

一种基于PCA和相关向量机的锂电池在线寿命预测方法研究

王国良, 狄心莹

(辽宁石油化工大学 信息与控制工程学院, 辽宁 抚顺 113001)

摘 要: 针对已有基于相关向量机对锂离子电池进行在线寿命预测因考虑因素单一而导致预测精度不理想这一问题,提出了一种基于主元分析(PCA)的特征因素变量加权建构的方法。该方法首先将多种特征因素变量作为研究对象,找到其线性变换后的得分向量所构矩阵;分析其不同得分向量对原变量数据矩阵特征覆盖程度,进一步加权构建融合得到相应特征向量。将所得向量作为输入,经相关向量机建立预测模型并进行锂离子电池寿命在线预测,最终得到预测结果。采用国际公用电池数据作为研究对象,通过MATLAB软件验证了有多变量预测电池寿命的可行性,结果表明预测效果较好。

关键词: 主元分析; 相关向量机; 线性相关性; 寿命预测

中图分类号: TM912.9

文献标志码: A

doi:10.3969/j.issn.1672-6952.2022.06.014

A Method for Online Life Prediction of Lithium Batteries Based on PCA and Relevance Vector Machine

Wang Guoliang, Di Xinying

(School of Information and Control Engineering, Liaoning Petrochemical University, Fushun Liaoning 113001, China)

Abstract: Aiming at the problem that the existing online life prediction of lithium-ion batteries based on the correlation vector machine has a single consideration factor, which results in unsatisfactory prediction accuracy, a method based on principal component analysis (PCA) for weighted construction of characteristic factor variables was proposed. In this method, a variety of characteristic factor variables are taken as the research object to find the matrix of the score vector after the linear transformation. The feature coverage degree of different score vectors to the original variable data matrix is analyzed, and the corresponding feature vectors are constructed by weighted fusion. Using the vector as input, a prediction model is established by the relevance vector machine and the online prediction of lithium-ion battery life is performed, and the prediction results are finally obtained. International public battery data was used as the research object, and MATLAB experiments were used to verify that the method has the feasibility of multivariate prediction of battery life, and the prediction effect is better.

Keywords: Principal component analysis; Relevance vector machine; Linear correlation; Life prediction

锂离子电池作为电源系统的重要组成部件,其具有小巧轻便、环保节能、高续航能力等优点^[1],被广泛应用于各大领域。为了促进新能源技术的发展,对锂离子电池的剩余寿命进行预测研究至关重要。以往对锂离子电池充放电过程的研究中发现,随着电池的使用时间和充放电循环次数的增加,其容量在不断减小^[2]。为了避免对重要设备造成永久性损坏,必须对电池容量进行预测,从而保证电池

在使用期间处于健康状态^[3]。

锂离子电池的剩余寿命预测可分为基于概率统计分布、模型驱动和数据驱动预测方法^[4]。数据驱动作为概率统计理论的主要预测方法,其结合神经网络和支持向量机(SVM)^[5]等学习方法,通过建立输入和输出变量之间的模型来预测锂离子电池的剩余寿命。解冰^[6]建立基于最小二乘支持向量机模型,引入遗传退火算法提高了预测精度。王瀛洲

收稿日期:2021-10-08 **修回日期:**2022-05-06

基金项目:国家自然科学基金项目(62073158)。

作者简介:王国良(1981-),男,博士,教授,从事故障诊断方向研究;E-mail:glwang@lnpu.edu.cn。

等^[7]利用蚁狮优化和支持向量回归的方法提高了电池寿命预测准确性。倪裕隆^[8]改进了蚁狮优化算法,为SVM方法提供最优参数,进一步提高了其预测精度。但是,支持向量机须选用满足一定约束条件的核函数,预测结果为无概率性表现形式,并且神经网络收敛速度慢,训练时间长,预测值为稀疏性的点估计值。引入贝叶斯思想后,弥补了该不足。但是,通常基于RVM相关的锂离子电池寿命预测方法只考虑单一变量的输入输出关系,对多变量输入与输出不能直接建立关系模型。对单一特征因素变量分别预测并加权融合得到的预测值,虽然在一定程度上提高了预测精度,但效果不够明显且步骤冗余。针对这一问题,本文提出了基于主元分析的特征因素变量加权的RVM锂离子电池剩余寿命预测方法,该方法的普适性更广,对锂离子电池的寿命预测精度更高。

