Power Engineering Letters

Data-Driven Online Health Estimation of Li-Ion Batteries Using A Novel Energy-Based Health Indicator

Wei Liu, *Student Member, IEEE*, and Yan Xu, *Senior Member, IEEE*

***Abstract*—Li-Ion batteries have been widely applied in power engineering. Aiming at online state of health (SOH) estimation of Li-Ion batteries, this letter develops a data-driven method using a novel energy-based health indicator (HI). The proposed HI is extracted from the discharge process considering that the discharge process is often less controllable than the charge process. Unlike previous works where only voltage sequences are considered, this HI incorporates both voltage sequences and discharge rates. Therefore, the developed HI enables online SOH estimation at different discharge rates from the offline training dataset. An open dataset is used for verification of the proposed method and very high accuracy is reported with an average RMSE of 1.23%.**

***Index Terms*—Online SOH estimation, Li-Ion batteries, health indicator (HI), discharge rate.**

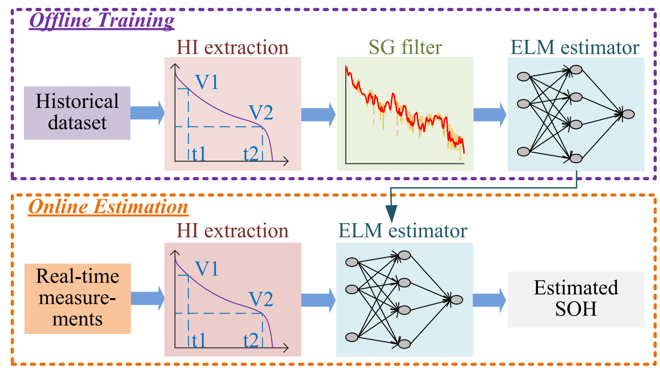

Fig. 1. Diagram of proposed SOH estimation method.

I. Introduction

Li-ion batteries have been widely utilized in electric vehicles (EVs) and power grid energy storage systems (ESS) [1]. The SOH, which can be defined as the ratio between the current capacity and initial capacity of the battery [2], is an important indicator for battery energy management and maintenance. To achieve online SOH estimation, researchers have proposed a variety of approaches [2], among which data-driven methods are drawing growing attention because they are model-free, flexible, and easy to apply.

For a data-driven online SOH estimation method, HI is firstly extracted from real-time signals that can be directly measured (e.g., voltage, current and temperature). The HI is then used as the input of a prediction model which was trained by the dataset generated from the offline aging test. However, existing HIs [3]–[6] only consider the voltage sequence while assuming that the current of online batteries is identical to that of the training dataset. Such presumptions greatly constrain the application of HIs because they can only be used when the battery operates at a specified rate. It is therefore challenging to apply existing methods in certain applications (e.g., power grid) where the charge/discharge profiles are highly variable.

Towards a more comprehensive approach, this letter combines both voltage and current and develops a novel energy-based HI called *energy of equal discharge voltage difference (EEDVD)*. The charge process is not studied in this letter because it is often controllable (i.e., in EVs), making it less necessary to consider different operating currents. With the discharge current incorporated, the requirement of an identical rate is no longer necessary. To this end, the proposed method is applicable to batteries operating at different discharge rates. Based on the developed HI, Savitzky – Golay (SG) filter [4] is then applied to reduce measurement noise and a randomized machine learning algorithm named extreme learning machine (ELM) [7] is used to estimate SOH with the extracted HI.

II. Proposed Method

The framework of the proposed data-driven online SOH estimation method is illustrated in Fig. 1, which consists of two stages: 1) offline training stage and 2) online estimation stage. In the offline training stage, HI is firstly extracted from the training dataset and then smoothed with the SG filter. Then, the ELM is trained to extract the mapping relationship between the HI and SOH. In the online estimation stage, the extracted HI from real-time measurement is imported to the trained ELM to predict the SOH. Typically, the training dataset is generated from offline aging test and real-time measurement is obtained online.

Manuscript received August 20, 2019; revised April 23, 2020; accepted May 10, 2020. Date of publication May 15, 2020; date of current version August 20, 2020. Y. Xu's work is supported by Nanyang Assistant Professorship from Nanynag Technological University, Singapore. Paper no. PESL-00198-2019. *(Corresponding author: Yan Xu.)*

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wei.liu@ntu.edu.sg; xuyan@ntu.edu.sg).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEC.2020.2995112

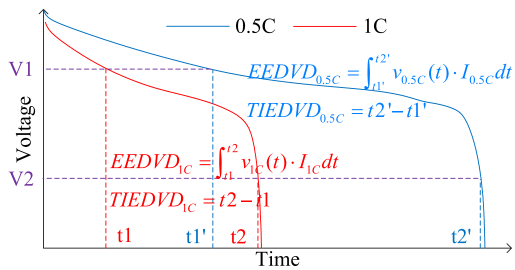

Fig. 2. Extraction of EEDVD and TIEDVD with discharge rates of 0.5C and 1C.

