

# 基于数据驱动的电动汽车动力电池故障预测算法

李晓杰<sup>1</sup>, 苏振洋<sup>1</sup>, 丁技峰<sup>2</sup>

(1. 中北大学 能源与动力工程学院, 山西 太原 030051; 2. 中国北方发动机研究所(天津), 天津 300400)

**摘要:** 针对现在电动汽车电池系统实车故障预测的难题, 提出了一种利用贝叶斯算法优化长短期记忆神经网络(Long Short-Term Memory)的实车故障预测算法。首先通过皮尔逊相关系数法确定了LSTM的输入与输出特征, 有效解决了现有工程数据量大、输入输出特征过多导致深度学习模型过拟合或欠拟合的问题。然后利用贝叶斯算法优化LSTM的超参数, 重点解决了LSTM的超参数设置困难和导致误报的问题, 经过贝叶斯算法优化后确定了最终的超参数组合, 提出了建立单体电池电压预测模型来预测整车电池电压的方法, 节约了模型训练时间, 接着对单体选择的不同方式进行了测试, 确定将每帧时间下所有单体电池电压的中位数作为新的单体电池电压来进行模型训练, 进而建立了单体电池电压的预测模型。经过实车数据验证, 相对于LSTM的整车电池电压预测模型, 基于LSTM的单体电池电压预测模型的RMSE、MAE以及MRE分别下降了61.59%, 61.31%和60.94%, 有效提高了实车电池电压预测精度, 最终验证了所建立的电压预测模型的优越性、可靠性以及鲁棒性。

**关键词:** 电动汽车; 故障预测; 皮尔逊相关系数; 贝叶斯算法; 单体预测模型

**中图分类号:** TM912 **文献标识码:** A **doi:** 10.62756/jnuc.issn.1673-3193.2024.11.0018

**引用格式:** 李晓杰, 苏振洋, 丁技峰. 基于数据驱动的电动汽车动力电池故障预测算法[J]. 中北大学学报(自然科学版), 2025, 46(3): 293-305.

LI Xiaojie, SU Zhenyang, Ding Jifeng. Real vehicle fault prediction algorithm of electric vehicle based on data drive[J]. Journal of North University of China(Natural Science Edition), 2025, 46(3): 293-305.

## Real Vehicle Fault Prediction Algorithm of Electric Vehicle Based on Data Drive

LI Xiaojie<sup>1</sup>, SU Zhenyang<sup>1</sup>, Ding Jifeng<sup>2</sup>

(1. School of Energy and Power Engineering, North University of China, Taiyuan 030051, China;

2. China North Engine Research Institute(Tianjin), Tianjin 300400, China)

**Abstract:** To solve the problem of real vehicle fault prediction of electric vehicle battery system, a new algorithm for predicting real vehicle fault using Bayesian algorithm to optimize Long Short-Term Memory neural network (LSTM) was proposed. Firstly, the input and output features of LSTM were determined by Pearson correlation coefficient method, which effectively solved the problem of overfitting or underfitting of deep learning model caused by large amount of existing engineering data and excessive input and output features. Then, Bayes algorithm was used to optimize the hyperparameters of LSTM, it focused on solving the problem of difficult setting of LSTM hyperparameters, which led to false positives. The final combination of hyperparameters was determined after optimization by Bayes algorithm, and the

收稿日期: 2024-11-30

基金项目: 山西省重点实验室建设项目(GDZBKKX-10)

作者简介: 李晓杰(1979-), 男, 高级实验师, 博士, 主要从事动力电池BMS设计与开发方面的研究。E-mail: lixiaojie16@nuc.edu.cn.

method of establishing a single battery voltage prediction model to predict the vehicle battery voltage was proposed, which saved the model training time. The median of all single battery voltages under each frame time was determined as a new single battery voltage to train the model, and then the single battery voltage prediction model was established. Through the verification of real vehicle data, RMSE, MAE and MRE of the single unit voltage prediction model based on LSTM decreased by 61.59%, 61.31% and 60.94%, respectively, compared with the vehicle voltage prediction model based on LSTM, effectively improving the accuracy of real vehicle voltage prediction. Finally, the superiority, reliability and robustness of the proposed voltage prediction model were verified.

**Key words:** electric vehicle; fault prediction; pearson correlation coefficient; Bayesian algorithm; cell prediction model

## 0 引言

锂离子电池因其高能量密度和超长的使用寿命,在交通运输领域中具有重要的地位<sup>[1-2]</sup>,随着新能源汽车的不断发展,其安全性能也备受关注<sup>[3]</sup>。锂离子电池作为电池储能系统和电动汽车的核心部件,其故障约占车辆故障的60%,且具有故障蔓延快、逃逸时间短、易引发火灾等特点<sup>[4]</sup>,提升锂离子电池的安全性迫在眉睫,准确预测和诊断动力电池故障是提升电动汽车的动力性、安全性以及可靠性的重要保障措施<sup>[5]</sup>。为此国内外学者提出了许多锂离子电池故障诊断算法,根据研究锂电池故障数据可以分为故障注入实验、软件仿真、以及实车数据三类,前两类数据难以模拟实车环境下的运行工况,导致许多故障诊断算法很难在实车得到验证<sup>[6]</sup>。

现有的锂电池故障诊断算法可以分为基于知识、基于模型以及基于数据驱动的故障诊断算法,基于知识的故障诊断算法大致可分为故障决策树法<sup>[7]</sup>、专家系统法<sup>[8-9]</sup>以及模糊逻辑法<sup>[10]</sup>等,该类方法利用已有的领域知识和经验进行故障诊断,具有较高的准确性,但对于未知的或复杂的故障模式,可能无法提供准确的诊断结果。基于模型的故障诊断算法通过建立锂电池的等效电路模型、热模型、电热耦合模型以及电化学模型等来预测电池的外特性参数<sup>[11-13]</sup>,通过实际值与预测值产生的残差来进行故障诊断,但其受限于精确的电池模型,且模型的复杂性和精度会对算法的实时性和准确性产生影响。基于数据驱动的故障诊断算法可以从大量的实际数据中学习和挖掘电池的故障模式,从而能够适应各种复杂情况,由于全国各地建立了许多新能源汽车大数据平台,因此基于数据驱动的电池故障诊断算法得到

了应用广泛。

深度学习由于其良好的性能而在电动汽车故障诊断领域得到了广泛应用,虞婧<sup>[14]</sup>将BP神经网络与离群点算法进行结合,有效识别了电动汽车充电过程中的故障特征;Hong等<sup>[15]</sup>考虑天气和驾驶员行为对电池系统性能的影响,提出了一种利用长短期记忆神经网络(LSTM)进行电池系统多步电压预测的电池故障诊断算法,有效识别了电动汽车充电与行驶过程中的电压异常;张朝龙等<sup>[16]</sup>提取锂离子电池组充电状态下各单体电压的信息熵和平均温度,基于粒子群算法优化长短期记忆神经网络(PSO-LSTM)构建了锂电池组健康状态估计模型,在锂电池组以及锂电池单体的老化数据中都得到了很好的验证。王天城<sup>[17]</sup>通过小波分解重构提取了事故车的电池故障特征,利用LSTM实现了热失控车辆数据的准确识别;胡杰等<sup>[18]</sup>将随机森林(RF)和SHAP相结合来筛选特征,将电池安全风险预测问题进行解耦,然后分别采用XGBoost模型以及RF模型进行了传统机器学习和时间序列的预测,最后进行了Stacking集成,提高了电池安全风险预测精度;Li等<sup>[19]</sup>将长短期记忆神经网络(LSTM)与等效电路模型(ECM)相结合,并采用改进的自适应增强方法提高了故障诊断精度,并采用预判模型来减少计算时间,提高诊断可靠性,实现了对电池单体潜在故障的准确诊断和对热失控电池单体的精确定位。