1 基本原理和算法

基于相关向量机的锂离子电池寿命预测模型数据处理方法,对单一特征变量预测输出值加权和基于PCA的特征因素变量加权融合构建得到预测输出值。首先,对不同特征变量,通过相关向量机得到相应预测模型,考虑输出的各预测值向量之间线性相关性得到相应权值。然后,基于PCA思想,直接将不同特征变量对电池寿命衰减的影响程度作为给出权值的重要依据。

1.1 PCA的数学意义

假设对系统的 n 个采样点进行 m 次独立采样,则系统可以由一个 $n \times m$ 维的随机矩阵 $G \in R^{m \times n}$ 描述。

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & & \vdots \\ g_{m1} & \cdots & g_{mn} \end{bmatrix} = [G_1 \ G_2 \ G_3 \ \cdots \ G_n] \quad (1)$$

式中, g_{ij} 为采样数据变量, $i=1, \dots, m; j=1, \dots, n$ 。

由于所采集的数据中包含不同量纲,为消除带有不同量纲的数据所组成的数据矩阵对PCA的运行结果有所影响,先对 G 进行归一化处理,去除系统数据的量纲。

1.1.1 归一化处理 通过式(2)对数据进行归一化处理。

$$\bar{G}_i = \frac{G_i - \mu_i}{\sigma_i} \quad (2)$$

式中, \bar{G}_i 为对向量 G_i 归一化处理后的得到的新数据向量; μ_i 和 σ_i 分别为 G_i 的均值和方差。通过处理,

可得随机矩阵 \bar{G} :

$$\bar{G} = \begin{bmatrix} \bar{g}_{11} & \cdots & \bar{g}_{1n} \\ \vdots & & \vdots \\ \bar{g}_{m1} & \cdots & \bar{g}_{mn} \end{bmatrix} = [\bar{G}_1 \ \bar{G}_2 \ \bar{G}_3 \ \cdots \ \bar{G}_n] \quad (3)$$

随机矩阵 \bar{G} 可以分解为 n 个向量的外积之和,即:

$$\bar{G} = U_{m \times n} V_{n \times n}^T = \sum_{i=1}^n u_i v_i^T \quad (4)$$

式中, $U_{m \times n} = [u_1, u_2, \dots, u_n]$ 为得分矩阵,由得分主元 $u_i \in R^{m \times 1}$ 构成; $V_{n \times n} = [v_1, v_2, \dots, v_n]$ 为负荷矩阵,由负荷向量 $v_i \in R^{n \times 1}$ 构成,且由矩阵 \bar{G} 的协方差矩阵给出。

1.1.2 特征值与特征向量 记上述随机矩阵 \bar{G} 的协方差矩阵为 $\Sigma_{\bar{G}}$,则:

$$\Sigma_{\bar{G}} = \frac{\bar{G}^T \bar{G}}{n-1} \quad (5)$$

对式(5)求解:

$$\begin{cases} |\lambda_i I - \Sigma_{\bar{G}}| = 0 \\ (\lambda_i I - \Sigma_{\bar{G}}) v_i = 0 \end{cases} \quad i=1, 2, \dots, n \quad (6)$$

式中, λ_i 为特征值; I 为特征向量。

求解后可以分别得到协方差矩阵 $\Sigma_{\bar{G}}$ 的特征值 $\lambda_i (i=1, \dots, n)$ 及相对应的特征向量 $v_i (i=1, \dots, n)$ 。通常假设 n 个特征值都不相等,并且有 $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n$ 。为证明式(4)成立,两边同乘以 v_i 得:

$$\begin{aligned} \bar{G} v_i &= U V^T v_i \\ &= u_1 v_1^T v_i + u_2 v_2^T v_i + \cdots + u_n v_n^T v_i \end{aligned} \quad (7)$$

由 $\begin{cases} v_i^T v_j = 0 & i \neq j \\ v_i^T v_j = 1 & i = j \end{cases}$ 可知, $\bar{G} v_i = u_i$,则原始数据

