This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation in Electric Vehicles

Xiaosong Hu, *Member, IEEE*, Shengbo Eben Li, *Member, IEEE*, Yalian Yang

***Abstract*—To fulfill reliable battery management in electric vehicles, an advanced State-of-Charge (SOC) estimator is developed via machine learning methodology. A novel genetic algorithm-based fuzzy C-means clustering technique is first used to partition the training data sampled in the driving cycle-based test of a lithium-ion battery. The clustering result is applied to learn the topology and antecedent parameters of the model. Recursive least-squares algorithm is then employed to extract its consequent parameters. To ensure good accuracy and resilience, the backpropagation learning algorithm is finally adopted to simultaneously optimize both the antecedent and consequent parts. Experimental results verify that the proposed estimator exhibits sufficient accuracy and outperforms those built by conventional fuzzy modeling methods.**

***Index Terms*—Energy Storage, Electric Vehicle, Battery Management, State Estimation, Machine Learning**

I. Introduction

The limited fossil fuels, increasing demand for individual mobility, serious pollutant emissions, and global warming urgently call for profound innovations in ground transportation system to evade a looming energy crisis. Automotive industry and academic community are devoted to advancing high-efficiency, clean, and affordable mobility concepts [1]. Electric vehicles (EVs) are one of the most promising technologies, owing to their remarkable energy-saving capabilities and potential interactions with renewable power grid [2], [3]. Traction battery packs are currently the most common electric energy carrier onboard and thus play an imperative role in the performance, economy, and acceptance of EVs [4]-[7]. To make costly batteries safe, efficient, and durable in complex vehicle environment, meticulous monitoring and control of internal battery states, e.g., State-of-Charge (SOC), is required. The uncertainty of states may thwart vehicle energy routing and exacerbate battery safety/durability problems [8]-[10]. Limited sensing and actuation, nevertheless, constitute a daunting technological challenge holding back accurate SOC tracking. They also constitute a major incentive to harness advanced control and artificial intelligence algorithms for combating such a challenge. In this paper, we design a novel SOC indicator through advanced machine learning tools.

X. Hu is with the State Key Laboratory of Mechanical Transmissions and also with the Department of Automotive Engineering, Chongqing University, Chongqing 400044, China (e-mail: xiaosonghu@ieee.org, corresponding authors are X. Hu amd S. E. Li who equally contributed to this work).

S. E. Li is with the State Key Laboratory of Automotive Safety and Energy, Department of Automotive Engineering, Tsinghua University, Beijing 100084, China (e-mail: lisb04@gmail.com).

Y. Yang is with the State Key Laboratory of Mechanical Transmissions and also with the Department of Automotive Engineering, Chongqing University, Chongqing 400044, China (email: yyl@cqu.edu.cn)

There are various methods reported to estimate the battery SOC, each with its strengths and limitations [11], [12]. For example, Coulomb counting and its modified versions were often used [13], [14]. This approach is simple, online, but open-loop and extremely sensitive to current-sensor precision and disturbances, which tends to incur large accumulated errors. Regular recalibration is hence inevitable in practice. Additionally, the initial SOC is very difficult to be determined precisely, leading to a long-lasting initial deviation. Some impedance-based methods were also developed [15], [16]. The relationship between the measured impedance and SOC, however, may be unstable in the case of demanding battery load in realistic EVs operations. Moreover, the impedance-based models are temperature sensitive and cost intensive [11]. Kalman filtering-based techniques have been exploited to implement the battery SOC indication as well [10], [17]. This class of method is online and closed-loop, thereby circumventing the deficiencies pertaining to Coulomb counting. However, the associated disadvantages are threefold. First, the estimation precision is highly dependent on the validity of battery models that Kalman filters use. Extensive battery characterization/parameterization tests are, therefore, necessary for constructing battery models with good robustness and extrapolation. Second, a relatively small variation of the filter-tuning parameters probably induces diverged SOC estimates. Third, the computational burden resulting from multiple recursive equations is relatively heavy. Several observers, such as sliding-mode observer [18], [19], H-infinity observer [20], and adaptive Luenberger observer [21], have also been deployed to estimate the battery SOC. Their pros and cons are similar to those attached to Kalman filtering.

As an intriguing machine learning approach, fuzzy logic provides a powerful means of modeling complex and nonlinear dynamic systems, which was often used to resolve the battery SOC indication problem. While the offline training is computationally intensive, the established fuzzy models can readily estimate the battery SOC in real time. For instance, a fuzzy Mamdani reasoning system for estimating the SOC of a lead-acid battery was synthesized [22]. A neuro-fuzzy system was devised to model a lithium-ion battery for the SOC estimation under constant-current discharge profiles [23]. An adaptive neuro-fuzzy inference system (ANFIS) was introduced to implement the capacity evaluation for a lithium-ion battery under four typical electric vehicle driving cycles [24]. Cai et al. trained an ANFIS to estimate the SOC of a high power Nickel Metal Hydride (NiMH) battery, based on datasets of four different constant-current discharge profiles [25]. Design and implementation of a lithium-ion battery SOC meter for portable defibrillators was also reported. The SOC meter was based on an ANFIS model whose structure was determined by grid partition [26]. Unfortunately, the topologies

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

of all the aforementioned fuzzy logic models — the prior number of membership functions and fuzzy rules — were learned by intuition or experiential knowledge, therefore engendering heuristic mistakes. In order to alleviate the reliance on trial and error, a first-order Sugeno fuzzy system was leveraged for the SOC monitoring of a Li/SO2 battery [27], where subtractive clustering technique was employed to decide the model topology and antecedent parameters, and the least-squares algorithm for extracting the consequent parameters. Nevertheless, the consequent parameters were calibrated when keeping the antecedent ones fixed without a simultaneous optimization of both parts. Even if the simultaneous optimization could be exerted in this scenario, trapping into local minima would be highly possible, because of poor initialization prescribed by the non-optimization-based subtractive clustering.

An important original contribution of this article is to develop a powerful learning-based SOC estimation model that sufficiently absorbs the advantages of fuzzy clustering, subtractive clustering, genetic algorithm, direct search algorithm, and artificial neural network. The model is able to efficiently address the foregoing two issues about the model topology determination and parameter optimization, with the intention to accomplish a high-fidelity SOC indicator for a lithium-ion battery used in EVs. The modeling procedure comprises two steps. In the first step, a novel genetic algorithm based fuzzy C-means (FCM) clustering algorithm that reduces the possibility of trapping into local minima is applied to the identification of the topology and antecedent parameters of the model, in which subtractive clustering is adopted to derive the number of fuzzy clusters. The consequent parameters of the model are then calibrated by virtue of the recursive least-squares algorithm. The backpropagation learning algorithm is, in the second step, utilized to further optimize both the antecedent and consequent parameters of the model obtained in the first step. The efficacy of the learning-based SOC estimator is confirmed by experimental analysis and a variety of comparative studies with conventional approaches.