LSTM(Long Short-Term Memory)是一种循环神经网络,在处理时间序列数据时具有很好的表现。但是这种模型有大量的超参数,为了使LSTM模型能够更好地拟合数据并提高其性能,本文利用贝叶斯优化算法优化长短期记忆神经网络,确定了最终的超参数组合,提出了通过建立单体电压模型来预

测整车电压的方法,减少了离线模型的训练时间,最终通过实车测试结果验证了所建立的单体电压预测模型的优越性、可靠性以及鲁棒性。

### 1 数据预处理

#### 1.1 实车数据

为了实现新能源汽车故障诊断及安全风险预测,我国于2016年在北京建立了新能源汽车国家监测与管理平台,该平台的数据类型主要包括静态数据和动态数据两种。静态数据包含了车辆的基本信息,如生产厂家、车辆类型、车牌VIN码、车牌号以及销售区域等。动态数据分为实时在线运行数据和历史存储数据,研究人员一般利用历史存储数据开发故障诊断算法,经过不断更新算法,最后将其应用于新能源汽车大数据平台来实时监控新能源汽车的安全风险。《电动汽车远程服务与管理系统技术规范》(GB/T 32960)规定采集的系统数据如表1所示。

表1 数据采集来源及名称

Tab. 1 The source and name of the data collection

数据采集来源	数据项
动力电池系统	总电压、总电流、SOC、单体电压、单体温度等
电机驱动系统	电机电压、电流、转速、扭矩、温度等
车辆控制系统	车辆速度、挡位信息、加速踏板行程、GPS位置等
其他部分	空调信息和轮胎胎压状态等其他信息数据

本文所使用的车辆数据如表2所示,1~4号、18号车的车载电池为三元锂电池,各车有96块单体电池,车型为电动公交车,额定工作电压为345.6V。5~17号车的车载电池为磷酸铁锂电池,各车有16块单体电池,为2s8p形式,车型为农用工程机械车辆,额定工作电压为51.2V。1~16号车的数据采样时间段均为2022/1/1 00:00:00—2023/8/31 23:59:59,17号车为故障车辆,数据采样时间段为2022/1/1 00:00:00—2023/3/18 16:49:18,18号车为热失控车辆,数据采样时间段为2020/1/1 00:00:00—2020/12/1 22:22:57。

表2 车辆数据统计表

Tab. 2 Vehicle statistics sheet

电池类型	车辆编号	数据采样时间段
三元锂电池	1~4	2022/1/100:00:00—2023/8/31 23:59:59
磷酸铁锂电池	5~16	2021/1/1 00:00:00—2023/8/31 23:59:59
磷酸铁锂电池	17	2022/1/1 00:00:00—2023/3/18 16:49:18
三元锂电池	18	2020/1/1 00:00:00—2020/12/1 22:22:57

#### 1.2 数据预处理

在数据传输过程中,原始数据可能会发生一些不可避免的错误,包括数据重复、数据缺失、格式不统一等,并且原始数据特征较多,数据维度较大,直接进行数据分析与建模会消耗大量的计算资源和时间成本,因此进行一些必要的预处理,可以保证数据质量,减少冗余数据,加快数据分析速度,提高计算效率。

本文采用以下数据预处理方式:

1)对大数据平台后台下载的原始数据进行初步清洗,删掉缺失值以及重复值,利用箱型图去除异常值。

2)对清洗后的数据按照时间序列进行排序。

3)分析数据特征,筛选与单体电压相关的特征,剔除无关特征,降低数据处理维度。

##### 1.2.1 特征选择

实车数据工况复杂,受到驾驶员行为习惯、行驶工况以及环境因素等多方面的影响,单体电池电压波动较大,本文采用LSTM网络的时间序列功能来预测单体电压,为了确定输入特征,利用皮尔逊相关系数来确定各种数据与单体电压相关系数的大小,选择了总电流、总电压、SOC、单体电压、行驶里程、绝缘电阻、充放电状态、探针最高温度这些特征参数,利用1号车的数据计算相关系数并绘制了热力图,相关系数的计算公式为

$$r_{x,y} = \frac{cov_{x,y}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

式中:  $r_{x,y}$  为  $x$  和  $y$  的相关系数的值;  $cov_{x,y}$  为  $x$  和  $y$  的协方差;  $n$  为样本数量的大小;  $\bar{x}$  为  $x$  的均值;  $\bar{y}$  为  $y$  的均值。

皮尔逊相关系数关联度<sup>[20]</sup>如表3所示。

表3 皮尔逊相关系数关联度标准

Tab. 3 Pearson correlation coefficient correlation degree standard

相关性	相关系数绝对值
无相关性	0~0.1
弱相关性	0.1~0.5
较强相关性	0.5~0.8
强相关性	0.8~1.0

一般认为相关系数绝对值大于0.8具有强相关性,由图1可以看出,单体电压与总电压、SOC的相

关系数均大于0.9,因此本文选择总电压、SOC以及单体电压作为输入特征,输出特征为时间序列下的单体电压。

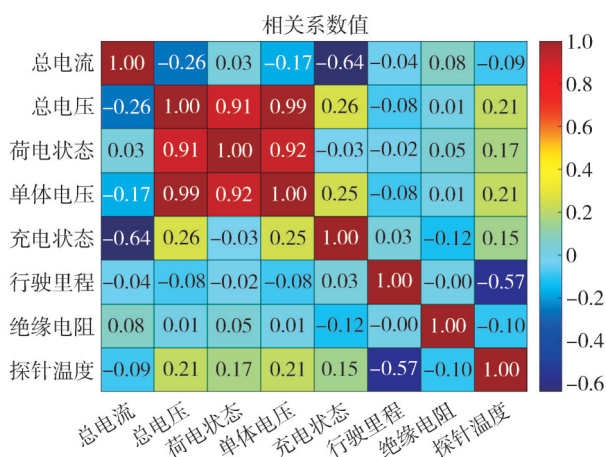


图1 1号车特征相关系数

Fig. 1 Feature correlation coefficient for the vehicle no. 1

### 1.2.2 单体选择

由于从大数据平台下载的数据庞大,优化整车数据时LSTM进行数据转换的步骤运行内存量过大,会存在内存溢出现象,并且此时会出现输入量以及输出量过多的情况,导致运行缓慢,浪费了很大的时间成本,因此本文采用优化一个单体来预测整车所有单体电压,在选择需要优化的目标单体时,有两种方式:

1) 选择每帧数据下所有单体电压的平均数作为一个新的单体,以此单体建立的模型为基础来预测整车电压。

2) 选择每帧数据下所有单体电压的中位数作为一个新的单体,以此单体建立的模型为基础来预测整车电压。

后续将会对这两种单体选择方式进行测试对比,进而确定最终的目标单体选择。

## 2 算法介绍

### 2.1 长短期记忆神经网络

相对于传统的循环神经网络(RNN),长短期记忆神经网络(LSTM)由于其计算复杂度相对较低,可以避免梯度消失或爆炸,以及更好的长期记忆<sup>[21]</sup>等优点而被广泛应用于需要处理长序列数据的场景中。LSTM可以同时拥有两种类型的记忆:短期记忆(short-term memory)和长期记忆(long-term memory)。短期记忆由当前输入决定,而长期记忆则由历史输入决定。LSTM通过单元来处理信息,

每个LSTM单元有3个门控单元<sup>[22]</sup>:输入门、输出门和遗忘门。这些门被用来控制记忆细胞中的信息流,并且允许LSTM保留、删除或添加新数据到记忆细胞中。LSTM的结构如图2所示,其中涉及的门控机制操作,有助于LSTM在迭代过程中保存重要的信息以及从原始输入中提取出最优信息。