矩阵 \bar{G} 可改写为

$$\bar{G} = u_1 v_1^T + u_2 v_2^T + \cdots + u_n v_n^T = \sum_{i=1}^n u_i v_i^T \quad (8)$$

1.1.3 确定主成分个数 一般情况下,设主成分个数为 $k (k \leq n)$,且 k 满足:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \geq 0.85 \quad (9)$$

求得 k ,那么矩阵 \bar{G} 则可表示为:

$$\begin{aligned} \bar{G} &= (u_1 v_1 + u_2 v_2 + \cdots + u_k v_k) + (u_{k+1} v_{k+1} + \cdots + u_n v_n) \\ &= \sum_{i=1}^k u_i v_i + \sum_{i=k+1}^n u_i v_i \\ &= \hat{G} + \tilde{G} \end{aligned} \quad (10)$$

式中, \hat{G} 为主成分空间(主元空间); \tilde{G} 为残差空间。

在现代工业控制研究中,PCA常与其他方法结合应用于各领域控制系统的在线故障检测问题研究^[9-10]及其他方面的数据分析中。 T^2 和SPE是PCA中两个常用的性能指标,用于故障诊断。

1.2 相关向量机

相关向量机(RVM)^[11]是在SVM的基础上做相应的改进,由E.M.Tipping^[12]在2001年提出的。

假设存在训练样本集 $\{x_k, y_k\}, (k=1, 2, \dots, n)$, 其中 x_k 为输入向量, y_k 为输出向量,并认为 y_k 是关于 x_k 的函数 $\phi(x_k)$ 的输出,即:

$$y_k = \phi(x_k) + \varepsilon_k \quad (11)$$

式中, ε_k 为方差 σ^2 的高斯噪声,服从均值为0。常见的 $\phi(x_k)$ 形式为:

$$\phi(x|\omega) = \sum_{i=1}^n \omega_i K(x, x_i) + \omega_0 \quad (12)$$

式中, $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ 为权重向量; $K(x, x_i)$ 为核函数; ω_0 为偏差^[13]。E.M.Tipping从贝叶斯理论的角度出发,认为 ω 服从先验分布,即

$$p(\omega|\theta) = \prod_{i=0}^n p(\omega_i|0, \theta_i^{-1}) \quad (13)$$

可以得到权重向量的后验分布:

$$p(\omega|y, \theta, \sigma^2) = \frac{p(y|\omega, \theta) p(\omega, \theta)}{p(y|\theta, \sigma^2)} \quad (14)$$

$$\Sigma = (\beta \Phi^T \Phi + A)^{-1} \quad (15)$$

$$\mu = \beta \Sigma \Phi^T y \quad (16)$$

式中, Σ 为后验分布方差; μ 为后验分布均值; y 为训练样本的输出量; $A = \text{diag}(\theta_1, \theta_2, \dots, \theta_n)$, $\Phi = [\phi(x_1), \dots, \phi(x_n)]^T$ 为设计矩阵, $\phi(x) = [1, K(x, x_1), \dots, K(x, x_n)]^T$ 。超参数的 θ 和方差 σ^2 优化可以通过式(17)和式(18)实现。

$$\theta_i^{\text{new}} = \frac{1 - \theta_i \sum_{ii} \mu_i^2}{\mu_i^2} \quad (17)$$

$$\sigma^2 = \frac{\|y - \Phi \mu\|^2}{n - \sum_{ii} (1 - \theta_i \sum_{ii} \mu_i^2)} \quad (18)$$

式中, \sum_{ii} 为矩阵 Σ 的第 i 个对角线元素。当迭代过程满足一定收敛条件时,所得的 θ 和 σ^2 是最优估计值,记作 θ_{MP} 和 σ_{MP}^2 。若重新给定一个输入向量,则对应的点估计值和方差为:

$$\phi_*(x_*) = \mu^T \phi(x_*) \quad (19)$$

$$\sigma_*^2 = \sigma_{\text{MP}}^2 + \phi(x_*)^T \Sigma \phi(x_*) \quad (20)$$

选取RBF核函数^[14],该核函数具有使特征输入和输出样本呈现非线性的能力,仅在测试数据点的附近较小领域内产生影响,非线性逼近能力很强,

其表达式为:

$$K_{\text{RBF}}(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{r^2}\right) \quad (21)$$

