

Bayesian Network Based State-of-Health Estimation for Battery on Electric Vehicle Application and Its Validation Through Real-World Data

QIAN HUO¹, ZHIKAI MA¹, XIAOSHUN ZHAO¹, TAO ZHANG², AND YULONG ZHANG¹

¹College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding 071001, China

²China North Vehicle Research Institute, Beijing 100072, China

Corresponding authors: Qian Huo (huoqian234@163.com) and Tao Zhang (ztao1208@126.com)

ABSTRACT State-of-health (SOH) estimation is crucial for ensuring efficient, reliable and safe operation of power battery in electric vehicle (EV) application. However, due to the complicated physicochemical reactions happened in battery cells, it is extremely difficult to accurately estimate SOH, especially in real-world EV application scenarios. Traditional SOH estimation methods, including both model-based and data-driven ones, are deterministic, which cannot capture the stochastic property of battery aging process aroused from the inherent inconsistency during battery production. In this paper, Bayesian network (BN), which is a probabilistic graphical modeling method for indeterministic process, is used to battery degradation modeling. Its structure is derived from existing knowledge about battery aging mechanism. Two-year operational data and capacity calibration results of 16 electric taxis are collected for model training and validation. Specifically, a systematic data filling procedure is proposed to predict the missing values of variables necessary for SOH estimation. Markov Chain Monte Carlo method is adopted to generate the samples from parameterized BN for SOH estimation. Results show that the estimation result is very close to the calibrated SOH with mean absolute error below 4%. The proposed method is promising to be applied online for SOH estimation in real-world EV application.


INDEX TERMS Electric vehicle, battery aging, state-of-health estimation, real-world data.

I. INTRODUCTION

In the face of severe energy crisis and environmental problems, governments all over the world are actively promoting the development of electric vehicles (EVs) to reduce carbon emissions and reduce fossil energy consumption [1], [2]. Lithium-ion battery has become the preferred battery type for EVs due to its advantages of high energy density, high working voltage platform, no memory effect, low self-discharge rate and long service life [3], [4]. However, the aging of lithium-ion batteries is inevitable. When the degradation of battery accumulates to a certain extent, the performance of EVs will greatly deteriorate. Capacity decrease is one of the most dominant phenomena during battery aging [5]. The decrease of battery capacity will lead to the decrease of driving range of vehicles, which will increase the mileage anxiety of users. Therefore, to ensure efficient, reliable

and safe operation of power battery, accurate estimation of state-of-health (SOH) is one of the most important functions of battery management system (BMS) [6], [7]. However, the battery capacity, as a landmark parameter of the health status, cannot be directly measured. How to accurately evaluate the battery aging under the complex and changeable operating conditions of the real vehicle becomes the key to the realization of elaborate management of BMS [8].

Electrochemical methods are commonly used to analyze the battery degradation mechanism and to reveal the relationship between SOH and the material change inside the battery. For example, Wohlfahrt-Mehrens *et al.* [9] and Yang *et al.* [10] respectively showed the change of cathode and anode materials as the battery ages using electrochemical methods like X-ray diffraction, scanning electron microscopy and so on. However, these electrochemical methods are intrusive or destructive and cannot be applied in engineering applications like EVs. An important electrochemical method that can be conducted without destroying the battery cell is

The associate editor coordinating the review of this manuscript and approving it for publication was Bernardo Tellini .

electrochemical impedance spectroscopy (EIS). A considerable amount of work has been done on the application of EIS in recognition of battery degradation [11]. However, EIS requires the battery to be excited using sinusoidal current of different frequencies, which may not be applicable in EVs.

Model-based methods are another type of approach used for SOH estimation. Most of them use equivalent circuit model (ECM) rather than electrochemical model as ECM has reasonable tradeoff between accuracy and complexity [12]. For example, Zou *et al.* [13] incorporated the capacity into the state equation of the battery ECM and estimated SOH using Extend Kalman Filter (EKF). Li *et al.* [14] used three different state estimation methods (particle filter, EKF and least squares) to estimate state-of-charge (SOC) first, and then obtained the battery capacity through dividing the accumulated charge by the SOC variation. Unfortunately, accuracy of this kind of methods heavily relies on the parameters of ECM [15], thus adaptive identification method is required to update the parameters online [16], which increases the computational burden greatly. Moreover, the employment of state estimation methods like EKF involves large matrix operations, which makes it difficult to be implemented in BMS for real-world application [17].

The third type of SOH estimation methods are data-driven ones. For example, Käbitz *et al.* [18], [19] derived an empirical model to describe SOH fade of a LiNMC cell under different combinations of current rate, temperature and depth-of-discharge (DOD) through accelerating life test. Andre *et al.* [20] constructed a SOH regression model based on support vector machine (SVM) using maximum/minimum temperature, SOC variation, mean current rate, etc. as model inputs. Other data-driven models like neural networks [7], regression tree [21] are also applied for SOH estimation. In essence, such models exploit the sophisticated and inherent relationship between SOH and the battery features like terminal voltage, current, temperature and so on from a large amount of data. As the big-data technology has been widely promoted in China EV industry, abundant data are available to train the data-driven model, which ensures that the training data has relatively wide coverage so the derived model has desirable prediction performance. For example, Wang *et al.* [22] presented a cell inconsistency evaluation model for series-connected battery systems based on real-world EV operational data, which is helpful to accurately access the battery SOH. She *et al.* [23] presented a novel battery aging assessment model based on incremental capacity analysis and neural network through analysis of real-world EVs' operational data. Tian *et al.* [24] applied box-plot method to explore battery degradation based on a large-scale electric taxi GPS and deal data. Existing researches imply that data-driven method has great potential in revealing the intrinsic relationship between battery aging and external factors.

Despite above achievements, there still exist some shortcomings in current researches. Firstly, most real-world EV data recorded in existing researches do not have capacity item

because battery capacity cannot be measured directly from on-board sensors. Therefore, when evaluating battery degradation based on real-world data, most researches used driving mileage to represent battery health status like in Ref. [23]. Despite driving mileage is proportional to capacity degradation to some extent, their relationship may not be completely linear, thus the derived battery aging model or revealed degradation mechanism may not be accurate. Secondly, data missing and distortion are inevitable when recoding operational data of real-world EVs. Most existing researches simply drop out these low-quality data. However, at one time stamp, usually not all data items are missing and missing data item may be filled based on its relationship between other observed data. Filling such missing data may increase the utilization rate and maximize the potential of data-driven model [25]. Thirdly, most existing battery aging models are deterministic while the truth is that many battery aging influence factors are uncertain [26]. In addition, even two batteries undergo the exactly same aging stress loading process, their capacity still may be different because of inherent inconsistency in production. Therefore, the deterministic model cannot capture the probabilistic properties in battery degradation.

In order to mitigate existing research gaps, this paper analyzes battery degradation for real-world electric taxis using Bayesian network (BN). The main contributions of this work include following three aspects. Firstly, by conducting capacity calibration periodically, real-world operational EV data together with capacity at specific stage are collected, which provides accurate degradation evaluation benchmark for in-depth analysis. Secondly, to fully exploit the data value and provide high-quality data source for subsequent model training, one-stage Markov chain and radius basis function neural network (RBF-NN) are adopted for vehicle-related and battery-related missing data filling respectively. Thirdly, to incorporate the indeterministic characteristic existing in battery aging process, this paper applies BN to capture the probabilistic causality between different influencing factors and battery capacity. Because in BN, each node represents a variable and a specific distribution is associated with each node, this kind of probabilistic description inherited in BN provides the capability to consider the uncertainties of the variables and provide probability distributions instead of point value estimation.

The rest of the paper is organized as follows. Section 2 describes the data collection and data cleaning procedure. Section 3 gives a brief introduction of BN. Section 4 provides a detailed construction procedure of the hierarchical BN model for battery degradation analysis. Section 5 introduces the training and validation process, followed by the key conclusions summarized in Section 6.

II. DATA COLLECTION AND CLEANING

A. ELECTRIC VEHICLE TYPE

In this research, 16 EVs are used as the investigated targets. All of these vehicles are BAIC EU260. The detailed specification of this vehicle model is listed in Table 1. It needs to

be mentioned here that all these vehicles are used as taxis. According to Ref. [27] and Ref. [28], the average daily driving distance is 49.8 km for Beijing private electric vehicles while the value is 117.98 km for Beijing electric taxis. Because taxis have relatively longer daily driving distance compared with private vehicles, their battery degradation is more obvious after short period of usage. So using taxis as investigated targets can greatly shorten the necessary data span while ensuring battery aging still be tangible.

