指标:

$$E_{j_1, j_2, \dots, j_k} = \frac{1}{m} \sum_{i=1}^m E_{j_1, j_2, \dots, j_k}^{(i)} \quad (4)$$

e. 模型预测。使用训练好的SVR模型在测试集上进行预测,假设测试集包含 Q 个样本,第 j 个样本的特征为 x_j ,则模型的预测输出为:

$$\hat{y}_j = \mathbf{w}^T \varphi(x_j) + b \quad (5)$$

f. 模型评估。采用均方误差(Mean-Square Error, MSE) M 和决定系数 R^2 对模型的预测性能进行评估:

$$M = \frac{1}{Q} \sum_{j=1}^Q (\hat{y}_j - y_j)^2 \quad (6)$$

$$R^2 = 1 - \frac{\sum_{j=1}^Q (\hat{y}_j - y_j)^2}{\sum_{j=1}^Q (y_j - \bar{y})^2} \quad (7)$$

式中: y_j 为真实标签, \bar{y} 为标签的均值。

由于行驶至40 000 km的车辆数量较多,数据丰富,满足拟合条件,最后利用SVR模型对总体数据和区域

数据拟合(0~40 000 km),并通过拟合的数据对已行驶40 000~50 000 km里程的车辆进行预测,表3所示为数据拟合结果,图3所示为对总体数据拟合示意,车辆行驶50 000 km时能量总体衰减3.01%。

表3 SVR对总体数据拟合

总里程/km	等效总能量均值/kW·h
10 000	51.741 8
20 000	51.335 8
30 000	50.778 5
40 000	50.313 3
50 000	50.183 5

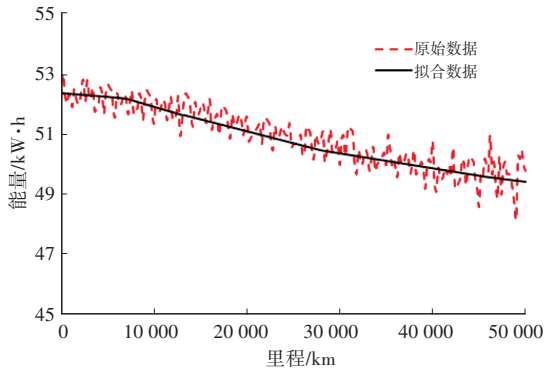


图3 SVR对总体数据拟合示意

考虑不同地区气候因素对电池能量衰减的影响,按照区域对数据进行划分,对吉林省、江苏省、广东省3个地区进行数据拟合和预测,衰减情况与总体统计结果基本一致,分别衰减2.64%、3.42%、2.01%。

3.2 基于长短时记忆神经网络的数据分析

LSTM能够捕捉电池能量变化的长期模式,并能对时序数据进行有效建模,从而实现准确的预测。本文采用该算法进行数据拟合和预测,验证前文分析趋势的准确性。

3.2.1 数据准备

首先对输入和输出数据进行处理,将其转化为适合LSTM模型训练的格式。将输入数据进行最大最小值归一化,将里程、环境温度分别归一化为 x_o 、 x_T ,输出数据电池充入能量 E 归一化为 y :

$$x_o = \frac{O - O_{\min}}{O_{\max} - O_{\min}} \quad (8)$$

$$x_T = \frac{T - T_{\min}}{T_{\max} - T_{\min}} \quad (9)$$

$$y = \frac{E_{\text{total}} - E_{\min}}{E_{\max} - E_{\min}} \quad (10)$$

式中: O 为里程, O_{\min} 、 O_{\max} 分别为里程的最小值、最大值, T 为电池温度, T_{\min} 、 T_{\max} 分别为电池温度的最小值、最大值, E_{\min} 、 E_{\max} 分别为电池等效总能量的最小值、最大值。

3.2.2 模型建立

LSTM模型的建立过程可以简化表示为:

$$\begin{cases} h_t = LSTM(x_t, h_{t-1}) \\ h_0 = \text{zeros}(128) \\ y = \text{Dense}(\text{Dropout}(\text{ReLU}(h_t)), W) + b \end{cases} \quad (11)$$

式中: h_t 为LSTM层在时间步 t 的输出, x_t 为时间步 t 的输入, W 为全连接层的权重, $LSTM$ 为基于长短时记忆神经网络计算, zeros 为创建由零组成的数组, Dense 为神经网络中的全连接层函数, Dropout 为神经网络正则化方法, ReLU 为线性整流函数。

在预测过程中,将输入序列 x_1, x_2, \dots, x_K 输入模型,得到输出序列 y_1, y_2, \dots, y_K ,其中 K 为序列长度。

3.2.3 模型训练

首先定义损失函数:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

式中: θ 为模型参数。

计算模型参数的梯度并通过训练进行更新:

$$\nabla_{\theta} L(\theta) = (1/S_{\text{batch}}) \cdot \sum_{i=1}^{S_{\text{batch}}} (y_i - \hat{y}_i) \cdot \frac{\partial f(x; \theta)}{\partial \theta} \quad (13)$$

