

# 纯电动客车荷电状态低故障预测

方卫东,陈汉林,陈子标

(福建工程学院 信息科学与工程学院,福建 福州 350118)

**摘要:** 针对纯电动客车荷电状态(SOC)低故障预测问题,在分析其开始充电 SOC 序列的周期性波动规律及变化趋势的基础上,构建差分自回归滑动平均模型(ARIMA)进行车辆 SOC 的短期预测,最后比较预测结果与故障阈值以判断是否发生故障。以某车企的纯电动客车为例进行实证分析,研究表明:差分自回归滑动平均模型的故障预测真阳率为 96.4%,误诊率为 10.3%,说明方法具有良好的预测可行性。

**关键词:** SOC 低;故障预测;ARIMA;纯电动客车

**中图分类号:** TP206+3

**文献标志码:** A

**文章编号:** 1672-4348(2020)04-0353-05

## Research on fault prediction of low state of charge for pure electric buses

FANG Weidong, CHEN Hanlin, CHEN Zibiao

(School of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China)

**Abstract:** Aiming at the problem of fault prediction of the low state of charge (SOC) of electric buses, a differential autoregressive moving average model (ARIMA) was constructed for the short-term prediction of the SOC of pure electric buses. The construction of the model was based on the analysis of the periodic fluctuation law and change trend of the SOC sequence when the charge started. Finally, the prediction results were compared with the fault threshold value to determine whether a fault occurs. The pure electric bus from a certain automobile enterprise was taken for case analysis. Research results show that the true positive rate of the fault prediction of the differential autoregressive moving average model is 96.4%, and the misdiagnosis rate is 10.3%, which shows that the method has high prediction feasibility.

**Keywords:** low SOC; fault prediction; ARIMA; pure electric bus

SOC (state of charge, 荷电状态, 用来反映电池的剩余容量) 是电池管理系统的重要参数之一, SOC 低故障是剩余容量占电池容量过低, 在国标《电车汽车远程服务与管理系统技术规范》<sup>[1]</sup> 中标注为重点故障。国内纯电动客车的 SOC 低故障常常发生在每日行车的最后时段, 与充电后的 SOC 值、行驶里程、路况、载客量、天气等因素相关<sup>[2]</sup>。

在纯电动客车频繁发生的 20 项故障统计中, SOC 低报警故障占总故障的 4.63%<sup>[3]</sup>。SOC 越低

意味着放电深度越深, 会导致电池寿命缩短及电压和电流输出不稳定, 严重影响交通及行车安全<sup>[4]</sup>。纯电动客车 SOC 低故障预测主要是车辆能耗预测。当前国内对车辆能耗预测模型一般以续航里程作为指标, 且着重分析车辆电池参数对里程的影响<sup>[5]</sup>, 或者是通过历史的平均能耗和电池剩余能量预测未来的行驶能量损耗<sup>[6-7]</sup> 或者通过 GPS 轨迹信息预测未来一段时间的能耗<sup>[8]</sup>。这些方法需要建立复杂的路况模型、电池模型、车辆运行模型, 计算量和估计难度较大。针对上述

收稿日期: 2020-05-08  
基金项目: 教育部协同育人项目(201901189009); 福建工程学院引进人才启动基金项目(GY-13102); 福建省中青年教师教育科研项目(JB14067/GY-Z14052)  
第一作者简介: 方卫东(1969—), 男, 安徽黄山人, 副教授, 硕士, 研究方向: 电动汽车融入现代电网等。

问题,本研究以随机选取某车企 50 辆纯电动公交车,近 8 400 万条运行样本数据为研究对象,提出基于 ARIMA 方法的 SOC 低故障预测模型,并采用混淆矩阵的真阳率与误诊率为评估指标,验证方法的有效性。

1 ARIMA 模型的理论知识

ARIMA 模型为差分自回归移动平均模型,是时间序列分析中最常见的一种模型,实际是差分运算与自回归移动平均模型(ARMA)的结合<sup>[9]</sup>。

设  $x_t$  为时间序列, $B$  为后移算子,即  $Bx_t = x_{t-1}$ ,记  $\nabla$  为差分算子,则:

$$\nabla^1 x_t = (1 - B) x_t = x_t - x_{t-1} \tag{1}$$

一般有:

$$\nabla^d x_t = \sum_{i=0}^d (-1)^i C_{i,d} x_{t-i} \tag{2}$$

式中  $C_{i,d} = d! / i! (d - i)!$ ,表明差分后序列等于原序列的若干序列值的加权和。若  $x_t$  在  $d$  阶差分后平稳,则满足式(3)的可称为差分自回归移动平均模型,记为 ARMA( $p, d, q$ ) 模型。

$$\begin{cases} \Phi(B) \nabla^d x_t = \Theta(B) \varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2_{\varepsilon}, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(\varepsilon_t \varepsilon_s) = 0, \forall_s < t \end{cases} \tag{3}$$