The remainder of the paper proceeds as follows. Section II introduces the lithium-ion battery experimental tests and data preparation for the learning of the SOC estimation model. The two-step model identification and optimization procedure is detailed in Section III. The model is evaluated experimentally and compared with conventional methods in Section IV. Principle conclusions are finally summarized in Section V.

II. Lithium-ion Battery Test and Data Preparation

A. Battery Test

A lithium-ion ($LiMn_2O_4$/graphite) battery module composed of sixteen serially connected cells was used for experimentation. Each healthy cell has a nominal voltage of 3.6 V and a nominal capacity of 100 Ah. The actual capacity of the battery was, however, 86.5 Ah, due to aging and deviant cell behaviors inside the battery module (there was no cell-balancing functionality). As shown in Fig. 1 [9], the whole battery testing system mainly encompasses a Digatron Battery Testing System

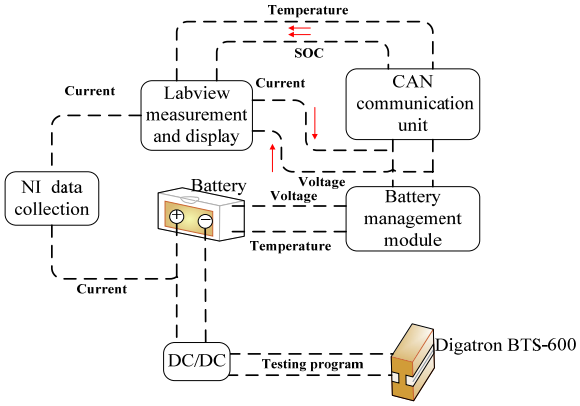


Fig. 1 Architecture of lithium-ion battery test bench [9].

(BTS-600) for cycling the battery and recording quantities, e.g., current, voltage, power, accumulative Ampere-hours, and Watt-hours, a battery management module for status monitoring, a controller area network (CAN) communication unit for signal transmission, and a Labview-based virtual measurement unit. The measured current by a high-precision Hall current sensor is transmitted to the battery management module through the CAN bus driven by the Labview program and CAN communication unit. The battery SOC estimates can be attained by the devised algorithm in the battery management module. The measured voltage, temperature, and estimated SOC are then transmitted through the CAN bus to the Labview for a real-time display. As in [9], [17], [28], [29], the Federal Urban Driving Schedule (FUDS) test was often picked as a benchmarking dataset to assess diverse SOC estimation algorithms. Therefore, the lithium-ion battery module was herein excited by concatenated FUDS cycles, where the battery temperature was controlled at 24°C with a tolerance of 2°C. A variety of signals, such as the current, voltage, and power, were sampled and recorded.

B. Data Acquirement and Preprocessing

The input variables of the predictive model should be easily measurable from a practical perspective. The battery voltage, current, power, and the discharged capacity might be good candidates. In addition, since the battery is subject to hysteresis effect, it is challenging to detect the SOC from the synchronous voltage and current values [30]. For this reason, historical current and voltage information is needed. The moving-averaged battery voltage and current were accordingly calculated as follows (moving horizon is five seconds):

$$\begin{aligned}\bar{U}(k) &= \sum_{i=k-10}^{i=k} U(i)/11, \\ \bar{I}(k) &= \sum_{i=k-10}^{i=k} I(i)/11, (k=11,12,13\cdots)\end{aligned} \tag{1}$$

where $\bar{U}, \bar{I}$, U, and I are the averaged voltage, averaged current, instantaneous voltage, and instantaneous current, respectively. The battery terminal power, averaged voltage, and averaged current are considered as the input variables, while the battery

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

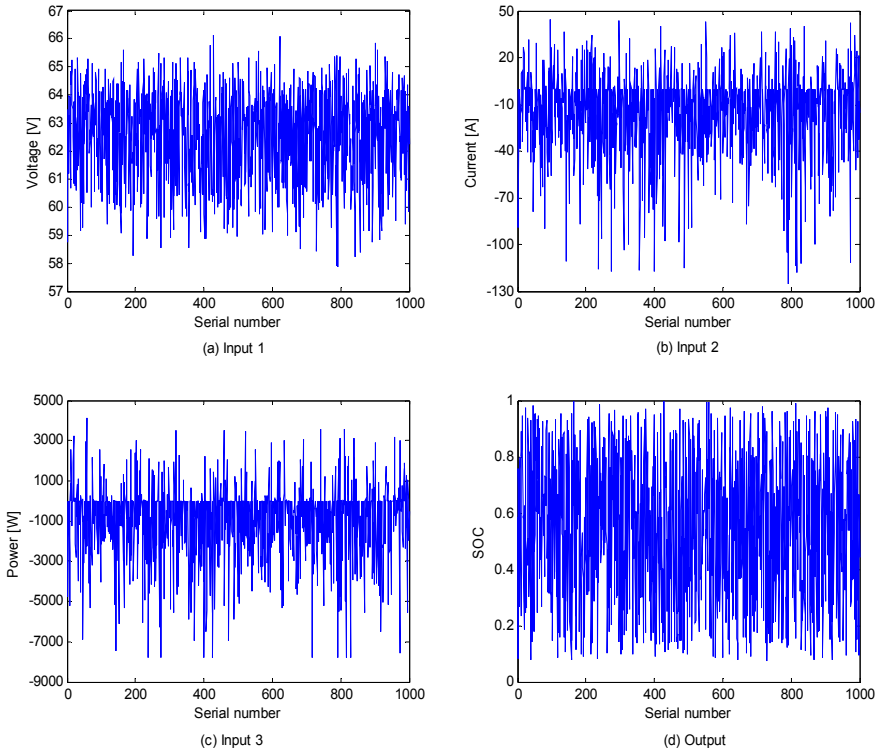


Fig. 2 Training data example: (a) Input 1: average voltage; (b) Input 2: average current: (c) Input 3: power; (d) Output: SOC.

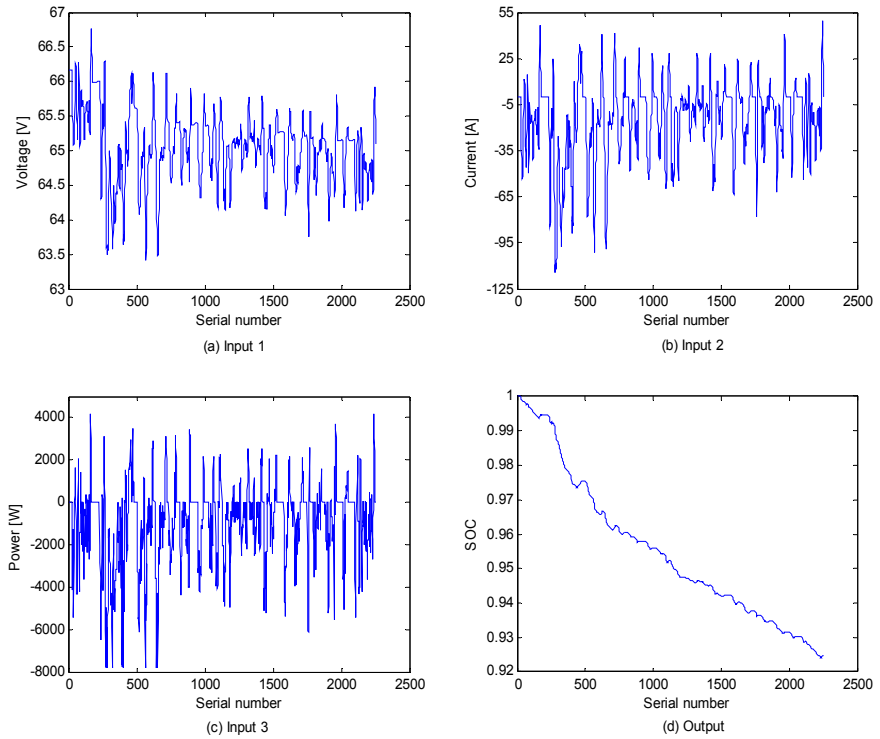


Fig. 3 Validation data in the first FUDS cycle: (a) Input 1: average voltage; (b) Input 2: average current: (c) Input 3: power; (d) Output: SOC.

