BATTERY STATE ESTIMATION AND ANOMALY DETECTION

USING MACHINE LEARNING

A Thesis in
Computer Science and Engineering
by
Mayank Garg

© 2017 Mayank Garg

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2017
The thesis of Mayank Garg was reviewed and approved* by the following:

Daniel Kifer  
Associate Professor of Computer Science and Engineering  
Thesis Co-Advisor

Vasant Honavar  
Professor of Computer Science and Engineering  
Thesis Co-Adviser

Mahmut Kandemir  
Professor in Charge of the Graduate Program

*Signatures are on file in the Graduate School.
Abstract

This thesis studies the battery state of charge (SOC) estimation and battery anomaly detection using machine learning technique. Classically, battery SOC is estimated using a model-based approach which requires extensive modeling, experimentation, and validation before accurate SOC estimation. The technique developed in this thesis for SOC estimation is unique and does not require battery temperature or capacity as inputs which are essential for model-based estimation.

Machine learning model is initially developed and validated using the simulated battery data. Later, the model is validated experimentally at different temperatures for NCM (Lithium nickel manganese cobalt oxide) battery. Finally, unsupervised machine learning algorithm is developed for the battery anomaly detection. Most of the battery anomalies reflect either change in battery resistance or diffusivity. Both types of changes are included in the simulated data. Anomaly detection algorithm can cluster out the faulty batteries. The methods developed in the thesis are suitable for on-line and real-time applications as it does not require a battery to be physically present in the laboratory.
# Table of Contents

List of Figures vi

Acknowledgments ix

Chapter 1
Introduction 1
  1.1 Terminology And Definition 1
  1.2 Literature Review 5
    1.2.1 SOC Estimation 5
    1.2.2 Health Monitoring And Safety 8
    1.2.3 Data Driven Approach 9
  1.3 Research Objective And Contribution 10

Chapter 2
Battery SOC Estimation Using Machine Learning 12
  2.1 Introduction 12
  2.2 Battery Data 13
    2.2.1 Data Source And Nature 13
    2.2.2 Data Size 15
  2.3 Machine Learning Model 18
    2.3.1 Objective 18
    2.3.2 Feature Selection 19
  2.4 SOC Estimation 22

Chapter 3
SOC Estimation Under Varying Conditions 26
  3.1 Introduction 26
  3.2 SOC Estimation Without Temperature Information 27
  3.3 SOC Estimation Without Capacity Information 27
List of Figures

1.1 Typical battery response: (a) Battery voltage profile, (b) Battery current profile. Charging current is negative and discharging current is positive ........................................... 6

1.2 Black box representation of the battery system. Current is input and voltage is output. SOC and SOH are battery internal states and not observable directly. .............................. 7

1.3 Battery open circuit potential vs state of charge for NCM chemistry at 25°C .................................................. 8

2.1 Experimental battery voltage (Red solid) and simulated battery voltage (Blue dotted) of a 10Ah NCM battery at 25°C. Battery response is simulated using physics based model. The same model is used to generate simulated battery data under different conditions. ......................... 13

2.2 Battery voltage response in different conditions; (a) Battery voltage at different SOC, (b) Battery voltage at different SOH, (c) Battery voltage at different temperature, and (d) Battery input current profile. ... 14

2.3 Profile 1: (a) The current profile used to generate training data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 50%. .................................................. 16

2.4 Profile 2: (a) The current profile used to generate testing data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 50%. Profile 2 is different from profile 1 to avoid overfitting. ............................................... 17

2.5 Machine learning model schematic for SOC estimation using supervised learning. Time, current, and voltage are input parameters, SOC is the target variable. ................. 21
2.6 Procedure for feature selection for SOC estimation using time series data. .................................................. 22
2.7 Battery SOC estimation using machine learning model in SOC range 50% to 90%. Machine learning model training and testing data were for 10Ah NCM battery at 25°C. . . . . 23
2.8 Battery SOC estimation using machine learning model in SOC range 20% to 50%. Machine learning model training and testing data were for 10Ah NCM battery at 25°C. . . . . 24
2.9 Battery Open circuit potential at different state of charge for NCM chemistry at 25°C. ................................. 25
3.1 Battery SOC estimation: Machine learning model was trained at 25°C, but the testing data was at 35°C. .......... 28
3.2 Battery SOC estimation: Machine learning model was trained at 25°C, but the testing data was at 45°C. ........ 29
3.3 Battery SOC estimation: Machine learning model was trained for 10Ah battery, but the testing data was for 12Ah battery. ................................................................. 30
3.4 Battery SOC estimation: Machine learning model was trained for 10Ah battery, but the testing data was for 8Ah battery. ................................................................. 31
4.1 Experimental set up: Arbin and thermal chamber. Arbin is for battery charging and discharging. Thermal chamber is to maintain battery temperature. ............................. 33
4.2 Experimental set up: Battery in a battery holder with terminal connections for voltage measurement and current input/output. ................................................................. 34
4.3 Experimental Profile 1: (a) The current profile used to generate training data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 50%. ................................. 35
4.4 Experimental Profile 2: (a) The current profile used to generate testing data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 60%. ................................. 36
4.5 Experimental Validation: Battery train data and test data at 25°C for 2.6 Ah NCM battery. ............................ 37
4.6 Experimental Validation: Battery train data at 25°C for 2.6 Ah NCM battery, but battery test data at 15°C for the same battery. ................................................................. 39
4.7 Experimental Validation: Battery train data at 25°C for 2.6 Ah NCM battery, but battery test data at 35°C for the same battery. ................................................................. 40
4.8 Experimental Validation: Battery train data at 25°C for 2.6 Ah NCM battery, but battery test data at 45°C for the same battery. ................................................................. 41

