Random forest regression for lithium ion battery capacity estimation *

Yi Li

1. Background and motivation

Why do we need to know battery state of health (SOH)?

Battery Ageing

SOH = 100%

SOH = \frac{Q_t}{Q_0} \times 100\%

SOH = 80%

Battery Replacement
1. Background and motivation

- Capacity test: fully charge and then discharge

- Model Driven
- Data Driven
  - Empirical/Semi-empirical model
  - Machine learning
  - .....

Capacity Diagnosis

Off-line

On-line
2. Data & Model

Machine Learning for SOH estimation

Step 1

Data collection

- Historical data base
- Data collection from sensors (e.g. $I, V, T$)

Data pre-processing

- Reducing noise in data
- Data reduction/ transformation
- Feature extraction

Predictive model development

- Model creation
- Parameter tuning/ optimization
- Model validation

Integrate analytics with systems

- Embedded in BMS
Step 1.
Data collection

Start Test

Capacity test at beginning of life

Cycling ageing test

CC: \( \frac{1}{3} \) charge, 1 I, discharge
Group 1: 35°C, 100% DOD, 50% Mid-SOC
Group 2: 35°C, 80% DOD, 50% Mid-SOC
Group 3: 35°C, 50% DOD, 50% Mid-SOC
Group 4: 35°C, 50% DOD, 65% Mid-SOC
Group 5: 45°C, 100% DOD, 50% Mid-SOC
Group 6: 45°C, 80% DOD, 50% Mid-SOC

Capacity test after every 100 FECs

25°C, \( \frac{1}{3} \) charge and discharge

Cell Capacity ≤ 80%

No

Yes

End of Test
2. Data & Model

Capacity fade result:

- Different cycling temperature: 35, 45°C
- Varied cycling depth: 100%, 80%, 50%
- Different cycling mid-SOC: 50%, 65%
2. Data & Model

Step 2

Access and explore data
- Historical data base
- Data collection from sensors (e.g. $I, V, T$)

Data pre-processing
- Reducing noise in data
- Data reduction/ transformation
- Feature extraction

Predictive model development
- Model creation
- Parameter tuning/ optimization
- Model validation

Integrate analytics with systems
- Embedded in BMS
2. Data & Model

Step 2. feature selection

Step 1: Defining lower and higher voltage bounds

\[ V_l, V_h \]

\[ (Q_k, V_h) \]

\[ (Q_0, V_l) \]

\[ (Q_{t+1}, V_{t+1}) \]

\[ (Q_t, V_t) \]

\[ (Q_{t-1}, V_{t-1}) \]

\[ \Delta V \]

\[ \{X_t, Y_t\} = \{(x_0, x_1, \ldots, x_k), Y_t\} = \{(Q_0, Q_1, \ldots, Q_t, \ldots, Q_h), Y_t\} \]
2. Data & Model

Step 3

Access and explore data

- Historical data base
- Data collection from sensors (e.g. $I, V, T$)

Data pre-processing

- Reducing noise in data
- Data reduction/transformation
- Feature extraction

Predictive model development

- Model creation
- Parameter tuning/optimization
- Model validation

Integrate analytics with systems

- Embedded in BMS
2. Data & Model

Step 3. Algorithm selection

Random Forest Regression: An ensembled method

Extracted Features

\[ S_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\} \]

Bagging: Randomly sampled

Bootstrap Samples, \( S_{n1}^{\Theta_1} \)

Tree 1

Prediction, \( \hat{Y}_1 = \hat{h}(X, S_{n1}^{\Theta_1}) \)

Bootstrap Samples, \( S_{n2}^{\Theta_2} \)

Tree 2

Prediction, \( \hat{Y}_2 = \hat{h}(X, S_{n2}^{\Theta_2}) \)

Bootstrap Samples, \( S_{nq}^{\Theta_q} \)

Tree q

Prediction, \( \hat{Y}_q = \hat{h}(X, S_{nq}^{\Theta_q}) \)

Output: \( \hat{Y} = \frac{1}{q} \sum_{i=1}^{q} \hat{h}(X, S_{ni}^{\Theta_i}) \)
2. Data & Model

Offline parameter tuning

- \( n_{\text{tree}} \): the number of trees
- \( m_{\text{try}} \): the number of random features for each split in the forest to build (default number)

\[
\text{Loss} = \text{MSE} = \sum (y_i - y_i^p)^2
\]

- \( y_i \): \( i \)th target value
- \( y_i^p \): \( i \)th prediction
2. Data & Model

Leave n-out cross validation

\[ \{X, Y\} = \{[X_{c1}, Y_{c1}], [X_{c2}, Y_{c2}], \ldots, [X_{c17}, Y_{c17}]\} \]
2. Data & Model

Experimental Results

Estimation Results

- [Graph showing experimental results with SoH (%) on the y-axis and full equivalent cycle number on the x-axis.
- Graph showing estimation results with tested SoH and estimated SoH by RF.

12
3. Results & Conclusion

Group 1

Group 2

Group 3

Group 4

Group 5

Group 6
3. Results & Conclusion

Not OK

Good

BAD

\((X_1, Y_1), (X_2, Y_2), \ldots, (X_M, Y_M)\)

\((X_1, Y_1), (X_2, Y_2), \ldots, (X_M, Y_M)\)

\((X_1, Y_1), (X_2, Y_2), \ldots, (X_M, Y_M)\)
3. Results & Conclusion

Incremental Capacity analysis

\[ IC = \frac{dQ}{dV} \approx \frac{\Delta Q}{\Delta V} = \frac{Q_t - Q_{t-1}}{V_t - V_{t-1}} \]
3. Results & Conclusion

Conclusion
• A online capacity estimation method with random forest regression was proposed
• Low effort for input feature collection
• IC analysis was used for input feature selection

Limits:
• Only charging voltage-capacity curves at 25°C are used for input feature selection
• Require constant current rate
• Low charging C-rate (C/3)
Outlook

- Dynamic using conditions
- Battery pack
- Health prediction
Thank you!

Yi Li
y.li82@lancaster.ac.uk

Research Associate
Energy Lancaster
Lancaster University
https://www.lancaster.ac.uk/energy-lancaster/

LANCASTER BATTERY LAB

THE FARADAY INSTITUTION

COLLABORATIVE RESEARCH
Tuesday, 3rd
Shahin Nikman - Parallel Session 1b | 3 pm

Wed, 4th - Parallel Session 5a | 4pm
Michael Mercer
Beatrice Wolff
Robert Burrell