TABLE 1. Specification of BAIC EU260.

Parameters	Value
Size	4582mm×1794mm×1515mm
Curb Weight	1583kg
Battery Capacity	41.4kWh
Battery Material	LiNMC
Motor Nominal Power	50kW
Motor Maximum Torque	260Nm
Driving Range	≥260km
Maximum Speed (km/h)	≥140km/h

B. DATA COLLECTION

In order to obtain high-frequency and high-quality EVs’ battery operational data, we installed data collector in each investigated EV and connected them to the vehicle’s Controller Area Network (CAN) bus through on-board diagnostic (OBD) port to get signals related to battery, such as current, voltage, SOC, temperature, *et al.* The GPS module inside the data collector can record the position, speed and acceleration of the vehicle while collecting the power battery data. The data collector starts to record data after the vehicle is started and stops working after the vehicle is switched off. The collected data are stored in the memory card in the message format defined by the vehicle manufacturer in the unit of a trip or a charge event. The specification of the data collector is listed in Table 2 and the installation of the data collector on the investigated vehicles is shown in Fig.1.

The memory card of the data collector is taken out every month to copy the original data. According to the vehicle manufacturer’s settings, the original data of BAIC EU260 is in *.inr format. They need to be decoded and converted to *.xls files according to the communication protocol provided by the manufacturer. In order to facilitate the subsequent usage of MATLAB for data analysis, all files are converted to *.mat format finally.

At the same time, when the investigated vehicle is brought back for data copy, it is also subjected to capacity calibration test. The test procedure is as follows: first, the vehicle is put on the dynamometer for discharging. When the vehicle’s SOC is higher than 10%, the vehicle is forced to discharge at speed of 100km/h. When the SOC decreases below 10%, the

TABLE 2. Specification of the data collector.

Parameters	Value
Size	138mm×100mm×34mm
Weight	415g
Voltage	8-32V DC
Communication Rate	50kbps, 500kbps, 1000kbps
Power	<3W
GPS Resolution	2.5m
CAN Bus	High Rate CAN Bus
Memory Capacity	32G
Working Condition	-40~+80°C, 10-80% RH
Protection Level	IP54

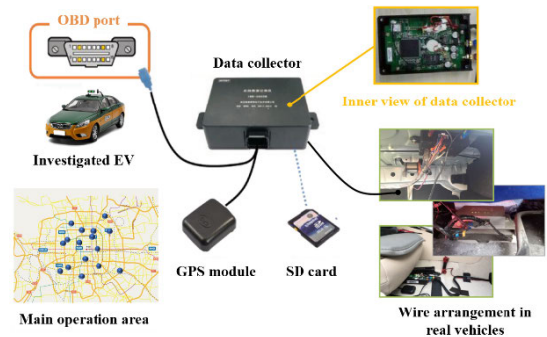


FIGURE 1. Installation of the data collector on the investigated vehicles.

vehicle’s speed is adjusted to 10km/h. The reason to use high speed when SOC is high is to shorten the discharge time while using low speed when SOC is low is to ensure the battery is discharged with small current near the end of discharge so that the battery can be close to fully discharged when SOC decreases to zero. The discharge process stops when the dashboard of the vehicle demonstrates the SOC is zero. Then the vehicle is connected to the charging pile and a device, which can record the current exchanged, is installed between the vehicle and charging pile. The device can do ampere-hour counting calculation, so when the vehicle is fully charged, the demonstrated value on the device is the calibrated capacity of the vehicle’s battery pack. It needs to be mentioned that the whole test procedure is conducted in the climatic chamber with fixed temperature 25°C. The fixed temperature is meant to avoid the influence of temperature on capacity calibration and make the calibrated SOH comparable. Before the test, the vehicle will be placed in the climatic chamber for at least 12 hours to make the vehicle close to the temperature equilibrium state.

The holistic data collection procedure is shown in Fig.2. The above data collection process is continued from June 2017 to June 2019. Finally, dataset combing vehicle operational data and calibrated capacity at specific time with timespan of 2 years is obtained.

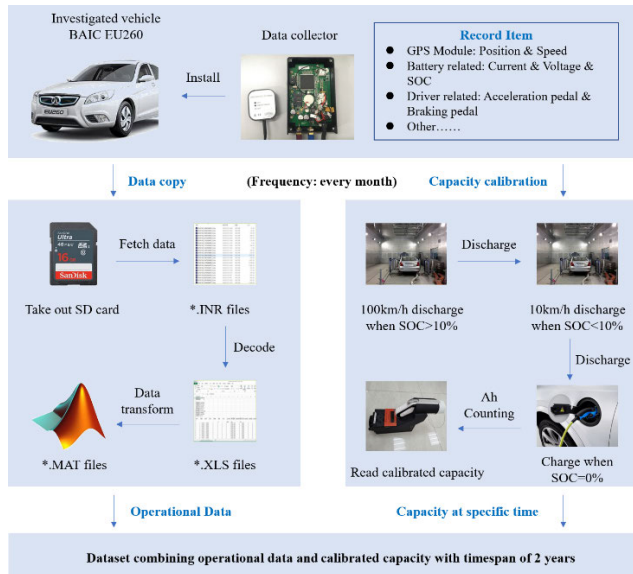


FIGURE 2. Holistic data collection procedure.

Fig.3 shows the battery degradation trajectory for three vehicles. It can be seen that at the beginning of the experiment, vehicle #4 is relatively new while vehicle #7 and #13 have experienced aging to some extent because SOH (see Section 4 for its definition) for vehicle #4 is close to 1 while for vehicle #7 and #13, SOH is around 98%. In addition, the aging process for vehicle #4 is relatively even, while vehicle #7 encounters faster battery degradation in later stage. For vehicle #13, its degradation is fast at early and final experiment stage, while slow in the middle stage. Different battery aging trajectory may be related with the different driving pattern and usage of EVs. Such observations ensure that the dataset is relatively abundant to cover different usage styles of power battery so the developed model will have much more powerful generalization capacity.

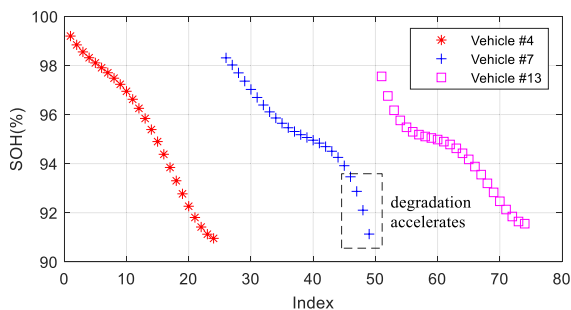


FIGURE 3. Battery degradation for three investigated vehicles.

C. DATA CLEANING

The data cleaning process is mainly for the operational data. For the calibrated capacity, it is obtained from standard experiment thus having no necessity for data processing.

Data distortion and data missing are inevitable in reality due to unreliable wire connection and unexpected conditions. Therefore, a systematic and effective data preprocessing method, which can distinguish the erroneous data and fill

the missing data and is necessary. For erroneous data identification, data items beyond the technically feasible range are firstly removed. For example, in some cases, the recorded speed may increase to over 150km/h abruptly while the adjacent speed is mainly below 80km/h, which is impossible for real-world scenarios. In addition, if the acceleration is greater than $4.5m/s^2$, we will also consider such records are unreliable like in Ref. [29]. Fig.4(a) gives an example of speed and acceleration distortion record. Some other erroneous data like Nan or discontinuity are also removed for subsequent data filling. Fig.4(b) provides such an example in demonstration of discontinuity.

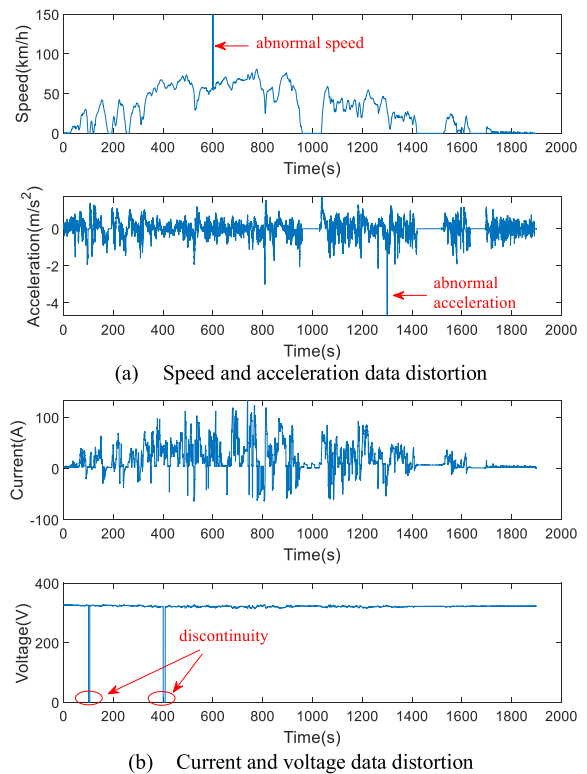


FIGURE 4. Example of erroneous data.

