Energy 254 (2022) 124415

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Towards efficient state of charge estimation of lithium-ion batteries using canonical correlation analysis

Zichuan Ni [a], Xianchao Xiu [b], Ying Yang [a, *]

[a] State Key Laboratory for Turbulence and Complex Systems, Department of Mechanics and Engineering Science, College of Engineering, Peking University, Beijing, 100871, China
[b] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China

ARTICLE INFO

Article history:
Received 26 January 2022
Received in revised form
29 May 2022
Accepted 29 May 2022
Available online 3 June 2022

Keywords:
Canonical correlation analysis
Denoising method
Lithium-ion battery
State of charge

ABSTRACT

State of charge (SOC) estimation plays an important role for lithium-ion batteries indicating the remaining charge during a cycle. The deep networks adopt the complicated network structure with a large number of parameters, which are sophisticated and lack generality. This paper presents a novel and facile data-driven method based on canonical correlation analysis (CCA) for battery SOC estimation. Firstly, CCA is demonstrated in a regression form and given with an optimizing algorithm for battery SOC estimation. Then the offline training results are followed by the Kalman filter (KF) for online error correction. Finally, a robust canonical correlation analysis (RCCA) is proposed for noise corruption on the input data. Simulation results on different dynamic profiles show the effectiveness of RCCA compared with CCA with improved accuracy by 40% for input noise, and the final results of RCCA with KF achieve root mean squared error (RMSE) of 0.71%. The proposed method achieves superior results in accuracy under input noise and is also computationally efficient with less training time compared with other methods.

1. Introduction

Lithium-ion batteries have attracted considerable attention nowadays with a wide range of applications from portable electronic products like cellphones and laptops to electric vehicles. The lithium-ion batteries stand out due to their higher energy density, lighter weight, and safety [1]. The operation of batteries in electric vehicles is accompanied by the battery management system, which will basically monitor the battery status like the current, voltage, and temperature. On the other hand, it will also perform management tasks including the SOC estimation, remaining useful life prediction, thermal regulation [2], *etc*. Among these, the SOC estimation is of vital significance, which is a percentage number indicating the remaining capacity before the battery needs to get recharged. However, unlike the fuel gauge in a fuel car, the SOC is not a physical quantity to be directly measured.

Various methods have been attempted to estimate the SOC. A typical method is the Coulomb counting method [3] which estimates SOC by subtracting the current integral through a period from initial SOC. However, the estimation relies on initial value which is usually unknown. Also, the estimation error will accumulate due to an open loop calculation. Another classical method, the open circuit voltage (OCV) [4] look up tables, also lack potential in real time application due to the difficulty to rest the battery for several hours before estimation. Other methods can be generally divided into model-based and data-driven methods. Model-based methods, like the equivalent circuit models, establish state space systems and build the inner battery dynamics by resistors and capacitors. The parameters can be determined offline by experimental methods or least square estimation [5,6]. The SOC is then defined in the state vector and can be estimated by observers like extended Kalman filter (EKF) [7], sigma point Kalman filter (SPKF) [8], particle filter [9], Luenberger observer [10], sliding mode observer [11], H-infinity observer [12], *etc*. In order to improve the parameter adaption, online recursive least square with forgetting factors [6,13] can be used for adaptive parameter estimation. The capacity can also be estimated as a varying parameter as degrading factor for improved accuracy [14]. However, the accurate parameter estimation is still difficult to achieve. Also, model-based methods suffer the complicated models, where the model order

* Corresponding author.
E-mail addresses: nizc@pku.edu.cn (Z. Ni), yy@pku.edu.cn (Y. Yang).

https://doi.org/10.1016/j.energy.2022.124415
0360-5442/

determination and observer design can influence estimation accuracy [8,12].

Data-driven methods have also attracted substantial attention because they can map input features to SOC directly without the knowledge of physical models [15]. The support vector machine (SVM) has been a useful tool for regression and can be applied to SOC estimation [16]. The neural network (NN)-based approaches [17] show more powerful capability of nonlinear mapping in SOC estimation. Further deep networks like long short term memory (LSTM) networks and convolutional neural networks (CNN) [18–20] can capture the time series dynamics and achieve better performance in SOC estimation. Other variations of such methods include extreme learning machines (ELM) with gravitational search algorithm (GSA), bidirectional LSTM networks, and dynamically driven recurrent networks [21–24]. Another kind of NN-based method uses the estimated SOC by networks together with Coulomb counting for error correction. The formed process scheme can be realized by filters like EKF, unscented Kalman filter (UKF) [25–27]. Besides, the Coulomb counting can be merged into NN loss objective for physical constraints to reduce error [28]. However, the data-driven methods suffer the complex structures of deep networks with multiple layers and neurons, which make them difficult for generalization. The same type of networks used in different studies differ in structure and parameters in each case. The network design with a large number of parameters makes the process complicated. For instance, we have to make difficult effort to determine the layers, neurons, all of the forgetting gates and updating gates and other parameters of a gated recurrent network. Another problem is that the noise corruption on input data has not been directly addressed in data-driven methods. The uncertainty on input feature data in practical applications should be considered for the training algorithm. Based on the above analysis, challenges still remain to find both an accurate and simpler way for battery SOC estimation.

The CCA is adopted in this study to provide an efficient and simple method for SOC estimation, which is free of the complicated layers and neurons of a network. As a local linear model, CCA can be utilized to capture nonlinear features, which has been widely applied in fault detection, face recognition [29,30], *etc*. It is the first time to apply CCA for battery SOC estimation. In this paper, CCA is first derived in a regression form to map input feature data for SOC estimation and an optimizing algorithm is given. The CCA estimation result is then followed by KF for online error correction. In practical applications, the noise corruption on the input data should also be considered. It is shown the estimation error will increase if noise is added on input data directly using CCA. Thus a RCCA scheme is further presented to eliminate the error when the input is corrupted with noise. The experimental results show the effectiveness of RCCA against CCA with input noise, and the overall results achieve competitive performance compared with conventional methods. The main contributions of the work can be summarized as follows:

- The scheme of CCA is adopted in a regression form for battery SOC estimation together with an optimizing algorithm. It is free of complicated layers and neurons or gates in a deep network.
- A RCCA method is presented when the input is corrupted with noise. It is used as a denoising algorithm to reduce the estimation error and has been validated on dynamic working conditions.
- The proposed method is trained offline with CCA or RCCA and online updated with KF, which is model-free and provides a feasible way for battery SOC estimation that can be generalized to other types.

The rest of the paper is organized as follows. Section 2 first introduces the CCA method with its use for SOC estimation, and also the problem solving algorithm. Then Section 3 presents the training process and estimation results with CCA. Also, the KF results for the noise on output and the RCCA for noise corrupted input are addressed. Finally Section 4 concludes the paper.

2. CCA for battery SOC estimation

This section first introduces the basic idea of CCA. Then a regularized CCA with ℓ_2-norm regularization to prevent overfitting is applied in a regression sense for battery SOC estimation. Finally the optimization algorithm for solving CCA is also illustrated.

2.1. CCA formulation

CCA is one of the multivariate statistical methods which has been extensively investigated [31,32]. The scheme of CCA is to find the combinations of two sets of variables, where the two combined variables are most correlated.