式中  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ,为 ARMA( $p, q$ ) 模型的自回归系数多项式; $\Phi(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ,为 ARMA( $p, q$ ) 模型的移动平滑系数多项式。 $\{\varepsilon_t\}, \{\varepsilon_s\}$  为零均值白噪声序列。

2 数据分析

2.1 数据基本特征

研究数据符合《电车汽车远程服务与管理系统技术规范》,具有驱动电机数据、车辆位置数据、发动机数据、整车数据、报警数据等<sup>[9]</sup>。其中,与 SOC 低故障相关的字段具有 4 项:充电开始时间、充电开始 SOC、截止下次充电前消耗 SOC、截止下次充电前运行里程。

纯电动客车每轮充放电的 SOC 变化有:

$$\text{SOC}_t = \text{SOC}_{t-1} + \text{SOC}_{\text{Charging}t-1} - \text{SOC}_{\text{Consumed}t} \tag{4}$$

$\text{SOC}_t$  为纯电动客车  $t$  轮开始充电的 SOC 值,

$\text{SOC}_{t-1}$  是纯电动客车  $t-1$  轮开始充电的 SOC 值,  $\text{SOC}_{\text{Charging}t-1}$  是  $t-1$  轮充电时刻的总充电 SOC,  $\text{SOC}_{\text{Consumed}t}$  为  $t-1$  轮充电后车辆运行消耗的 SOC。

由于纯电动客车在营运结束后会充电较长的时间,在结束充电的时候 SOC 一般为 100%,并且由于路线较为固定的原因,每日的运行里程也较为固定。如图 1 部分展示某辆车截止下次开始充电前消耗的 SOC 与截止下次开始充电前运行里程之间的关系,实线为消耗的 SOC,虚线为运行里程,星形标志为该过程发生 SOC 低故障。从图可以看出,即使在里程变化较小的情况下,不同运行周期 SOC 的消耗依旧不一致,这主要是由载客量、天气、路况、驾驶习惯等因素导致。如果需要定性的分析这些因素就需要建立复杂的路况模型、电池模型、车辆运行模型,而采用时间序列的分析方法可以更为简单直接地预测结果。

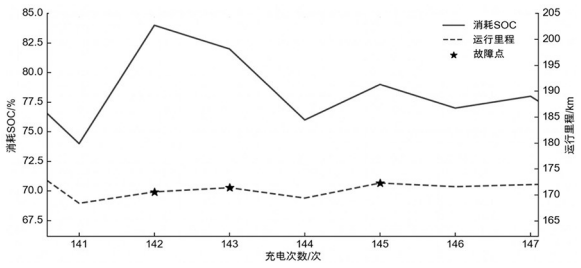


图 1 消耗 SOC 与运行里程关系

Fig.1 Relationship between SOC consumption and mileage

2.2 SOC 低故障阈值

将 50 辆纯电动客车从 2019 年 11 月到 2019 年 12 月的报 SOC 低故障时的 SOC 值拼接如图 2 显示,共计 27 218 次故障,有效故障 56 次,横坐标为充电次数,纵坐标为开始充电 SOC 值。从图 2 中明显可以发现故障时刻的 SOC 值均小于 20%,因此可以确定 SOC 低故障的阈值为 20%。

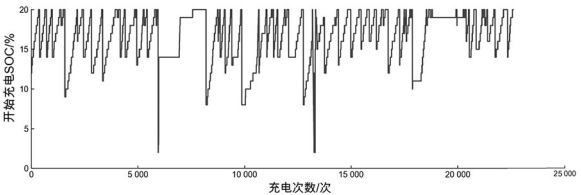


图 2 SOC 低故障情况

Fig.2 Low SOC fault condition

### 3 ARIMA 模型预测

利用 VIN (vehicle identification number, 车辆识别号) 为 L \* 76407 的纯电动客车 2018 年 6 月 1 日~2020 年 4 月 1 日共计 346 次充电过程的起始充电 SOC 数据标定模型和训练样本, 然后对 50 辆纯电动客车 2020 年 4 月 1 日~2020 年 5 月 1 日的开始充电 SOC 进行故障预测。