SOC is the output variable. As justified and treated in [9], [16], [29], in a well-controlled experiment (i.e., with a correct initial SOC, a highly accurate current sensor, and a timely recalibration), the SOC values procured by Coulombic efficiency-perceptive current integration can be used for the supervised learning and benchmarking purposes of SOC inference models. According to a uniform distribution, 10000 input/output data pairs randomly selected in the FUDS test served as the training dataset, while the remaining "unseen" data pairs were for model verification and validation. The first 1000 training data pairs are illustrated in Fig. 2, and the validation data in the first FUDS cycle in Fig. 3. Note that only

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

a portion of the battery dataset is given for better readability, and the horizontal axis is the serial number of data pair randomly chosen rather than the time.

III. Supervised Learning of the SOC Predictive Model

A. Novel Genetic-fuzzy Clustering of Training Data

The identification of model topology poses a considerable challenge in the supervised learning process. Intuitive and experiential ideas were predominantly attempted to cope with this difficulty. The related heuristic errors may, however, yield unfavorable/non-optimal model topologies. In order to mitigate or eliminate negative effects of subjective judgment, clustering alternatives were used. Clustering data based on a measure of similarity is of great significance in engineering data analysis. Classic partition-based clustering approaches include conventional K-means algorithm, FCM algorithm, maximum entropy method, and so forth, in which the FCM algorithm is the most well-known. However, its standard formulation suffers from three intrinsic shortcomings: 1) a prior knowledge of the number of clusters must be assigned; 2) clustering result is highly sensitive to initialization; 3) there is a great possibility of getting stuck in local minima. These drawbacks do have an obviously undesirable influence on the accuracy of the neuro-fuzzy model derived by clustering. To address the drawbacks 2) and 3), we employ genetic algorithm to minimize the objective function rather than the gradient-descent method in the standard FCM algorithm. In order to compensate for the relatively poor local-optimization capability of genetic algorithm, the direct search method is further incorporated. To reduce the degree of arbitrariness when determining the number of clusters, i.e., drawback 1), subtractive clustering algorithm, which is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a dataset [31], is applied to provide the initial parameters for the genetic algorithm-based FCM clustering. The steps of the proposed genetic-fuzzy clustering algorithm are summarized in Table I, and the main tuning hyperparameters in Table II. Please refer to [32] and [33] for more theoretic properties of genetic algorithm and direct search algorithm, respectively.

The optimization outcome of the fuzzy partition is indicated in Fig. 4. The clustering result in the three dimensional input space is displayed in Fig. 5. It can be seen that six clusters are determined. The elements of the achieved best individual can be easily decoded to get the cluster centers—each four elements constitute a center. It is not very useful to show the fuzzy partition matrix, due to large amounts of data points. Instead, several validity measure metrics are calculated to examine the goodness of the partitioning result, including Partition Coefficient (PC) [34], Xie and Beni's Index (XB) [35], and PBMF [36], [37]. The corresponding results are contrasted with those of the standard FCM algorithm in Table III. The smaller XB or the larger PC and PBMF indicate a more effective partition. Our algorithm exhibits a noticeably superior clustering performance.

B. Model Topology

TABLE I
GENETIC-FUZZY CLUSTERING ALGORITHM.

Step 1, Data normalization: The features of all patterns in the training dataset are normalized between [0, 1] in linear operation so as to avoid numerical difficulties during computation and the undesirable situation that features with large numerical ranges unduly dominate those with small numerical ranges.

Step 2, Population initialization: Subtractive clustering algorithm is firstly used to partition the dataset, which generates the number of clusters and the initial cluster centers. Then, randomly select multiple individuals to form a population. Each individual is a vector whose elements are the values of cluster centers, representing a clustering structure. Ensure that at least one individual is composed of the cluster centers produced by subtractive clustering.

Step 3, Fitness function evaluation: The fitness function of each individual is calculated by using the *C*-means functional

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m \left\|\boldsymbol{x}_k - \boldsymbol{v}_i\right\|_2^2$$

where $\boldsymbol{v}_i$ is the i^{th} cluster center, m is the weight, c is the number of clusters, and μ_{ik} is the membership degree of the k^{th} pattern $\boldsymbol{x}_k$ belonging to the i^{th} cluster. The fuzzy partition matrix is constituted by

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\left\|\boldsymbol{x}_k - \boldsymbol{v}_i\right\|}{\left\|\boldsymbol{x}_k - \boldsymbol{v}_j\right\|}\right)^{\frac{2}{m-1}}} \quad (i = 1...c, k = 1...N).$$

Step 4, Population update: In light of the result from Step 3, perform selection-crossover-mutation manipulation to yield new individuals for the next population.

Step 5, Genetic algorithm termination judgment: If the weighted average change in the fitness function over 50 generations is less than the specified tolerance or the maximum number of iterations reaches, go to Step 6. Otherwise, come back to Step 3.

Step 6, Direct search: Taking the solution obtained from the above steps as the initial value, direct search algorithm further reduces the fitness function iteratively until the termination condition occurs.

TABLE II
HYPERPARAMETERS OF THE GENETIC-FUZZY CLUSTERING ALGORITHM.

Subtractive clustering: Influence range of cluster centers in all the four dimensions: 0.4; Squash factor: 1.25; Accept ratio: 0.6; Reject ratio: 0.2.
Fuzzy clustering: Weight *m*: 2.
Genetic algorithm: Population size: 200; Fitness scaling function: Proportional; Selection: Reminder; Crossover: Heuristic; Mutation: Adaptive feasible; Change tolerance over 50 generations: 0.000001; Maximum number of generations: 200.
Direct Search: Mesh initial size: 1; Expansion factor: 2.0; Contraction factor: 0.5; Minimum tolerances for mesh size: 0.000001; Maximum iteration: 300.