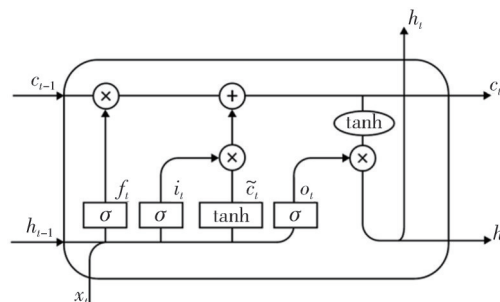


图2 LSTM网络结构图

Fig. 2 The network structure diagram of LSTM

1) 遗忘门:从单元状态中删除信息

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f). \quad (2)$$

2) 输入门:向单元状态添加信息

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i), \quad (3)$$

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c). \quad (4)$$

更新单元状态:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t. \quad (5)$$

3) 输出门:从单元状态中选择有用的信息,并将其作为输出进行传输,

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o), \quad (6)$$

$$h_t = o_t \tanh(c_t), \quad (7)$$

式中: $w$ 为权重矩阵; $b$ 为偏置项; $f$ 为遗忘门; $i$ 为输入门; $o$ 为输出门; $\sigma$ 为sigmoid函数; $\tanh$ 为双曲正切函数。

### 2.2 贝叶斯优化算法

贝叶斯优化算法是一种基于贝叶斯理论的全局优化方法,它在优化超参数时充分利用了历史信息,从而能够在较少评估次数的代价下取得最优效果<sup>[23]</sup>,其通过不断地估计超参数的后验分布来寻找最佳超参数组合,并使用高斯过程模型对超参数空间进行建模以明确超参数之间的相关性。贝叶斯优化算法在超参数搜索空间比较大和比较复杂时的效果特别突出,通过贝叶斯优化算法可以加快搜索速度,减少试错次数,从而提高优化效率和准确度。

贝叶斯优化算法基于贝叶斯理论<sup>[24]</sup>。

$$p(f|D_{1:t}) = \frac{p(D_{1:t}|f)p(f)}{p(D_{1:t})}, \quad (8)$$

式中:  $f$  为未知的目标函数;  $D_{1:t}$  为已经观测到的超参数以及对观测值的集合;  $p(f)$  为目标函数的先验概率分布;  $p(D_{1:t})$  为目标函数的边际似然分布;  $p(f|D_{1:t})$  为目标函数的后验概率分布。

贝叶斯优化算法包括概率代理模型以及采集函数两个核心, 概率代理模型用来代理未知的目标函数, 通过不断修正先验概率使得代理模型更加准确, 采集函数依据后验概率, 通过最大化采集函数来选择下一个最可能出现最优解的点, 同时保证损失函数值最小<sup>[25]</sup>。

### 2.2.1 概率代理模型

概率代理模型用来代理未知的目标函数, 按照模型参数是否固定可以分为参数模型和非参数模型, 参数模型有线性模型、广义线性模型以及贝塔-伯努利模型等, 非参数模型有随机森林、深度神经网络以及高斯过程等<sup>[24]</sup>, 非参数模型由于其较好的预测性能以及不容易发生过拟合而被广泛应用, 其中高斯过程(Gaussian Processes, GPs)可以生成多维高斯分布, 理论上可代理任何目标函数, 并广泛应用于黑箱函数的领域中。

高斯过程回归假设了高斯先验分布来推断未知的目标函数,  $f: R^D \rightarrow R$ , 则目标函数可由均值函数  $\mu(x)$  与协方差函数  $k(x, x')$  唯一确定<sup>[26]</sup>

$$y = f(x) \sim G_p(\mu(x), k(x, x')), \quad (9)$$

$$\mu(x) = E(f(x)), \quad (10)$$

$$k(x, x') = E((f(x) - \mu(x))(f(x') - \mu(x))). \quad (11)$$

首先假设一个均值为 0 的先验分布

$$p(f|X, \theta) = N(0, \Sigma), \quad (12)$$

式中:  $X$  为训练集样本;  $f$  为未知的目标函数组成的函数值集合;  $\Sigma$  为  $k(x, x')$  组成的协方差矩阵;  $\theta$  为超参数。

假设观测噪声  $\epsilon$  满足独立同分布的高斯分布, 可得似然分布

$$p(y|f) = N(f, \sigma^2 I), \quad (13)$$

$$y = f(x) + \epsilon. \quad (14)$$

根据以上公式, 可得边际似然分布

$$p(y|X, \theta) = N(0, \Sigma + \sigma^2 I). \quad (15)$$

存在如下的联合分布

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma + \sigma^2 I & K_* \\ K_*^T & K_{**} \end{bmatrix}\right), \quad (16)$$

$$K_*^T = (k(x_1, X_*), k(x_2, X_*), \dots, k(x_t, X_*)), \quad (17)$$

$$K_{**} = k(X_*, X_*), \quad (18)$$

式中:  $f_*$  为函数的预测值;  $X_*$  为预测输入。

可得预测分布

$$p(f_*|X, y, X_*) = N(\langle f_* \rangle, cov(f_*)), \quad (19)$$

$$\langle f_* \rangle = K_*^T [\Sigma + \sigma^2 I]^{-1} y, \quad (20)$$

$$cov(f_*) = K_{**} - K_*^T [\Sigma + \sigma^2 I]^{-1} K_*, \quad (21)$$

式中:  $\langle f_* \rangle$  为  $f_*$  的均值;  $cov(f_*)$  为  $f_*$  的协方差。

### 2.2.2 采集函数 EI

期望改进采集函数(Expected Improvement, EI)是一种在高斯过程代理模型基础上的优化方法, 目标是寻找最优解。EI采用观察到的最小函数值作为目标函数值, 根据高斯过程模型计算采样点的期望改进值, 用改进值作为采样点的评估指标。EI函数充分挖掘了贝叶斯优化的优势, 可以高效地找到全局最优解, EI采集函数为

$$EI(x) = \begin{cases} (\mu(x) - f(x^*))\Phi(z) + \\ \sigma(x)\Phi(z), \sigma(x) > 0, \\ 0, \sigma(x) = 0, \end{cases} \quad (22)$$

$$z = \frac{\mu(x) - f(x^*)}{\sigma(x)}, \quad (23)$$

式中:  $f(x^*)$  为函数的最优值;  $\mu(x)$  为均值;  $\sigma(x)$  为标准差;  $\Phi$  为标准正态分布概率密度函数。

## 2.3 BO-LSTM

### 2.3.1 算法参数设定

一般来说, LSTM的超参数有学习率、最小批次、神经元个数、正则化系数以及丢弃层等, 正则化系数和丢弃层这两个超参数对于LSTM神经网络的性能和泛化能力至关重要, 合理调整这两个超参数可以帮助模型达到更好的性能, 提高其对新数据的预测能力, 并减轻过拟合问题的影响, 因此, 本文对正则化系数以及丢弃层这两个超参数进行优化, 根据测试经验以及尽可能地减少离线时间训练成本的要求, 最终的算法参数设定如表 4 所示。

表4 BO-LSTM算法参数设定值  
Tab. 4 Algorithm parameter settings of BO-LSTM

参数名称(输入样本数据)	参数设定(一年的运行数据)
训练窗口	360
预测窗口	6
滑动窗口	6
隐藏层神经元个数	100
LSTM层	2
最小批次	32
最大迭代次数	50
初始学习率	$1 \times 10^{-3}$
学习率下降周期	20
学习率调整因子	0.1
丢弃层	[0.2, 0.5]
L2正则化系数	$[1 \times 10^{-5}, 1 \times 10^{-1}]$
贝叶斯优化最大迭代次数	30

### 2.3.2 BO-LSTM流程

超参数优化的目的是通过优化算法找到一组合适的超参数,使得这组超参数应用于该组数据的验证集上时效果最佳,贝叶斯优化长短期记忆神经网络的超参数优化可以表示为

$$x^* = \operatorname{argmin}_f(x), x \in X, \quad (24)$$

式中: $x$ 为需要优化的超参数, $x = \{x_1, x_2, \dots, x_n\}$ ;  $X$ 为决策空间; $f(x)$ 为目标函数; $x^*$ 表示使目标函数 $f(x)$ 取得最优解的一组超参数。

为准确评估模型的泛化能力、稳定性和对未见过数据的表现,并减少由单次数据集划分带来的随机性误差,本文对原始数据进行10折交叉验证,流程如图3所示,其步骤如下:

- 1) 将原始数据集划分为10个大小相等的子集(称为折)。
- 2) 在模型评估过程中,每次选取其中9个折作为训练集,剩下的1个折作为测试集。
- 3) 使用训练集进行LSTM模型的训练,并在测试集上进行性能评估。
- 4) 按照这种方式依次轮换选择不同的训练集和测试集,直到所有的折都被用作测试集。
- 5) 对10次评估结果取平均值,得到最终的模型性能评估。

本文以验证集中所有单体进行10折交叉验证后预测的电压的均方根误差作为适应度值,其计算公式为

$$E_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^n (T_{\text{predict}} - T_{\text{real}})^2}{n}}, \quad (25)$$

$$\text{fitness} = \frac{E_{\text{RMS}}}{10}, \quad (26)$$

式中: $T_{\text{real}}$ 为测试集的真实值; $T_{\text{predict}}$ 为测试集的预测值; $n$ 为样本个数; $\text{fitness}$ 为适应度值。

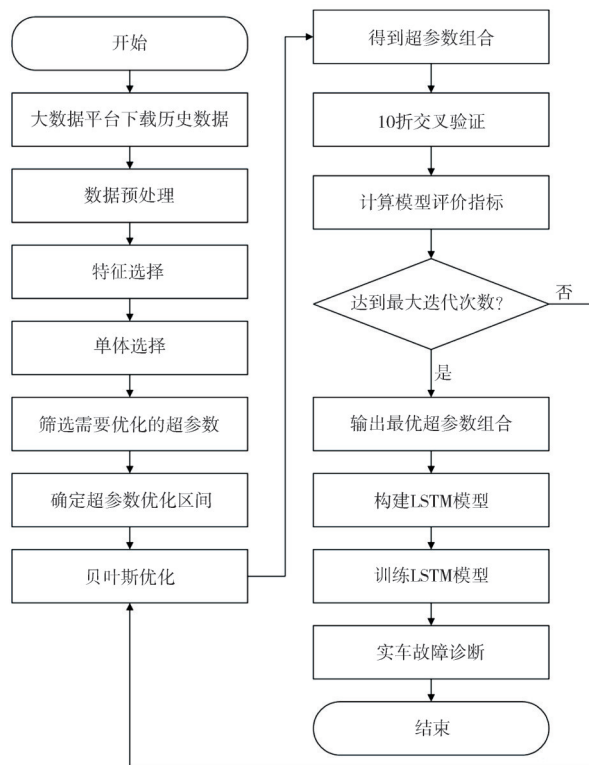


图3 BO-LSTM流程

Fig. 3 The process of BO-LSTM

## 3 实车数据验证

### 3.1 单体确定

在进行实车测试之前,首先对单体选择的不同方式进行了测试,基于控制变量法将正则化系数定为 $1 \times 10^{-3}$ ,丢弃层定为0.3,其他参数如2.3.1所示,以1~14号车2022/1/1 00:00:00—2022/12/31 23:59:59的数据为训练集,2023/1/1 00:00:00—2023/8/31 23:59:59的数据为测试集,模型评价指标为测试集所有单体的均方根误差(RMSE),平均绝对误差(MAE),平均相对误差(MRE)以及平均绝对百分比误差(MAPE)等指标,测试结果如图4所示。

$$E_{\text{MA}} = \frac{1}{n} \sum_{i=1}^n |T_{\text{predict}} - T_{\text{real}}|, \quad (27)$$

$$E_{\text{MR}} = \frac{1}{n} \sum_{i=1}^n \frac{|T_{\text{predict}} - T_{\text{real}}|}{T_{\text{real}}}, \quad (28)$$

$$E_{\text{MAP}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{T_{\text{predict}} - T_{\text{real}}}{T_{\text{real}}} \right| \times 100\%。 \quad (29)$$

由1~14号车的测试结果可以看出,以每帧时

间单体电压的中位数作为一个新的单体时各项指标较为稳定, 因此本文的单体选择对象为每帧时

间单体电压的中位数单体, 以此单体经过贝叶斯优化后建立的模型为基础来预测整车电压。

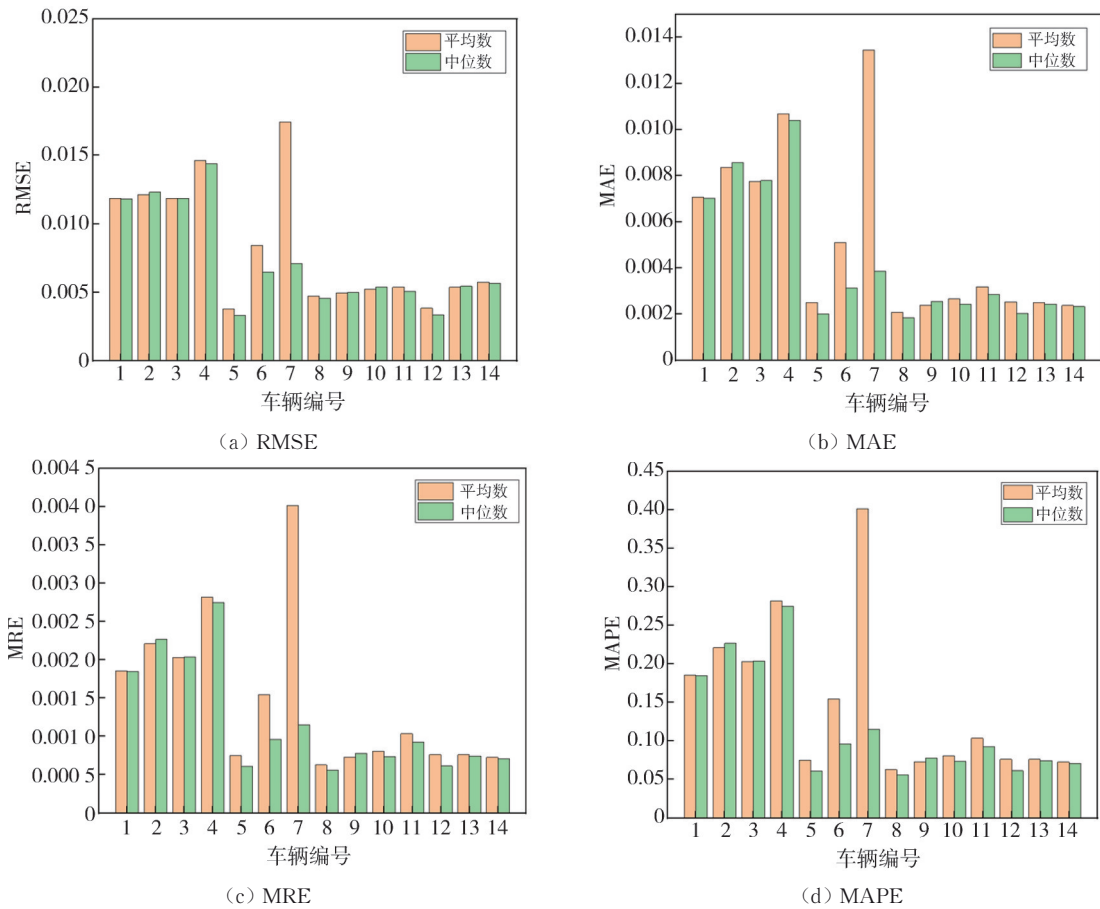


图 4 模型评价指标

Fig. 4 Model evaluation index

基于 LSTM 的整车电压预测模型输入特征为

$$\begin{bmatrix} totalv_i & SOC_i & V_{1,i} & V_{2,i} & \dots & V_{m,i} \\ totalv_{i+1} & SOC_{i+1} & V_{1,i+1} & V_{2,i+1} & \dots & V_{m,i+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ totalv_{i+tw-1} & SOC_{i+tw-1} & V_{1,i+tw-1} & V_{2,i+tw-1} & \dots & V_{m,i+tw-1} \end{bmatrix} \circ$$