出于数据的安全性考虑, 车辆 L \* 76407 的 VIN 只显示第一位和最后五位作为标识, 该车积累数据近 2 年, 总数据量为 3.9 GB, 有效 SOC 低故障数量为 25 次, 总行驶里程为 7.5 万 km, 数据的时间跨度为 2018 年 6 月 1 日至 2020 年 4 月 1 日, 每日运行里程维持在 140 ~ 175 km, 单趟行程为 34 km, 固定行驶在湖南省永州市东安县。将车辆数据根据比例 0.66 分割为训练集和测试集如图 3, 训练集为实线, 测试集为虚线, 星形标志为该过程发生 SOC 低故障。

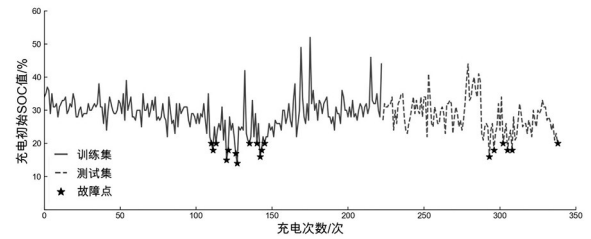


图 3 开始充电 SOC 序列的训练集和测试集

Fig.3 Training set and test set of the SOC sequence when the charge started

#### 3.1 序列季节性分解

一个时间序列是几种变化形式的叠加或耦合: 趋势序列  $T_t$ 、季节序列  $S_t$ 、循环序列  $C_t$ 、残差序列  $I_t$ 。趋势序列是原始序列在较长时期内持续发展变化的一种趋向或状态, 循环序列是不具严格规则的周期性连续变动, 残差序列是由于众多偶然因素对时间序列造成的影响。时间序列  $Y$  可以表示为以上四个因素的函数, 即:  $Y_t = F(T_t, S_t, C_t, I_t)$ 。其中  $F()$  常用的模型分为加法模型和乘法模型, 从图 3 发现开始充电 SOC 序列的季节性并不明显, 故而采用加法模型:

$$Y_t = T_t + S_t + C_t + I_t \quad (5)$$

对开始充电 SOC 序列的训练集分解如图 4 所示。其中, 季节序列具备明显的规律, 趋势序列较为平滑, 有利于后续的预测分析。残差序列围绕着横

坐标轴不断上下波动, 可以判断为白噪声序列, 对其从小到大排序并取 5% (-5.540) 为预测值叠加量。

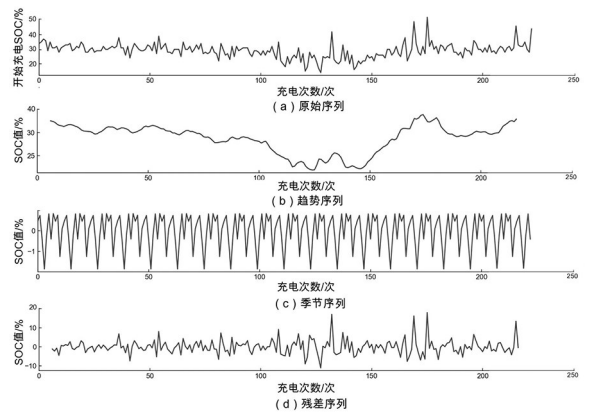


图 4 季节性分解

Fig.4 Seasonal decomposition

#### 3.2 趋势序列平稳性检验与参数标定

##### 3.2.1 平稳性检测

对给定的时间序列进行 ARIMA 模型的分析与预测需要序列符合平稳性, 否则会出现伪回归问题。ADF 检测序列是否存在单位根可判断平稳性, 存在单位根则是非平稳序列。对季节性分解后的趋势序列进行分析, 检测结果表明: ADF 值 -1.529 404, 大于 1%、5%、10% 的 3 个显著性水平下的临界值, 统计量  $P$  值为 0.518 919, 大于临界值 0.05, 接受原假设, 所以该序列为非平稳序列, 需要进行差分处理。

##### 3.2.2 参数标定

对趋势序列进行一阶差分处理如图 5, 结果基本对称且在 0 值附近波动, 初步判断属于平稳状态。通过 ADF 平稳性检测结果表明: ADF 值 -3.045 144 大于 10% 的显著性水平下的临界值, 统计量  $P$  值为 0.038 881, 小于 0.05, 不满足存在单位根的原假设, 故而一阶差分后的序列稳定, 可对其进行 ARIMA 建模分析。