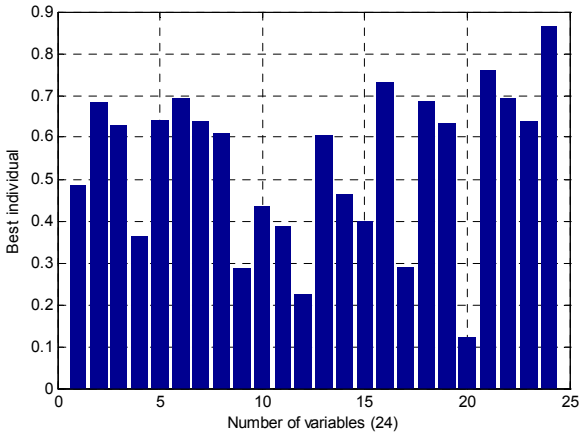


Fig. 4 The best individual.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

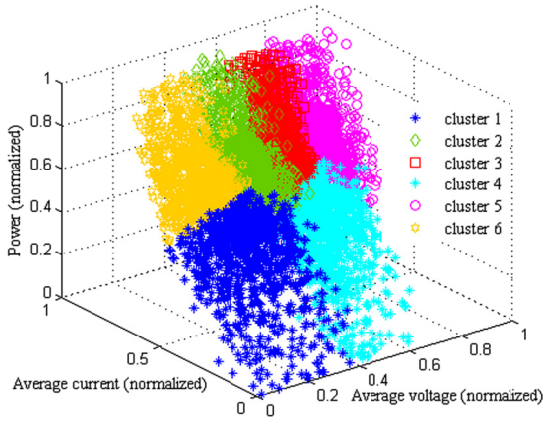


Fig. 5 Clustering result in the input space.

TABLE III
VALIDITY MEASURES.

Metric	Proposed algorithm	Standard FCM
PC	0.6547	0.5479
XB	0.2113	0.3616
PBMF	3.8655	3.2446

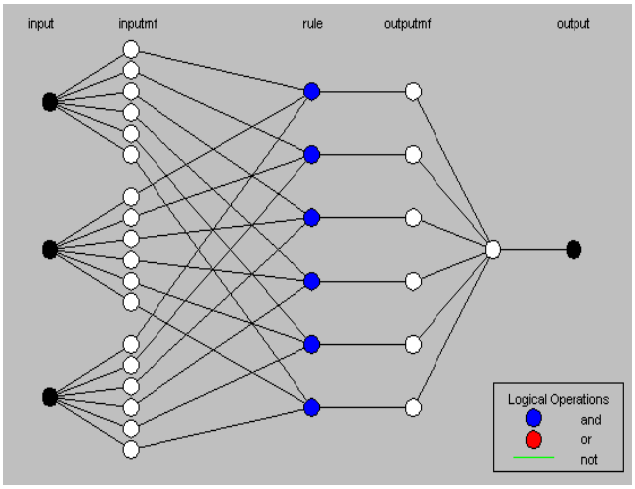


Fig. 6. Model topology identified by the genetic-fuzzy clustering. "inputmf" and "outputmf" mean the membership functions of the input and output, respectively.

According to the clustering result above, each model input has as many membership functions as the number of clusters, and the number of inference rules equals the number of clusters. Each rule maps the cluster in the input space to the corresponding cluster in the output space. For example, the rule 1 maps the cluster 1 in the input space to the cluster 1 in the output space. The resultant model topology is shown in Fig. 6.

C. *Antecedent Parameterization*

For each model input, the membership function type is two-sided Gaussian, as described by

$$M=\begin{cases} e^{-\frac{(x-\beta_1)^2}{\sigma_1^2}}, x<\beta_1 \\ e^{-\frac{(x-\beta_2)^2}{\sigma_2^2}}, x\geq\beta_2 \\ 1, \beta_1\leq x<\beta_2 \end{cases} \quad (2)$$

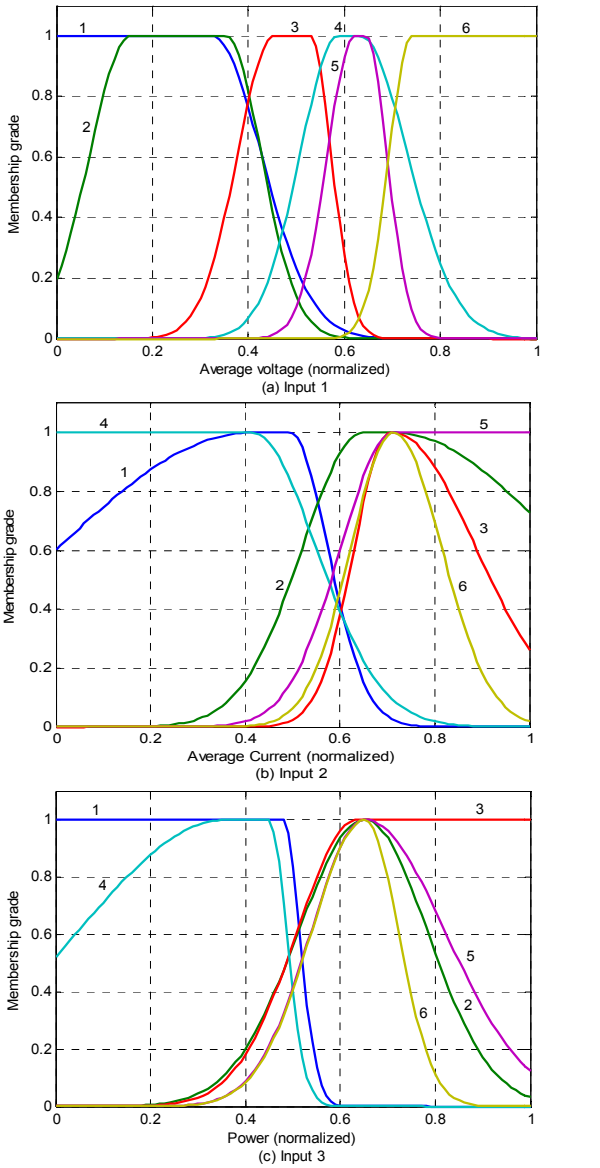


Fig. 7 Membership functions of input variables: (a) Input 1: average voltage; (b) Input 2: average current; (c) Input 3: power. Note that each input has six membership functions as labeled by the numbers.

where β_1 and σ_1 denote the width and mean value of the left half Gaussian function, respectively; β_2 and σ_2 are the width and mean value of the right half Gaussian function. M represents the membership degree of the input x. The parameters of the six two-sided Gaussian functions for each model input are derived through the projection of the fuzzy partition matrix onto the input space. The antecedent membership functions are demonstrated in Fig. 7.