In this research, the measured data items that are necessary for the following battery SOH estimation include vehicle acceleration a , vehicle velocity v , battery current I , battery voltage V , battery SOC, and temperature T . Therefore, we focus on the filling of such indispensable data items. Generally, the above measurement data can be classified into two categories, namely vehicle-related data including velocity and acceleration, and battery-related data including current, voltage, SOC and temperature.

For the vehicle-related data, if only velocity is missing, its value can be filled according to acceleration data and it is vice versa for the scenario when only acceleration is missing. When both velocity and acceleration data are missing, a one-stage Markov-chain velocity predictor is constructed to predict the missing value because velocity is generally regarded to satisfy the Markov property [30]. The proposed one-stage

Markov-chain velocity predictor describes the probability of appearance of specific acceleration at the next time step when given the velocity at current time step. Fig.5 gives an example of the velocity predictor based on one-stage Markov chain. For example, when the current velocity is 15m/s, the next acceleration with probability 49% equals to 0m/s², as labelled with red star in the figure. It can also be distinguished from Fig.5 that next acceleration has higher probability around value 0m/s, which means the vehicle has more time in constant-velocity driving state. In this paper, we assume that the weekdays share the similar velocity transition style while the driving style on weekends basically is the same. Therefore, when filling the lost velocity on weekdays, velocity predictor calculated by velocity data of Monday to Friday on the last week is used. It is also similar for the missing velocity on weekends. Once the velocity predictor is obtained, it can be used to generate the missing velocity on the next step according to current existing velocity value based on Monte Carlo method [31] and this process can be continued until all continuous missing data are filled.

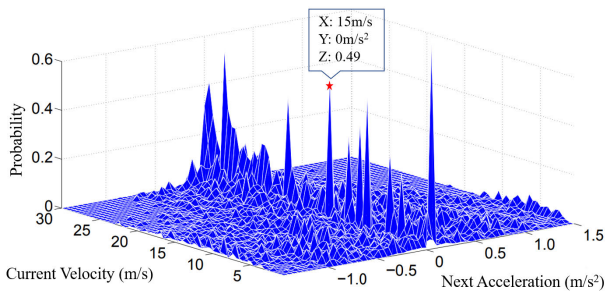


FIGURE 5. Velocity predictor based on one-stage Markov chain.

It also needs to be mentioned that in order to construct the Markov Chain based velocity predictor, velocity and acceleration need to be discretized. Nearest neighborhood method is used to attribute the velocity to one specific value when using the velocity predictor. For example, in Fig.5, velocity range (0,30m/s) is divided into 60 intervals and acceleration range (-0.5m/s², 0.5m/s²) is divided into 50 intervals, which means when using the Markov Chain, only discrete velocity values like 0, 0.5m/s, 1m/s, . . . , 29.5m/s, 30m/s and discrete acceleration values like -0.5m/s², -0.48m/s², . . . 0.48m/s², 0.5m/s² have corresponding probability. If the real velocity is 0.7m/s, then it will be regarded as 0.5m/s when looking up the probability matrix because its value is closer to 0.5m/s than 1m/s.

For the battery-related data, if temperature or SOC is missing, they can be filled through linear interpolation because these two variables change very slowly. Because battery current and voltage come from different kind of sensors, in our database, the situation where both current and voltage data are missing does not exist. Therefore, when coping with the battery-related data missing, it finally boils down to two scenarios. One scenario is current data needs to be filled with voltage, temperature and SOC are known and

the other scenario is voltage data needs to be filled with current, temperature and SOC are known. In order to fill the current or voltage missing data with other three variables observed, two regression models are developed based on RBF-NN with current and voltage as outputs respectively, as shown in Fig.6. Compared with traditional neural network, RBF-NN can provide more flexibility and reduce the dependence on the quality of datasets [32].

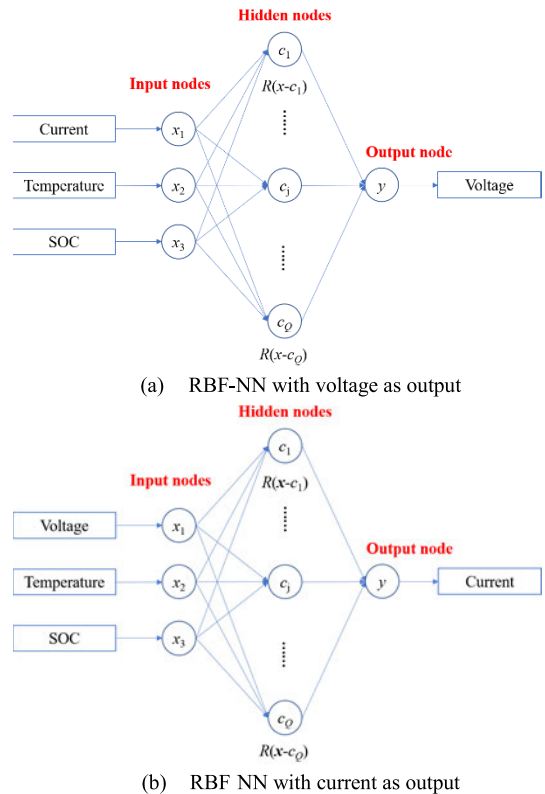


FIGURE 6. RBF-NN structure for battery voltage/current prediction.

RBF-NN is composed of input layer, hidden layer and output layer. When the input signal of RBF artificial neural network is $\mathbf{x}^p = [x_1^p, x_2^p, \dots, x_i^p, \dots, x_m^p]$, output signal is $\mathbf{y}^p = [y_1^p, y_2^p, \dots, y_k^p, \dots, y_n^p]$, its activation function of Gaussian radial basis function R can be expressed as:

$$R(\mathbf{x}^p - c_j) = \exp\left(-\frac{1}{2\sigma_j^2} |\mathbf{x}^p - c_j|^2\right) \quad (1)$$

where p is the index of the sample, i and k are the index number of input and output node respectively, m and n are the number of input and output nodes respectively, c is the center of Gaussian radial basis function, j is the index of hidden layer nodes, σ is the variance of Gaussian radial basis function.

Then the prediction output y of RBF-NN can be expressed as:

$$y_k^p = \sum_{j=1}^Q w_{jk} \exp\left(-\frac{1}{2\sigma_j^2} |\mathbf{x}^p - c_j|^2\right) \quad (2)$$

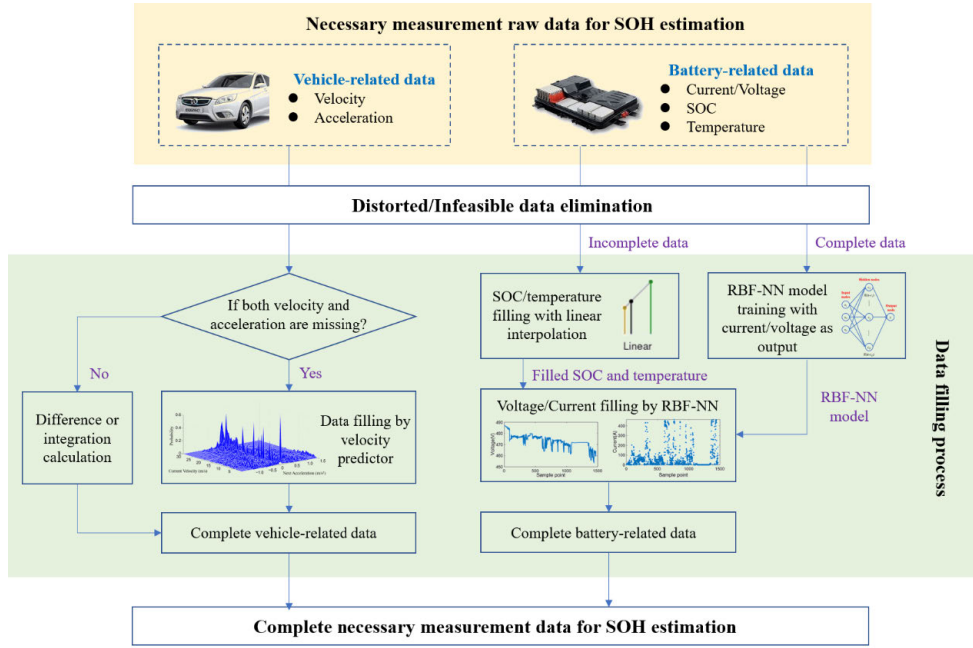


FIGURE 7. Holistic data filling procedure.

where Q is the number of hidden layer nodes, w_{jk} is the weight between the hidden layer and output layer.

