Applied Energy 232 (2018) 197–210

Contents lists available at ScienceDirect

Applied Energy

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

Random forest regression for online capacity estimation of lithium-ion batteries

Yi Li[a,*], Changfu Zou[b,*], Maitane Berecibar[a], Elise Nanini-Maury[c], Jonathan C.-W. Chan[d], Peter van den Bossche[a], Joeri Van Mierlo[a], Noshin Omar[a]

[a] Department of Mobility, Logistics and Automotive Technology Research Centre, Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium
[b] Department of Electrical Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden
[c] ENGIE LAB Laborelec, Rodestraat 125, B-1630 Linkebeek, Belgium
[d] Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

HIGHLIGHTS

- Random forest regression is proposed for on-line battery capacity estimation.
- The estimation is developed from partial charging voltage-capacity data.
- Two features indicative of battery capacity fade are extracted from charging curves.
- An incremental capacity analysis is used for assisting battery feature selection.

ARTICLE INFO

Keywords:
Lithium-ion battery
On-line capacity estimation
State of health
Random forest regression
Incremental capacity analysis

ABSTRACT

Machine-learning based methods have been widely used for battery health state monitoring. However, the existing studies require sophisticated data processing for feature extraction, thereby complicating the implementation in battery management systems. This paper proposes a machine-learning technique, random forest regression, for battery capacity estimation. The proposed technique is able to learn the dependency of the battery capacity on the features that are extracted from the charging voltage and capacity measurements. The random forest regression is solely based on signals, such as the measured current, voltage and time, that are available onboard during typical battery operation. The collected raw data can be directly fed into the trained model without any pre-processing, leading to a low computational cost. The incremental capacity analysis is employed for the feature selection. The developed method is applied and validated on lithium nickel manganese cobalt oxide batteries with different ageing patterns. Experimental results show that the proposed technique is able to evaluate the health states of different batteries under varied cycling conditions with a root-mean-square error of less than 1.3% and a low computational requirement. Therefore, the proposed method is promising for online battery capacity estimation.

1. Introduction

Lithium-ion batteries (LIBs) have been widely applied as energy storage systems, such as the fields of electrified vehicles and power grids. The biggest concern about these batteries is their limited lifetime, as their performance deteriorates with usage. To prolong a battery's longevity while ensuring reliability over the entire service life, accurate diagnosis of the state of health (SOH) in real-time is essential. The SOH reflects the current capability of a battery to store and supply energy relative to that at the beginning of its life and is an indicator to evaluate the degradation level of batteries. Quantitatively, it can be calculated by the ratio of the actual cell capacity to its initial value.

Extensive research efforts have been dedicated to SOH monitoring since the last decades, resulting in different online estimation methods. These SOH monitoring techniques can be categorized into two types, namely electrical model-based and data-driven approaches. Electrical models use passive electrical components, such as resistors and capacitors, to simulate the behavior of a battery. Enabled by these models, recursive Bayesian state estimation algorithms (such as the extended Kalman filter, EKF) [1,2] and particle filter (PF) [3,4] have been

* Corresponding authors.
E-mail addresses: li.yi@vub.be (Y. Li), changfu.zou@chalmers.se (C. Zou).

https://doi.org/10.1016/j.apenergy.2018.09.182
Received 27 February 2018; Received in revised form 18 September 2018; Accepted 23 September 2018

adopted to identify and update SOH related model parameters, e.g., capacity and internal resistance, according to data acquired during operation. Most of these filters were implemented in a joint/dual configuration for the co-estimation problem of the state of charge (SOC) and capacity. For example, Plett [1] pioneered the concurrent use of two EKFs for SOC and SOH estimation. At each time step, the results of each EKF were calculated separately and then fed into its counterpart. Zou et al. [2] proposed a general multi-timescale estimation algorithm with rigorous stability analysis and then applied it to estimate battery SOC and SOH. To further improve the estimation accuracy, Schwunk et al. [3] introduced PF for battery state estimation. Although the class of electrical model-based methods can effectively estimate the capacity under certain conditions, these techniques suffer from intensive computation required by a large number of matrix operations, thereby hindering their real-time application in battery management systems (BMSs). Data-driven methods rely on a significant amount of statistical data to predict battery ageing behavior. Because physical insights and mathematical models with a set of parameters are not needed, these methods have potential to significantly reduce the computational time in comparison with electrical model-based approaches.

One of the most widely used data-driven techniques is incremental capacity/differential voltage (IC/DV) analysis. The IC/DV analysis has proven to be a powerful tool for battery capacity estimation [5]. IC is calculated by differentiating the change in battery capacity to the change in terminal voltage during either charging or discharging, while DV is defined as the inverse of IC. With this method, the voltage plateaus on charging/discharging curves can be transformed into clearly identifiable peaks on IC/DV curves. Each peak of the curve represents a specific electrochemical process taking place in the cell and can be characterized by features such as the intensity and position [6]. These peak features are closely related to battery capacity fade and can therefore be used as indicators for the SOH estimation. Weng et al. [7] estimated the battery SOH by relating it to the peak intensity of IC curves. Li et al. [8] established a linear regression relationship between battery capacity and the peak position on IC curves. However, IC/DV curves are sensitive to measurement noise inherent in battery systems [8,9]. Accordingly, proper smoothing methods have to be proposed for obtaining smooth curves that facilitate identification and evaluation of IC/DV curve features.