D. *Consequent Parameterization*

The consequent membership functions are linear, as depicted by the following equation:

$$Y_i = a_i\bar{U} + b_i\bar{I} + c_iP + d_i \quad (3)$$

where P is the battery power, Y_i is the output of the i^{th} rule, a_i

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

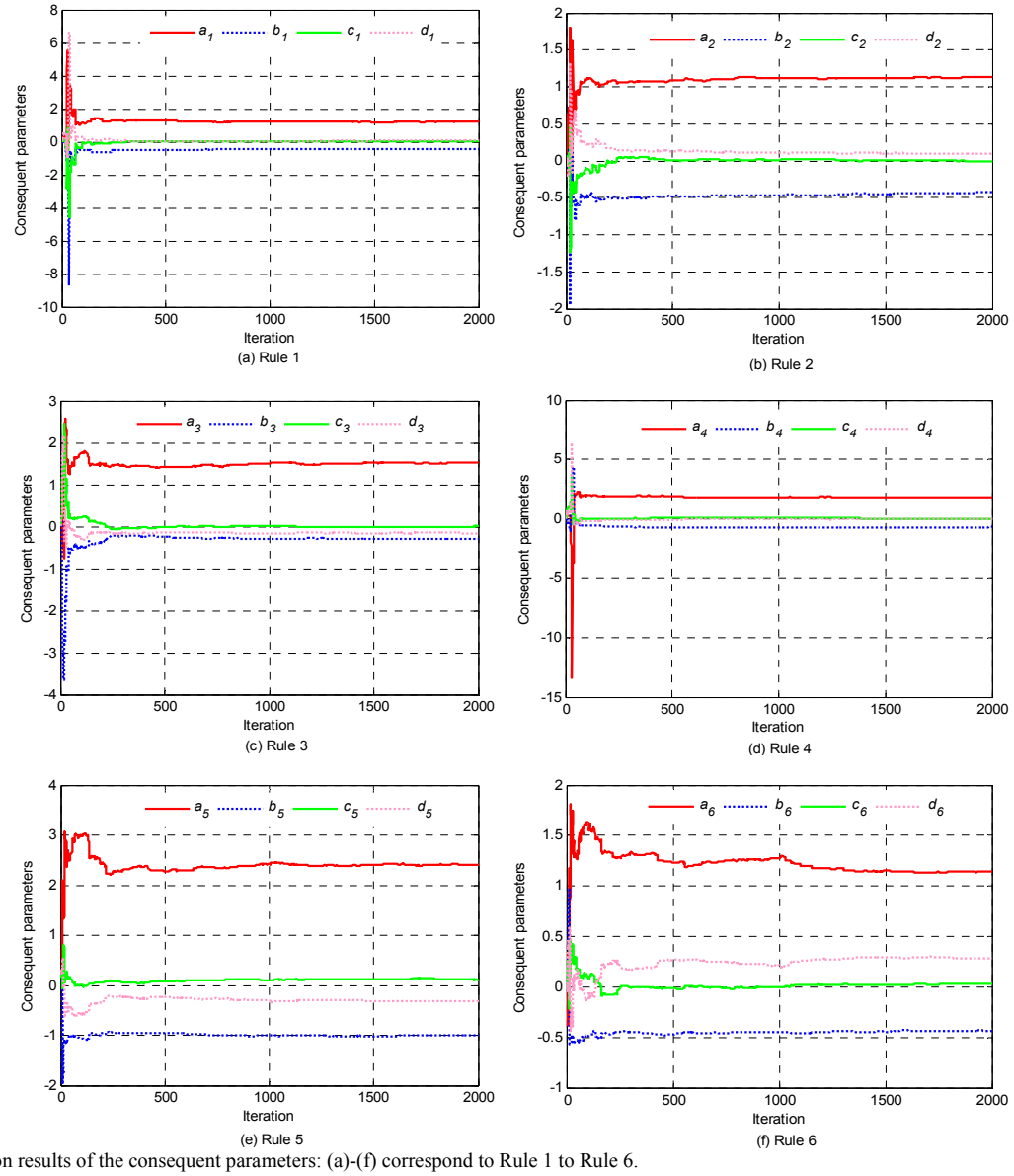


Fig. 8 Identification results of the consequent parameters: (a)-(f) correspond to Rule 1 to Rule 6.

b_i, c_i, and d_i are the consequent parameters of the i^{th} rule. Given the training data and antecedent parameters calibrated above, the consequent parameters can be identified through the recursive least-squares algorithm. The associated results are given in Fig. 8. It is clear that almost all the consequent parameters become convergent within 2000 recursions, thus proving the effectiveness of parameterization.

E. Simultaneous Optimization of Antecedent and Consequent Parameters

The nonlinear input-output surface of the achieved model is plotted in Fig. 9, which shows how the model responds to moving average current and voltage in the training dataset. In the two small corners of the surface marked by ellipses are anomalous SOC estimates that are larger than 1 or less than 0, largely deteriorating the model. There is hence a necessity and potential for simultaneously refining the antecedent and consequent parameters to decrease the spurious effect and to improve the extrapolation capability of the model.

The backpropagation algorithm of the artificial neural network [38] is used to implement such a simultaneous optimization, with the foregoing identified parameters being the initial guess. Three thousand data pairs randomly selected from the validation dataset serve as the checking dataset to reduce the likelihood of overfitting. The errors for the training and checking data over 200 epochs are shown in Fig. 10. Both the training and checking errors decrease, even though fluctuations occur in the beginning. This means that the antecedent and consequent parameters continue to be optimized iteratively. There is no steady increase of the checking error, implying that the simultaneous optimization process is not at risk of overfitting.

IV. Experimental Validation

The validation dataset was employed to examine the optimized model. For instance, the experimental and estimated

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

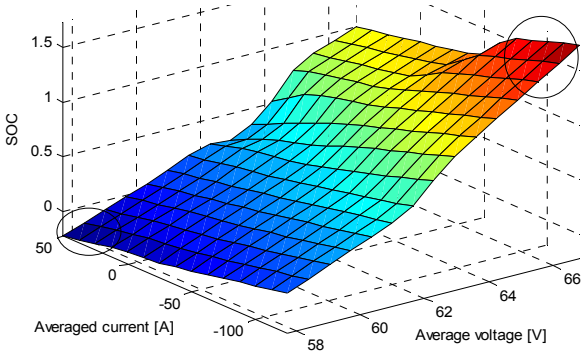


Fig. 9 Response surface of the learned model in the training data.

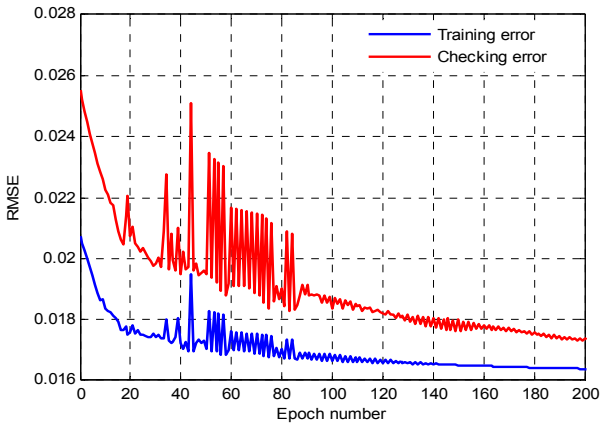


Fig. 10 Simultaneous optimization by the backpropagation learning.