5.1 Machine learning model schematic for anomaly detection using unsupervised learning. Time, battery current, and battery voltage are the only inputs. The goal is to cluster out the faulty batteries. ................................................................. 44
5.2 Battery anomaly detection with resistance related anomalies using DBSCAN. ................................................................. 47
5.3 Battery anomaly detection with diffusion related anomalies using DBSCAN. ................................................................. 48
5.4 Battery anomaly detection with {(a)resistance (b) diffusion} related anomalies using K-means clustering. Number of clusters is given as input. ................................................................. 49
5.5 Battery anomaly detection with {(a)resistance (b) diffusion} related anomalies using Agglomerative clustering. Number of clusters is given as input. ................................................................. 50
5.6 Battery anomaly detection with {(a)resistance (b) diffusion} related anomalies using Meanshift clustering. Number of clusters is not given as input. ................................................................. 51
5.7 Battery anomaly detection with {(a)resistance (b) diffusion} related anomalies using Spectral clustering. Number of clusters is given as input. ................................................................. 52
5.8 Battery anomaly detection with {(a)resistance (b) diffusion} related anomalies using Spectral clustering. Number of clusters is not given as input. ................................................................. 53
Acknowledgments

I would like to heartily thank my advisors Prof. Daniel Kifer and Prof. Vasant Honavar for their guidance and support in my research. I extend my warm regards to my family and friends for their unwavering support. I would also like to thank Prof. Christopher Rahn for providing the battery model and allowing me to use his lab for experimental validation of my thesis.
Chapter 1  |  Introduction

Batteries are ubiquitous in our daily life. They are used everywhere from cell phone to electric vehicles, medical devices to power tools, home appliances to grid-level energy storage. Battery technology is also critical for the reduction of the greenhouse emission and renewable energy storage [1]. Car manufacturers are using lithium ion batteries for electric energy storage in hybrid electric vehicles (HEVs) and electric vehicles (EVs). Battery state of charge, state of health, and anomaly detection are necessary for safe, efficient, and reliable battery operation. The goal of the thesis is to develop new methods for battery state estimation and anomaly detection using machine learning techniques.

1.1 Terminology And Definition

This section will define basic battery system terminologies [1–3].

C-rates: Battery discharge current is often expressed as a C-rate in order to normalize against battery capacity, which is often very different between batteries. A C-rate is a measure of the rate at which a battery is discharged relative to its maximum capacity. A 1C rate means that the discharge current will discharge
the entire battery in 1 hour. For a battery with a capacity of 100 Amp-hrs, this equates to a discharge current of 100 Amps. A 5C rate for this battery would be 500 Amps, and a C/2 rate would be 50 Amps.

**Terminal Voltage (V):** Terminal voltage is the voltage between the battery terminals with the load applied. Terminal voltage varies with battery’s remaining energy and discharge/charge current.

**Open-circuit voltage (V):** Open-circuit voltage (OCV) is the voltage between the battery terminals with no load applied. Battery OCV depends on the battery state of charge, increasing with the state of charge. Figure 1.3 shows the OCV of the NCM chemistry.

**Cut-off Voltage:** Cut-off voltage is the minimum allowable battery voltage. It is this voltage that generally defines the *empty* state of the battery.

**Coulometric Capacity (Ah for a specific C-rate):** The coulometric capacity is the total Amp-hours available when the battery is discharged at a certain discharge current (specified as a C-rate) from 100 percent charged to the cut-off voltage. Capacity is calculated by multiplying the discharge current (in Amps) by the discharge time (in hours). Coulometric capacity decreases with increasing C-rate.

**Nominal Capacity, C:** Nominal capacity, C, of a cell is the maximum number of ampere-hours that can be drawn from the fully charged cell at room temperature and a C/30 rate. Manufacturers provide nominal capacity for a fresh cell. If
capacity is not known, it may take few recursive steps to get the exact nominal capacity.

**Remaining Capacity, $C_r$:** The remaining capacity $C_r(t)$ is defined as the number of ampere-hours that can be drawn from the cell at time $t$, at room temperature, and at a $C/30$ rate.

**State of Charge (SOC):** SOC is defined as:

$$SOC = \frac{\text{Available energy in the battery at time } t}{\text{Total possible energy in the battery}} = \frac{C_r(t)}{C}$$  \hspace{1cm} (1.1)

Assuming the initial SOC at $t = 0$ is 100%, and $I(t)$ is the applied current with $I > 0$ during discharge.

$$SOC = 1 - \frac{1}{C} \int_0^t I(\tau)d\tau$$  \hspace{1cm} (1.2)

Battery SOC is an indication of the amount of the energy available in the battery compare to the total battery pack capacity. It is important for any application to know the correct SOC of the battery pack.

**Depth of discharge, (DOD):**

$$DOD = 1 - SOC$$  \hspace{1cm} (1.3)

DOD indicates the amount of the battery energy used so far.
**State of Health (SOH):** Let a new battery be put into service at time $t_0$. SOH(t) of the (possibly used) battery at the current time $t$, where $t \geq t_0$, is defined to be the ratio of the battery capacities at time $t$ and $t_0$, i.e.

\[
SOH = \frac{C(t)}{C(t_0)} \quad \text{for all} \quad t \geq t_0
\] 

Battery SOH is an indication of the battery life. For a new battery, SOH is one. SOH determines whether there is a need to change a battery pack. Different applications have different SOH requirements for the battery end of life. For example, electric vehicles typically use a battery till 80% SOH [4], while a cell phone may use it for a longer time. HEVs use a battery as long as it can serve the designed HEV cycle [5]. Estimation of SOH is essential for proactive health monitoring and uninterrupted service of the battery system.