To be specific, suppose that we are given two random variables $X\in\mathbb{R}^p,\ Y\in\mathbb{R}^q$. CCA tries to find two vectors $\beta\in\mathbb{R}^p,\theta\in\mathbb{R}^q$, such that

$$\begin{aligned} U &= \beta^T X = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p, \\ V &= \theta^T Y = \theta_1 y_1 + \theta_2 y_2 + \cdots + \theta_q y_q. \end{aligned} \tag{1}$$

The U and V here are linear combinations of the two random variable vectors. The goal of CCA is to maximize the Pearson correlation ρ between U and V. That is

$$\max_{\beta,\theta}\ \rho = corr(U,V) = \frac{\beta^T \Sigma_{XY}\theta}{\sqrt{\beta^T\Sigma_{XX}\beta}\sqrt{\theta^T\Sigma_{YY}\theta}}, \tag{2}$$

where Σ is the covariance matrix between two variables to the corresponding subscripts, and β and θ are the parameters to be optimized. Unlike the covariance matrix which describes the variance between variables but is not scaled, the Pearson coefficient is divided by the normalized covariance with value from -1 to 1 and can be thus used to describe the correlation between variables [33]. The two sets of variables are input testing features and output SOC for estimation purpose described in the next section. To find the most correlated relation, the value of ρ should be as close to as 1. In the following, the data matrices of observation are denoted as $X\in\mathbb{R}^{m\times p},\ Y\in\mathbb{R}^{m\times q}$, where each row in m is one observation. For simplicity, suppose that both X and Y are zero mean. Then the covariance matrices here are given by

$$\Sigma_{XY} = \frac{X^TY}{m}, \Sigma_{XX} = \frac{X^TX}{m}, \Sigma_{YY} = \frac{Y^TY}{m}, \tag{3}$$

and formulation (2) can be rewritten as

$$\max_{\beta,\theta}\ \frac{\beta^T X^T Y\theta}{\sqrt{\beta^T X^T X\beta}\sqrt{\theta^T Y^T Y\theta}}. \tag{4}$$

Since a scaling in both denominator and nominator does not change the value, the denominator is constrained to one, which derives the following minimization problem

$$\begin{aligned} &\min_{\beta,\theta}\ -\beta^T X^T Y\theta \\ &\text{s.t. } \|X\beta\|^2 = 1,\ \|Y\theta\|^2 = 1. \end{aligned} \tag{5}$$

The optimal values of canonical coefficients β and θ can be

derived by solving problem (5). Then the correlation between the corresponding canonical variables U and V is thus maximized.

2.2. SOC estimation using CCA

In order to apply CCA to battery SOC estimation, one can readily assume that X is the input testing features and y is the output SOC, where the correlation between them is maximized. Then a mapping between testing features and SOC is established. Such a regression purpose can be formulated as follows.

Assume that the dimension of Y vector $q = 1$, i.e. $y \in \mathbb{R}^1$. Then minimization (5) can be equivalently written as

$$\begin{aligned} &\min_{\beta,\theta} \ \|X\beta - y\theta\|^2 \\ &\text{s.t. } \|X\beta\|^2 = 1, \ \|y\theta\|^2 = 1. \end{aligned} \tag{6}$$

From the above formulation, we see that the minimization objective is the loss between two terms, which can be used as regression between X and y. By solving problem (6), we can derive the CCA parameters β and θ. Then with new input data X, we can predict y as $y = X\beta/\theta$.

Further, the ℓ_2-norm regularization term is added to prevent overfitting. The final CCA optimization problem is as follows:

$$\begin{aligned} &\min_{\beta,\theta} \ \|X\beta - y\theta\|^2 + \lambda_1\|\beta\|^2 + \lambda_2\|\theta\|^2 \\ &\text{s.t. } \|X\beta\|^2 = 1, \ \|y\theta\|^2 = 1, \end{aligned} \tag{7}$$

where λ_1 and λ_2 are regularization parameters.

The battery testing system directly records current, voltage, and temperature at each time step. These measured data I, V, T are treated as input features and are correlated to output SOC which is trained by the CCA process. That means $X = [I, V, T]$. The output $y = [SOC]$, and the true SOC is calculated by the Coulomb counting method during training.

Although the CCA is a local linear model, it can achieve nonlinearity by the mapping of the input X to get nonlinear terms. The polynomial mapping is adopted here as Φ: $X = [x_1, x_2, x_3] \mapsto \Phi(X) = [\varphi_1(X), \varphi_2(X), \ldots, \varphi_n(X)]$. It maps features to high dimension space that contains all the exponential terms and cross terms of a certain degree p and also lower. For instance, if polynomial degree p is 2, then polynomial mapping maps elements of the input vector into all its 1st and 2nd order terms: $[x_1, x_2, x_3] \mapsto [x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3]$.

Then the mapped data are also normalized element-wise:

$$X = \frac{X - \mu_1}{\sigma_1}, \quad y = \frac{y - \mu_2}{\sigma_2}, \tag{8}$$

before the CCA training process. The algorithm flowchart is shown in Fig. 1. Here the $\mu_{1,2}$ and $\sigma_{1,2}$ are the mean and standard deviation vectors for data matrices X and y along the samples. In the training process, 75% and 15% of all data are used as the training set and the cross-validation set, respectively.

There are various methods to solve the optimization problem (7). A typical method is to calculate the eigenvalues or perform singular value decomposition [31]. Such a method results in an analytical solution. However, it can sometimes fail especially when the matrix size is large as it requires the calculation of the inverse of a matrix. Here, in this paper, the problem solving algorithm is illustrated in Subsection 2.3.

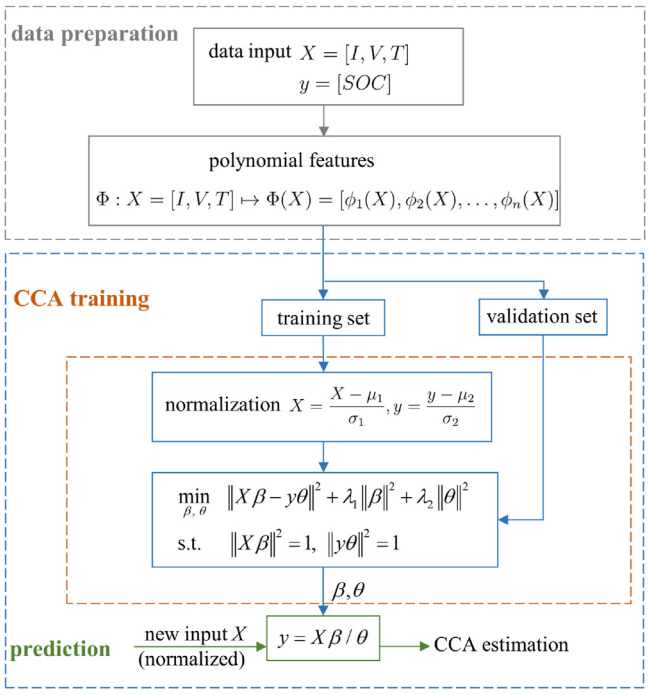

Fig. 1. The scheme of CCA for battery SOC estimation.