SOC values in the first FUDS cycle are contrasted in Fig. 11, and the associated error is given in Fig. 12. The results in the seventh FUDS cycle are presented in Fig. 13 and Fig. 14. It is noted that the proposed SOC estimator has excellent performance, i.e., the maximal absolute error in both cycles is less than 5%. The probabilistic distribution of the estimation error in the whole validation dataset is portrayed in Fig. 15. Apparently, most of the errors are less than 3%.

For better showcasing the credibility of the proposed model, we conducted a comparative analysis with models developed by two standard fuzzy modeling methods. One is to use the conventional gradient-descent FCM to calibrate the antecedent parameters (i.e., "genfis3" function in Fuzzy Logic Toolbox in MATLAB environment); the other relies on the subtractive clustering (i.e., "genfis2" function). The consequent parameters are identified by least-squares algorithm in both standard methods. To guarantee a fair comparison, all the models have the same topology. Root mean square error (RMSE) is chosen to measure the average performance, while the probability that the absolute error is less than 5% is considered to scrutinize the worst-case performance (i.e., the robustness of the estimator). The comparison outcome is demonstrated in Table IV. Because we observed that the subtractive clustering-based model with Gaussian membership function is substantially better than that with two-side Gaussian one, two types of membership functions are considered for the standard models. It is evident that our model is the best in terms of both the average and worst-case behaviors. In addition, we can find that for the FCM-based models, the two types of membership functions

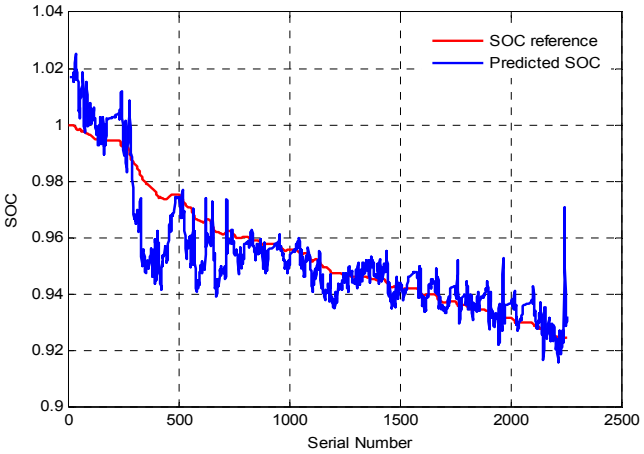


Fig. 11 Experimental and estimated SOC in the first FUDS cycle.

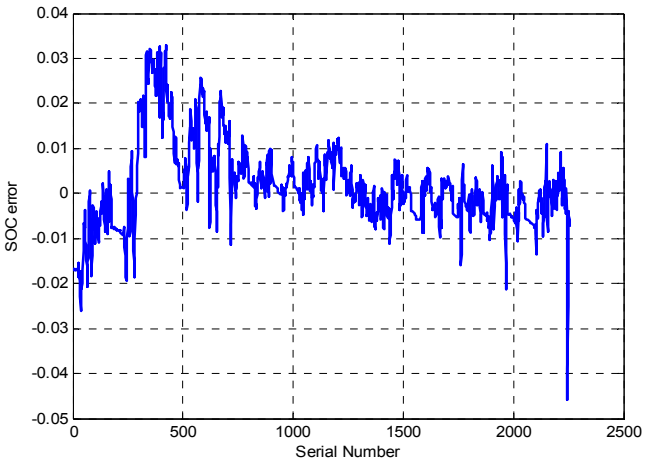


Fig. 12 SOC estimation error in the first FUDS cycle.

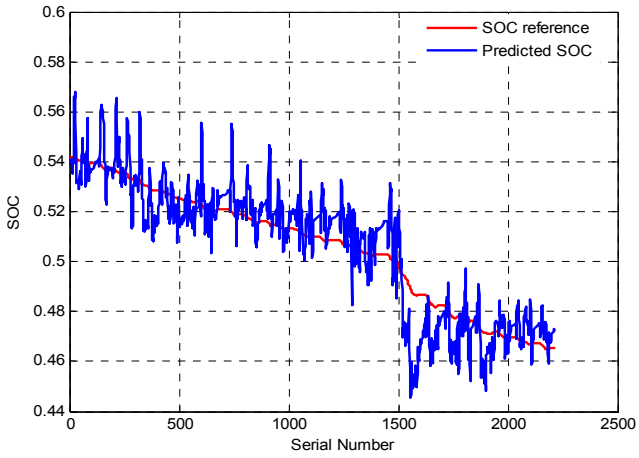


Fig. 13 Experimental and estimated SOC in the seventh FUDS cycle.

make no difference. The subtractive clustering-based model with Gaussian membership function is superior to other standard models. Further, we consider integrating the second-step optimization process into the standard models trained by "genfis2" and "genfis3" functions. The comparison outcome in this scenario is exhibited in Table V. The second-step simultaneous antecedent and consequent parameter optimization is definitely beneficial to enhancing the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

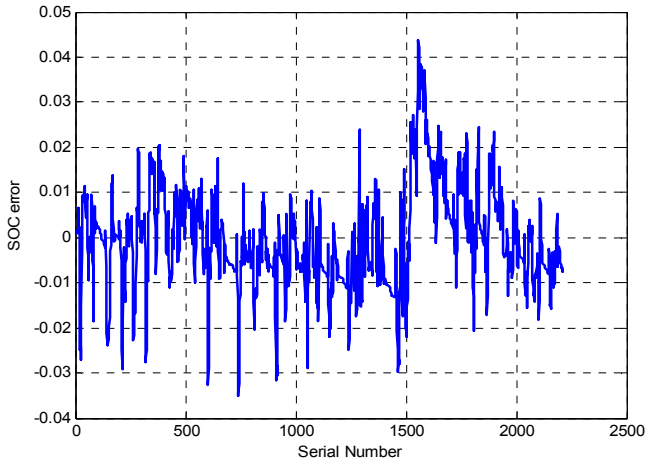


Fig. 14 SOC estimation error in the seventh FUDS cycle.

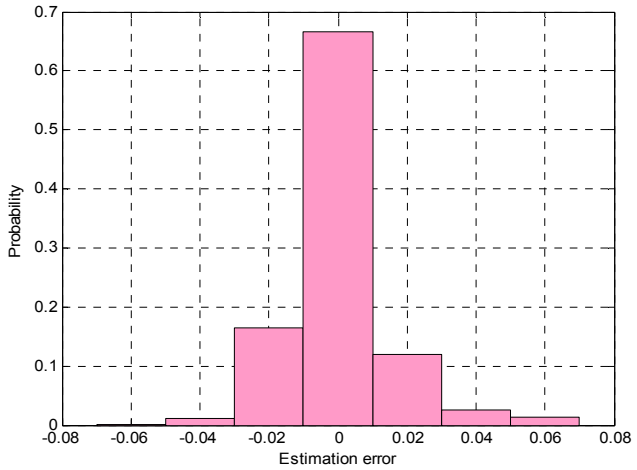


Fig. 15 Error distribution.

TABLE IV
COMPARISON BETWEEN THE PROPOSED MODEL AND STANDARD MODELS WITHOUT THE SECOND-STEP OPTIMIZATION.