由于一阶差分后, 序列已经维持稳定, 差分阶数  $d=1$ , 进而对趋势序列建立 ARIMA( $p, 1, q$ ) 模型。根据一阶差分的 ACF 图 (autocorrelation function, 自相关函数), PACF 图 (partial autocorrelation function, 偏自相关函数) 的拖尾及截尾状况, 可以得到相关的  $p, q$  参数。

对差分序列做 ACF 和 PACF 如图 6 所示, 自相关 5 阶拖尾, 偏相关 6 阶拖尾, 即  $p=5, q=6$ 。

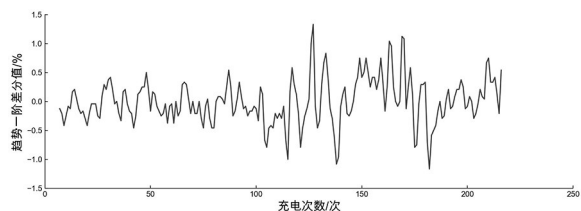


图 5 趋势序列一阶差分

Fig.5 First order difference of the trend series

此时模型 ARIMA(5,1,6),表达式为:

$$\begin{aligned} (1 - \varnothing_1 B - \varnothing_2 B_2 - \varnothing_3 B_3 - \varnothing_4 B_4 - \varnothing_5 B_5) \Delta^1 X_t = \\ (1 - \theta_1 B - \theta_2 B_2 - \theta_3 B_3 - \theta_4 B_4 - \theta_5 B_5 - \theta_6 B_6) \mu_t \end{aligned} \tag{6}$$

则

最后计算出:

$$\Delta^1 X_t = \frac{(1 - 0.310\ 7B - 0.348\ 4B_2 - 0.337\ 3B_3 - 0.320\ 2B_4 - 0.362\ 8B_5 + 0.657\ 7B_6)}{(1 - 0.919\ 1B + 0.933\ 1B_2 - 0.707\ 4B_3 + 0.686\ 3B_4 - 0.476\ 3B_5)} \mu_t \tag{8}$$

3.3 预测结果

将上述获取到的模型用于测试集中进行拟合数据,结果如图 7(a),均方误差为 0.129。预测结果叠加季节性序列和残差叠加量-5.540,显示结果如图 7(b),实线为测试集,虚线为预测结果,可以发现测试集中属于阈值线以下的故障点基本和预测结果符合。

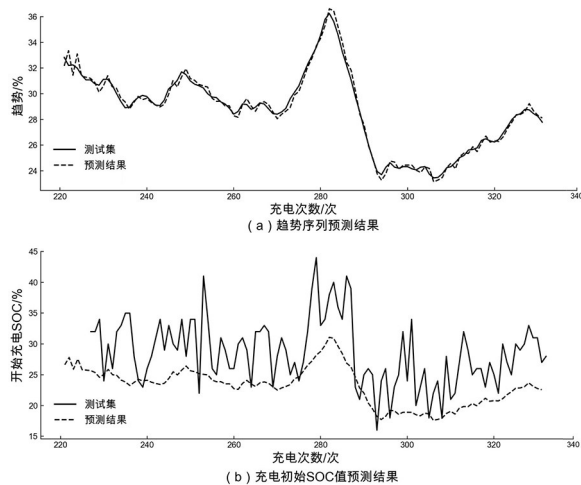


图 7 SOC 低故障预测结果

Fig.7 Fault prediction results of low SOC

3.4 典型纯电动客车故障预测

以 50 辆纯电动客车 2020 年 4 月 1 日~2020 年 5 月 1 日的数据的 SOC 建立 ARIMA(5,1,6)模

$$\begin{aligned} \Delta^1 X_t = \\ \frac{(1 - \theta_1 B - \theta_2 B_2 - \theta_3 B_3 - \theta_4 B_4 - \theta_5 B_5 - \theta_6 B_6)}{(1 - \varnothing_1 B - \varnothing_2 B_2 - \varnothing_3 B_3 - \varnothing_4 B_4 - \varnothing_5 B_5)} \mu_t \end{aligned} \tag{7}$$