式中, r 为核函数的参数。 r 越小,RBF对原数据的影响范围越小。

1.3 基于PCA特征因素变量加权建构概述

1.3.1 单因素变量RVM预测值的加权处理 根据RVM理论基础,分别对可能影响锂离子电池容量的不同特征变量测量值与电池容量真实值建立RVM模型,得到所对应的预测输出 y^* 。由于电池容量的变化非单独某个特征变量引起,不同 y^* 和实际容量存在相应的线性关系:

$$Y^* = t_1 y_1^* + t_2 y_2^* + \dots + t_n y_n^* = \sum_{i=1}^n t_i y_i^* \quad (22)$$

式中, $y_1^*, y_2^*, \dots, y_n^*$ 分别对应 n 个特征变量; x_1, x_2, \dots, x_n 为RVM模型下的预测容量值; Y^* 为预测值线性融合向量,由预测值线性融合所得; t_i 为相应的权重系数,由向量相关系数给出:

$$t_i = \frac{\rho_{y_i^* y}}{\sum_{i=1}^n \rho_{y_i^* y}} \quad (23)$$

式中, $\rho_{y_i^* y}$ 为表征电池容量预测值向量 y^* 与真实值向量 y 之间线性关系的紧密程度指标($|\rho| \leq 1$),

$$\rho_{y_i^* y} = \frac{\text{cov}(y_i^*, y)}{\sqrt{D(y_i^*)} \sqrt{D(y)}} \text{ 其绝对值决定二者之间的}$$

紧密程度, $\text{cov}(\cdot)$ 为协方差; $\sqrt{D(\cdot)}$ 为标准差。 $|\rho_{y_i^* y}|$ 较大,说明 y_i^* 与 y 的线性相关程度较强; $|\rho_{y_i^* y}|$ 较小,说明 y_i^* 与 y 的线性相关程度较弱; $|\rho_{y_i^* y}|=0$,则说明 y_i^* 与 y 不相关。

1.3.2 基于PCA的特征变量加权建构方法 根据PCA的数学意义可知,对采样特征变量原始数据矩阵 G 归一化去量纲处理后得到矩阵 \bar{G} ,通过 \bar{G} 求取协方差矩阵 $\Sigma_{\bar{G}}$,并计算该协方差矩阵特征值及与之对应的特征向量,可以找到表征变量主要特征的主元向量 $u_i (i=1, 2, \dots, m)$ 。特征值决定主元的重要程度。

$$\bar{G}_n = r_1 u_1 + r_2 u_2 + \dots + r_n u_n \quad (24)$$

式中, r_i 为权重系数由协方差矩阵特征值给出:

$$r_i = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad (25)$$

该协方差矩阵特征值同样决定主元的重要程度,即主元向量对原变量向量的覆盖程度。因此,可以得到相关主元的权重系数,从而得到各特征因

素变量加权构建以后的特征变量。将此变量作为锂离子电池在线剩余寿命预测相关向量机的输入,与真实电容值建立相应预测模型。对新采集的电池数据进行PCA分析,并对处理后的特征变量通过相关向量机进行在线预测,得到预测结果。具体流程如图1所示。