**Battery management system (BMS):** BMS is responsible for the safe and efficient battery operation. Many control and estimation algorithms are implemented in the BMS. The primary tasks of BMS are:

1. Accurate SOC and SOH estimation
2. Avoid overcharge/over discharge
3. Avoid underutilization
4. Control battery temperature within the safe limits
5. Balance individual batteries in the battery pack
Battery SOC changes when a battery pack is charged or discharged. Battery SOC change depends on the input current and the capacity of the battery pack. SOC estimation for low current and slow dynamic applications such as cell phone batteries are relatively simpler, but SOC estimation for high current and fast dynamics applications such as HEVs require sophisticated models, experimental data, and estimation algorithms. Battery SOC information is necessary for the controller to decide whether the battery will be able to provide the demanded current pulse.

Figure 1.1 shows a typical battery current and voltage profile for HEV application. A negative current is charging, and a positive current is discharging. In HEVs, charging pulse is generated either due to regenerative braking or extra engine power. For this research, the battery is assumed as a black box (Figure 1.2) where current and voltages are the input and output of the system, respectively. Battery SOC, SOH, and other properties are internal parameters which are not observable directly.

1.2 Literature Review

1.2.1 SOC Estimation

Battery SOC gives an indication how long battery will last before it needs to be recharged. It is like a fuel gauge for an electric vehicle that makes SOC estimation crucial for real-life application. Battery SOC estimation is also important for battery safety. An inaccurate SOC estimation can lead to battery overcharge or over discharge that can damage the battery and create unsafe conditions [6–9]. There are many SOC estimation techniques reported in the literature. Most of these techniques depend on some battery models. Battery models can be physics
Figure 1.1. Typical battery response: (a) Battery voltage profile, (b) Battery current profile. Charging current is negative and discharging current is positive.

Based full order models [10–13], single particle reduced order models [14–19], or equivalent circuit models [20–23]. There are advantages and disadvantages of each approach. Full order models are very accurate, but they are very slow for real-time implementation. On the other hand, reduced order and equivalent circuit models are fast for real-time implementation, but they are not very accurate. All model-based approaches require extensive experimental data for model development and validation. They infer the battery SOC from the battery current, voltage, and temperature as these are the most easily available parameters. Battery SOC estimation also depends on the battery chemistry and region of operation. For few chemistries, SOC estimation is easier and more accurate depending on their OCV.
curves.

The most common method of SOC calculation is Coulomb-counting where:

$$SOC(t) = SOC(t_0) - \frac{1}{C} \int_{t_0}^{t} I(\tau) d\tau$$  \hspace{1cm} (1.5)

Here, the discharge current is assumed positive. A major drawback of the Coulomb-counting method is biased system noise. As it integrates the input current, it will also integrate the noise. If the noise is biased, the SOC estimation by this method can be far from the actual value. Coulomb counting also needs the exact value of the battery capacity. In real life, a battery degrades over time, and battery degradation models will be necessary for the battery capacity estimation.

Another common technique for SOC estimation is voltage look up. In this method, SOC estimation is performed by creating extensive and costly tables comparing SOC and OCV under different temperatures. Battery SOC has one to one relationship with the battery OCV (Figure 1.3), so it is possible to know the exact battery SOC if the battery is in steady state. A major drawback of this method is that it will work only when a battery is in steady state. Different chemistries reach to steady state in different time. It can be as high as few hours for
some chemistries. In real life situation, such steady state condition is not feasible, so dynamic battery SOC estimation methods are desired. Plett et al. [24–26] developed the methods to use Kalman filter for SOC estimation in the model-based approach. Salkind et al. [27] applied the fuzzy logic to estimate SOC using EIS (Electrochemical impedance spectroscopy) data.

### 1.2.2 Health Monitoring And Safety

Battery health monitoring is a challenging process due to complex electrochemical processes within the battery pack. Battery SOH and health conditions are different. It is possible that battery SOH is high, but it is not safe to operate the battery due to a dendrite growth or mild internal short. Battery SOH represents the battery
capacity fade. Battery capacity fade is irreversible loss of the cyclable lithium. Different mechanisms for lithium loss are SEI layer growth, lithium plating, and dendrite formation. SEI layer growth is relatively slow and a continuous process, but lithium plating is fast at low temperature and high current [28–33]. Safari et al. [34] have shown that battery temperature and usage affect the battery capacity fade. Other than the capacity loss, lithium plating and dendrite growth can create unsafe conditions such as battery internal short. Charging/discharging current rate and depth of discharge also affect the battery life.

Battery overcharge and over discharge severely degrade battery pack and can cause unsafe conditions for battery operation. To avoid battery overcharge or over discharge, manufactures provide the maximum and minimum voltage limits. For NCM chemistry, these limits are 4.2V and 2.75 V, and for LFP chemistry, 3.6V and 2V. An accurate SOC estimation also helps to avoid battery overcharge and over discharge.

1.2.3 Data Driven Approach

Data-driven techniques learn from the historical data. They can be used to estimate battery state or detect battery anomaly. Data-driven modeling can be categorized as supervised, semi-supervised and unsupervised. Three categories are based on what type of labeled data is available: supervised learning when both healthy and faulty data is accessible; semi-supervised when just one of the classes is known; and unsupervised when no labeled data is available [35,36]. Raw battery data (battery voltage, current, and temperature) may contain specific patterns that can be useful for the feature extraction related to the battery degradation, performance, and other properties. Data-driven techniques assume battery as a black box and infer internal battery states using observed data. They do not require any chemical or
Researchers have used neural network, support vector machine, and RNN for SOC and capacity prediction. Yamazaki et al. [37] used four neurons in the input layer and ten neurons in the output layer. Four input neurons represented physical parameters: temperature, current, voltage, and impedance. Ten output neurons were battery SOC from 0 to 100% with a step of 10%. Shen [38] used seven neurons in the input layer and one neuron in the output layer. Seven neurons in the input layer were related to different charge capacities and temperature, and one output neuron was battery state of charge. Shen [38] was able to achieve SOC estimation error within 1.6%. Capizzi et al. [39] implemented RNN to predict battery terminal voltage and SOC for lead acid batteries and used them for tuning battery model parameters. Monfared et al. [40] implemented ANN to estimate the parameters of the equivalent battery model for lead acid battery. Anton et al. [41] developed SVR based model to predict SOC for high capacity lithium iron phosphate batteries. Battery current, temperature, and voltage were model input parameters and SOC was the predicted variable. The model predicted SOC within 3% error. Nuhic et al. [42] implemented SVM based SOH and RUL (remaining useful life) estimation model and was able to achieve mean square error less than $8 \times 10^{-4}$.