The prediction error for the k th output node of the p th sample then can be expressed as:

$$MSE = \frac{1}{Pn} \sum_{p=1}^P \sum_{k=1}^n e_k^2 = \frac{1}{Pn} \sum_{p=1}^P \sum_{k=1}^n (y_p^k - \hat{y}_p^k)^2 \quad (3)$$

where P is the number of training samples, e represents the absolute error between prediction output \hat{y} and real output y .

RBF-NN uses gradient descent method to update the radial basis function center c_j and the weights between the hidden layer and output layer w_{jk} to make the predicted output approach the expected output. The update process can be expressed as:

$$c_j(t+1) = c_j(t) - \eta_1 \frac{\partial MSE(t)}{\partial c_j(t)} \quad (4)$$

$$w_{jk}(t+1) = w_{jk}(t) - \eta_2 \frac{\partial MSE(t)}{\partial w_{jk}(t)} \quad (5)$$

where $\eta_1 \in [0, 1]$ and $\eta_2 \in [0, 1]$ represent the learning rate for c_j and w_{jk} respectively. t is the iteration index.

In summary, the holistic data cleaning procedure is shown in Fig.7. Firstly, the collected data are subjected to pre-checking process to remove distorted and infeasible data. Then for the vehicle-related data, one-stage Markov Chain predictor will be used to fill the missing data if both velocity and acceleration are missing. Difference or integration calculation will be executed if only acceleration or velocity is missing. For the battery-related data, the complete part will be used to train the RBF-NN model with current/voltage as output first, then together with filled SOC and temperature by

linear interpolation, the missing current or voltage in incomplete dataset will be filled using trained RBF-NN model. Finally, combining the filled vehicle-related and battery-related data, the complete measurement dataset can be constructed for subsequent SOH model training and validation.

III. BASIC KNOWLEDGE OF BAYESIAN NETWORK

Bayesian network, also known as belief network, is an extension of Bayes method and is one of the most effective theoretical models in the field of uncertain knowledge representation and reasoning. Since it was proposed by pearl in 1988 [33], it has become a hot topic in recent years. Bayesian network uses graphical network structure to express the joint probability distribution and conditional independence of the variables intuitively, which can greatly lighten the computational burden of probabilistic reasoning.

A. CONSTRUCTION OF BAYESIAN NETWORK

BN is a directed acyclic graph (DAG), which is composed of nodes representing variables and directed edges connecting these nodes. The directed edges are represented by the arrow directing from the parent node to the child node. Fig.8 demonstrates an example of BN. Usually denotation $B \langle G, P \rangle$ is used to represent BN, where B is short for Bayesian, G stands for graph and P is short for probability. Therefore, it can be seen that a BN is comprised of two parts:

(1) A directed acyclic graph G with n nodes. The nodes in the graph represent random variables, and the edges directing from the *parent* node to *child* or *descent* node represent their dependent relationship. For example, in Fig.8, there is an edge directing from v_1 to v_3 , so node v_3 is v_1 's child or descent node, while v_1 is v_3 's parent node. Node variables can be abstractions of any problem, such as fault hypothesis, test

value, observation phenomenon, opinion consultation, etc. It is generally believed that directed edges express a causal relationship, so BNs are sometimes called causal networks. It is very important that the directed graph implies the conditional independence hypothesis, namely every variable is independent of any subset of its non-descendants conditioned on its parents. For any variable v_i , let $p(v_i)$ represent the parent nodes of v_i , $A(v_i)$ denote the non-descendants of v_i , then BN implies the following conditional independence hypothesis:

$$P(v_i | A(v_i), p(v_i)) = P(v_i | p(v_i)) \quad (6)$$

(2) A conditional probability distribution (CPD) P associated with each node. CPD for each node can be described by $P(v_i|p(v_i))$, which expresses the conditional probability relationship between the node and its parent nodes, as shown in Fig.8 for nodes v_2 to v_6 . The conditional probability of a node without any parent node is called prior probability, like node v_1 in Fig.8.

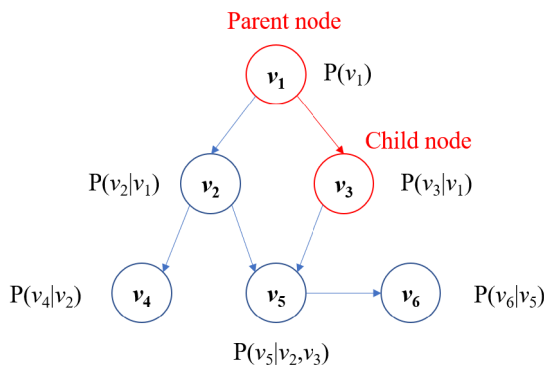


FIGURE 8. Example of Bayesian network.

With nodes and their relationships (directed edges) and CPD, BN can express the joint probability distribution of all nodes (variables) in the network:

$$P(v_1, v_2, \dots, v_n) = \prod_{i=1}^n P(v_i | p(v_i)) \quad (7)$$

Take Fig.8 as the example, the joint distribution can be expressed as:

$$\begin{aligned} P(v_1, v_2, v_3, v_4, v_5, v_6) &= P(v_1) P(v_2 | v_1) P(v_3 | v_1) P(v_4 | v_2) \\ &\quad \times P(v_5 | v_2, v_3) P(v_6 | v_5) \end{aligned} \quad (8)$$

It needs to be mentioned that the detailed BN structure changes with specific investigated problem. The detailed BN development process for battery SOH estimation will be introduced in Section IV.

B. APPROXIMATION REASONING ALGORITHM

BN reasoning is a process to calculate the probability by using the conditional independence of its expression. At present, the main accurate reasoning algorithms include polytree propagation, clique tree propagation, graph reduction and combinatorial optimization [34].

Although BN is considered as one of the best uncertain reasoning methods because of its effectiveness and solid probability theory foundation, the research focus of BN reasoning turns to approximate reasoning algorithm because for complicated structure, accurate reasoning of BN is an NP hard problem. Markov Chain Monte Carlo (MCMC) is one of the most commonly used approximation reasoning methods for BN [35], which uses a random number generator to generate a set of samples according to the CPD of BN. Then the approximate value of the probability to be calculated is obtained by processing the samples instead of directly using the joint probability distribution. Because for the developed BN in this paper, there are 20 nodes, which makes accurate reasoning extremely time-consuming. Therefore, MCMC is adopted to calculate the posterior distribution for the estimated capacity value.

The basic idea of MCMC is to construct a Markov chain so that its stable distribution is a posterior distribution of the parameters to be estimated. Samples of posterior distribution are generated through the Markov Chain, and Monte Carlo integration is performed based on the effective samples when the Markov chain reaches the stable distribution. Let ϕ denote the sampling space, n is the total number of samples generated and M is the number of samples when the chain reaches a stable state, the flowchart of MCMC algorithm is shown in Table 3.

TABLE 3. Flowchart of MCMC algorithm.

MCMC algorithm
Step 1: Construct Markov chain A Markov chain whose stationary distribution is converged to $\pi(x)$ is constructed.
Step 2: Generate samples Starting from a point $x^{(0)}$ in ϕ , the Markov Chain in step 1 is used for sampling simulation, and the point sequence $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ is generated.
Step 3: Monte Carlo integration The expected estimation of any function $f(x)$ is $E[f(x)] = \frac{1}{n-m} \sum_{t=m+1}^n f(x^{(t)})$.

Metropolis-Hasting algorithm [36] is one of the most commonly used and effective MCMC methods. It was initially proposed by Metropolis etc. in 1953 and then refined by Hasting. Let $q(x; x^{(i-1)})$ denote the transition function and $x^{(0)}$ denote the initial value. Then the i th iteration with the variable value before i th iteration $x^{(i-1)}$ is listed in Table 4.

IV. BAYESIAN NETWORK CONSTRUCTION AND VARIABLE PROBABILITY DENSITY FUNCTION TYPE DETERMINATION

In this section, firstly we constructed the BN for battery degradation analysis. The BN structure is determined based on physical equations from existing research findings about battery aging mechanisms and vehicle system dynamics. After establishing the BN structure, proper distribution type for each variable is selected based on the analysis of their unique characteristics.