2.3. Optimization algorithm

It is worth noting that the final CCA optimization objective (7) is biconvex, meaning if we fix θ, the resulting objective is convex respect to β, and if we fix β, the objective is convex respect to θ. However, the constraints are nonconvex, which cannot guarantee a global solution. This motivates us to separate and relax the problem into two stages:

- Fix θ, optimization problem (7) becomes

$$\begin{aligned} &\min_{\beta} \ \|X\beta - y\theta\|^2 + \lambda_1\|\beta\|^2 \\ &\text{s.t. } \|X\beta\|^2 \le 1. \end{aligned} \tag{9}$$

- Fix β, optimization problem (7) becomes

$$\begin{aligned} &\min_{\theta} \ \|X\beta - y\theta\|^2 + \lambda_2\|\theta\|^2 \\ &\text{s.t. } \|X\theta\|^2 \le 1. \end{aligned} \tag{10}$$

The following shows how to solve the minimization problem (9), and similarly (10). It is worth pointing out that these two subproblems could be solved by an existing optimization solver, however, the special structure cannot be explored. For optimizing β, by introducing a new variable α, problem (9) can be rewritten as

$$\begin{aligned} &\min_{\alpha,\beta} \ \|X\beta - y\theta\|^2 + \lambda\|\beta\|^2 + \mathbb{I}\{\|a\|^2 \le 1\} \\ &\text{s.t. } X\beta = \alpha, \end{aligned} \tag{11}$$

where $\mathbb{I}(\cdot)$ is an indicator function, defined as

$$\mathbb{I}\{\|a\|^2 \le 1\} = \begin{cases} 0, & \|a\|^2 \le 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

The augmented Lagrangian is defined as

$$\mathcal{L}(\alpha,\beta,\gamma) = \|X\beta - y\theta\|^2 + \lambda\|\beta\|^2 + \mathbb{I}\{\|\alpha\|^2 \le 1\} - \langle\gamma, \alpha - X\beta\rangle + \frac{\nu}{2}\|\alpha - X\beta\|^2,$$

where γ is the Lagrange multiplier, and $\nu > 0$ is the penalty parameter, which can be chosen by experience or 5-fold cross validation. According to the alternating direction method of multipliers, the variables can be updated one by one in a Gauss-Seidel manner. Next, the solutions with respect to each variable are as follows.

- Updating α can be obtained by minimizing

$$\min_{\alpha} \ \|\alpha - X\beta^k - \gamma^k/\nu\|^2 + \mathbb{I}\{\|\alpha\|^2 \le 1\}.$$

In fact, it is a projection onto $\mathrm{B} = \{\alpha \mid \|\alpha\|^2 \le 1\}$. Thus, the solution is

$$\alpha^{k+1} = \Pi_{\mathrm{B}}(X\beta^k + \gamma^k/\nu), \tag{12}$$

where $\Pi_{\cdot}(\cdot)$ is the projection onto B, which is defined as $\Pi_{\mathrm{B}}(\mathbf{z}) = \arg\min\{\|\mathbf{z} - \mathbf{x}\|_F, \mathbf{x} \in \mathrm{B}\}$.

- Updating β can be simplified to

$$\min_{\beta} \ \|X\beta - y\theta\|^2 + \lambda\|\beta\|^2 + \frac{\nu}{2}\|\alpha^{k+1} - X\beta - \gamma^k/\nu\|^2.$$

By simple calculation, one can obtain the minimizer β^{k+1} by solving the following linear equation:

$$((1+\nu)X^TX + 2\lambda)\beta^{k+1} = \nu X^T\alpha^{k+1} + X^Ty\theta - X^T\gamma^k. \tag{13}$$

- Updating γ can be written as

$$\gamma^{k+1} = \gamma^k - \nu(\alpha^{k+1} - X\beta^{k+1}). \tag{14}$$

Finally, the proposed algorithm for solving problem (9) can be summarized in Algorithm 1.

Algorithm 1
Solution of problem (9)

Input: Given X, y, θ, parameter $\lambda > 0$, penalty $\nu > 0$
Initialize: (α^0, β^0, w^0), set $k = 0$
Output: β^{k+1}
While not converged **do**
1:Compute α^{k+1} by (12)
2:Compute β^{k+1} by (13)
3:Compute γ^{k+1} by (14)
End While

The framework of CCA for SOC estimation is summarized in the following steps:

Step 1: The battery testing data are collected offline. The current, voltage, and temperature are selected as the input feature data. Then the input data are mapped into polynomial features to capture the intrinsic features.
Step 2: The model is prepared where the data sets are divided into training set and cross-validation set. The CCA loss function is calculated and the parameters are trained by the optimizing algorithm.
Step 3: When new data comes in, the estimation can be attained from the CCA parameters.

3. Results and discussion

3.1. Battery testing data description

The battery testing data are collected online from the CALCE [34]. The battery cells used in the experiments are INR-18650-20R. Each cell is cylindrical with a diameter of 18 mm and a length of 65 mm. The tested capacity is 1.8, 2.0, 2.0 Ah for 0, 25, and 45 °C, respectively. The charging and discharging cut off voltage are 4.2 V and 2.5 V, respectively. The charging cut off current is 0.01 C, which is correspondingly 0.02 A. In the experiment, each cell is discharged with a certain profile. The testing profiles are BJDST, DST, FUDS, and US06 [35,36], which are shown in Fig. 2. The discharging current of the four profiles is not constant and looks irregular just to simulate the real conditions of batteries used in electric vehicles on the road. Note that the current is mostly negative which means discharging. But there are some points where the current is positive indicating the charging process. This is because the vehicle will convert energy to the batteries during braking and thus the batteries will get charged. Fig. 3 shows the testing setup.

Each testing profile is carried out at three different temperatures 0, 25, and 45 °C. During the test, each cell is first charged with the constant current-constant voltage stage to 100% SOC. Then the battery is discharged with a constant current to 80% SOC. From there, it is discharged with one of the desired profiles in Fig. 2 to empty when the voltage reaches the discharge cut off limit. Data from 80% to empty are used in the study. The testing procedure is shown in Fig. 4. Fig. 5 shows the voltage response of the four dynamic profiles at 25 °C when the SOC is from 80% to 0%.

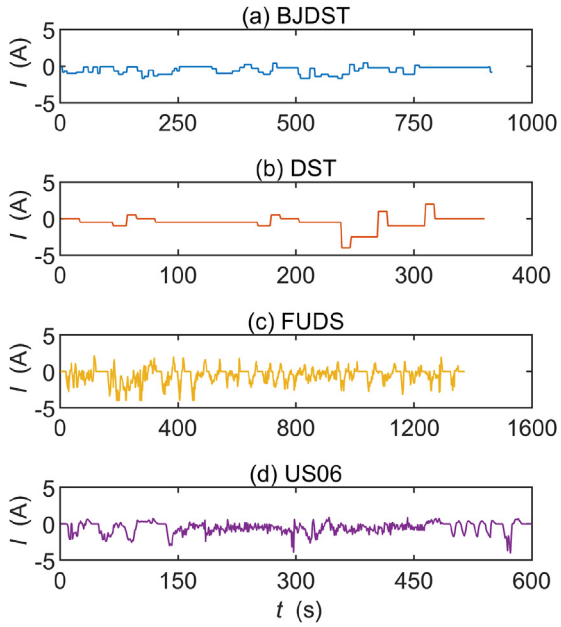

Fig. 2. The four different discharging profiles for the battery test: (a) BJDST, (b) DST, (c) FUDS, (d) US06.