1.3 Research Objective And Contribution

Lithium-ion batteries are safer and more durable compared to other battery chemistries which make them suitable for wide range of applications. Still, battery safety and efficient usage are critical for widespread acceptance of the battery technology. This study develops new methods to predict battery state of charge and to detect battery anomalies. Major contributions of the thesis are as follows:
1. **Battery state of charge estimation using data:** In a conventional setting, it is time-consuming to do the experiments and develop and validate models for the SOC estimation. In this steady, battery SOC is estimated using only battery data without any physical or chemical knowledge of the battery system.

2. **Experimental Validation:** The model is validated experimentally, and the experimental results match closely with the simulation.

3. **Effect of the temperature:** Machine learning model is further extended to varying temperature conditions. This model is the first such model which does not require battery temperature as an input for SOC estimation.

4. **Anomaly detection:** Machine learning model for anomaly detection is developed using unsupervised learning.
Chapter 2  
Battery SOC Estimation Using Machine Learning

2.1 Introduction

Battery SOC is an indication about the available energy in the battery system. SOC estimation is critical for power estimation, efficient operation, and overcharge and over discharge avoidance. The methodology developed in this chapter for SOC estimation is based on data-driven modeling. Simulated battery data is used to develop and validate the model. Later, the model is validated experimentally at different temperatures. The machine learning model uses only battery current and voltage for SOC estimation. It does not require battery temperature, capacity, SOH, and internal resistance for SOC estimation, though these parameters are essential for physics or equivalent circuit model based estimation.
Figure 2.1. Experimental battery voltage (Red solid) and simulated battery voltage (Blue dotted) of a 10Ah NCM battery at $25^\circ$C. Battery response is simulated using physics based model. The same model is used to generate simulated battery data under different conditions.

2.2 Battery Data

2.2.1 Data Source And Nature

In this research simulated battery data is used. Mechatronics Research Lab in Penn State has developed high fidelity physics based battery models [14–16,19]. These models are used to generate simulated data for this research. These models were validated experimentally, so the simulated data is very close to the actual battery data. Thus, any results developed using this simulated data will apply to the real life situation. Figure 2.1 shows the battery experimental and simulated voltage for a 10 Ah NCM battery at room temperature. The physics-based model, used to generate simulated data, is also validated for different temperature and health conditions, so diverse data sets are available as input for machine learning model. Simulated data sets have following advantages:

1. White Gaussian noise is added to simulate real-life conditions.

2. Diverse and plenty of data sets are possible which are essential for training and validating the model.
3. High accuracy of the physic based model makes the simulated data closer to the actual data, so once the model is developed, experimental validation of the model will be faster.

Figure 2.2 shows the nature of the battery data. Figure 2.2 (d) shows the input current profile. Charging current is negative and discharging current is positive. Figure 2.2 (a) shows the voltage response at different SOC for the given current profile. As battery SOC increases, battery voltage increases because battery OCV is related to the battery SOC (Figure 2.9). Battery voltage output can be simplified as:

\[ V_{\text{output}} = V_{\text{OCV}} + I r_{\text{battery}} + \text{Other dynamic terms} \]  

(2.1)
Where $V_{output}$ is battery output voltage, $V_{OCV}$ is battery open circuit voltage and a function of battery SOC, $I$ is battery current, and $r_{battery}$ is battery internal resistance. Other dynamic terms depend on the battery chemistry, temperature, and capacity. In the equivalent circuit modeling, there other dynamic terms are defined using a Laplace transfer function. Figure 2.2 (b) and (c) show the battery response at different SOH and temperature, respectively. Higher battery temperature reduces the battery voltage due to reduced internal resistance, and reduced SOH increases the battery voltage due to reduced battery capacity. Battery voltages are different for different SOC, temperature, and SOH even when the input current profile is same which make SOC estimation more difficult.

### 2.2.2 Data Size

One data point is time, current, voltage, and SOC value. 10 data points are collected per seconds. Two different current profiles are used. One current profile is for training data and other for testing data. Having different profiles for training and testing data will ensure that model is not overfitting the data.

Figure 2.3 shows the current, voltage, and SOC change of profile 1. Profile 1 is 4800 seconds long. Battery data is generated for different initial SOCs. Figure 2.3(c) shows the SOC change with starting SOC of 50%. Training data set is generated at each SOC from 25% to 90% with a step of 5% SOC to cover the whole SOC range. At each SOC, 48000 data points are collected. Thus total training data is $14 \times 48000 = 672000$ data points. Each data point is at 25°C for 10 Ah NCM battery.

Figure 2.4 shows the current, voltage, and SOC change of profile 2. Profile 2 is 8000 seconds long. Battery data is generated for different initial SOCs. Figure 2.4(c) shows the SOC change with starting SOC of 50%. Test data set is generated
Figure 2.3. Profile 1: (a) The current profile used to generate training data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 50%.
Figure 2.4. Profile 2: (a) The current profile used to generate testing data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 50%. Profile 2 is different from profile 1 to avoid overfitting.
at each SOC from 25% to 90% with a step of 5% SOC to cover the whole SOC range. At each SOC, 80000 data points are collected. Thus, one testing data set is $14 \times 80000 = 1120000$ data points. Total 5 testing set are generated:

1. 10Ah, 25°C, NCM chemistry, Profile 2
2. 10Ah, 35°C, NCM chemistry, Profile 2
3. 10Ah, 45°C, NCM chemistry, Profile 2
4. 8Ah, 25°C, NCM chemistry, Profile 2
5. 8Ah, 25°C, NCM chemistry, Profile 2

Total data points for testing are $5 \times 1120000 = 5600000$. A battery should be used at room temperature (25°C), but in practice, it is not possible to maintain the temperature either due to cost or availability of the cooling system. Thus, higher and lower temperature are practical, so datasets are generated at 25°C, 35°C and 45°C for 8Ah, 10Ah, and 12 Ah battery packs.