TABLE 4. Flowchart of Metropolis-Hasting algorithm.

Metropolis-Hasting algorithm
Step 1: Sample an alternative value x' according to current value $x^{(i-1)}$ and transition function $q(x; x^{(i-1)})$.
Step 2: Calculate acceptance probability $\alpha(x^{(i-1)}, x') = \min\{1, \frac{\pi(x^{(i-1)}, x')}{\pi(x' x^{(i-1)})}\}$.
Step 3: Set $x^{(i)} = x'$ with probability $\alpha(x^{(i-1)}, x')$ while setting $x^{(i)} = x^{(i-1)}$ with probability $1 - \alpha(x^{(i-1)}, x')$.
Step 4: Repeat (1)–(3) for n times, then posterior samples $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ can be obtained. Targeted statistical analysis can be conducted based on these samples.

A. BAYESIAN NETWORK CONSTRUCTION

In this paper, we use battery capacity as the index for degradation evaluation. Thus, SOH variable Q_N , which is defined as the ratio of battery current capacity Q to its initial capacity Q_0 , is used here to represent battery aging status:

$$Q_N = \frac{Q}{Q_0} \tag{9}$$

According to Ref. [26], battery capacity fade can be expressed by:

$$Q_N = 1 - [(\alpha \text{SOC}_i + \beta) e^{\frac{\eta I - E_a}{RT}} Ah^\zeta + \varepsilon] \tag{10}$$

where α and β are parameters describing battery aging relationship to SOC. E_a is activation energy. η is C-rate parameter. I is the battery current. T represents the temperature. R is Molar gas constant and is selected as 8.314J/(mol · K). Ah represents the total Ampere-hour throughput. ζ is the power factor, which describes battery aging dependency to Ah . ε represents residual error of capacity estimation. According to above description, following BN in Table. 5 Step 1 can be constructed.

In addition, Ah is related to battery current I as:

$$Ah = \int I dt \tag{11}$$

Besides, battery current I can be calculated by:

$$I = \frac{P_b}{V} \tag{12}$$

where P_b represents battery power and V is battery voltage. Thus, we can add another three directed edges to the BN in Table.5 Step 1, as labelled in red in Table.5 Step 2.

Battery power can be expressed as the sum of auxiliary power P_{aux} and driving power P_w when driving and the sum of auxiliary power P_{aux} and charging power P_g when charging. Here, a trigger parameter k_t is introduced to determine which one, the driving power P_w or the charging power P_g , is chosen. The BN corresponding to above description is shown in Table.5 Step 3.

TABLE 5. Bayesian network construction procedure.

Step index	Bayesian network	Description
Step 1:		Introducing battery degradation related factors
Step 2:		Supplementing relationship among basic battery parameters
Step 3:		Adding relationship to satisfy the law of conservation of energy
Step 4:		Expanding driving demand power

At last, we expand driving power P_w . According to the vehicle system dynamic theory:

$$P_w = mgf + \frac{C_D A v^2}{21.15} + mg\theta + \delta ma \tag{13}$$

where m is vehicle mass. g is the gravitational constant, namely 9.8m/s². f is the rolling resistance coefficient. C_D is aerodynamic drag coefficient. A is the frontal area. θ is the road incline. δ is the rotational mass conversion factor. To simplify representation, we use P_{vr} to represent the parameters including $m, g, f, C_D, A, \theta, \delta$ in Eq.(13). Based on above analysis, finally the complete BN in Table.5 Step 4

can be derived to fully describe the intrinsic relationship among different factors in the process of battery degradation.

B. PARAMETER DISTRIBUTION

After determining the structure of BN, it is necessary to set proper distribution type for all node variables.

For variables Q_N and SOC, their values are between 0 and 1. Therefore, Beta probability distribution function (PDF), which is a continuous distribution defined between 0 and 1, is suitable for these two variables.

For variables Ah , V and v , they are definitely positive. Therefore, Gamma PDF, which only has probability distribution on positive range, is proper for these variables.

For parameter I , it is mostly concentrated on positive small values while having slight chance distributing over large numbers. Therefore, Rayleigh PDF is selected for it.

For parameter k_t , it is a selector and only has probability distribution over two points, namely 1 for driving while 0 for charging. Thus, the discrete Bernoulli distribution is chosen for it.

For parameter P_{vr} , it includes vehicle-related parameters and environment-related parameters. Vehicle-related parameters can be deemed as constant because they only depend on the vehicle’s physical property. For the environment-related parameters, like road incline θ and rolling resistance coefficient f , because it is difficult to get accurate distribution of these parameters, for simplicity, we take them as constant values in this paper. For parameters α , β , ε , ζ , η and E_a , they only depend on the chemical property of the battery thus can be considered as constant values.

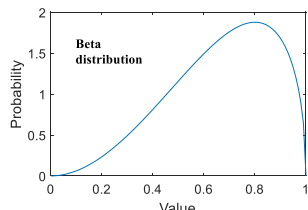
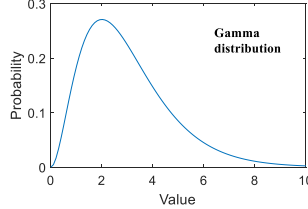
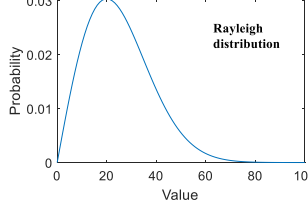
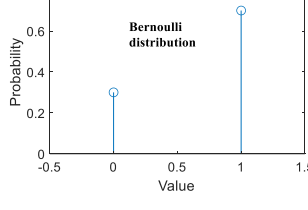
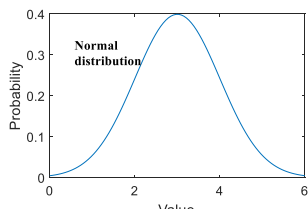
For variables T , P_g , P_b , P_w , a , their distributions have no distinct features, therefore Normal PDF is selected for these variables.

Following Table 6 concludes the distribution type for all variables and parameters.

V. MODEL TRAINING AND VALIDATION

Fig.9 demonstrates the framework of the proposed Bayesian network based capacity estimation method. Firstly, the real-world operational data of electric vehicles are collected and preprocessed to remove infeasible values. The systematic data filling procedure in Section 2 is used to predict the missing data and the complete dataset for later training and validation is obtained. Then the Bayesian network is constructed based on existing knowledge about battery degradation and vehicle system dynamics, generally the BN nodes comprise of four kinds of variables, namely measurable data, aging/driver-behavior/mechanical parameters and other variables. Proper distribution type is selected for all variables and parameters. Afterwards, the BN model is training with complete training data and the parameterized BN is obtained. Finally, MCMC method is applied to generate posterior distribution samples for capacity estimation. The used validation data is processed with sparsification for application of Metropolis-Hasting algorithm. After the number of samples satisfies the mix state criterion, the samples of capacity are

TABLE 6. Selected distribution type for all variables and parameters.

Variables and Parameters	Distribution type	Distribution
Q_N , SOC	Beta PDF	
Ah , V and v	Gamma PDF	
I	Rayleigh PDF	
k_t	Bernoulli PDF	
P_{vr} , α , β , ε , ζ , η and E_a	Constant	-
T , P_g , P_b , P_w , a	Normal PDF	

collected to fit the Beta distribution and the mode value is regarded as the estimation result.

The first step for the proposed framework is data cleaning, which mainly incorporates the filling of vehicle-related data and battery-related data. The most important part for filling of vehicle-related data is one-stage Markov chain velocity predictor. In order to verify its effectiveness, a velocity prediction test is conducted for a specific trip, as shown in Fig.10, where the prediction horizon is 5 steps ahead. It can be seen from Fig.10(a) that the predicted velocity is very close to the real value because the red line, which represents the estimated result, is basically overlapped with the blue line, which denotes the real velocity. Fig.10(b) gives a closer view of the velocity prediction result during 620s~740s. The root

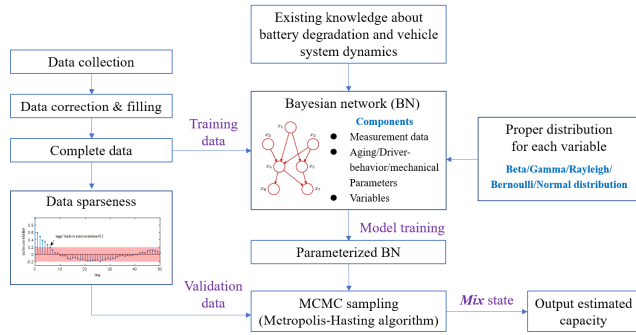


FIGURE 9. Framework of the proposed Bayesian network based capacity estimation method.