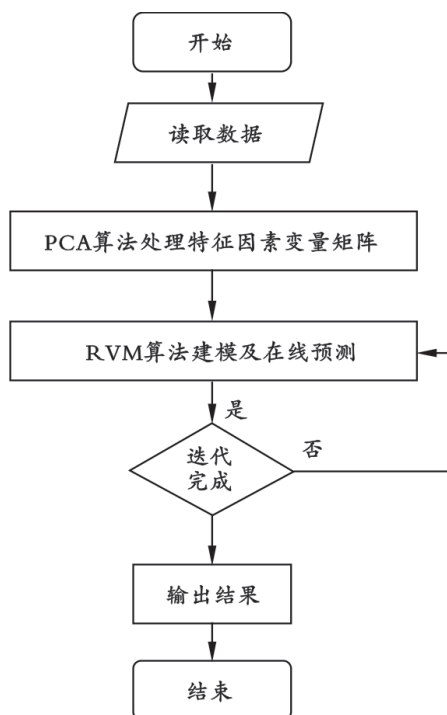


图1 PCA加权特征建构电池寿命预测流程

2 实验与分析

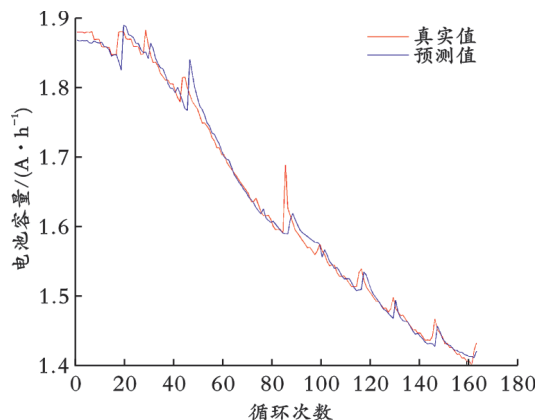
为了验证所提出的预测值加权构建及特征因素变量加权构建方法的性能,证明两种方法的有效性,选取B7型号锂离子电池实验所得数据,采用两种方法进行仿真实验。对已有的相关研究分析发现,锂离子电池在线运行存在不完全充放电的现象,因此需将数据中锂离子电池不完全充放电时数据全部剔除,对锂离子电池进行在线剩余寿命预测方法的研究才具有一定意义。等压升充电时间间隔、等流降充电时间间隔和等压降放电时间间隔三个特征因素变量可作为主要相关向量机建立模型的输入。提取可测量特征因素变量后,根据已有容量数据,通过相关向量机(RVM)建立特征因素变量与电池容量的关系模型;将在线采集的特征变量数据向量作为输入,输出得到相应的容量预测值向量。充电时,电压从2.7 V上升到4.2 V所需时间作为特征因素变量 x_1 ,即 x_1 为等压升充电时间间隔;电流从1.5 A下降到0.3 A所需时间作为特征因素变量 x_2 ,即 x_2 为等流降充电时间间隔。放电时,电压从3.7 V下降到2.7 V所需时间作为特征因素变量 x_3 ,即 x_3 为等压降放电时间间隔。

综上,提取等压升充电时间间隔等三个特征因素变量及相应的电池容量数据进行实验分析。实验所用数据均采用NASA公布的B7号锂离子电池的164组充放电循环真实测量数据。

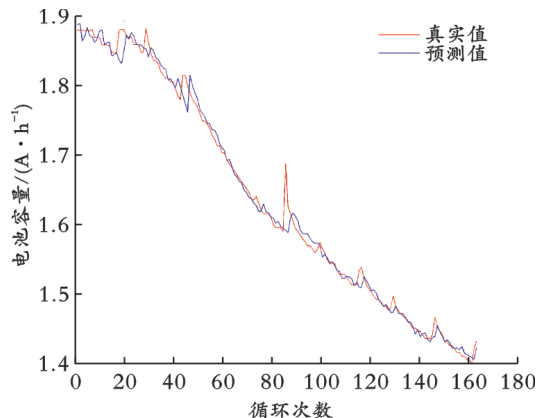
2.1 预测值加权构建

分别利用B7号锂离子电池的三个特征因素变量 x_1 、 x_2 、 x_3 进行预测,所得预测值由相关系数线性融合所得值与融合前的三组单一变量预测值进行对比,其中选取前100组建模数据作为静态数据,用于建立预测模型。模拟建模完成后,采集的新产生64组数据设为动态采集数据,用于在线预测,预测起点为101。

单一变量及预测值加权建构预测结果如图2所示。预测输出 Y^* 由预测值线性融合所得。预测值从101循环开始,64组预测数据实验分析结果为对不同单一特征因素变量预测结果线性拟合得到的加权预测值,考虑了多个特征因素变量对电池容量衰减的影响。对比单一特征因素变量建立的预测模型机制,由数学增长率算法计算得拟合程度分别比单个特征变量的预测值拟合程度提高2.3%、0.4%、0.7%,在一定程度上增加了其预测精度,预测值更加接近容量真实值,对提高锂离子电池在线寿命的预测性能具有一定意义。