In this chapter, train and test data for 10Ah NCM battery at 25°C are considered. In the later chapters, the model is extended for different temperatures and capacities.

### 2.3 Machine Learning Model

#### 2.3.1 Objective

A data-driven machine learning model is developed for battery SOC estimation. Figure 2.5 shows the model input and predicted parameters. It is a supervised learning problem. Battery current, voltage, and temperature are input parameters, and battery SOC is the target variable.
2.3.2 Feature Selection

Figure 2.6 summarizes the feature extraction process. Following steps are used for feature extraction:

1. 200 seconds long time series data is selected to create a feature set. Time series data includes time, current, voltage, and SOC. Time, current, and voltage are used to generate feature set. SOC is the target variable to be predicted.

2. Battery is a causal system. In any causal systems, output at any time instance will depend on the input at that time and previous time steps as well as output on previous time steps. So a voltage output function of the following form is assumed:

\[ V^n = aV^{n-1} + bV^{n-2} + cV^{n-3} + dV^{n-4} + eI^n + fI^{n-1} + gI^{n-2} + hI^{n-3} + iI^{n-4} + j \]  

(2.2)

Where \( V^n \) is the voltage at time step \( n \), \( V^{n-1} \) is the voltage at time step \( n-1 \), and so on. Similarly, \( I^{n-1} \) is current at time step \( n-1 \). \( a, b, c, d, e, f, g, h, i, j \) are the unknown parameters.

3. The assumed voltage function (Eq. 2.2) is best solved by a Laplace transfer function. Instead of using Equation 2.2, a corresponding Laplace transfer function is assumed as shown in Figure 2.6. Voltage is simulated using following equation:

\[ V_{out} = m + L^{-1}\left( r_1 + \frac{c_1}{s} + \frac{z_1}{s-p_1} + \frac{z_2}{s-p_2} \right) I(s) \]  

(2.3)

where \( s \) is Laplace parameter, and \( L^{-1} \) is Laplace inverse. With Equation
2.3, now parameters \( m, r_1, c_1, z_1, p_1, z_2, p_2 \) are fitted instead of parameters \( a, b, c, d, e, f, g, h, i, j \). There exists a relationship between Equation 2.2 and Equation 2.3, but it is not required here.

4. System is simulated for 200 seconds to fit the parameters \( m, r_1, c_1, z_1, p_1, z_2, p_2 \) in such a way the error between simulated and actual battery voltage is minimized.

\[
\text{minimize } \sum_{n=1}^{n=200} (V_{sim} - V_{actual})^2
\]

subject to

\[
V_{sim} = m + L^{-1} \left( \left( r_1 + \frac{c_1}{s} + \frac{z_1}{s-p_1} + \frac{z_2}{s-p_2} \right) I(s) \right)
\]

(2.4)

The parameters are discovered using MATLAB \textit{fmincon} and \textit{lsim} command.

5. This minimization will produce a value for parameters which will be one feature set, and SOC at time 0 second will be the target variable for machine learning algorithm. Such features and target variable will be defined for each time step.

6. KNN is applied for SOC estimation.

Thus, a continuous time series problem is converted into a discreet set problem where parameters are feature set, and SOC is the target variable. Such features and target variable will be defined for each time step.

As Equation 2.2 shows that voltage at next time step depends on the previous time step, a problem arises when the voltages \( V_{sim}^1, V_{sim}^2, \ldots \) are calculated. Voltage before time \( t=0 \) is not known for each sample. Here each sample of 200 seconds is considered independent from each other. As \( V_{sim} \) before \( t=0 \) is not known, it is assumed zero. Assuming it zero will induce an error in the voltage prediction. To minimize the effect of this assumption, Equation 2.4 is modified as follows:
Figure 2.5. Machine learning model schematic for SOC estimation using supervised learning. Time, current, and voltage are input parameters, SOC is the target variable.

\[
\begin{align*}
\text{minimize} & \quad \sum_{n=101}^{n=200} (V_{\text{sim}} - V_{\text{actual}})^2 \\
\text{subject to} & \quad V_{\text{sim}} = m + L^{-1}\left(\frac{c_1}{s} + \frac{z_1}{s-p_1} + \frac{z_2}{s-p_2}\right)I(s)
\end{align*}
\] (2.5)

The error is now minimized from time \(t=101\) to time \(t=200\). This works well because the error induced by the assumption dies down over time. It is found that 100 time steps are more than sufficient time for this error to die down.
Figure 2.6. Procedure for feature selection for SOC estimation using time series data.