TABLE I
PCC OF CS33 WITH DIFFERENT VOLTAGE RANGES

V1 \ V1-V2	0.1	0.2	0.3	0.4	0.5	0.6
3.5	-0.466	-0.419	-0.380	-0.363	-0.370	-0.402
3.7	0.529	0.538	0.545	0.551	0.568	0.653
3.9	0.977	0.978	0.980	**0.987**	0.977	0.914

A. Health Indicator Extraction And Evaluation

In the power grid, the charge/discharge current of batteries can be highly stochastic. Existing HIs can only be applied at an identical rate. This limitation can be explained with Fig. 2, where the voltage curves of two fresh batteries are presented. These two batteries are from two groups of LiCoO2 batteries. Group A consisting of two battery cells (labeled as CS33 and CS34) was discharged at 0.5C, while Group B comprising three battery cells (labeled as CS35, CS36 and CS37) was discharged at 1C. These batteries will be introduced in Section III.

As an example, the extraction of an existing HI called time interval of equal discharge voltage difference (TIEDVD) [3] is illustrated. Compared to 0.5C, 1C discharge leads to a faster voltage drop speed and TIEDVD of 1C discharge (t2 − t1) is much smaller than that of 0.5C discharge $(t2' - t1')$. However, the SOHs of two fresh batteries are deemed to be the same. Therefore, TIEDVD is not a robust HI to identify SOH at different discharge rates, which will be further verified later.

To eliminate such limitations and enable online SOH estimation at different discharge rates, this letter combines current with voltage sequence and develops as an energy-based HI named EEDVD. The extraction of EEDVD is also illustrated in Fig. 2 and EEDVD is calculated as follows:

$$EEDVD = \int_{t1}^{t2} v(t) \cdot I dt \tag{1}$$

where $t1$ and $t2$ are time spots corresponding to $V1$ and $V2$, I is the discharge current and $v(t)$ is the time-variant voltage.

The selection of $V1$ and $V2$ is based on correlation analysis between EEDVD and SOH. Taking CS33 as an example, the Pearson correlation coefficients (PCC) of different voltage ranges are presented in Table I. As is shown, the maximum PCC is 0.987 and corresponding voltages are $\mathrm{V1} = 3.9(\mathrm{V})$ and $\mathrm{V2} = 3.9 - 0.4 = 3.5(\mathrm{V})$.

To validate the robustness of EEDVD to indicate SOH at different rates, PCC between EEDVD/TIEDVD and SOH are presented in Table II. The correlations of each cell, each group and all cells are all given in Table II. It is shown that for both HIs, the correlation coefficients are higher than 0.91 at an identical rate. However, for TIEDVD at different rates, the overall correlation is significantly low (PCC = 0.41), meaning that TIEDVD is not valid to indicate SOH. Comparatively, the overall PCC between EEDVD and SOH is 0.95, implying that EEDVD is robust to measure SOH at different discharge rates. It is worth pointing out that EEDVD is extracted with constant current profiles, thus it can only be applied at a constant current.

TABLE II
PCC BETWEEN SOH AND TIEDVD AND EEDVD

HIs	Cells	CS33	CS34	CS35	CS36	CS37
TIEDVD	Individual	0.99	0.95	0.96	0.98	0.97
	By group	0.91		0.97		
	All cells	**0.41**				
EEDVD	Individual	0.99	0.96	0.96	0.98	0.97
	By group	0.94		0.97		
	All cells	**0.95**				

B. SG Filter

In practice, the measurements usually contain the noises, which will impact the estimation accuracy of the proposed method. Therefore, the SG filter is applied to smooth the extracted HI because it can increase the precision of the data without distorting the signal tendency.

Give polynomial order n and window length $L = 2k + 1$, matrix $\mathbf{H}$ can be represented as

$$\mathbf{H} = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -k & \cdots & -1 & 0 & 1 & \cdots & k \\ (-k)^2 & \cdots & (-1)^2 & 0 & 12 & \cdots & k^2 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (-k)^n & \cdots & (-1)^n & 0 & 1^2 & \cdots & k^n \end{vmatrix}^T \tag{2}$$

Vector $\hat{\mathbf{a}}$ is then calculated with least square minimization by solving the following equation shown as

$$\mathbf{Ha} = \mathbf{x} \tag{3}$$

where $\mathbf{x}$ denotes the original data vector.