输出特征为

$$\begin{bmatrix} V_{1,i+tw} & V_{2,i+tw} & \dots & V_{m,i+tw} \\ V_{1,i+tw+1} & V_{2,i+tw+1} & \dots & V_{m,i+tw+1} \\ \vdots & \vdots & \vdots & \vdots \\ V_{1,i+tw+pw-1} & V_{2,i+tw+pw-1} & \dots & V_{m,i+tw+pw-1} \end{bmatrix} \circ$$

基于 LSTM 的单体电压预测模型输入特征为

$$\begin{bmatrix} totalv_i & SOC_i & median(V_{1,i}, V_{2,i}, \dots, V_{m,i}) \\ totalv_{i+1} & SOC_{i+1} & median(V_{1,i+1}, V_{2,i+1}, \dots, V_{m,i+1}) \\ \vdots & \vdots & \vdots \\ totalv_{i+tw-1} & SOC_{i+tw-1} & median(V_{1,i+tw-1}, V_{2,i+tw-1}, \dots, V_{m,i+tw-1}) \end{bmatrix} \circ$$

输出特征为

$$\begin{bmatrix} median(V_{1,i+tw}, V_{2,i+tw}, \dots, V_{m,i+tw}) \\ median(V_{1,i+tw+1}, V_{2,i+tw+1}, \dots, V_{m,i+tw+1}) \\ \vdots \\ median(V_{1,i+tw+pw-1}, V_{2,i+tw+pw-1}, \dots, V_{m,i+tw+pw-1}) \end{bmatrix},$$

其中,  $totalv$  为总电压,  $V_{m,i}$  为单体  $m$  在  $i$  时刻的单体电压,  $m$  为单体个数,  $median(V_{1,i}, V_{2,i}, \dots, V_{m,i})$  为

$i$  时刻所有单体电压的中位数,  $tw$  为训练窗口,  $pw$  为预测窗口,  $sw$  为滑动窗口。

由此可以看出, 基于 LSTM 的单体电压预测模型的训练数据少于基于 LSTM 的整车电压预测模型, 可以快速建立电压预测模型并进行实车预测。

### 3.2 算法特性验证

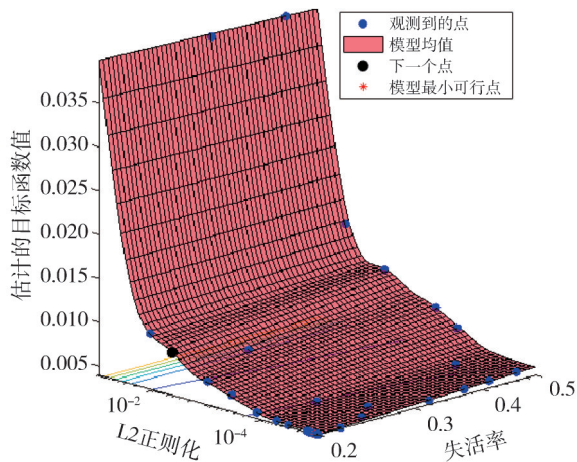
电动汽车实车电压故障预测算法应有以下特性:

- 1) 优越性: 该算法相对于其他算法预测精度更高, 模型评价指标优, 性能更好;
- 2) 可靠性: 该算法不会对正常运行车辆发出虚假报警;
- 3) 鲁棒性: 该算法在异常情况下, 仍可以有效识别电池系统故障并精确定位故障单体。

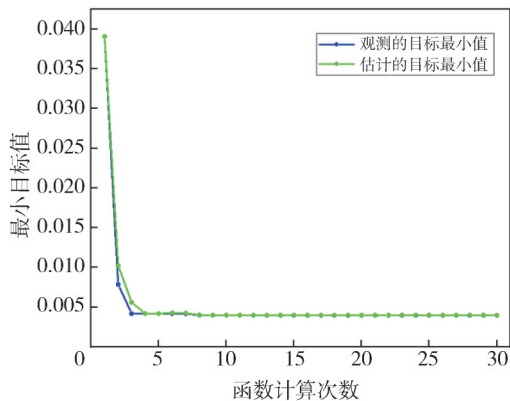
#### 3.2.1 优越性

以 15 号车 2022/1/1 00:00:00 - 2022/12/31 23:59:59 的数据为训练集, 2023/1/1 00:00:00 - 2023/1/3 23:59:59 的数据为预测集进行模型的训

练与预测,如图5(a)为15号车在进行贝叶斯优化过程的目标函数模型图,图5(b)为对应的适应度曲线图。可以看出,贝叶斯算法可以在较少的迭代次数下寻找到最优的超参数组合,在该组超参数优化后的最小目标值为0.003 979 1。



(a) 目标函数模型图



(b) 适应度曲线

图5 贝叶斯优化过程

Fig. 5 The process of Bayesian optimization

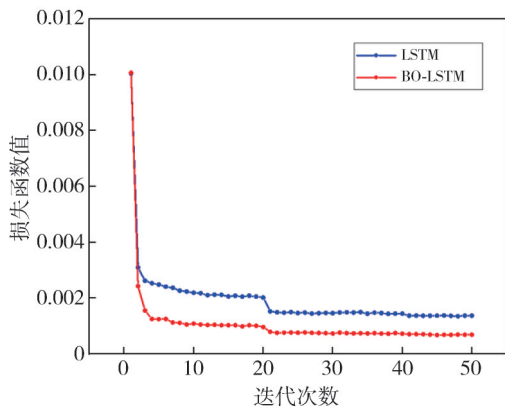


图6 损失函数值对比

Fig. 6 Comparison of loss function value

由图6可以看出,BO-LSTM可以将目标函数的损失值降到很小,第50次迭代的损失值为0.000 69,而LSTM第50次迭代的损失值为0.001 37。由图7可以看出,基于LSTM的单体电压预测模型拟合效果较好,预测误差控制在10%以内。

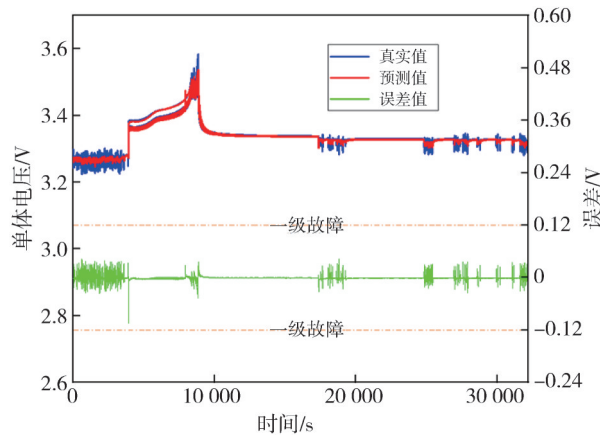


图7 15号车的BO-LSTM电压预测曲线

Fig. 7 Voltage prediction curve of BO-LSTM for the vehicle No. 15

在经过大量测试后,将最终的超参数组合定为0.2和 $1 \times 10^{-5}$ 。如表5所示,本文所提出的基于LSTM的单体电压预测模型的RMSE、MAE、MRE以及MAPE指标均优于基于LSTM的整车电压预测模型,基于LSTM的单体电压预测模型相对于LSTM的整车电压预测模型,其RMSE、MAE以及MRE指标分别下降了61.59%,61.31%和60.94%,由此验证了基于LSTM的单体电压预测模型的优越性。

表5 算法的测试数据

Tab. 5 Test data for different algorithms

评价指标	LSTM的单体预测模型	LSTM的整车预测模型
RMSE	0.003 969	0.010 334
MAE	0.002 327	0.006 015
MRE	0.000 698	0.001 787
MAPE/%	0.069 799	0.178 717

为了实现电动汽车运行过程中的电池故障诊断,在经过大量的实车数据测试之后,综合考虑正常车数据以及故障车数据的预测误差,从而设定报警阈值,如表6所示,其中, $V_{i,t}$ 为单体*i*在*t*时刻的真实电压, $\widehat{V}_{i,t}$ 为单体*i*在*t*时刻的预测电压。当故障等级达到1级时,驾驶员可根据1级故障预警次数的多少来考虑是否需要排查;当故障等级达到2级时,驾驶员需要立即停车排查故障原因;当故障等级达到3级时,车辆有热失控的风险,驾驶员需立即停车,远离车辆。