式中: S_{batch} 为批量化数据。

根据梯度更新模型参数:

$$\theta \leftarrow \theta - \alpha \nabla_{\theta} L(\theta) \quad (14)$$

式中: α 为学习率。

3.2.4 模型预测

对于给定的输入数据,使用已训练好的LSTM模型进行预测,得到输出为:

$$\hat{y}_{\text{pred}} = f(x_{\text{pred}}; \theta) \quad (15)$$

式中: f 为训练好的LSTM模型函数, x_{pred} 为预测值。

预测结果如图4所示,可以发现随里程增加,电池能量呈现下降趋势,总体下降幅度与前文分析结果基本一致。同时,引入环境温度作为另一个变量,将温度按照0℃、20℃分段,可见环境温度越低,电池等效总能量越少。

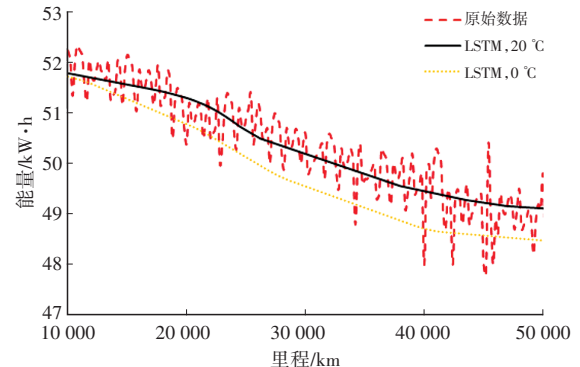


图4 基于LSTM的能量衰减-环境温度预测

4 试验对比

选取5台该车型试验车辆进行电池充电能量测试统计,各试验车辆已行驶里程分别为48 107 km、47 794 km、49 256 km、49 473 km、48 290 km,已行驶里程均分布在40 000~50 000 km范围内,且接近50 000 km,均为在吉林地区进行适应性测试的车辆,车辆运行工况基本为市区综合行驶工况。

记录2023年5月的30天内充电使用数据,每次使用至电池SOC低于15%后再进行充电,且每次充满后,利用式(1)计算等效总能量并按车辆取平均值,试验结果如表4所示,5台试验车辆等效总能量均值为50.35 kW·h,与拟合预测结果50.18 kW·h十分接近,误差为0.33%。

表4 试验数据记录

车辆序号	充电次数/次	等效总能量均值/kW·h
车辆1	12	50.24
车辆2	12	50.16
车辆3	10	50.60
车辆4	11	50.41
车辆5	14	50.33

5 结束语

本文通过某车型动力电池充电大数据,采用区间平均法和SVR模型对里程-总能量关系进行拟合,实现了对电池总能量衰减的高精度预测,并通过LSTM模型进行了数据分析趋势对比,实车测试结果证明了该结果准确、可靠。

该方法主要可应用于以下两个方面:企业监控电池性能变化,其结果比电池寿命测试更具普适性;车辆剩余续驶里程预估结果的修正,为整车能量管理提供准确的数据支持。

参 考 文 献

[1] 王思森,张英杰,王芳,等.基于大数据的动力电池健康度估算方法研究[J].传动技术,2021,35(4):8-12+23.

WANG S M, ZHANG Y J, WANG F, et al. Research on Estimation Method of Power Battery Health Based on Big Data[J]. Drive System Technique, 2021, 35(4): 8-12+23.

[2] 王震坡,孙逢春,刘鹏.电动汽车原理与应用技术[M].北京:机械工业出版社,2014:109-153.

WANG Z P, SUN F C, LIU P. Basic Principles and Application Technology for Electric Vehicles[M]. Beijing: China Machine Press, 2014: 109-153.

[3] 纪常伟,潘帅,汪硕峰,等.动力锂离子电池老化速率影响因素的实验研究[J].北京工业大学学报,2020,46(11):1272-1282.

JI C W, PAN S, WANG S F, et al. Experimental Study on Effect Factors of Aging Rate for Power Lithium-Ion Batteries [J]. Journal of Beijing University of Technology, 2020, 46 (11): 1272-1282.

[4] 韩雪冰.车用锂离子电池机理模型与状态估计研究[D].北京:清华大学,2014.

HAN X B. Study on Li-Ion Battery Mechanism Model and State Estimation for Electric Vehicles[D]. Beijing: Tsinghua University, 2014.

[5] YANG R X, XIONG R, HE H W, et al. A Novel Method on Estimating the Degradation and State of Charge of Lithium Ion Batteries Used for Electric Vehicles[J]. Applied Energy, 2017, 207: 336-345.

[6] ZHANG Y Z, XIONG R, HE H W, et al. Long Short Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium Ion Batteries[J]. IEEE Transactions on Vehicular Technology, 2018, 26(7): 5695-5705.

[7] 张凡,邢子轩,武明虎,等.数据驱动的车载动力电池使用行为分析[J].汽车技术,2023(3):49-55.

ZHANG F, XING Z X, WU M H, et al. Research on the Use Behaviors of Data-Driven Vehicle Power Battery[J]. Automobile Technology, 2023(3): 49-55.

(责任编辑 斛 畔)

修改稿收到日期为2023年9月19日。