By pseudoinverse operation, $\hat{\mathbf{a}}$ can be calculated as

$$\hat{\mathbf{a}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \tag{4}$$

Finally, the filtered data vector $\mathbf{y}$ can be computed as

$$\mathbf{y} = \mathbf{H}\hat{\mathbf{a}} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \tag{5}$$

C. ELM

ELM is used in this letter as the SOH estimation model and the structure of ELM is shown in Fig. 3. ELM is a randomized

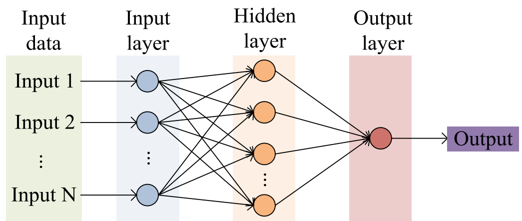

Fig. 3. Structure of ELM

TABLE III
DETAILS OF BATTERY AGING TEST

Group	Cells	Cut-off voltages	Rated capacity	Charge profile	Discharge rate
A	CS33-34	2.7V, 4.2V	1100 mAh	CC+CV	0.5C
B	CS35-37				1C

learning algorithm, which does not need the iterative network parameter optimization. Rather, it randomly selects the input weights and analytically determines the output weights. Therefore, it is much faster than traditional learning algorithms.

For an ELM with one output node, the mathematical expression of output can be presented as

$$\mathbf{Y} = g\left(\mathbf{W} \cdot \mathbf{X}^T + \mathbf{b}\right)^T \cdot \boldsymbol{\beta} \tag{6}$$

where $\mathbf{Y}$ is the output vector given input matrix $\mathbf{X}$; $\mathbf{W}$ is the randomized input weight matrix; $\mathbf{b}$ is the bias of hidden nodes; $g(\cdot)$ is the activation function and $\boldsymbol{\beta}$ is the output weight vector.

The optimal $\boldsymbol{\beta}$ can be analytically computed using least-square estimation, shown as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T \cdot \mathbf{T} \tag{7}$$

where $\mathbf{H} = g(\mathbf{W} \cdot \mathbf{X}^T + \mathbf{b})^T$ and $\mathbf{T}$ is the true target vector.

III. Testing Results

SOH is defined as the ratio between current capacity and initial capacity as expressed in (8) and the root of mean squared error (RMSE) calculated in (9) is used to quantify the error.

$$SOH = \frac{C_c}{C_i} \times 100\% \tag{8}$$

where C_c and C_i are current and initial capacities, respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{9}$$

where y_i and $\hat{y}_i$ are the true and estimated SOH, respectively.

A. Battery Aging Dataset

As abovementioned, this letter uses an open public dataset of LiCoO2 batteries [8]. More details are shown in Table III. Note that CC and CV represent constant current and constant voltage charge process, respectively.

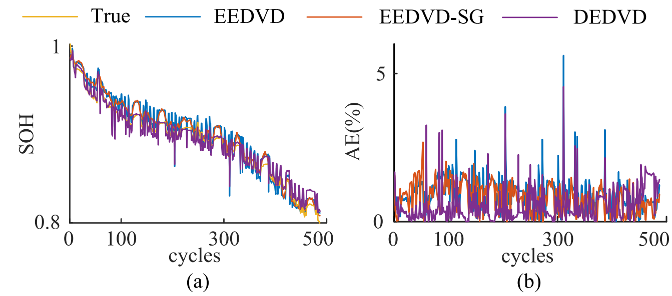

Fig. 4. Estimation results of CS36 at identical discharge rate. (a) estimated and true SOH. (b) absolute error (AE) at all the cycles.

TABLE IV
ERRORS IN RMSE (%) AT IDENTICAL DISCHARGE CURRENT

HI	CS35	CS36	CS37	Mean
TIEDVD	0.92	0.80	0.92	0.88
EEDVD	0.94	0.81	0.93	0.89
EEDVD-SG	0.74	0.73	0.71	0.73

B. Case A: Identical Discharge Rate

EEDVD is firstly tested at an identical discharge rate. In this scenario, Group B which was discharged at 1C is used for verification. TIEDVD is tested as the benchmark and the impact of SG filter on estimation accuracy is also studied. The selected estimation results of CS36 are presented in Fig. 4. In Fig. 4(a), the trajectories of true and estimated SOH are drawn. It is shown that at an identical discharge rate, the estimated SOH can generally trace the true SOH. Fig. 4(b) suggests that the absolute errors (AE) at all cycles are within 5% and at most of the cycles, AE is less than 2%, proving that both TIEDVD and EEDVD are valid HIs to indicate SOH.

The overall estimated accuracies are given in Table IV. It is shown that the average RMSE of EEDVD decreases from 0.89% to 0.73% by applying SG filter, verifying the effectiveness of SG filter. Comparing two HIs, it is shown that TIEDVD and EEDVD can both be used to achieve very high accuracy with a negligible difference (0.88% to 0.89%). It is therefore concluded that very accurate online SOH estimation can be realized with the proposed method at an identical discharge rate.