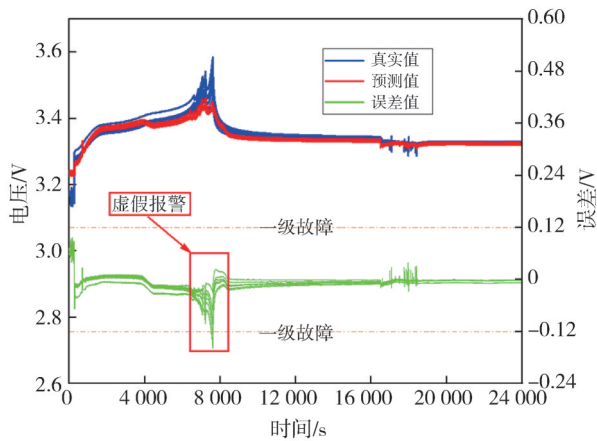
表 6 报警阈值设置

Tab. 6 Alarming levels and thresholds

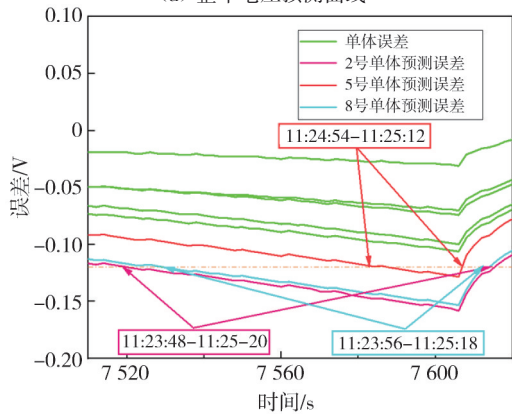
阈值设定	故障等级
$0 \leq  V_{i,t} - \widehat{V}_{i,t}  < 0.12 \text{ V}$	0
$0.12 \text{ V} \leq  V_{i,t} - \widehat{V}_{i,t}  < 0.24 \text{ V}$	1
$0.24 \text{ V} \leq  V_{i,t} - \widehat{V}_{i,t}  < 0.36 \text{ V}$	2
$ V_{i,t} - \widehat{V}_{i,t}  \geq 0.36 \text{ V}$	3

3.2.2 可靠性

以16号车2021/1/1 00:00:00—2021/12/31 23:59:59的数据为训练集,2022/1/1 00:00:00—2022/8/31 23:59:59的数据为预测集进行模型的预测,如图8(a)为所建立的整车电压预测模型在2022/8/3 00:00:00—2022/8/3 23:59:59的预测结果,如图8(b)。



(a) 整车电压预测曲线



(b) 局部放大图

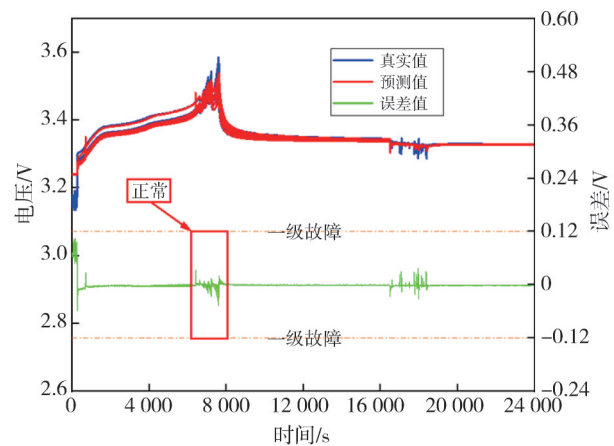
图 8 16号车整车电压预测结果

Fig. 8 Vehicle voltage prediction results of vehicle No. 16

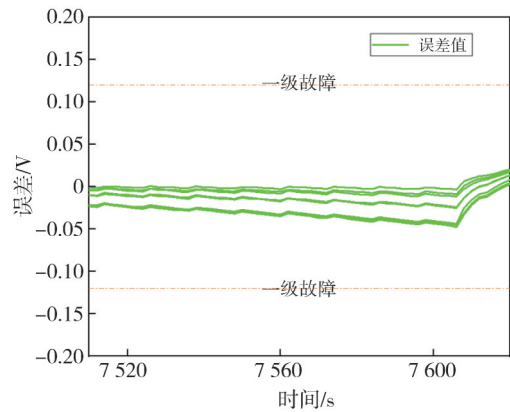
基于LSTM的整车电压预测模型的2号单体在2022/8/3 11:23:48—2022/8/3 11:25:20(第3761个采样点~第3807个采样点)累计触发了47次一级故障,5号单体在2022/8/3 11:24:54—2022/8/3 11:25:12(第3794个采样点~第3803个采样点)累计触

发了10次一级故障,8号单体在2022/8/3 11:23:56—2022/8/3 11:25:18(第3765个采样点~第3806个采样点)累计触发了42次一级故障,然而该车辆在该天运行时间内并未产生故障,属于正常运行车辆,属于误报。

如图9(a),基于LSTM的单体电压预测模型在该天运行期间预测良好,电压预测曲线拟合较好,如图9(b),在基于LSTM的整车电压预测模型产生误报的区间内,基于LSTM的单体电压预测模型没有发出故障预警,由此可以得出,基于LSTM的单体电压预测模型的可靠性高于基于LSTM的整车电压预测模型。



(a) 单体电压预测曲线



(b) 局部放大图

图 9 16号车单体电压预测结果

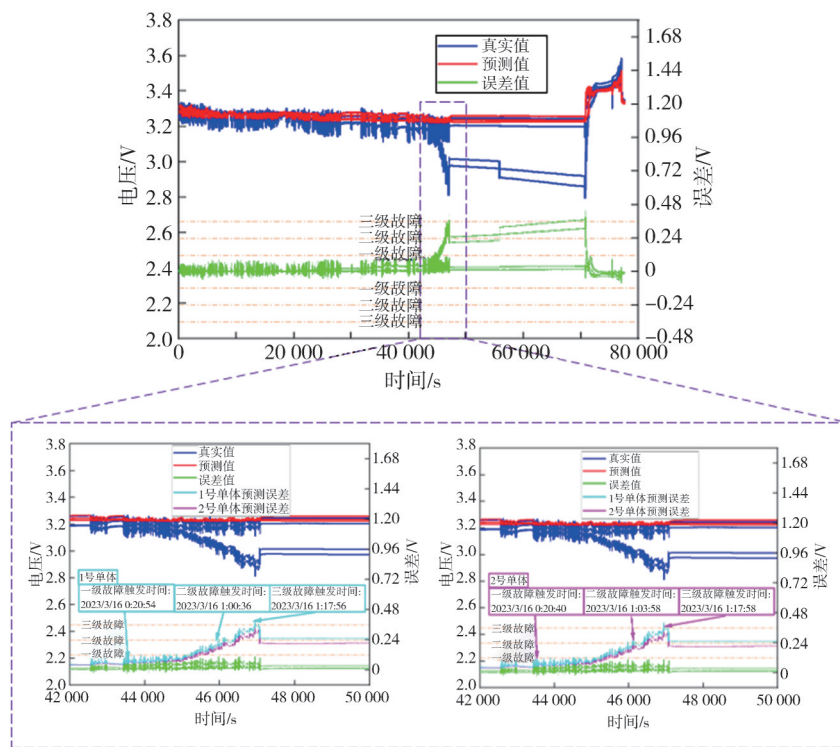
Fig. 9 Cell voltage prediction results of vehicle No. 16