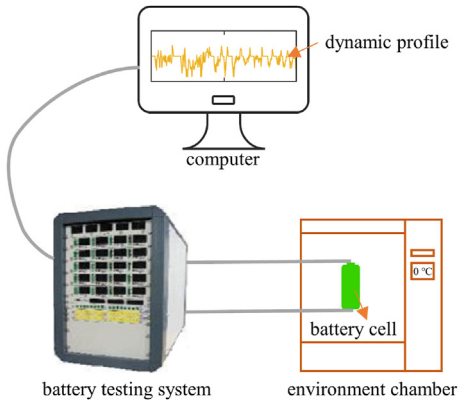

Fig. 3. Schematic of the experiment setup.

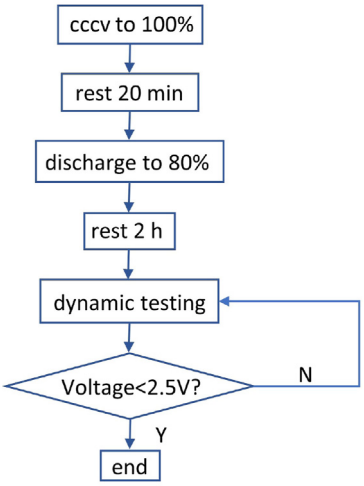

Fig. 4. The battery testing procedure.

Since the current can be precisely measured in the laboratory, the Coulomb counting method is adopted to calculate the true SOC as the reference, and the estimated SOC is then compared for performance evaluation. In fact, the full capacity depends on testing conditions and also the testing profiles. In the study, it is assumed that the full capacity is constant at a certain temperature and does not change for different testing profiles, which is a determined value derived from a prior characterization test.

3.2. Training process and parameter determinization

The testing data at different testing profiles and under different temperatures are used for the CCA training process. However, the data cannot be all used for training along because a smaller training error may not be a good result but the result of overfitting. It is hard to decide the parameters if only training data are used. As a result, 75% of all the data are randomly chosen as the training set and another 15% are used as the cross-validation set. In addition, the training process is repeated multiple times with each time the training and validation sets are randomly chosen. Also, the ratio of training and validation sets is set at other values like 65% for training and 20% for validation in order to prevent any incidental occurrence. The hyper parameters are tried on the training data set, and results are then applied on validation set to determine their values when both sets achieve the small error.

The testing features are mapped to their polynomial terms before training. A small polynomial degree may not capture the features well. On the other side, a large polynomial degree may generate redundant features and cause overfitting. To find the best polynomial degree, different polynomial degrees are tried as shown in Fig. 6. As can be seen here, both training and validation set errors decrease with the increasing polynomial degree. For the balance of model complexity and accuracy, the polynomial degree is chosen as 8. Fig. 7 shows the results of corresponding canonical correlation between the two canonical variables $X\beta$ and $y\theta$, where the Pearson coefficient is 0.99521. Another place of hyperparameters to be determined is the ℓ_2-norm regularization parameters λ_1 and λ_2. They are tried from $10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}$ which are finally determined as 0.01 for both in a similar way. The training process is performed on MATLAB R2020b on Windows 10 platform with 2.9G Core i5 processor and 16 GB RAM. The average training time takes 112 s, which is acceptable for the offline training process.

3.3. SOC estimation results

The battery testing data under four different testing profiles and three different temperatures have been trained by the proposed CCA method. Fig. 8 shows four cases of CCA estimation results for the BJDST, DST, FUDS, and US06 profiles all at 25 °C. The figures plot the curves of CCA predictions and the true value for reference. As can be seen from the plots, the SOC starts from 80% and gradually decreases to 0% for all the testing profiles. At first glance, the CCA prediction can roughly fit the true value well. But when looking at the error curves, we see that the estimation has fluctuations and is not continuous and smooth. The error is approximately within 5%. But there are some large errors at the end of estimation, which may be attributed to limited training data before the battery is near depleted. This can actually be ignored since in practical use the vehicle is not allowed for use in such a low state.

To quantitatively measure the estimation error of the proposed method, the RMSE and mean absolute error (MAE) are calculated:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{m}\sum_{m=1}^{n}(prediction - true)^2}, \\ \text{MAE} &= \frac{1}{m}\sum_{m=1}^{n}|prediction - true|. \end{aligned} \tag{15}$$

Table 1 shows the detailed errors RMSE and MAE for the four different testing profiles at three different temperatures. It can be seen from the table that the overall error is within 2.5% except for the cases of 0 °C under the BJDST and FUDS profiles. This is because the battery cell may degenerate under low temperatures. A special training set can be used for the low temperature cases to improve accuracy.

3.4. KF for output error correction

The Coulomb counting method can be adopted to calculate the true SOC in the laboratory, but in real applications it is corrupted by noise. Nevertheless, its form with noise corruption can be used in the KF [37] framework. The CCA estimation can be viewed as measurement of the SOC. So together with the Coulomb counting prediction model with noise corruption, the state-space model for KF is as follows.

State transition function:

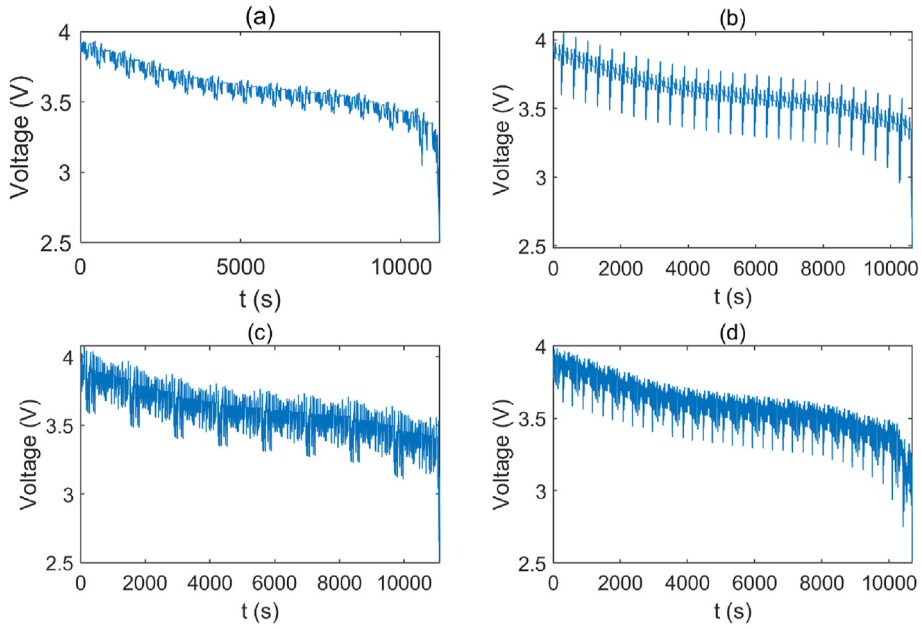

Fig. 5. The terminal voltage response of the four dynamic profiles: (a) BJDST, (b) DST, (c) FUDS, (d) US06.

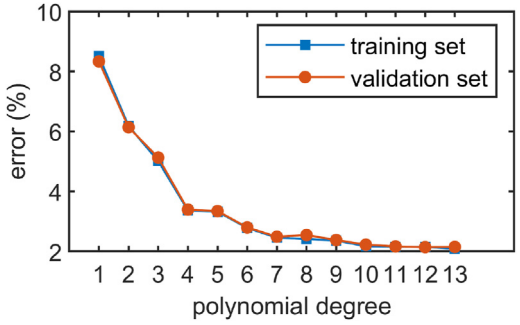

Fig. 6. The RMSE of estimation error for training set and validation set with respect to the polynomial degree.