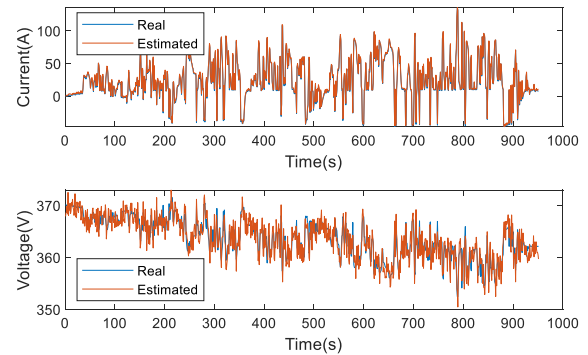
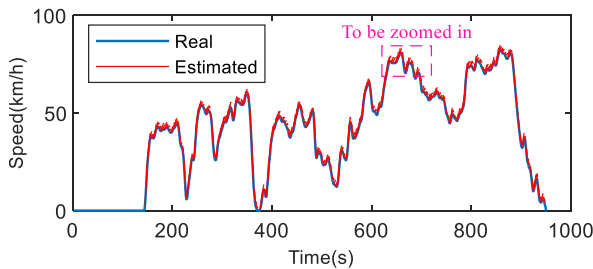
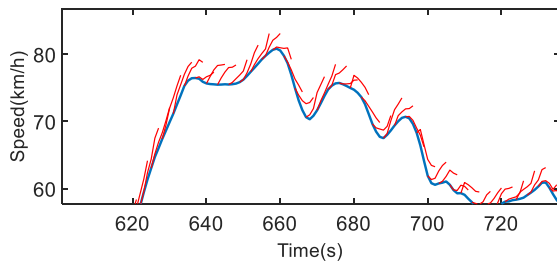


FIGURE 11. Current/voltage estimation using RBF-NN.



(a) Speed prediction for a specific trip



(b) Speed prediction for part of the trip

FIGURE 10. Speed prediction using one-stage Markov Chain.

mean square error of the prediction result over the whole trip is 2.2km/h, which means the velocity prediction performance is desirable.

In order to verify the effectiveness of the proposed RBF-NN based battery current and voltage filling method, Fig.11 shows the current and voltage estimation results for the same trip in Fig.10. It can be seen from Fig.11 that the estimated value is very close to the real value. The root mean square error of the current and voltage estimation results are 2.3A and 3.7V respectively, which means the RBF-NN can capture the intrinsic nonlinear relationship involved in the battery’s complicated physicochemical relationship.

The complete data after data cleaning are used to train and test the developed model. In the training stage, the model gets the temperature, velocity, acceleration, current, voltage and calibrated SOH as inputs and calculates the posterior distributions for the intermediate variables like Ah , SOC, required power and model parameters α , β , ε , ζ , η and E_a . During this stage, maximum likelihood estimation method [37] is used to find the optimal model parameters that maximize the occurrence of all the train data collections. Then, in the

validation stage, it uses the parameters and inputs to estimate the capacity fade without measured SOH data by MCMC method.

However, when implementing the MCMC method to get the posterior distribution for estimated variables, it is necessary to dedicate initial values for subsequent iterations. For constant variables related to battery aging involving α , β , ε , ζ , η and E_a , they are determined based on engineering experience and existing research results [38]. Their initial values need to ensure that the capacity fade estimation on the sample is within feasible range. For variables SOC, P_b , P_w and P_{aux} , they are decided according to the usage pattern of EVs. It is more suitable to initialize these variables with values of high appearance probability. Parameter P_{vr} is chosen according to the vehicle’s basic property parameters and parameter k_t is set as 1 because in most cases the vehicle is in driving state.

Because the data collector recorded the operational data of EV with 10Hz, adjacent data has strongly correlation with each other, thus the multiplication of probabilities will lead to a significantly small quantity, which makes it difficult for the Metropolis-Hasting algorithm to accept new generating samples. To address this problem, we sparse the input observation data by taking average value of adjacent data to represent the whole dataset. Here, it is necessary to dedicate the length of average operation window. Because autocorrelation of sequence data can be calculated according to:

$$A(lag) = \frac{E[(y_s - \mu)(y_{s+lag} - \mu)]}{\sigma^2} \quad (14)$$

where E is the mathematical expectation operator, s is the sampling index, y refers to any observed data, μ and σ represent average value and standard deviation of the data respectively. It is generally acknowledged that when autocorrelation is smaller than 0.2, the data points can be regarded as uncorrelated. In the developed BN, the observation data mainly refer to the vehicle’s speed, acceleration and measured ambient temperature. Fig.12 shows the autocorrelation for the acceleration samples with different lags. It can be seen that when the data interval lag is equal to or greater than 70, the data autocorrelation can be smaller than 0.2. This observation also applies to the velocity and temperature data. Thus, in this paper, we take the average of every 70 data

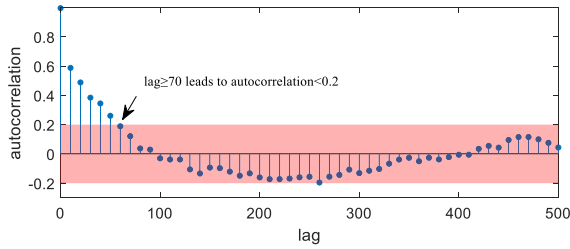


FIGURE 12. Autocorrelation for acceleration data with different lags.

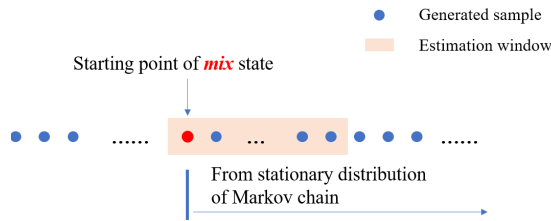


FIGURE 13. Illustration of mix state of generated samples.

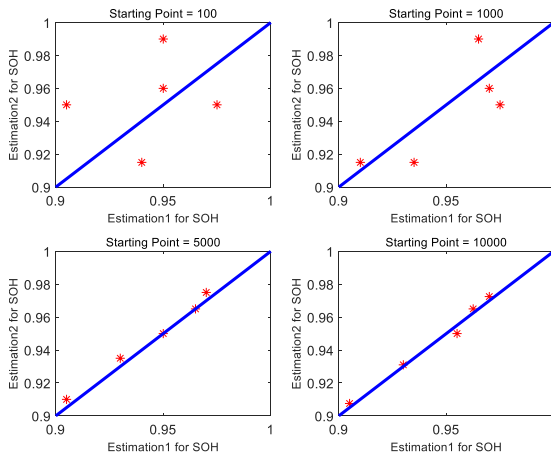


FIGURE 14. Determining starting point of mix state.

points to sparse the original dataset for Metropolis-Hasting algorithm implementation.

Another important thing when applying MCMC is to decide whether enough samples have been generated that the samples have subjected to the stationary distribution of Markov Chain, which is also called that the generated samples has come into “mix” state. Fig.13 illustrates the mix state of generated samples. Because initial state may deviate from the true value severely, if only few samples have been generated and they are used to estimate the variables, the estimation result may be inaccurate. Therefore, it is necessary to determine from which sample you begin to take as from the stationary distribution.

Fig.14 shows the mean estimation for the variable state-of-health Q_N at different aging states with two different sets of initial parameters. It can be seen that when starting point equals to 100 or 1000, estimations with different initial values are quite distinct (they deviate from the diagonal line), which means that mix state has not been reached because if the samples have entered mix state, all samples are from the same stationary distribution, so the estimation should be the same.

however, when the starting point equals to 5000 or 10000, estimations with parameters initialized differently give out similar result, which means that the generated samples have mixed. Therefore, in the following analysis, we take 5000 as the starting point index. It also needs to be mentioned that the estimation window length used here (denoted in pink in Fig.13) is set as 1000. Its length is also determined through similar method as for the starting point for mix state.