In addition to the IC/DV analysis, a wealth of machine learning techniques have been devised for battery SOH estimation, such as artificial neural network (ANN) [10,11], support vector machine (SVM) [12,13], regressive vector machine (RVM) [14] and Gaussian process regression (GPR) [15,16]. These SOH estimators are trained until they learn the complex mapping from the feature space to the capacity measurement space. To estimate battery SOH accurately, a critical step in machine learning algorithms is to process the data, such as measured current, voltage, and temperature, to effectively extract representative and necessary features of the battery ageing process. In general, these features can be categorized into: internal features, processed external features, and direct external features.

In details, the internal features, like battery internal resistance, capacitance and battery SOC, cannot be measured directly from BMS sensors and must resort to parameter/state estimation algorithms. Pan et al. [17] developed an extreme learning-machine-based method for battery capacity estimation, in which parameters of an electrical model, i.e., internal ohmic resistance and polarization resistance, were considered as the input data. Then, they identified these model parameters in real-time using a recursive least square algorithm. In comparison, the processed external features, e.g., peak position and intensity, are extracted from differential charging curves, like IC/DV curves [18,19] and voltage gradient curves [10,16]. Berecibar et al. [19] estimated cell capacity with a selection of features from IC/DV curves by using three different regression methods, namely linear regression, ANN and SVM. Similar work has been conducted by Wang et al. [18] with the aid of GPR. Wu et al. [10] trained a polynomial neural network based on the arc length and curvature from voltage velocity curves, and established the relation between the geometric properties of charging curves and the battery capacity. Different to the above two types of features, the direct external features are directly recorded in BMSs during battery operation, typically including the measured terminal voltage, current, and surface temperature. Hu et al. [14] applied an RVM algorithm to learn the complex dependency of the battery capacity on characteristic features extracted from voltage and current measurements during charging operation. Recently, Richardson et al. [15] proposed a capacity estimation algorithm by using GPR based on a small portion of charging voltage–time data under a constant current. They selected a subset of the smoothed data from the charging voltage curve as the model input to reduce the computational cost. Among all these input features used for model training, the direct external features are the easiest to obtain.

Due to the limited computational capability of the present BMSs, many features of the batteries are hard to obtain during the actual operation. A state monitoring method which can directly utilize the measurable features from the BMS for battery SOH estimation is highly desired. Ideally, the battery modeling and data pre-processing steps should be avoided to reduce computational efforts. Motivated by the correlation between the battery capacity and selected features of IC curves established in our previous work [8], we seek a method capable of estimating the battery capacity accurately by directly using partial charging curves without any pre-derivation or pre-smoothing steps. Driven by this purpose, this paper proposes a novel statistical learning method, random forest (RF) regression, to diagnose the SOH of LIB based on the voltage, current and time measurements during the charging process. The RF regression, initially presented by Breiman [20], is one of the most popular supervised machine learning algorithms and has been successfully applied to both classification and regression in many different fields, such as wind power forecast [21], wheat biomass estimation [22], and spatial prediction of soil organic carbon [23]. This method has been demonstrated to have the ability of well approximating variables with nonlinear relationships and also have high robustness performance against outliers. Despite the excellent predictive performance and reliable identification of relevant variables and interactions, few has employed the RF regression for SOH monitoring of lithium-ion batteries. The present work aims to fill this gap by proposing an RF regression-based estimation algorithm for online battery capacity estimation. This proposed approach has several salient characteristics desired for SOH estimation in BMSs. First, it is able to maintain high accuracy in the absence of any pre-selection of features, although confronted by significant noise in the predictive variables. Furthermore, while overfitting can cause inaccurate estimation with new testing data and thus negatively affect the model generality, the proposed algorithm is sufficiently robust against the overfitting phenomena.[1] In addition, compared to other machine learning techniques, e.g., ANN and SVM, it only needs a few tunable parameters and therefore requires low effort for offline model tuning [20].

The remainder of this article is organized as follows. Section 2 specifies the experiments, including implementation procedures, testing cells, and equipment. The proposed RF regression technique, feature selection, and model validation tools are presented in Section 3, followed by experimental implementation and discussion of the results in Section 4. A comparative study of the proposed SOH estimation algorithm and its two benchmarking methods is conducted in Section 5. Section 6 completes the present work with a concluding summary.

[1] A model that over-fits the data means that it is too flexible so that the isolated structures (e.g., noise) that are specific to the learning set can be captured erroneously.