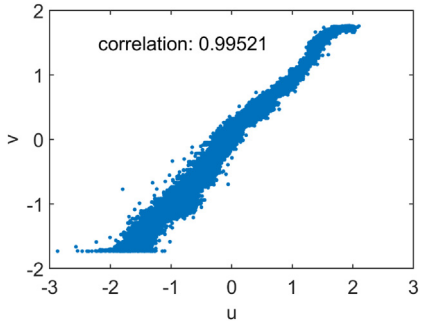

Fig. 7. The canonical variables of input and output.

$$SOC_k = SOC_{k-1} + \frac{I \cdot \Delta t}{capacity} \times 100\% + w, \tag{16}$$

Measurement function:

$$CCA_k = SOC_k + v. \tag{17}$$

Here, I is the current, and Δt is the time interval between two steps which is 1 s. The capacity can be calculated from a characterization test. The process noise and measurement noise are assumed as Gaussian noise with zero mean and variance Q and R, respectively, where $w \sim \mathcal{N}(0, Q)$ is from the current sensor noise, and $v \sim \mathcal{N}(0, R)$ is from the CCA estimation error. The process is shown in Fig. 9. Notations are referenced in Ref. [37]. Fig. 10 shows the KF results after CCA. The estimation results become smooth with reduced average RMSE from 2.4% to 1.7%.

3.5. RCCA for noise corrupted input

Although the above KF deals with uncertainty for error reduction, the noise is assumed on the system output. The input noise corruption should also be considered. In other words, we may come up with cases when the CCA input X in (7) is measured with sensor noise, where the measured data are:

$$\begin{aligned} \tilde{I} &= I + e \\ \tilde{V} &= V + e. \end{aligned} \tag{18}$$

The $e \sim \mathcal{N}(0, \sigma^2)$ is assumed as Gaussian noise with standard deviation σ. The measured temperature noise is ignored. For such cases, direct training CCA on the noise corrupted data using (7) may cause errors. It is tested separately for noise on current and voltage.

Fig. 11 shows the estimation results directly using CCA when the input voltage is added with noise with different standard deviation σ tested on the US06 profile at 25 °C. It can be seen that with the noise increases, the estimated SOC deviates from the true value. Fig. 12 shows the RMSE when different noise standard deviation is added on current and voltage separately. It is found the noise on current has hardly effect on CCA training results while noise on voltage shows more bias on estimation. This is because that the

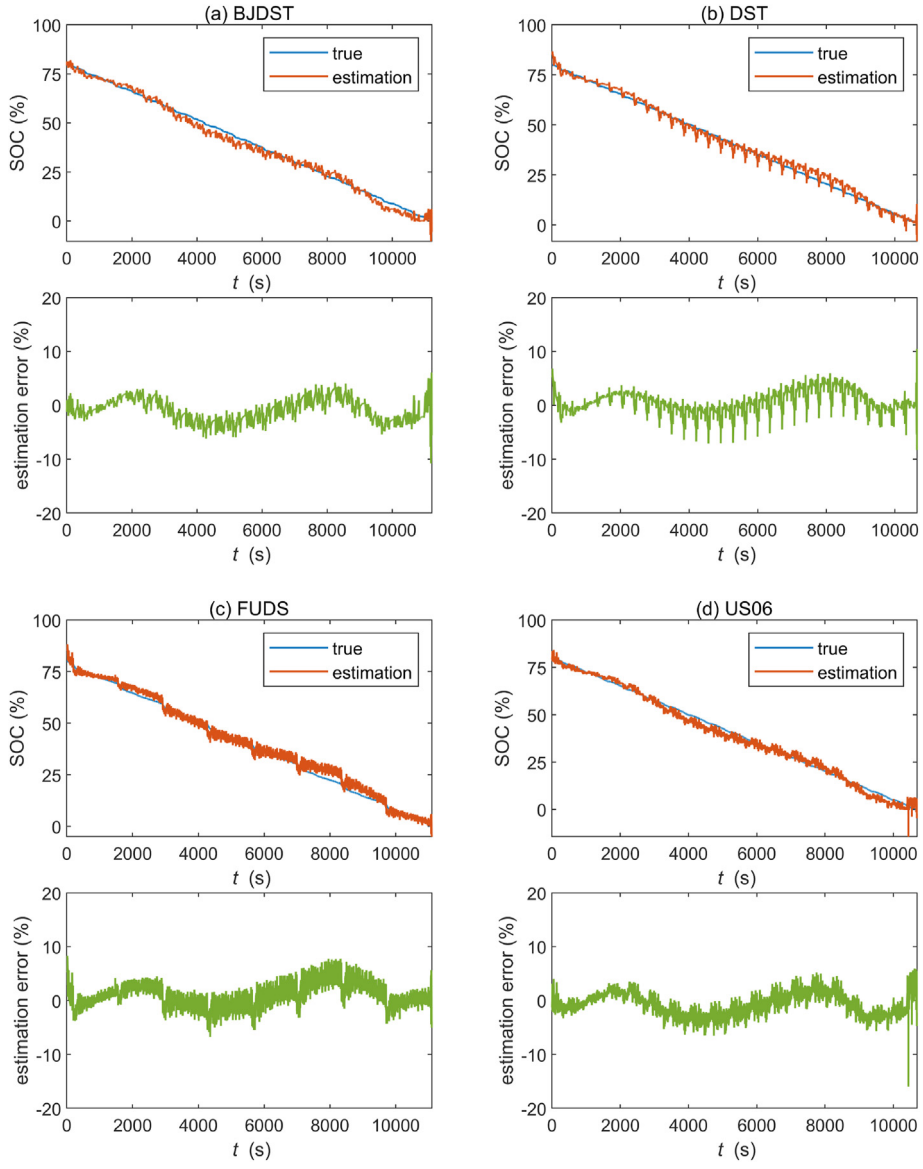

Fig. 8. SOC estimation results using CCA and their errors under four different testing profiles: BJDST, DST, FUDS, and US06 all at 25 °C.

Table 1
Estimation error using CCA at different testing conditions.

Temperature (°C)	RMSE/MAE (%)		
	0	25	45
BJDST	3.4/2.7	2.1/1.8	1.8/1.5
DST	2.7/2.2	1.9/1.5	1.7/1.4
FUDS	3.7/3.0	2.3/1.8	1.8/1.5
US06	2.5/2.0	2.1/1.8	1.7/1.4

voltage reflects more on the SOC estimation as it indicates a direct relationship with SOC in equivalent circuit models.

To address the issue, a RCCA scheme is introduced for training data corrupted with noise, as:

$$\min_{\beta,\theta,E} \ \|X\beta - y\theta - E\|^2 + \lambda_1\|\beta\|^2 + \lambda_2\|\theta\|^2 + \mu\|E\|^2 \quad \text{s.t. } \|X\beta\|^2 = 1,\ \|y\theta\|^2 = 1, \tag{19}$$

It is interpreted that the input and output on projection space is deviated with error due to the data noise. Thus a vector E is introduced in the minimization objective to eliminate the effect. μ

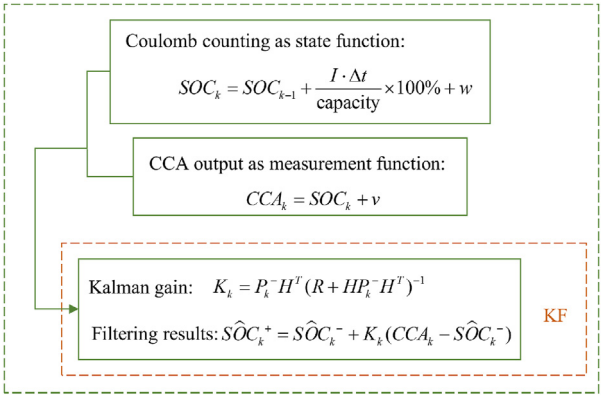

Fig. 9. The KF for CCA estimation error correction.