In order to verify the effectiveness of the proposed method, first we use the data to train the model and then use MCMC method to generate 1000 samples to estimate the SOH of the vehicle’s battery to be investigated. In needs to be mentioned here that the data used for model training incorporate the data of all electric taxis except the investigated vehicle, which means to verify the method’s generalization ability. Fig.15 gives an example of Beta distribution fitting for the generated samples of one specific vehicle at a specific aging state. It can be seen that the values distributed below 80% is rather rare because none training data contains SOH below 80%. Thus the estimation given by the BN only contains scarce points with SOH less than 80%, which may be caused by the generalization function of BN when traversing the Markov Chain. Because Beta distribution is a set of continuous probability distribution defined in (0, 1) interval. The PDF of Beta distribution is:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du} \quad (15)$$

Mode value M and standard deviation σ for the Beta distribution are:

$$M = \frac{\alpha - 1}{\alpha + \beta - 2} \quad (16)$$

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} \quad (17)$$

According to the fitting result, the characteristic parameters for Beta distribution here are $\alpha = 7.36$, $\beta = 1.45$, then the mode value and standard deviation of Beta distribution can be calculated and they are 93.4% and 11.8%. In this case, the true SOH of the battery is 93.9%, which is very close to the estimated result.

Fig.16 demonstrates the SOH estimation result during the 2-year experimental period for the same specific vehicle in Fig.15. In the figure, the red star represents the SOH value given by the capacity calibration experiment, while the blue line denotes the estimated result. The blue circle correspond to the mode value of the fitted Beta distribution, while the upper and lower cut-off lines deviate from the mode value with distance of $1/10$ variance. It can be seen from the figure that all experimental calibration values are distributed within the estimation range $(M - \frac{1}{10}\sigma, M + \frac{1}{10}\sigma)$. This observation indicates that the model’s training and parameters’ using are successful during the whole battery degradation period.

It is also interesting to note that as the battery degrades, the estimated variance is increasing simultaneously. The reason

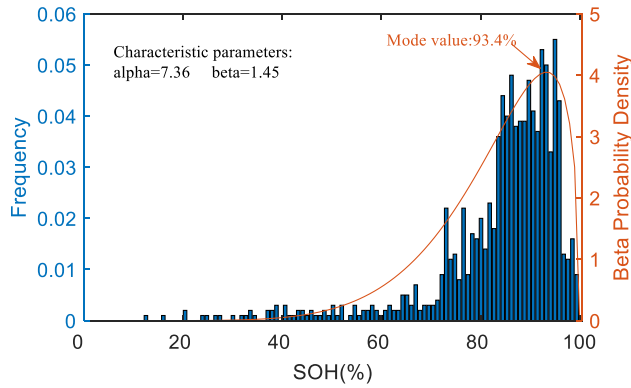


FIGURE 15. Example of Beta distribution fitting for the generated samples.

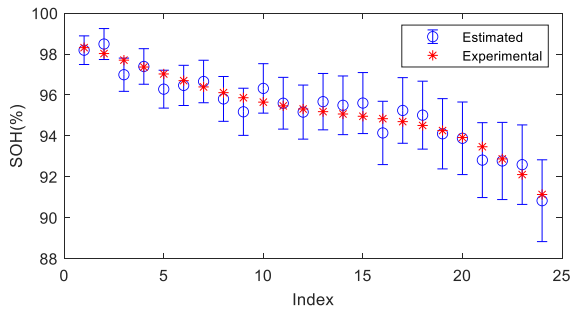


FIGURE 16. Estimated distribution of SOH compared with experimental results.

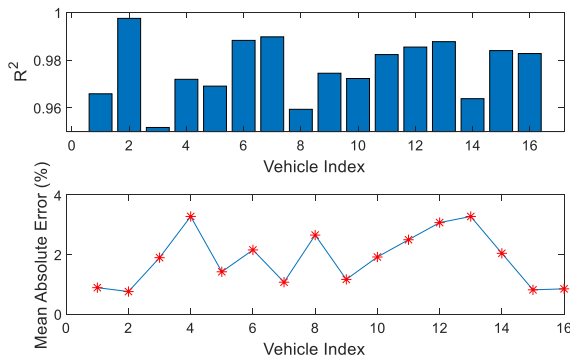


FIGURE 17. Estimation evaluation on different vehicles.

is that as the battery ages, the inconsistency of the battery will gradually become distinct, which means the SOH dispersion will enlarge. This property is also hidden in the collected data for training and in the training process, the model captures this intrinsic characteristic successfully and demonstrates it in the prediction process.

Fig.17 shows the performance of the proposed BN based capacity estimation model on different vehicles. The coefficient of determination R^2 and mean absolute error (MAE) are used to evaluate the prediction accuracy and their definitions are as follows:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (18)$$

$$MAE = \frac{1}{n} \sum_i |y_i - f_i| \quad (19)$$

where y_i represents the real value, here it refers to the SOH calibrated by experiment. f_i is the estimation value and refers to the estimated SOH given by BN, namely the mode value of the fitted Beta distribution. \bar{y} is the mean value of all y_i . n is the number of points. In this paper, the experimental period for each vehicle is 2 years and the SOH is calibrated every month, therefore n is 24 here.

Theoretically, R^2 is distributed between 0 and 1. A higher R^2 generally corresponds to a more accurate prediction model. It can be seen from Fig.17 that for all vehicles, R^2 is larger than 0.95 and MAE is lower than 4%, which verifies the accuracy of the proposed method. In addition, when estimating the battery SOH, the investigated vehicle’s operational data is unseen for the model as the data is excluded during model training process. Therefore, the accurate estimations shown in Fig.17 also verify the generalization of the proposed method.

VI. CONCLUSION

In this paper, a BN based SOH estimation framework is proposed to depict the inherent stochastic property in battery degradation process and validated by real-world operation datasets of electric taxis. For the model part, existing findings about battery aging mechanism are used to construct the BN structure and proper distribution type is selected for all variables and parameters. For the data part, to cope with the inevitable data missing problem caused by unexpected situations like unreliable wire connection, a systematic data filling procedure, which applies one-stage Markov chain to fill vehicle-related data and adopts RBF-NN to fill battery-related data, is put forward. Then the BN model is trained with complete data. Data sparsity and mix state determination are adopted to apply the MCMC method, which generates samples subjected to posterior distribution for SOH estimation. Results show that the model can capture the increasing SOH dispersion trend during battery aging. The estimated SOH range can fully cover the calibrated SOH at different aging stages.

Currently, the proposed method can only be used to estimate the battery SOH at current time based on historical operational data. It does not have the capacity to predict the future SOH change, thus cannot be used for remaining useful life (RUL) prediction directly. Another model which can predict the future working conditions of the battery needs to be incorporated into the current proposed BN-based SOH estimation framework if it means to be used for RUL prediction, which is one of the most important directions of our future work.

REFERENCES

- [1] T. Campi, S. Cruciani, F. Maradei, and M. Feliziani, “Magnetic field mitigation by multicoil active shielding in electric vehicles equipped with wireless power charging system,” *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 4, pp. 1398–1405, Aug. 2020.
- [2] S. Fan, J. Liu, Q. Wu, M. Cui, H. Zhou, and G. He, “Optimal coordination of virtual power plant with photovoltaics and electric vehicles: A temporally coupled distributed online algorithm,” *Appl. Energy*, vol. 277, Nov. 2020, Art. no. 115583.