2.4 SOC Estimation

SOC is estimated using k nearest neighbor algorithm. The results showed a good match between actual and predicted SOC. Figure 2.7 shows the SOC prediction between .55 to .9 SOC range. The mean error in the SOC estimation for this range is 1.04 %. Figure 2.8 shows the SOC prediction between .2 to .5 SOC range. The mean error in the SOC estimation for this range is 2.03 %. In range .2 to .5, the error is relatively higher because the prediction error depends on the relationship between the battery OCV and SOC. If the relationship has a higher slope, the estimation error will be less. Figure 2.9 shows the relationship between OCV and SOC for NCM chemistry, the OCV curve has a higher slope in .55 to .9 SOC range.
Figure 2.7. Battery SOC estimation using machine learning model in SOC range 50% to 90%. Machine learning model training and testing data were for 10Ah NCM battery at 25°C.
Figure 2.8. Battery SOC estimation using machine learning model in SOC range 20% to 50%. Machine learning model training and testing data were for 10Ah NCM battery at 25°C.
Figure 2.9. Battery Open circuit potential at different state of charge for NCM chemistry at 25°C.
Chapter 3  
SOC Estimation Under Varying Conditions

3.1 Introduction

In the previous chapter, the data-driven model is developed for battery SOC estimation. Fitting the parameters of the voltage equation (Eq. 2.2) using \textit{fmincon} is critical for the model. These parameters have some physical significance, and it is found out that few of these parameters are not affected by the change in battery temperature or capacity. Those parameters are helpful to estimate the battery SOC at different temperature and capacity, so SOC estimation model is extended for different temperature and capacity batteries. It is shown in this chapter that battery temperature and capacity are not required for the SOC estimation using this method. This is one of the most important advantages of the method as information about battery capacity and temperature are essential for any model based SOC estimation.
3.2 SOC Estimation Without Temperature Information

Machine learning model is trained only once using the battery data at 25°C for 10Ah NCM battery, but testing data is generated at different temperature for different battery profile. Figure 3.1 shows the SOC estimation at 35°C battery temperature. Here the model can predict the correct battery SOC at the temperature 10°C higher than the temperature of the training data with mean SOC error around 1%. Figure 3.2 shows the SOC estimation at 45°C battery temperature. Here the model can predict the correct battery SOC at the temperature 20°C higher than the temperature of the training data with mean SOC error around 1.25%. Both figures clearly show that the SOC estimation using this technique does not require the knowledge of the battery temperature.

3.3 SOC Estimation Without Capacity Information

Machine learning model is trained only once using the battery data at 25°C for 10Ah NCM battery, but testing data is generated at different battery capacities for different battery profile. Figure 3.3 shows the SOC estimation for 12Ah battery pack. Here the model can predict the correct battery SOC for 20% higher capacity battery pack compared to the training data with mean SOC error around 1%. Figure 3.4 shows the SOC estimation for 8Ah battery pack. Here the model can predict the correct battery SOC for 20% lower capacity battery pack compared to the training data with mean SOC error around 1.5%. Both figures clearly show that the SOC estimation using this technique does not require the knowledge of the battery capacity.
Figure 3.1. Battery SOC estimation: Machine learning model was trained at 25°C, but the testing data was at 35°C.
Battery SOC prediction at 45°C using battery data of 25°C

Figure 3.2. Battery SOC estimation: Machine learning model was trained at 25°C, but the testing data was at 45°C.
Figure 3.3. Battery SOC estimation: Machine learning model was trained for 10Ah battery, but the testing data was for 12Ah battery.
Figure 3.4. Battery SOC estimation: Machine learning model was trained for 10Ah battery, but the testing data was for 8Ah battery.
Chapter 4  
Experimental Validation Of SOC Estimation

4.1 Introduction

In the previous chapters, data-driven model for SOC estimation is developed and validated with simulation data for different temperatures and capacities. In this chapter, the model is validated experimentally. The model is trained using the experimental data for a 2.6Ah NCM battery at 25°C, and tested at the same temperature but for the different current profile. Later, the experiment is performed at different temperatures, and SOC is estimated using the model trained at 25°C.

4.2 Experimental Setup

Figure 4.1 shows the experimental set up for the battery SOC estimation. Arbin [43] is used for charging and discharging of the battery pack. Arbin interacts with a PC and charges/discharges the battery pack as per the predefined HEV cycle. Arbin can provide a required charge/discharge current as well as can measure the...
battery voltage and current. Battery temperature plays an important role for the battery voltage output, so the battery is placed inside a thermal chamber to control the battery temperature efficiently. Figure 4.2 shows a 2.6 Ah NCM battery in a battery holder with terminal connections for voltage measurement and current input/output.
4.3 Experimental Data

One data point is time, current, voltage, and SOC value. 10 data points are collected per second. Two different current profiles are used. One current profile is for training data and other for testing data. Having different profiles for training and testing data will ensure that model is not overfitting the data as mentioned before also.

Figure 4.3 shows the current, voltage, and SOC change of profile 1. Profile 1 is 3400 seconds long, so total data points are 34000. Battery data is generated with initial SOC of 50% for 2.6Ah NCM battery at 25°C.

Figure 4.4 shows the current, voltage, and SOC change of profile 2. Profile 2 is 3600 seconds long, so total data points from one experiment are 36000. Battery data is generated with initial SOC of 60% for 2.6Ah NCM battery at 15°C, 25°C, 35°C, and 45°C. Total data points for testing are $4 \times 36000 = 144000$. 

Figure 4.2. Experimental set up: Battery in a battery holder with terminal connections for voltage measurement and current input/output.
Figure 4.3. Experimental Profile 1: (a) The current profile used to generate training data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 50%.
Figure 4.4. Experimental Profile 2: (a) The current profile used to generate testing data, (b) Voltage response for the input current profile with added Gaussian noise, and (c) Battery SOC change if the initial SOC is 60%.
4.4 Experimental Validation

Battery SOC is initially determined using voltage lookup method, and during the cycling, battery SOC is calculated using current integration method. These methods work well in laboratory conditions and short experiments, but cannot be used in real life. Training data is collected at 25°C, and machine learning model is trained as explained before. Another experiment is performed at the same temperature but for a different current profile to generate test data for the model. Test data SOC is estimated using data-driven model. Figure 4.5 shows the results of the SOC estimation for the test data. It is clear that machine learning algorithms can estimate battery SOC experimentally. The mean error in estimation is around 2%.
4.5 Experimental Validation At Different Temperature