3.2.3 鲁棒性

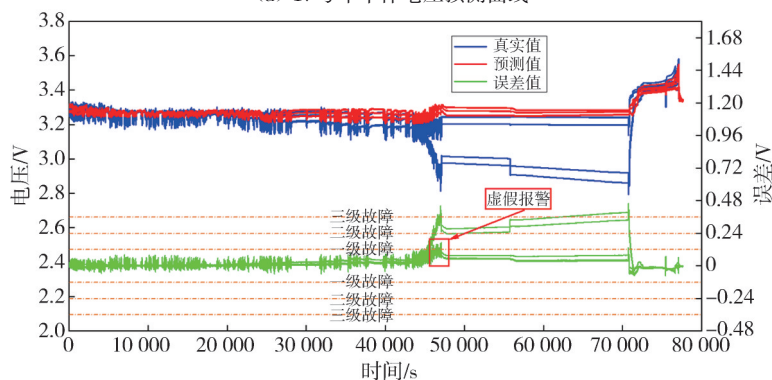
图10(a)为17号车基于LSTM的单体电压预测模型结果图。由局部放大图可以看出:1号单体在2023/3/16 0:20:54首次触发一级故障,2023/3/16 1:00:36首次触发二级故障,2023/3/16 1:17:56首次触发三级故障;2号单体在2023/3/16 0:20:40首次触发一级故障,2023/3/16 1:03:

58首次触发二级故障,2023/3/16 1:17:58首次触发三级故障。通过后期数据的反馈分析得出:在整个电压预测过程中,1号单体的一级故障累计触发648次,二级故障累计触发10441次,三级故障累计触发1818次;2号单体的一级故障累计触发4991次,二级故障累计触发7803次,三级故障累计触发5次;在整个过程中,均是针对于1号单体

与2号单体的预警,最终经故障反馈,发现是由于1号单体与2号单体的欠压故障未及时处理,导致电池发生过放,产生了局部燃烧。图10(b)为基于LSTM的整车电压预测模型结果图,由图可以看出,在整个预测时间内该模型中不同单体均触发了一级故障,虽然最终也识别出了单体故障,但是定位不够精确,产生了很多虚假报警。



(a) 17号车单体电压预测曲线



(b) 17号车整车电压预测曲线

图10 17号车电压预测结果

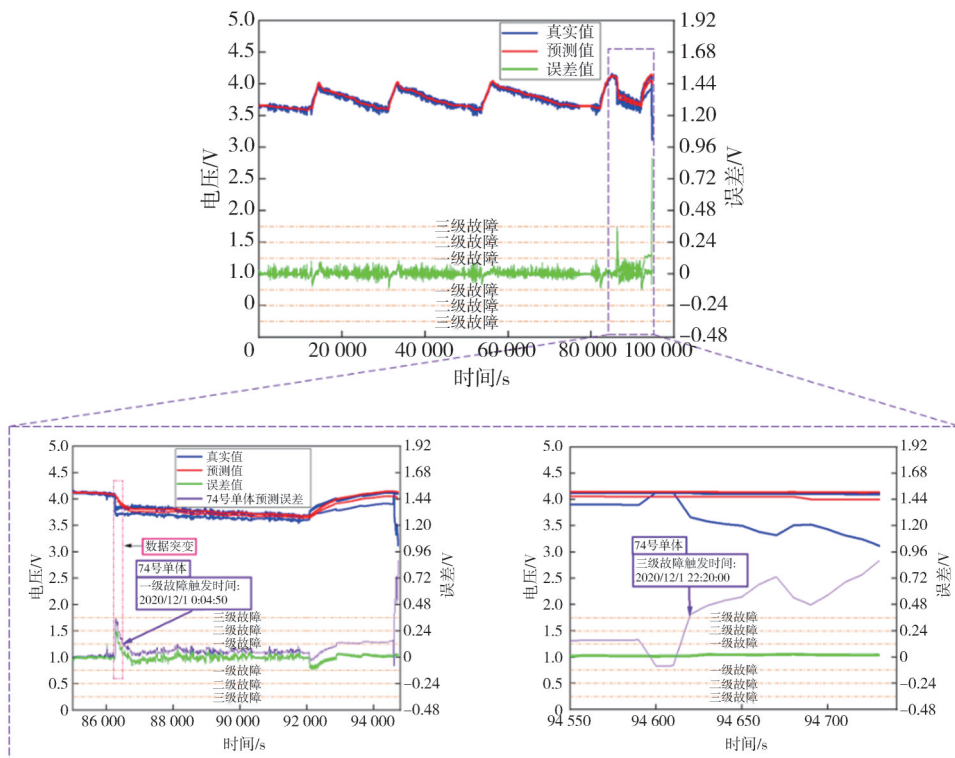
Fig. 10 Voltage prediction results of vehicle No. 17

图11(a)为18号车基于LSTM的单体电压预测模型结果图,由图可以看出,在第8650个采样点左右,所有单体都发生了不同程度的故障预警,后期经过数据溯源可知,该段时间内的车辆数据没有反馈回大数据平台,造成了数据突变,因此,在进行实车运行时,在数据处理这一步设置若单体电压跳变

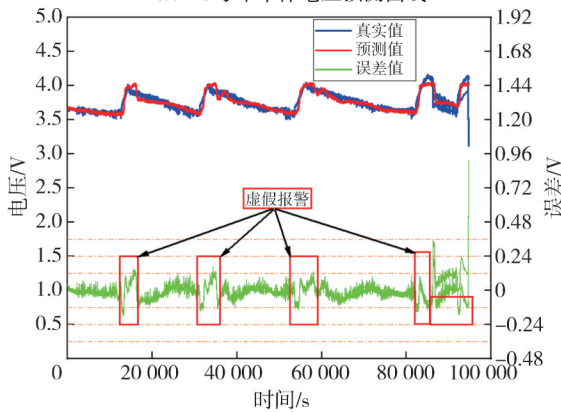
大于0.2V, SOC变化大于5%,则删除整个窗口内的数据,这样可以排除数据突变导致的虚假报警。排除数据突变后,经过数据分析发现,74号单体在2020/12/1 0:04:50首次触发一级故障,直到2020/12/1 22:19:30最后一次触发一级故障,一共触发了174次;在2020/12/1 22:20:00首次触发三级故障,

直到 2020/12/1 22:21:50 触发最后一次三级故障, 一共触发了 12 次三级故障, 随后 74 号单体产生热失控, 蔓延至整个模组, 最后整车发生燃烧, 经过数据分析可知, 针对该热失控车辆, 预警时间至少提前了 22 h。图 11(b) 为 18 号车基于 LSTM 的整车电压

预测模型, 由图可以看出, 虽然在最后即将热失控阶段识别出了电池故障, 但前期时不同单体触发了很多虚假报警, 算法的拟合效果明显下降, 这是由于整车电压模型输入数据过多, 且算法预测窗口过长, 导致输出效果下降。



(a) 18号车单体电压预测曲线



(b) 18号车整车电压预测曲线

图 11 18号车电压预测结果

Fig. 11 Voltage prediction results of vehicle No. 18

由 17 号车和 18 号车的不同电压预测模型可以看出, 基于 LSTM 的单体电压预测模型在不同异常情况下均可以准确识别并定位故障单体, 基于 LSTM 的整车电压预测模型虽然最终也可以识别出异常单体, 但在预测过程中会产生许多虚假预警, 在面对异常数据时的鲁棒性低于基于 LSTM 的单体电压预测模型。

## 4 结论与展望

### 4.1 结论

1) 电动汽车由多个电池单体串并联, 在一致性良好的情况下, 每个单体的电压基本一致, 受到电流的激励趋势一样, 这为建立单体电压模型

预测整车电压提供了理论依据,经过测试,选取每帧时间电压的中位数作为一个新的单体来进行建模会使得模型的评价指标达到最优,更加能反映出电池组的整体性能。

2) 经过正常车辆数据验证,利用贝叶斯算法优化后的长短期记忆神经网络模型可以很好地预测电动汽车实车电压,且贝叶斯优化算法会很快将损失函数值降到最低,最后一次迭代时 BO-LSTM 相对于 LSTM 损失值降低了 49.64%,利用贝叶斯优化算法确定最终的超参数组合后,通过与 LSTM 的整车电压预测模型进行对比,验证了基于 LSTM 的单体预测模型的优越性与可靠性。