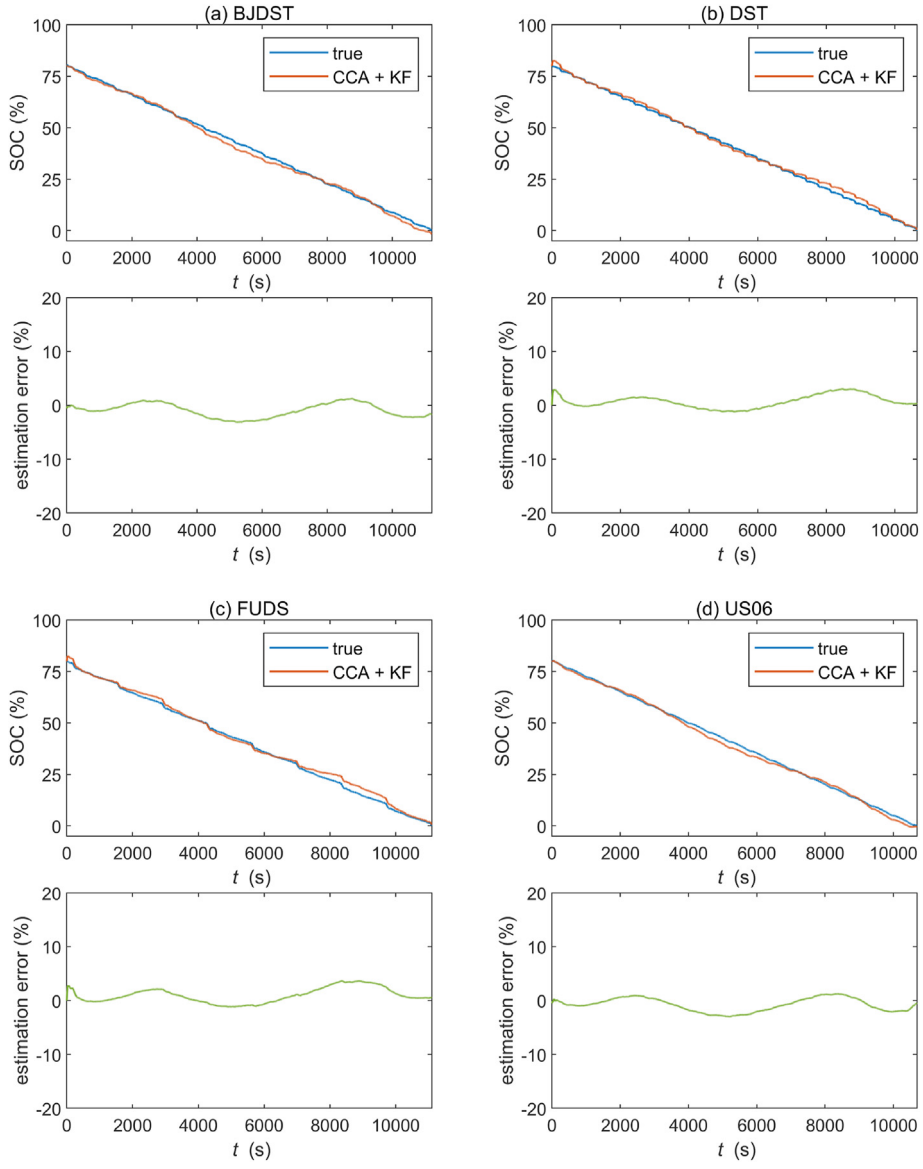

Fig. 10. SOC estimation results using CCA with KF under four different testing profiles: BJDST, DST, FUDS, and US06 all at 25 °C.

is a hyperparameter for balancing the error penalty. The estimation results on the US06 profile at 25 °C are shown in Fig. 13(a). The noise standard deviation is 10 mV. Compared with Fig. 11(c), the RMSE of RCCA estimation results is reduced from 2.4% to 1.96%. Then Fig. 13(b) shows the estimation after KF. It is assumed that the final estimation achieves better results due to both the RCCA and also the filtering process. Fig. 14 (a) shows the comparison of directly using CCA and RCCA with input voltage noise at different standard deviations. It can be seen that RCCA is effective for noise corruption with improved accuracy up to 40%.

Fig. 14 (b) shows the comparison of the model complexity indicated by the number of hyperparameters of CCA with LSTM, CNN, and SVM. CCA has three hyperparameters including λ_1, λ_2, and the polynomial degree. It is shown that CCA has fewer hyperparameters than other methods. Then Table 2 shows the training cost time for the four methods. All methods are run on the same dataset with the MATLAB. The average computation time for CCA is the smallest among the methods.

Finally Table 3 shows the comparison results using RCCA and KF with some of the state of the art methods in the literature. The proposed RCCA with KF scheme achieves average RMSE of 0.71% when the standard deviation of input voltage noise is set at 10 mV. The proposed method shows competitive results though the different batteries and test conditions are used among these methods. Moreover, the proposed method has the ability of treating input noise using RCCA which is not considered in other methods.

To further illustrate the effectiveness of the proposed method, experiments are further tested in the laboratory with the Neware battery testing system on the Panasonic NCR18650PF cell and the Samsung 18650-26J cell. The battery data are collected at 25 °C with SOC from 90% to 0% under the four profiles in Fig. 2 and also the UDDS, LA92 and HWFET profiles in practical applications. Fig. 15

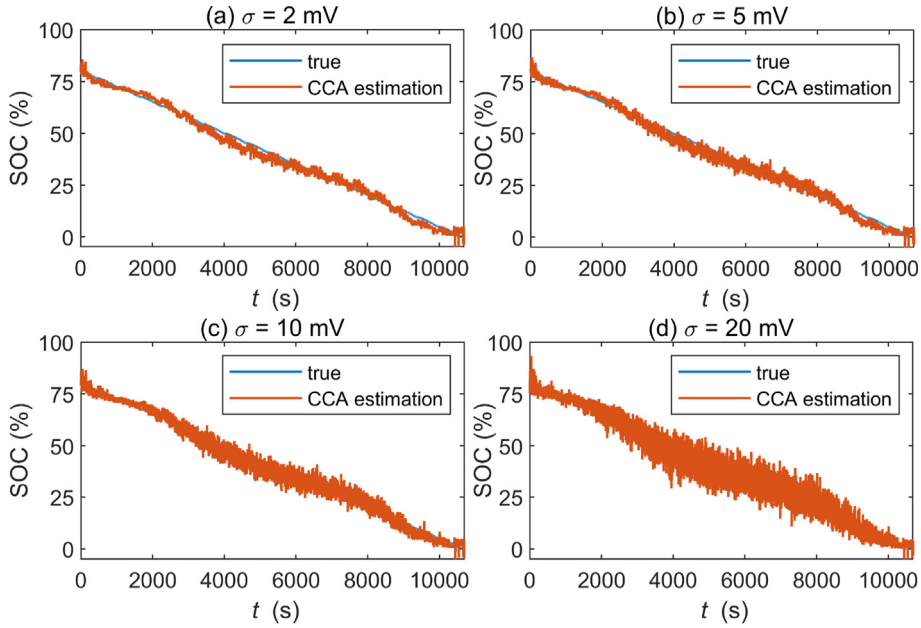

Fig. 11. CCA training results with noise on measured voltage for different standard deviation σ: (a) 2 mV, (b) 5 mV, (c) 10 mV, (d) 20 mV.