One of the main advantages of the method developed in this research is that the battery SOC estimation is independent of the battery temperature. So far, it has been shown in the simulation that the model trained at 25°C can predict the SOC for a battery at 35°C and 45°C. In this section, it is experimentally proved that the battery temperature is not a requirement for the model. The model is trained using the experimental data of a 2.6 NCM battery at 25°C. Now, the battery is cycled with a different profile at different temperatures. Specifically, the battery is cycled at 15°C, 35°C, and 45°C. Figure 4.6 shows the battery SOC estimation at 15°C. The result shows that the mean error in SOC estimation is less than 2%. Similarly, Figure 4.7 and 4.8 shows that battery SOC estimation at 35°C and 45°C, respectively. In these cases, the mean error in SOC estimation is less than 2%. Thus, it is experimentally demonstrated that the machine learning model, developed in this thesis, can predict SOC at different temperatures without any information about the battery temperature.
Figure 4.6. Experimental Validation: Battery train data at 25°C for 2.6 Ah NCM battery, but battery test data at 15°C for the same battery.
Figure 4.7. Experimental Validation: Battery train data at 25°C for 2.6 Ah NCM battery, but battery test data at 35°C for the same battery.
Figure 4.8. Experimental Validation: Battery train data at 25°C for 2.6 Ah NCM battery, but battery test data at 45°C for the same battery.
Chapter 5 | Battery Anomaly Detection

5.1 Introduction

One of the challenges for widespread acceptance of the battery technology is battery safety. Recent events of the fire in cell phones and electric vehicles hinder the growth of the battery technology [44, 45]. This research develops a method to predict the faulty batteries in real-time using machine learning techniques. The main idea for anomaly detection is based on the SOC estimation technique. For SOC estimation different parameters are fit and used as features. Similarly, those features can be used for anomaly detection. It is unsupervised learning problem where the goal is to cluster out the unhealthy batteries.

5.2 Anomalous Battery Data

Simulated battery data is generated for different batteries including faulty batteries. A battery can have different kinds of anomalies, some can cause fire and thermal runaway instantly, while others develop slowly. Batteries can also have some manufacturing defects which can grow over time and create unsafe conditions.
These slow growing anomalies may lead to a benign failure or catastrophic thermal runaway and fire. Due to their nature, they cannot be detected at the time of initial battery testing. Researchers have shown when a battery goes through any mild abuse its chemical parameters changes [46,47]. Two major parameters that changes are battery resistance and diffusivity. In this research, the battery is assumed to be one of the two states: either it is healthy or it has some anomalies which it had developed over time. The time period when it is developing the anomaly is not considered here. The anomalous data sets represent the batteries which have developed sufficient anomaly over time to be detected by the on-line algorithm.

Total data is generated for 1200 batteries, 1000 batteries are healthy without any anomaly, and rest 200 have anomalies. 100 batteries have different internal resistance, and 100 batteries have different diffusivity. It is easier to find of faulty battery if each battery has a same current profile and same initial SOC. One can simply find out the batteries with different voltage responses. To make the problem more practical, each battery is initialized with different initial SOC and cycled with the different current profile. Each profile is 1000 seconds long, and 10000 data points are collected for each test. Total data points are $1,200 \times 10,000 = 12,000,000$.

### 5.3 Anomaly Detection: Unsupervised Learning

Figure 5.1 shows the schematic for the unsupervised learning problem. The objective is to cluster out the faulty batteries. The first step is to generate different features using time series data and then use some clustering algorithm to cluster out the faulty batteries. Feature generation algorithm is similar as it was for SOC estimation except for few changes. There is no need to select the feature set for each time step because the anomalous battery data is for the batteries who have already developed the anomaly. If the batteries are developing the anomaly slowly, then time series
data at different time step should be considered. Here, one feature set for each battery is sufficient. Process for feature selection is as follows:

1. 200 seconds long time series data is selected to create the feature set for the given battery. Time series data includes time, current, and voltage. SOC is not required here as it is unsupervised learning.

2. Battery is a causal system. In any causal systems, output at any time instance will depend on the input on that time and previous time steps as well as output on previous time steps. So a voltage output function of the following form is assumed:

\[ V^n = aV^{n-1} + bV^{n-2} + cV^{n-3} + dV^{n-4} + eI^n + fI^{n-1} + gI^{n-2} + hI^{n-3} + iI^{n-4} + j \]  

(5.1)
Where \( V^n \) is the voltage at time step \( n \), \( V^{n-1} \) is the voltage at time step \( n-1 \), and so on. Similarly, \( I^{n-1} \) is current at time step \( n-1 \). \( a, b, c, d, e, f, g, h, i, j \) are the unknown parameters.

3. The assumed voltage function (Eq. 5.1) is best solved by a Laplace transfer function. Instead of using Equation 5.1, a corresponding Laplace transfer function is assumed. Voltage is simulated using following equation:

\[
V_{\text{out}} = m + L^{-1}\left(\left(\frac{c_1}{s} + \frac{z_1}{s-p_1} + \frac{z_2}{s-p_2}\right)I(s)\right) \tag{5.2}
\]

where \( s \) is Laplace parameter, and \( L^{-1} \) is Laplace inverse. With Equation 5.2, now parameters \( m, r_1, c_1, z_1, p_1, z_2, p_2 \) are fitted instead of parameters \( a, b, c, d, e, f, g, h, i, j \). There exists a relationship between Equation 5.1 and Equation 5.2, but it is not required here.

4. System is simulated for 200 seconds to fit the parameters \( m, r_1, c_1, z_1, p_1, z_2, p_2 \) in such a way the error between simulated and actual battery voltage is minimized.

\[
\begin{align*}
\text{minimize} & \quad \sum_{n=1}^{n=200} (V_{\text{sim}} - V_{\text{actual}})^2 \\
\text{subject to} & \quad V_{\text{sim}} = m + L^{-1}\left(\left(\frac{c_1}{s} + \frac{z_1}{s-p_1} + \frac{z_2}{s-p_2}\right)I(s)\right) \tag{5.3}
\end{align*}
\]

The parameters are discovered using MATLAB `fmincon` and `lsim` command.