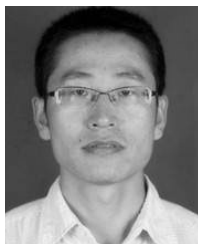
- [3] Y.-H. Chiang, W.-Y. Sean, and J.-C. Ke, "Online estimation of internal resistance and open-circuit voltage of lithium-ion batteries in electric vehicles," *J. Power Sources*, vol. 196, no. 8, pp. 3921–3932, Apr. 2011.
- [4] L. Zhang, W. Fan, Z. Wang, W. Li, and D. U. Sauer, "Battery heating for lithium-ion batteries based on multi-stage alternative currents," *J. Energy Storage*, vol. 32, Dec. 2020, Art. no. 101885.
- [5] M. Dubarry and B. Y. Liaw, "Identify capacity fading mechanism in a commercial LiFePO₄ cell," *J. Power Sources*, vol. 194, no. 1, pp. 541–549, Oct. 2009.
- [6] J. Remmlinger, M. Buchholz, M. Meiler, P. Bernreuter, and K. Dietmayer, "State-of-health monitoring of lithium-ion batteries in electric vehicles by on-board internal resistance estimation," *J. Power Sources*, vol. 196, no. 12, pp. 5357–5363, Jun. 2011.
- [7] G.-W. You, S. Park, and D. Oh, "Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach," *Appl. Energy*, vol. 176, pp. 92–103, Aug. 2016.
- [8] J. Bi, T. Zhang, H. Yu, and Y. Kang, "State-of-health estimation of lithium-ion battery packs in electric vehicles based on genetic resampling particle filter," *Appl. Energy*, vol. 182, pp. 558–568, Nov. 2016.
- [9] M. Wohlfahrt-Mehrens, C. Vogler, and J. Garche, "Aging mechanisms of lithium cathode materials," *J. Power Sources*, vol. 127, nos. 1–2, pp. 58–64, Mar. 2004.
- [10] L. Yang, M. Takahashi, and B. Wang, "A study on capacity fading of lithium-ion battery with manganese spinel positive electrode during cycling," *Electrochimica Acta*, vol. 51, no. 16, pp. 3228–3234, Apr. 2006.
- [11] R. Xiong, J. Tian, H. Mu, and C. Wang, "A systematic model-based degradation behavior recognition and health monitoring method for lithium-ion batteries," *Appl. Energy*, vol. 207, pp. 372–383, Dec. 2017.
- [12] X. Hu, S. Li, and H. Peng, "A comparative study of equivalent circuit models for li-ion batteries," *J. Power Sources*, vol. 198, pp. 359–367, Jan. 2012.
- [13] 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.
- [14] S. Li, S. Pischinger, C. He, L. Liang, and M. Stapelbroek, "A comparative study of model-based capacity estimation algorithms in dual estimation frameworks for lithium-ion batteries under an accelerated aging test," *Appl. Energy*, vol. 212, pp. 1522–1536, Feb. 2018.
- [15] X. Hu, S. Li, H. Peng, and F. Sun, "Robustness analysis of state-of-charge estimation methods for two types of li-ion batteries," *J. Power Sources*, vol. 217, pp. 209–219, Nov. 2012.
- [16] R. Xiong, F. Sun, X. Gong, and C. Gao, "A data-driven based adaptive state of charge estimator of lithium-ion polymer battery used in electric vehicles," *Appl. Energy*, vol. 113, pp. 1421–1433, Jan. 2014.
- [17] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs," *J. Power Sources*, vol. 134, no. 2, pp. 277–292, Aug. 2004.
- [18] S. Käbitz, J. B. Gerschler, M. Ecker, Y. Yurdagel, B. Emmermacher, D. André, T. Mitsch, and D. U. Sauer, "Cycle and calendar life study of a graphite/LiNi_{1/3}Mn_{1/3}Co_{1/3}O₂ li-ion high energy system. Part A: Full cell characterization," *J. Power Sources*, vol. 239, pp. 572–583, Oct. 2013.
- [19] M. Ecker, J. B. Gerschler, J. Vogel, S. Käbitz, F. Hust, P. Dechent, and D. U. Sauer, "Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data," *J. Power Sources*, vol. 215, pp. 248–257, Oct. 2012.
- [20] D. Andre, C. Appel, T. Soczka-Guth, and D. U. Sauer, "Advanced mathematical methods of SOC and SOH estimation for lithium-ion batteries," *J. Power Sources*, vol. 224, pp. 20–27, Feb. 2013.
- [21] Z. He, M. Gao, G. Ma, Y. Liu, and S. Chen, "Online state-of-health estimation of lithium-ion batteries using dynamic Bayesian networks," *J. Power Sources*, vol. 267, pp. 576–583, Dec. 2014.
- [22] Q. Wang, Z. Wang, L. Zhang, P. Liu, and Z. Zhang, "A novel consistency evaluation method for series-connected battery systems based on real-world operation data," *IEEE Trans. Transport. Electrific.*, early access, Aug. 20, 2020, doi: 10.1109/TTE.2020.3018143.
- [23] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, "Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3345–3354, May 2020.
- [24] Z. Tian, L. Tu, C. Tian, Y. Wang, and F. Zhang, "Understanding battery degradation phenomenon in real-life electric vehicle use based on big data," in *Proc. 3rd Int. Conf. Big Data Comput. Commun. (BIGCOM)*, Aug. 2017, pp. 334–339.
- [25] S. Li, H. He, and J. Li, "Big data driven lithium-ion battery modeling method based on SDAE-ELM algorithm and data pre-processing technology," *Appl. Energy*, vol. 242, pp. 1259–1273, May 2019.
- [26] M. Jafari, L. E. Brown, and L. Gauchia, "A Bayesian framework for EV battery capacity fade modeling," in *Proc. IEEE Transp. Electrific. Conf. Expo (ITEC)*, Jun. 2018, pp. 304–308.
- [27] X. Zhang, Y. Zou, J. Fan, and H. Guo, "Usage pattern analysis of Beijing private electric vehicles based on real-world data," *Energy*, vol. 167, pp. 1074–1085, Jan. 2019.
- [28] Y. Zou, S. Wei, F. Sun, X. Hu, and Y. Shiao, "Large-scale deployment of electric taxis in Beijing: A real-world analysis," *Energy*, vol. 100, pp. 25–39, Apr. 2016.
- [29] M. Tutuianu, P. Bonnel, B. Ciuffo, T. Haniu, N. Ichikawa, A. Marotta, J. Pavlovic, and H. Steven, "Development of the world-wide harmonized light duty test cycle (WLTC) and a possible pathway for its introduction in the European legislation," *Transp. Res. D, Transp. Environ.*, vol. 40, pp. 61–75, Oct. 2015.
- [30] Y. Zou, T. Liu, D. Liu, and F. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," *Appl. Energy*, vol. 171, pp. 372–382, Jun. 2016.
- [31] B. Arthur, J. Hansen, K. Malvin, and B. Kurt, "Applications of the Monte Carlo method in statistical physics," *Sov. Phys. Uspekhi*, vol. 1, no. 11, pp. 783–796, 1987.
- [32] S. P. Duvvuri and J. Anmala, "Fecal coliform predictive model using genetic algorithm-based radial basis function neural networks (GARBFNNs)," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8393–8409, Dec. 2019.
- [33] A. Prokhorchuk, J. Dauwels, and P. Jaillet, "Estimating travel time distributions by Bayesian network inference," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 5, pp. 1867–1876, May 2020.
- [34] A. Darwiche, *Modeling and Reasoning With Bayesian Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [35] K. Goswami and B.-G. Kim, "A design of fast high-efficiency video coding scheme based on Markov chain Monte Carlo model and Bayesian classifier," *IEEE Trans. Ind. Electron.*, vol. 65, no. 11, pp. 8861–8871, Nov. 2018.
- [36] J. Dhamala, H. J. Arevalo, J. Sapp, B. M. Horáček, K. C. Wu, N. A. Trayanova, and L. Wang, "Quantifying the uncertainty in model parameters using Gaussian process-based Markov chain Monte Carlo in cardiac electrophysiology," *Med. Image Anal.*, vol. 48, pp. 43–57, Aug. 2018.
- [37] D. Polani, *Probabilistic Graphical Model*. New York, NY, USA: Springer, 2013.
- [38] J. Shen, S. Dusmez, and A. Khaligh, "Optimization of sizing and battery cycle life in battery/ultracapacitor hybrid energy storage systems for electric vehicle applications," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2112–2121, Nov. 2014.



QIAN HUO received the M.S. degree from the Beijing Institute of Technology, in 2013. She is currently a Lecturer with Hebei Agricultural University. Her current research interests include automobile vibration, noise control, energy management, and machine learning.



ZHIKAI MA received the M.S. degree from the Hebei University of Technology, China, in 2013. He is currently a Lecturer with Hebei Agricultural University. His current research interests include vehicle dynamics controller, modeling and control for electrified vehicle, energy management strategy, and battery data safety analysis.



XIAOSHUN ZHAO received the Ph.D. degree from Hebei Agricultural University, in 2016. He was a Visiting Scholar with the Kansas State University. He is currently an Associate Professor with Hebei Agricultural University. His current research interests include tractor navigation systems, intelligent agricultural machinery, detection, and control for the planter.



YULONG ZHANG received the M.S. degree in vehicle engineering from the Nanjing University of Aeronautics and Astronautics, in 2008, and the Ph.D. degree in vehicle engineering from Beihang University, Beijing, China, in 2020. He is currently a Lecturer with the College of Mechanical and Electrical Engineering, Hebei Agricultural University, China. He is also familiar with the development process of model-based vehicle control strategy, master the design theory, and estimates of battery health in electric vehicles.

• • •



TAO ZHANG received the M.S. degree in mechanical engineering from the Beijing University of Technology, China, in 2015, and the Ph.D. degree in mechanical engineering from the Beijing Institute of Technology, in 2020. He is currently a Research Assistant with the China North Vehicle Research Institute. His current research interests include the hardware design of vehicle controller, vehicle dynamics control, machine learning, and reinforcement learning.