C. Case B: Different Discharge Rates

The proposed method is then tested in the case that the discharge current of the battery used for testing is different from that of the training dataset. In this case, CS33-34 are used for testing when CS35-37 are used as the training dataset, and vice versa. Some of the results are given in Fig. 5. In Fig. 5(a), the estimated and true SOHs of CS34 and CS37 are drawn. As is presented, the estimated SOH can trace the true SOH fairly well, indicating the good performance of the proposed method. Fig. 5(b) presents the distribution of AEs at different cycles. It is shown that CS33 generates relatively higher errors with a median error of around 2% and most of the errors are within 3%.

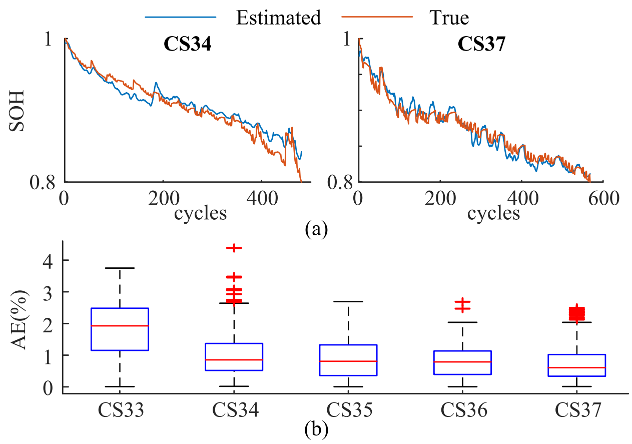

Fig. 5. Estimation results at different discharge rates. (a) estimated and true SOH of CS34 and CS37; (b) boxplot of AE (%) distributions.

TABLE V
ERRORS IN RMSE (%) AT DIFFERENT DISCHARGE CURRENTS

Testing cells	CS33	CS34	CS35	CS36	CS37	Mean
EEDVD	2.02	1.24	1.07	0.92	0.90	1.23
TIEDVD			>50000			Fail

For the other cells, the absolute errors are within 1.5% at most of the cycles and the median errors are all within 1%.

For benchmarking purposes, TIEDVD is also used for SOH estimation in this scenario. However, the results show that TIEDVD fails to estimate SOH at different discharge rates because the estimation errors are all exceptionally large. Comparatively, the proposed method can generate accurate SOH. The RMSEs obtained with EEDVD and TIEDVD are presented in Table V. It is shown that with EEDVD, the RMSEs are between 0.90% (CS37) and 2.02% (CS33) and the average RMSE is 1.23%. EEDVD is therefore verified to enable accurate SOH estimation at different discharge rates. To conclude, TIEDVD is only applicable to Case A while the proposed EEDVD can be applied to both Case A and Case B, highlighting the superiority of EEDVD at different discharge rates.

IV. CONCLUSION

This letter develops EEDVD as the HI and proposes a data-driven method for online SOH estimation. EEDVD is an energy-based HI because it combines the discharge rate with voltage sequence. The testing results on an open dataset verify that the proposed method enables accurate online SOH estimation at identical and different discharge rates. In the future, we will further incorporate other factors, i.e., temperature, into our model for online SOH estimation. Besides, we will investigate systematically on dynamic charging scenarios. Possible methods such as transfer learning will be used, which can adapt the trained model to unseen scenarios.

REFERENCES

[1] B. Xu, A. Oudalov, A. Ulbig, G. Andersson, and D. S. Kirschen, "Modeling of lithium-ion battery degradation for cell life assessment," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1131–1140, Mar. 2018.

[2] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016.

[3] D. Liu, H. Wang, Y. Peng, W. Xie, and H. Liao, "Satellite lithium-ion battery remaining cycle life prediction with novel indirect health indicator extraction," *Energies*, vol. 6, no. 8, pp. 3654–3668, Jul. 2013.

[4] R. R. Richardson, C. R. Birkl, M. A. Osborne, and D. A. Howey, "Gaussian process regression for in situ capacity estimation of lithium-ion batteries," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 127–138, Jan. 2019.

[5] D. Liu, J. Zhou, H. Liao, Y. Peng, and X. Peng, "A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 45, no. 6, pp. 915–928, Jun. 2015.

[6] Y. Zhu, F. Yan, J. Kang, and C. Du, "State of health estimation based on OS-ELM for lithium-ion batteries," *Int. J. Electrochem. Sci.*, pp. 6895–6907, Jul. 2017.

[7] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[8] Y. Xing, E. W. M. Ma, K.-L. Tsui, and M. Pecht, "An ensemble model for predicting the remaining useful performance of lithium-ion batteries," *Microelectron. Reliab.*, vol. 53, no. 6, pp. 811–820, Jun. 2013.