3) 经过故障车以及热失控车辆数据的验证,本文所提出的算法在合理设置故障阈值的情况下可以及时有效地定位故障单体和具有潜在热失控风险的单体,针对热失控车辆的预警最早提前了 22 h,因此可应用于电动汽车实车故障预测及热失控预警,与 LSTM 的整车电压预测模型的对比结果,验证了基于 LSTM 的单体预测模型的鲁棒性。

## 4.2 展望

1) 对于理论研究来说,在保证数据质量的前提下,机器学习算法数据量越多越能覆盖电池全生命周期的容量退化轨迹,实车故障诊断及热失控预警也就越精确;对于工程应用来说,随着新能源汽车大数据平台接入的车辆越来越多,模型离线训练时间成本也越来越高,在保证预测精度的同时,对训练数据量的评估成为了下一步需着重考虑的问题。

2) 本文目前的工作是为每一辆车建立一个单体预测模型,下一步将考虑模型的通用性,即该单体模型能够预测该车辆以及同类型所有车辆的单体电压,这将进一步减少模型的离线训练成本,更快地将模型投入在线使用。

## 参考文献:

- [1] LIU K, PENG Q, CHE Y, et al. Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects[J]. *Advances in Applied Energy*, 2023, 9: 100117.
- [2] WANG Q, YE M, CAI X, et al. Transferable data-driven capacity estimation for lithium-ion batteries with

deep learning: A case study from laboratory to field applications[J]. *Applied Energy*, 2023, 350: 121747.

- [3] 王震坡,袁昌贵,李晓宇. 新能源汽车动力电池安全管理技术挑战与发展趋势分析[J]. *汽车工程*, 2020, 42(12): 1606-1620.
- WANG Zhenpo, YUAN Changgui, LI Xiaoyu. An analysis on challenge and development trend of safety management technologies for traction battery in new energy vehicles [J]. *Automotive Engineering*, 2020, 42(12): 1606-1620. (in Chinese)
- [4] YANG Y, WANG R, SHEN Z, et al. Towards a safer lithium-ion batteries: A critical review on cause, characteristics, warning and disposal strategy for thermal runaway[J]. *Advances in Applied Energy*, 2023, 11: 100146.
- [5] 王震坡,李晓宇,袁昌贵,等. 大数据下电动汽车动力电池故障诊断技术挑战与发展趋势[J]. *机械工程学报*, 2021, 57(14): 52-63.
- WANG Zhenpo, LI Xiaoyu, YUAN Changgui, et al. Challenge and prospects for fault diagnosis of power battery system for electrical vehicles based on big-data [J]. *Journal of Mechanical Engineering*, 2021, 57(14): 52-63. (in Chinese)
- [6] ZHANG X, CHEN S, ZHU J, et al. A critical review of thermal runaway prediction and early-warning methods for lithium-ion batteries[J]. *Energy Material Advances*, 2023, 4: 8.
- [7] 孙振宇,王震坡,刘鹏,等. 新能源汽车动力电池系统故障诊断研究综述[J]. *机械工程学报*, 2021, 57(14): 87-104.
- SUN Zhenyu, WANG Zhenpo, LIU Peng, et al. Overview of fault diagnosis in new energy vehicle power battery system[J]. *Journal of Mechanical Engineering*, 2021, 57(14): 87-104. (in Chinese)
- [8] ZHANG Y, JIANG M, ZHOU Y, et al. Towards high-safety lithium-ion battery diagnosis methods[J]. *Batteries*, 2023, 9(1): 63-63.
- [9] XIONG R, SUN W, YU Q, et al. Research progress, challenges and prospects of fault diagnosis on battery system of electric vehicles [J]. *Applied Energy*, 2020, 279: 115855.
- [10] 张扬,李晓杰,马兹林,等. 锂离子电池故障诊断算法研究综述[J]. *重庆理工大学学报(自然科学)*, 2023, 37(9): 49-61.
- ZHANG Yang, LI Xiaojie, MA Zilin, et al. A review of fault diagnosis algorithms for lithium-ion batteries [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2023, 37(9): 49-61. (in Chinese)

- [11] KUMARA P A, CAHYADI A I, WAHYUNGGORO O. Fault detection algorithm on lithium-polymer (Li-Po) battery based on luenberger observer[C]//2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE). Jakarta: Indonesia, 2022: 108-113.
- [12] MARCICKI J, ONORI S, RIZZONI G. Nonlinear fault detection and isolation for a lithium-ion battery management system[C]//ASME 2010 Dynamic Systems and Control Conference, 2011: 607-614.
- [13] 冯旭宁. 车用锂离子动力电池热失控诱发与扩展机理、建模与防控[D]. 北京: 清华大学, 2016.
- [14] 虞婧. 基于数据挖掘技术的电动汽车充电安全监测与故障预警方法研究[D]. 北京: 华北电力大学, 2021.
- [15] HONG J, WANG Z, YAO Y. Fault prognosis of battery system based on accurate voltage abnormality prognosis using long short-term memory neural networks [J]. Applied Energy, 2019, 251: 113381.
- [16] 张朝龙, 赵筛筛, 何怡刚. 基于信息熵与 PSO-LSTM 的锂电池组健康状态估计方法[J]. 机械工程学报, 2022, 58(10): 180-190.  
ZHANG Chaolong, ZHAO Shaishai, HE Yigang. State-of-health estimate for lithium-ion battery using information entropy and PSO-LSTM[J]. Journal of Mechanical Engineering, 2022, 58(10): 180-190. (in Chinese)
- [17] 王天城. 电动汽车动力电池热失控状态评估方法及应用研究[D]. 哈尔滨: 哈尔滨工业大学, 2021.
- [18] 胡杰, 余海, 杨博闻, 等. 基于数据驱动的电动汽车电池安全风险预测[J]. 汽车工程, 2023, 45(5): 814-824.  
HU Jie, YU Hai, YANG Bowen, et al. Battery safety risk prediction for data-driven electric vehicles. [J]. Automotive Engineering, 2023, 45(5): 814-824. (in Chinese)
- [19] LI D, ZHANG Z, LIU P, et al. Battery fault diagnosis for electric vehicles based on voltage abnormality by combining the long short-term memory neural network and the equivalent circuit model [J]. IEEE Transactions on Power Electronics, 2021, 36(2): 1303-1315.
- [20] 刘树鑫, 高士珍, 刘洋, 等. 基于 LSTM 的交流接触器剩余寿命预测[J]. 高电压技术, 2022, 48(8): 3210-3220.  
LIU Shuxin, GAO Shizhen, LIU Yang, et al. Residual life prediction of AC contactor based on long short-term memory [J]. High Voltage Engineering, 2022, 48(8): 3210-3220. (in Chinese)
- [21] FELTUS C. Learning algorithm recommendation framework for IS and CPS security: Analysis of the RNN, LSTM, and GRU contributions [J]. International Journal of Systems and Software Security and Protection (IJSSSP), 2022, 13(1): 1-23.
- [22] 董添. 基于深度学习的电力负荷模式识别与预测方法研究[D]. 长春: 吉林大学, 2022.
- [23] 魏佳恒. 基于贝叶斯优化的 BiLSTM 模型输电塔损伤识别研究[D]. 重庆: 重庆大学, 2021.
- [24] 马梓程. 基于贝叶斯优化的 LSTM 模型在动力电池 SoC 估算中的应用[D]. 镇江: 江苏大学, 2020.
- [25] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. 软件学报, 2018, 29(10): 3068-3090.  
CUI Jiaxu, YANG Bo. Survey on Bayesian optimization methodology and applications [J]. Journal of Software, 2018, 29(10): 3068-3090. (in Chinese)
- [26] 周昌凯. 大数据高斯过程回归模型研究[D]. 杭州: 杭州电子科技大学, 2022.