Model	RMSE (%)	Probability (%)
Our model	1.68	98.12
Standard FCM-based model (two-side Gauss.)	4.78	64.65
Standard FCM-based model (Gaussian)	4.78	64.65
Subtractive clustering model (two-side Gauss.)	6.59	52.23
Subtractive clustering model (Gaussian)	2.38	95.15

TABLE V
COMPARISON BETWEEN THE PROPOSED MODEL AND STANDARD MODELS WITH THE SECOND-STEP OPTIMIZATION.

Model	RMSE (%)	Probability (%)
Our model	1.68	98.12
Standard FCM-based model (two-side Gauss.)	2.20	96.41
Standard FCM-based model (Gaussian)	2.11	96.48
Subtractive clustering model (two-side Gauss.)	2.74	93.78
Subtractive clustering model (Gaussian)	2.38	95.15

model prediction capability, precluding the last model in the table. This exception is because the checking error indicates overfitting to end the optimization, upon starting the second-step optimization. Likewise, our model is the winner when taking the second step into account. It can also be observed that the FCM-based models have a higher precision than do the subtractive clustering-based ones, as the FCM clustering can offer better initial values to the second-step optimization, evoking fewer occurrences of local minima.

V. CONCLUSION

A novel machine learning enabled SOC estimator is established for a lithium-ion battery module used in EVs. The learning mechanism includes two steps. In the first step, the genetic-fuzzy clustering technique into which the subtractive clustering and direct search algorithm are incorporated is leveraged to learn the model topology and antecedent parameters. The consequent parameters are subsequently extracted by the recursive lease-squares approach. In the second step, the backpropagation learning algorithm is capitalized to perform the simultaneous optimization of the antecedent and consequent parameters, with good initialization provided by the first step. The learning method can overcome the two challenges of the model topology determination and parameter optimization. The experimental validation confirms that the model has a satisfactory SOC tracking precision (the RMSE is merely 1.68%). A comparison with models built by the standard learning methods shows the superiority of our model in both the average and worst-case behaviors.

It is worth mentioning that the training of this machine-learning-based SOC estimator is off-line, thanks to relatively heavy optimization processes. However, after training, the established model can be readily used to real-time applications, since the model inputs, i.e., terminal power, moving average voltage and current, are practically available. Given such inputs, the trained model can rapidly and efficiently forecast the battery SOC via simple fuzzy-logic manipulation (In fact, fuzzy control has been widely employed for real-time industrial control. Of course, fuzzy controllers are typically synthesized offline).

Future work could extend the proposed machine learning approach to battery State-of-Health (SOH) monitoring, where how to construct effective signatures symptomatic of battery degradation poses a great challenge. Another technical challenge is to achieve abundant amounts of aging data taking various stressing factors into account, which is critical to the development, simulation, and verification of a robust, efficient SOH monitoring scenario.

ACKNOWLEDGMENT

Dr. Y. Yang would like to acknowledge the support of the National Natural Science Foundation of China (Project No. 51575064).

REFERENCES

[1] B. Bilgin, P. Magne, P. Malysz, Y. Yang, V. Pantelic, M. Preindl, A. Korobkine, W. Jiang, M. Lawford, and A. Emadi, "Making the case for electrified transportation," *IEEE Trans. Transport. Electrific.*, vol. 1, no. 1, pp. 4–17, Jun. 2015.

[2] M. Yilmaz and P. T. Krein, "Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles," *IEEE Trans. Power Electron.*, vol. 28, no. 5, pp. 2151–2169, May 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification

[3] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Optimal dimensioning and power management of a fuel cell/battery hybrid bus via convex programming," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 1, pp. 457–468, Feb. 2015.

[4] M. Jafari, A. Gauchia, K. Zhang, and L. Gauchia, "Simulation and analysis of the effect of real-world driving styles in an EV battery performance and ageing," *IEEE Trans. Transport. Electrific.*, 2015.

[5] L. Tang, G. Rizzoni, and S. Onori, "Energy management strategy for HEVs including battery life optimization," *IEEE Trans. Transport. Electrific.*, vol. 1, no. 3, pp. 211–222, 2015.

[6] X. Hu, S. Li, H. Peng, and F. Sun, "Charging time and loss optimization for LiNMC and $LiFePO_4$ batteries based on equivalent circuit models," *J. Power Sources*, vol. 239, pp. 449–457, Oct. 2013.

[7] L. Long and P. Bauer, "Practical capacity fading model for Li-ion battery cells in electric vehicles," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5910–5918, Dec. 2013.

[8] J. Meng, G Luo, and F. Gao, "Lithium Polymer Battery State-of-Charge Estimation Based on Adaptive Unscented Kalman Filter and Support Vector Machine," *IEEE Trans. Power Electron.*, vol. 31, no. 3, pp. 2226–2238, Mar. 2016.

[9] F. Sun, X. Hu, Y. Zou, and S. Li, "Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles," *Energy*, vol.36, no. 5, pp. 3531–3540, May 2011.

[10] Y. Zou, X. Hu, H. Ma, and S. E. Li, "Combined State of Charge and State of Health estimation over lithium-ion battery cell cycle lifespan for electric vehicles," *J. Power Sources*, vol. 273, pp. 793–803, Jan. 2015.

[11] S. Piller, M. Perrin, and A. Jossen, "Methods for state-of-charge determination and their applications," *J. Power Sources*, vol. 96, no. 1, pp. 113–120, Jun. 2001.

[12] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion battery," *J. Power Sources*, vol. 196, no. 15, pp. 6007–6014, Aug. 2011.

[13] J. H. Aylor and B. W. Johnson, "A battery state-of-charge indicator for electric wheelchairs," *IEEE Trans. Ind. Electron.*, vol. 39, no. 5, pp. 398–409, Oct. 1992.

[14] Y. Çadirci and Y. Özkazanç, "Microcontroller-based on-line state-of-charge estimator for sealed lead-acid batteries," *J. Power Sources*, vol. 129, no. 2, pp. 330–342, Apr. 2004.

[15] F. Heut, "A review of impedance measurements for determination of the state-of-charge or state-of-health of secondary batteries," *J. Power Sources*, vol. 70, no. 1, pp. 59–69, Jan. 1998.

[16] S. Rodrigues, N. Munichandraiah, and A. K. Shukla, "Review of state of charge indication of batteries by means of a. c. impedance measurements," *J. Power Sources*, vol. 87, no. 1-2, pp. 12–20, Apr. 2000.

[17] X, Hu, F. Sun, and Y. Zou, "Comparison between two model-based algorithms for Li-ion battery SOC estimation in electric vehicles," *Simulation Modelling Practice and Theory*, vol. 34, pp. 1–11, May 2013.

[18] I. S. Kim, "Nonlinear state of charge estimator for hybrid electric vehicle battery," *IEEE Trans. Power Electron.*, vol. 23, no. 4, pp. 2027–2034, Jul. 2008.

[19] I. S. Kim, "A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer," *IEEE Trans. Power Electron.*, vol. 25, no. 4, pp. 1013–1022, Apr. 2010.