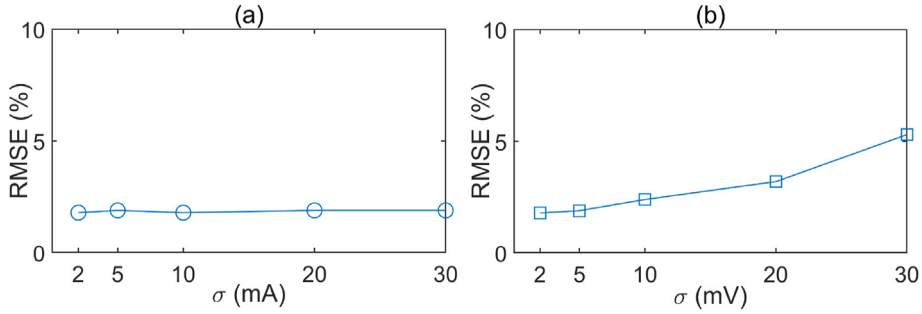

Fig. 12. (a) The RMSE of estimation with different noise standard deviation on current. (b) The RMSE of estimation with different noise standard deviation on voltage.

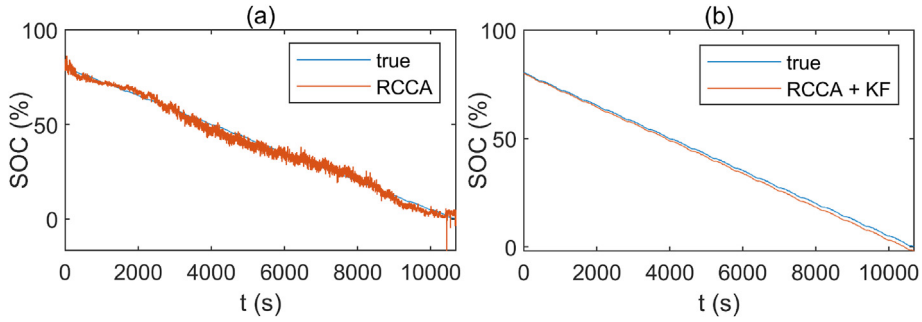

Fig. 13. (a) The RCCA for SOC estimation when the input is corrupted with noise tested on the US06 profile at 25 °C. (b) The estimation results using KF with RCCA results in (a).

shows two examples where (a) is the estimation results for Panasonic cell under the UDDS profile and (b) is the Samsung cell under the LA92 profile. The RCCA sets the input voltage noise with standard deviation of 10 mV. The estimation results using RCCA with KF achieve accuracy with RMSE of 0.86% and 0.7%, respectively.

One thing to mention is that the proposed method is applied assuming that the degrading state is unchanged. In real applications, the voltage response may change with the fresh state of the battery and also affects the SOC estimation. One possible solution is to retrain the battery data after a period during the degrading process.

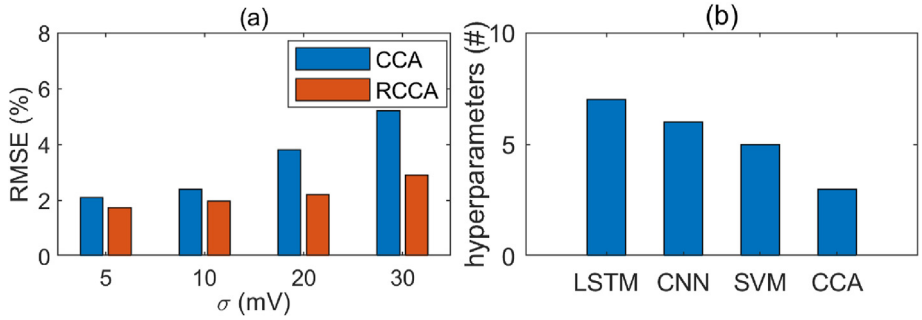

Fig. 14. (a) The comparison of RCCA with CCA at different noise standard deviations. (b) The comparison of the number of hyperparameters of CCA with other methods.

Table 2
Comparison of Training cost time for different methods.

	LSTM	CNN	SVM	CCA
Mean (minutes)	4.6	2.8	5.5	1.9

Moreover, the proposed method for SOC estimation could also be further improved in the accuracy and robustness to noise since the RCCA faces the difficulty of the increasing number of minimization variables.

Table 3
Comparison of proposed method with other methods in the literature on SOC estimation.

Methods	RMSE	Input noise	Battery	Test case
LSTM [18]	0.7%	N	NCR18650PF	HWFET, UDDS, US06, LA92.
LSTM-UKF [26]	1.1%	N	A123 18650	DST, US06, FUDS.
ELM-GSA [22]	0.68%	N	NMC 18650	BJDST, US06.
CNN [20]	0.87%	N	NCR18650PF	HWFET, UDDS, US06, LA92.
RCCA-KF	0.71%	Y	INR18650-20R	BJDST, DST, FUDS, US06.

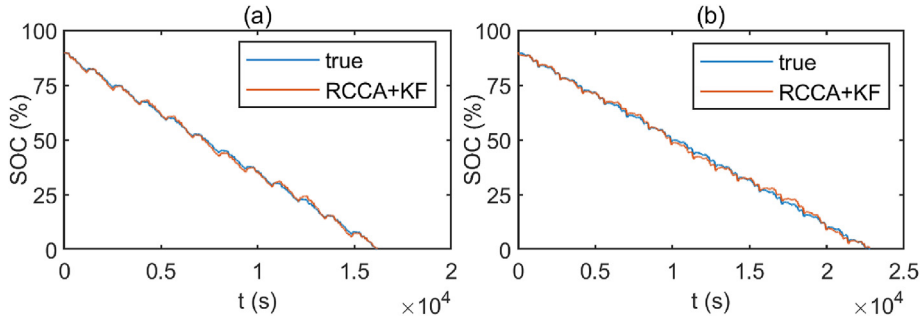

Fig. 15. The estimation results of proposed method tested on other batteries: (a) the Panasonic cell under the UDDS testing profile. (b) The Samsung cell under the LA92 testing profile.

4. Conclusion

Current deep learning-based networks suffer a complicated network structure and a large number of parameters. In this paper, the CCA-based approach is proposed for SOC estimation, which presents a simple and feasible way without the complicated layers and neurons. The KF corrects the error points and smoothes the estimation curve of CCA estimation outputs. For the input data corrupted with noise, a RCCA is presented which aims to eliminate the error on the projection space.

Naturally, there are some limitations of the proposed method. The proposed method does not take battery degradation into account. To address the issue, the offline training model can be retrained after a period or updated online during the estimation.