(a) 等压升充电时间间隔



(b) 等流降充电时间间隔

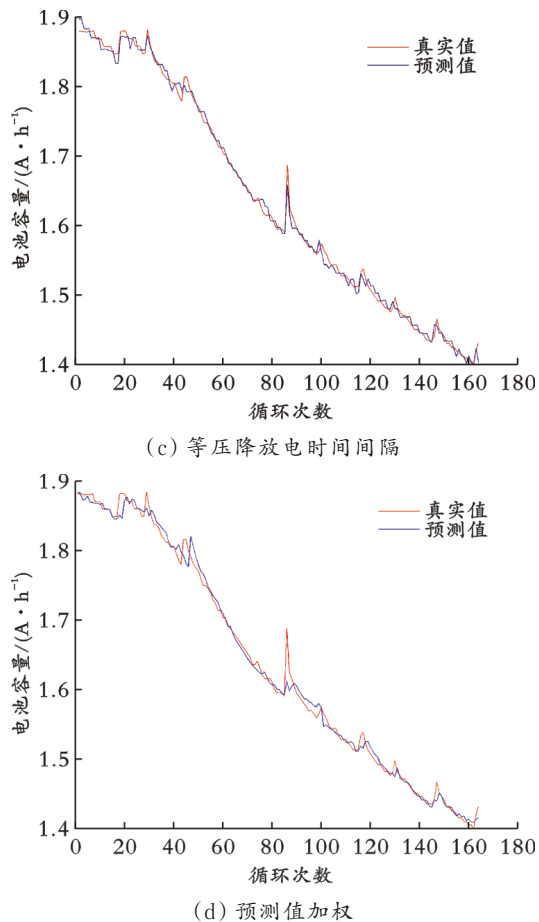


图 2 单一变量及预测值加权构建预测结果

单一变量加权构建方法的预测效果如表 1 所示。

表 1 单一变量加权构建方法的预测效果

变量	预测 起点	拟合程度	偏差度	标准差
x_1	101	0.949 979	-0.011 734	0.010 092 8
x_2	101	0.964 823	-0.009 022	0.008 463 7
x_3	101	0.962 039	-0.015 385	0.008 792 4
Y^*	101	0.968 937	-0.016 051	0.007 953 5

2.2 基于 PCA 的特征因素变量加权建构

基于 PCA 的特征因素变量加权建构预测实验分析,采用 B7 号锂离子电池 164 组充放电循环真实数据进行实验分析。其中,选取前 100 组中 3 个特征因素变量数据作为静态数据,经 PCA 分析用于相关向量机建立电池寿命预测模型,得到其相应核函数参数,即通用程序里可调节参数 $a=3, J=1.8$; 取后 64 组设为动态采集数据,用于电池寿命的在线预测。在线寿命预测过程如下:

预测建模确定相关向量机核函数参数后,开始进行在线预测实验;读取第 101 组数据中 3 个变量值为第一组测量值,将所构成的 1×3 维向量型矩阵作为 PCA 的输入得到加权建构特征变量值,由相关

向量机得到对应预测电池容量值;继而读取第 102 组数据,重复上述运算步骤,得到对应预测值;以此类推,分析计算第 103, 104, ..., 164 组数据,依次得到相应电容的预测值,进而完成锂离子电池的在线寿命预测实验。

基于 PCA 的特征变量加权构建预测结果如图 3 所示。

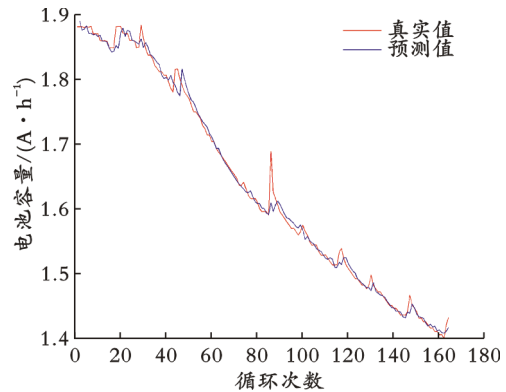


图 3 基于 PCA 的特征变量加权构建预测结果

由 PCA 加权建构所得特征变量作为输入得到的电池容量预测值,避免忽略多个特征变量输入对锂离子电池容量衰减的影响,从根本上解决了多个不同特征因素变量对电池容量衰减影响程度不同而导致存在预测精准度差异的现象,拟合程度较比单个特征变量的预测值拟合程度分别提高 2.3%、0.7%、0.9%,在一定程度上提高了其预测精度,使预测值更加接近容量真实值,有效地提高了锂离子电池寿命的在线预测性能。