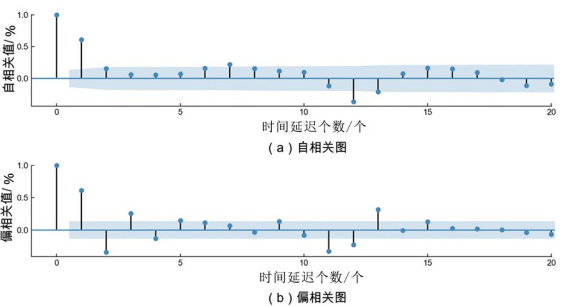


图 6 差分序列定阶

Fig.6 Order determination of difference sequences

型,从而验证模型的有效性与可行性。数据共 948 次充电过程,其中发生 SOC 低故障数为 56 次,由于故障样本占总数的十分之一以下,属于分类不均衡问题,故而采用混淆矩阵的准确率 (ACC)、真阳率 (TPR)、误诊率 (FNR) 作为评估指标。准确率提供了实验结果的全局信息,真阳率是正确预测为正的样本占全部正样本的比例,误诊率是实际为反例的样本中,预测为正例的占比,预测结果的二元分类混淆矩阵如表 1。

表 1 基于 ARIMA 的 SOC 低预测结果混淆矩阵

Tab.1 Low prediction result confusion matrix of SOC Based on ARIMA

故障评判	预测为故障	预测为正常
本身为故障	54	2
本身为正常	92	800

从表 1 可以观察到,TP(本身为故障,预测结果也为故障)是 54 个,TN(本身为正常,预测结果也为正常)是 800 个,FN(本身是故障,预测结果为正常)是 2 个,FP(本身是正常,预测结果为故障)是 92 个。

准确率:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} =$$



$$\frac{54 + 800}{54 + 800 + 92 + 2} \approx 90.1\%$$

真阳率:

$$TPR = \frac{TP}{TP + FN} = \frac{54}{54 + 2} \approx 96.4\%$$

误诊率:

$$FNR = \frac{FP}{FP + TN} = \frac{92}{800 + 92} \approx 10.3\%$$

从结果可知,2020 年 4 月 1 日~2020 年 5 月 1 日故障预测的准确率为 90.1%,真阳率为 96.4%,误诊率为 10.3%,表明 ARIMA 模型具有良好的预测效果与实用性。

5 结论

针对于纯电动客车 SOC 低故障预测问题,首

先分析了车辆运行周期的能耗特征,然后结合充电开始 SOC 特性,通过差分获得平稳序列,构建差分自回归滑动平均模型以拟合数据的趋势性和周期性,最后以湖南某运营公司的纯电动客车开始充电 SOC 数据进行模型参数标定,利用构建的模型预测与诊断故障,将结果与实际故障比较进行验证。结果显示,针对于不同的纯电动客车,真阳率在 96.4%,误诊率在 10.3%,验证了模型的故障预测有效性。

基于 ARIMA 的 SOC 低故障预测结果为公交运营公司的车辆排班与出行计划提供重要参考依据,对提高运营管理具有显现实意义,但是故障预测结果有 10.3%的故障误诊率,如何提高真阳率的同时减低误诊率将作为未来的研究方向。

参考文献:

[1] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 电动汽车远程服务与管理系统技术规范:GB/T 32960.3-2016 [S]. 北京:中国标准出版社,2016: 10.

[2] 刘光明,欧阳明高,卢兰光,等. 基于电池能量状态估计和车辆能耗预测的电动汽车续驶里程估计方法研究[J]. 汽车工程,2014, 36(11): 1302-1309, 1301.

[3] 中国汽车技术研究中心,重庆长安新能源汽车科技有限公司,新能源汽车国家大数据联盟. 中国新能源汽车大数据研究报告[M]. 北京:社会科学文献出版社,2018: 1-50.

[4] 孟仙雅,刘立炳,惠怀兵,等. 放电深度对电池使用寿命的影响[J]. 汽车科技,2016(3): 47-51.

[5] 赵侃,朱聪,曾诚,等. 基于电化学模型的电动公交车续驶里程预测及分析[J]. 公路交通科技,2013, 30(8): 153-158.

[6] 王欲进. 基于电池循环寿命的纯电动汽车续驶里程估算[J]. 科学技术与工程,2019, 19(20): 363-368.

[7] 李雷,高鹏,吴广. 纯电动汽车续驶里程计算研究[J]. 汽车电器,2018(12): 4-5.

[8] 鲍博. 基于最优速度轨迹控制的电动汽车续驶里程研究[C]//2019 中国汽车工程学会年会论文集(2). 北京:中国汽车工程学会,2019: 259-265.

[9] 单伟,何群. 基于非线性时间序列的预测模型检验与优化的研究[J]. 电子学报,2008, 36(12): 2485-2489.

(责任编辑:方素华)