Credit author statement

Zichuan Ni: Methodology, Investigation, Formal Analysis, Writing - Original Draft. Xianchao Xiu: Methodology, Investigation, Writing - Review & Editing. Ying Yang: Conceptualization, Writing - Review & Editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Key R&D Program of China (No. 2021YFB3301204) and the National Natural Science Foundation of China under Grants 62173003 and T2121002.

References

[1] Zubi G, Dufo-López R, Carvalho M, Pasaoglu G. The lithium-ion battery: state of the art and future perspectives,. Renew Sustain Energy Rev 2018;89:292–308.
[2] Jiao R, Peng K, Dong J. Remaining useful life prediction of lithium-ion batteries based on conditional variational autoencoders-particle filter. IEEE Trans Instrum Meas 2020;69(11):8831–43.
[3] Zhang S, Guo X, Dou X, Zhang X. A rapid online calculation method for state of health of lithium-ion battery based on coulomb counting method and differential voltage analysis,. J Power Sources 2020;479:228740.
[4] Zheng F, Xing Y, Jiang J, Sun B, Kim J, Pecht M. Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries. Appl Energy 2016;183:513–25.
[5] Chen X, Shen W, Cao Z, Kapoor A. A novel approach for state of charge estimation based on adaptive switching gain sliding mode observer in electric vehicles,. J Power Sources 2014;246:667–78.
[6] He Z, Yang Z, Cui X, Li E. A method of state-of-charge estimation for EV power lithium-ion battery using a novel adaptive extended Kalman filter. IEEE Trans Veh Technol 2020;69(12):14618–30.
[7] Xiong R, He H, Sun F, Zhao K. Evaluation on state of charge estimation of batteries with adaptive extended Kalman filter by experiment approach. IEEE Trans Veh Technol 2012;62(1):108–17.
[8] Barillas JK, Li J, Günther C, Danzer MA. A comparative study and validation of state estimation algorithms for Li-ion batteries in battery management systems,. Appl Energy 2015;155:455–62.
[9] Wang Y, Zhang C, Chen Z. A method for state-of-charge estimation of LiFePO4 batteries at dynamic currents and temperatures using particle filter,. J Power Sources 2015;279:306–11.
[10] Tang X, Liu B, Gao F, Lv Z. State-of-charge estimation for Li-ion power batteries based on a tuning free observer. Energies 2016;9(9):675.
[11] Chen X, Shen W, Dai M, Cao Z, Jin J, Kapoor A. Robust adaptive sliding-mode observer using RBF neural network for lithium-ion battery state of charge estimation in electric vehicles. IEEE Trans Veh Technol 2015;65(4):1936–47.
[12] Lin C, Mu H, Xiong R, Shen W. A novel multi-model probability battery state of charge estimation approach for electric vehicles using H-infinity algorithm,. Appl Energy 2016;166:76–83.
[13] Duong V-H, Bastawrous HA, Lim K, See KW, Zhang P, Dou SX. Online state of charge and model parameters estimation of the LiFePO4 battery in electric vehicles using multiple adaptive forgetting factors recursive least-squares. J Power Sources 2015;296:215–24.
[14] Wei Z, Zhao J, Ji D, Tseng KJ. A multi-timescale estimator for battery state of charge and capacity dual estimation based on an online identified model,. Appl Energy 2017;204:1264–74.
[15] Vidal C, Malysz P, Kollmeyer P, Emadi A. Machine learning applied to electrified vehicle battery state of charge and state of health estimation: state-of-the-art. IEEE Access 2020;8:52796–814.
[16] Hu J, Hu J, Lin H, Li X, Jiang C, Qiu X, Li W. State-of-charge estimation for battery management system using optimized support vector machine for regression. J Power Sources 2014;269:682–93.
[17] Kang L, Zhao X, Ma J. A new neural network model for the state-of-charge estimation in the battery degradation process,. Appl Energy 2014;121:20–7.
[18] Chemali E, Kollmeyer PJ, Preindl M, Ahmed R, Emadi A. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. IEEE Trans Ind Electron 2017;65(8):6730–9.
[19] Song X, Yang F, Wang D, Tsui K-L. Combined CNN-LSTM network for state-of-charge estimation of lithium-ion batteries. IEEE Access 2019;7:88894–902.
[20] Liu Y, Li J, Zhang G, Hua B, Xiong N. State of charge estimation of lithium-ion batteries based on temporal convolutional network and transfer learning. IEEE Access 2021;9:34177–87.
[21] Bhattacharjee A, Verma A, Mishra S, Saha TK. Estimating state of charge for xEV batteries using 1D convolutional neural networks and transfer learning. IEEE Trans Veh Technol 2021;70(4):3123–35.
[22] Lipu MSH, Hannan MA, Hussain A, Saad MH, Ayob A, Uddin MN. Extreme learning machine model for state-of-charge estimation of lithium-ion battery using gravitational search algorithm. IEEE Trans Ind Appl 2019;55(4):4225–34.
[23] Bian C, He H, Yang S. Stacked bidirectional long short-term memory networks for state-of-charge estimation of lithium-ion batteries. Energy 2020;191:116538.
[24] Chaoui H, Ibe-Ekeocha CC. State of charge and state of health estimation for lithium batteries using recurrent neural networks. IEEE Trans Veh Technol 2017;66(10):8773–83.
[25] He W, Williard N, Chen C, Pecht M. State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation. Int J Electr Power Energy Syst 2014;62:783–91.
[26] Yang F, Zhang S, Li W, Miao Q. State-of-charge estimation of lithium-ion batteries using LSTM and UKF. Energy 2020;201:117664.
[27] Tian Y, Lai R, Li X, Xiang L, Tian J. A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter,. Appl Energy 2020;265:114789.
[28] Ni Z, Yang Y. A combined data-model method for state-of-charge estimation of lithium-ion batteries,. IEEE Trans Instrum Meas 2022;71:1–11.
[29] Zheng W, Zhou X, Zou C, Zhao L. Facial expression recognition using kernel canonical correlation analysis (KCCA). IEEE Trans Neural Network 2006;17(1):233–8.
[30] Xiu X, Yang Y, Kong L, Liu W. Data-driven process monitoring using structured joint sparse canonical correlation analysis. IEEE Trans Circ Syst II Exp Briefs 2020;68(1):361–5.
[31] Yang X, Weifeng L, Liu W, Tao D. A survey on canonical correlation analysis. IEEE Trans Knowl Data Eng 2019;33(6):2349–68.
[32] Uurtio V, Monteiro JM, Kandola J, Shawe-Taylor J, Fernandez-Reyes D, Rousu J. A tutorial on canonical correlation methods. ACM Comput Surv 2017;50(6):1–33.
[33] Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. Neural Comput. 2004;16(12):2639–64.
[34] Zheng F, Xing Y, Jiang J, Sun B, Kim J, Pecht M. Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries. Appl Energy 2016;183:513–25.
[35] G. Hunt, et al., USABC electric vehicle battery test procedures manual, Washington, DC, USA: United States Department of Energy.
[36] Ma Z, Jiang J, Shi W, Zhang W, Mi CC. Investigation of path dependence in commercial lithium-ion cells for pure electric bus applications: aging mechanism identification. J Power Sources 2015;274:29–40.
[37] Y. Kim, H. Bang, Introduction to Kalman filter and its applications, introduction and implementations of the Kalman filter, F. Govaers, Ed. IntechOpen.