PCA 特征变量加权构建方法的预测效果如表 2 所示。其中, X^* 为 PCA 加权建构所得特征变量。

表 2 PCA 特征变量加权构建方法的预测效果

变量	起点	拟合程度	偏差度	标准差
X^*	101	0.970 628	-0.015 385	0.007 733 9

2.3 两种建构方法预测结果对比

预测值加权建构与基于 PCA 特征变量加权建构方法对锂离子电池容量预测结果(见图 4)。

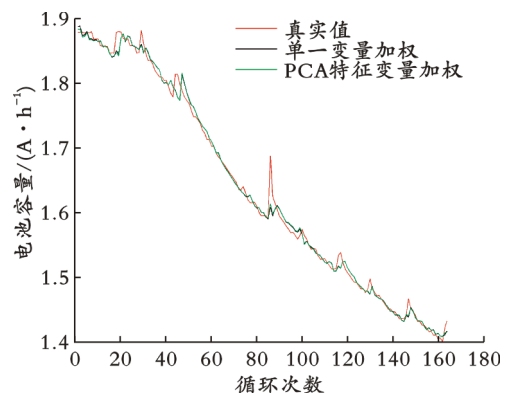


图 4 两种预测结果对比

从图4可以看出,与单一变量加权构建方法相比基于PCA的特征变量加权构建方法的预测效果更好。

3 结 论

多个单一变量预测值加权线性融合的方法,虽考虑到多个特征因素变量对电池容量衰减的影响,在一定程度上增加了预测精度,但仅适用于变量少的相关预测实验研究,普适性低。由PCA加权构建

所得特征变量的方法进行预测,可避免不同特征变量对锂离子电池容量衰减的影响,同时从根本上解决因多个不同特征因素变量对电池容量衰减影响程度不同而存在预测精度差异的现象,增加其预测精度,使预测值更加接近容量真实值,普适性更高。验证实验结果表明,所提变量加权建构处理方法的预测结果与电池真实容量值拟合程度更好,证明所提方法的可行性和有效性。

参 考 文 献

- [1] 张洋.基于相关向量机的锂离子电池在线剩余寿命预测方法研究[D].长沙:国防科技大学,2016.
- [2] Schmalstieg J, Kbitz S, Ecker M, et al. From accelerated aging tests to a lifetime prediction model: Analyzing lithium-ion batteries[C]// Electric Vehicle Symposium & Exhibition. Barcelona: IEEE, 2014.
- [3] 张卓识.锂离子电池建模与故障预测方法研究[D].大连:大连海事大学,2016.
- [4] 朱亮标.基于数据驱动的锂离子电池剩余寿命预测模型及软件实现[D].广州:华南理工大学,2014.
- [5] 李赛.基于数据驱动的锂离子电池剩余寿命预测研究[D].太原:中北大学,2018.
- [6] 解冰.基于支持向量机的锂离子电池寿命预测方法研究[D].武汉:华中科技大学,2012.
- [7] 王瀛洲,倪裕隆,郑宇清,等.基于ALO-SVR的锂离子电池剩余使用寿命预测[J].中国电机工程学报,2021,41(4): 1445-1457.
- [8] 倪裕隆.基于SVR的锂离子电池剩余有效寿命预测方法[D].吉林:东北电力大学,2020.
- [9] 闫翔.无人机PCA故障检测与诊断技术研究[J].电子世界,2020(14): 184-185.
- [10] 苗菟.基于PCA-LSSVM算法的矿用空压机故障诊断模型[J].山东化工,2020,49(14): 120-123.
- [11] 赵倩倩.相关向量机的改进及应用研究[D].天津:河北工业大学,2014.
- [12] Tipping M E. Sparse bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research, 2001, 1(3):211-244.
- [13] 何畏,罗潇,曾珍,等.利用QPSO改进相关向量机的电池寿命预测[J].电子测量与仪器学报,2020,34(6): 18-24.
- [14] 李亚滨,林硕,袁学庆,等.基于新容量退化模型的锂电池RUL预测研究[J].计算机仿真,2020,37(2): 120-124.

(编辑 陈 雷)