5. This minimization will produce a value for parameters which will be the feature set

A feature set is generated for each battery as per the procedure. PCA is performed on the selected features for better data visualization. Once features are
selected, different clustering algorithms can be used for clustering out the faulty batteries.

5.3.1 DBSCAN Clustering

Many clustering algorithms are used, but DBSCAN performed the best. In Figure 5.2, batteries which have higher or lower resistance are separated from the healthy batteries. The DBSCAN algorithm created the clusters perfectly. In Figure 5.3, batteries which have higher or lower diffusivity are separated from the healthy batteries. The DBSCAN algorithm can cluster out most of them. One more advantage of DBSCAN algorithm is that it does not need the number of clusters apriori. As there can be a different number of clusters based on different type of anomalies, algorithms that do not require a number of clusters as input are more suitable.

It is clear from the results that machine learning algorithm can cluster out the faulty batteries if they have different resistance and diffusivity. It has been shown in the literature that many unsafe conditions affect battery resistance and diffusivity. Thus, this algorithm makes battery safer.

5.3.2 Other Clustering Algorithms

Other clustering algorithms are also implemented, but the results are not as good as with DBSCAN.

K-means Clustering: Figure 5.4 shows the results for K-means clustering. K-means clustering needs the total number of clusters beforehand which may not be possible for all type of datasets. The algorithm is implemented for 3 clusters but it is not able to cluster out the faulty batteries. K-means clustering works well with the spherical cluster, and data clusters here are non-spherical.
DBSCAN Clustering

Figure 5.2. Battery anomaly detection with resistance related anomalies using DBSCAN.

Agglomerative Clustering: Figure 5.5 shows the results for agglomerative clustering. Agglomerative clustering also needs the total number of clusters beforehand. The algorithm is implemented for 3 clusters but it is not able to cluster out the faulty batteries.

Meanshift Clustering: Figure 5.6 shows the results for mean shift clustering. Meanshift algorithm does not require a number of clusters as input, but it is also not able to cluster out the faulty batteries.

Spectral Clustering: Figure 5.7 and 5.8 show the results for spectral clustering.
Figure 5.3. Battery anomaly detection with diffusion related anomalies using DBSCAN.

Number of clusters can be given as input to the algorithm, but it is not required. Figure 5.7 shows the results of clustering with number of clusters as input and Figure 5.8 shows the results of clustering without number of clusters as input. In both cases, algorithm is not able to cluster out the faulty batteries.
Figure 5.4. Battery anomaly detection with \{(a) resistance (b) diffusion\} related anomalies using K-means clustering. Number of clusters is given as input.
Figure 5.5. Battery anomaly detection with \{(a) resistance (b) diffusion\} related anomalies using Agglomerative clustering. Number of clusters is given as input.
Figure 5.6. Battery anomaly detection with \{(a)resistance (b) diffusion\} related anomalies using Mean-Shift clustering. Number of clusters is not given as input.
Figure 5.7. Battery anomaly detection with \{(a) resistance (b) diffusion\} related anomalies using Spectral clustering. Number of clusters is given as input.
Figure 5.8. Battery anomaly detection with {(a)resistance (b) diffusion} related anomalies using Spectral clustering. Number of clusters is not given as input.
Chapter 6 |
Conclusion And Future Work

The purpose of this thesis is to study the problem of SOC estimation and anomaly detection in the battery system. Specifically, we study the SOC estimation using machine learning technique without any knowledge of the chemical or physical process, and anomaly detection when anomaly changes battery resistance or battery diffusivity. The main contribution of the thesis are as follows:


2. Battery SOC estimation even when the battery temperature is not known.

3. Battery SOC estimation when battery capacity/SOH is not known.

4. Experimental validation of the machine learning model at different temperatures.

5. Developed algorithm for battery anomaly detection.

In future, the work can be extended for experimental validation of the SOC estimation for different capacity batteries without the information of the actual capacity. The simulation shows that only battery chemistry would be sufficient for SOC estimation. It is difficult to get experimental data of anomalous battery, but validating the battery anomaly detection can be another future task.
## Appendix

### Table of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Temperature in degree Celsius</td>
</tr>
<tr>
<td>Ah</td>
<td>Ampere-hour</td>
</tr>
<tr>
<td>BMS</td>
<td>Battery management system</td>
</tr>
<tr>
<td>C</td>
<td>Nominal capacity</td>
</tr>
<tr>
<td>C&lt;sub&gt;r&lt;/sub&gt;</td>
<td>Remaining capacity</td>
</tr>
<tr>
<td>DOD</td>
<td>Depth of discharge</td>
</tr>
<tr>
<td>EV</td>
<td>Electric vehicle</td>
</tr>
<tr>
<td>HEV</td>
<td>Hybrid electric vehicle</td>
</tr>
<tr>
<td>I</td>
<td>Current</td>
</tr>
<tr>
<td>KNN</td>
<td>k-nearest neighbors</td>
</tr>
<tr>
<td>LFP</td>
<td>Lithium iron phosphate</td>
</tr>
<tr>
<td>NCM</td>
<td>Lithium Nickel Cobalt Manganese Oxide</td>
</tr>
<tr>
<td>OCV</td>
<td>Open circuit voltage</td>
</tr>
</tbody>
</table>
\textit{RUL} Remaining useful life

\textit{RNN} recurrent Neural Network

\textit{SOC} State of charge

\textit{SOH} State of health

\textit{SVM} Support vector machine

\textit{SVR} Support vector regression

\textit{sim} simulated

\textit{T} Temperature

\textit{t} Time

\textit{V} Voltage