[20] F. Zhang, G. Liu, L. Fang, and H. Wang, "Estimation of battery state of charge with H_∞ observer: applied to a robot for inspecting power transmission lines," *IEEE Trans. Ind. Electron.*, vol. 59, no. 2, pp. 1086–1095, Feb. 2012.

[21] X. Hu, F. Sun, and Y. Zou, "Estimation of state of charge of a lithium-ion battery pack for electric vehicles using an adaptive Luenberger observer," *Energies*, vol. 3, no. 9, pp. 1586–1603, Sep. 2010.

[22] S. Han and W. Chen, "The algorithm of dynamic battery SOC based on Mamdani fuzzy reasoning," *In Proc. The 5th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 439–443, Jinan, Shandong, China, Oct. 18-20, 2008.

[23] Y. S. Lee, J. Wang, and T. Y. Kuo, "Lithium-ion battery model and fuzzy neural approach for estimating battery state-of-charge," *In Proc. The 19th International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium & Exhibition*, Busan, South Korea, Oct. 19-23, 2002.

[24] K. T. Chau, K. C. Wu, and C. C. Chan, "A new battery capacity indicator for lithium-ion battery powered electric vehicles using adaptive neuro-fuzzy inference system," *Energ. Convers. Manage.*, vol. 45, no. 11-12, pp. 1681–1692, Jul. 2004.

[25] C. Cai, D. Du, and Z. Liu, "Battery state-of-charge (SOC) estimation using adaptive neuro-fuzzy inference system (ANFIS)," *In Proc. The 12th IEEE International Conference on Fuzzy Systems*, pp. 1068–1073, St Louis, USA, May 25-28, 2003.

[26] P. Singh, R. Vinjamu, X. Wang, and D. Reisner, "Design and implementation of a fuzzy logic-based state-of-charge meter for Li-ion batteries used in portable defibrillators," *J. Power Sources*, vol. 162, no. 2, pp. 829–836, Nov. 2006.

[27] A. J. Salkind, C. Fennie, P. Singh, T. Atwater, and D. Reisner, "Determination of state-of-charge and state-of-health of batteries by fuzzy logic methodology," *J. Power Sources*, vol. 80, no. 1-2 , pp. 293–300, Jul. 1999.

[28] J. Wang, B. Cao, Q. Chen, and F. Wang, "Combined state of charge estimator for electric vehicle battery pack," *Control Eng. Pract.*, vol. 15, no. 12, pp. 1569–1576, Dec. 2007.

[29] B. Cheng, Y. Zhou, J. Zhang, J. Wang, and B. Cao, "Ni-MH batteries state-of-charge prediction based on immune evolutionary network," *Energ. Convers. Manage.*, vol. 50, no. 12, 3078–3086, Dec. 2009.

[30] Y. Morita, S. Yamamoto, S. H. Lee, and N. Mizuno, "On-line detection of state-of-charge in lead acid battery using radial basis function neural network," *Asian J. Control*, vol. 8, no. 3, pp. 268–273, Sep. 2006.

[31] S. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 267–278, 1994.

[32] E. Falkenauer, Genetic algorithms and grouping problems, Chichester, Wiley, 1997.

[33] M. Bartholomew-Bigg, Nonlinear optimization with engineering applications, Berlin: Springer, 2008.

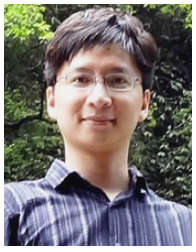
[34] J. C. Bezdek, "Mathematical models for systematics and taxonomy," *In Proc. The 8th International Conference on Numerical Taxonomy*, pp. 143–165, San Francisco, USA, 1975.

[35] X. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

[36] U. Maulik, and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recogn.*, vol. 33, no. 9, pp. 1455–1465, Sep. 2000.

[37] M. K. Pakhira, and S. Bandyopadhyay, "A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification," *Fuzzy Set Syst.*, vol. 155, no. 2, pp. 191–214, Oct. 2005.

[38] S. Haykin, Neural network - a comprehensive foundation, Upper Saddle River: Prentice-Hall, 1999.

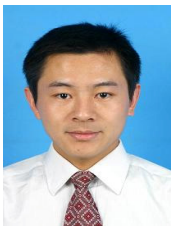


Xiaosong Hu (M'13) received the Ph.D. degree in Automotive Engineering from Beijing Institute of Technology, China, in 2012.

He did scientific research and completed the Ph.D. dissertation in Automotive Research Center at the University of Michigan, Ann Arbor, USA, between 2010 and 2012. He is currently a professor at the Department of Automotive Engineering, Chongqing University, Chongqing, China. He was a postdoctoral researcher at the Department of Civil and Environmental Engineering, University of California, Berkeley, USA, between 2014 and 2015, as well as at the Swedish Hybrid Vehicle Center and the Department of Signals and Systems at Chalmers University of Technology, Gothenburg, Sweden, between 2012 and 2014. He was also a visiting postdoctoral researcher in the Institute for Dynamic systems and Control at Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 2014. His research interests include modeling and control of alternative-energy powertrains and energy storage systems.

Dr. Hu was a recipient of Beijing Best Ph.D. Dissertation Award in 2013 and Energy Systems Best Paper Award, Dynamic Systems & Control Division (DSCD), ASME, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2015.2512237, IEEE Transactions on Transportation Electrification



Shengbo Eben Li (M'13) received the M.S. and Ph.D. degrees from Tsinghua University in 2006 and 2009.

He worked as a visiting scholar at Stanford University in 2007, a postdoctoral research fellow in University of Michigan from 2009 to 2011, and a visiting professor in University of California, Berkeley, in 2015. He is currently an associate professor in Department of Automotive Engineering at Tsinghua University. His active research activities include autonomous vehicle control, driver behaviors and modeling, control topics of battery, optimal control and multi-agent control, etc. He is the author of more than 80 journal/conference papers, and the co-inventor of more than 20 patents.

Dr. Li was the recipient of the Distinguished Doctoral Dissertation of Tsinghua University (2009), Award for Science and Technology of China ITS Association (2012), Award for Technological Invention in Ministry of Education (2012), National Award for Technological Invention in China (2013), Honored Funding for Beijing Excellent Youth Researcher (2013), NSK Sino-Japan Outstanding Paper Prize in Mechanical Engineering (2014/2015), Best Student Paper Award in 2014 IEEE Intelligent Transportation System Symposium (as student advisor), Top 10 Distinguished Project Award of NSF China (2014), Best Paper Award in 14th ITS Asia Pacific Forum, 2015. He also served as the Associate editor of IEEE Intelligent Vehicle Symposium (2012/2013), Chairman of organization committee of China ADAS forum (2013), Chinese Mechanical Engineering Society-Senior Membership, etc.

Yalian Yang received the Ph.D. degree in Mechanical Engineering from Chongqing University, China, in 2002.

He is currently a professor at the State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China. His research interests include alternative-energy drivelines, virtual reality driving simulator, and vehicle dynamic simulation and control.