Differential Capacity Plot Visualization and Machine Learning Model for Battery Chemistry Classification

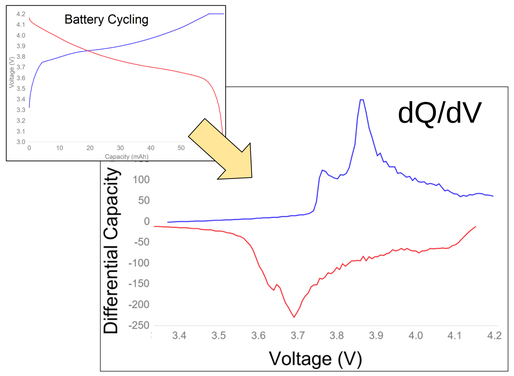
Sarah Alamdari[1], Theodore Cohen[1,2,3] Robert Masse[3], Nicole Thompson[1]

[1]Department of Chemical Engineering, [2] Molecular Engineering Department, [3]Material Science and Engineering Department

Introduction

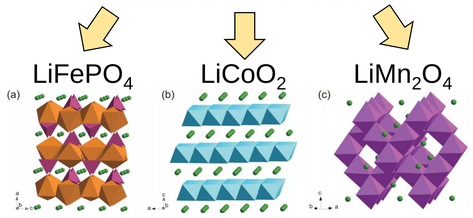
Motivation

Battery reliability testing requires months of charge/discharge cycling. More sophisticated analyses such as differential capacity (dQ/dV) analysis provides an underutilized means of digging deeper into battery performance data.



In the discovery of new materials for battery applications, these curves provide insight into the chemistries occurring at the molecular level. Currently the only way to analyze data from new materials is to perform extensive literature searches and compare the characteristics of the cycles manually. Features such as the position (voltage), number, intensity, and FWHM for these peaks can be used to classify electrochemical reactions inside the cell.

Classification Problem



Goals

1. **Data Mining and Cleaning**
   - Create an initial database of battery cycling data
2. **Machine Learning**
   - Define a set of descriptors which can accurately classify batteries
   - Test and Train an ML algorithm that can solve high dimensional classification problems
3. **Data Visualization**
   - Create an interface to visualize the data at each cycle

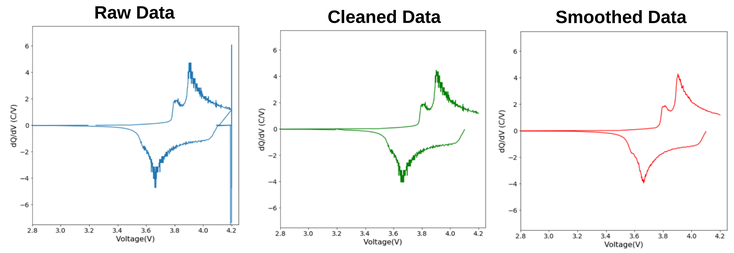
Battery Cycling Data → **Data Cleaning** → **Data Smoothing** → **Peak Fitting** → **Descriptors** → **Classify Battery**

Methods

Data Cleaning and Smoothing

To obtain the dataset used for the machine learning model, 1200 complete cycles were cleaned and processed using high performance computing resources.

1. Each dataset was split into individual, separated cycles
2. dQ/dV was calculated for each cycle using the raw capacity and voltage measurements
3. Each cycle was cleaned of large jumps in dQ/dV or voltage.
4. Each cleaned cycle underwent the Savitzky-Golay filter to smooth noise.



Peak Fitting to Obtain Descriptors

The LMFit package was utilized for polynomial fitting and PeakUtils was used for finding peaks. 32 descriptors were identified including variables from the polynomial fit, location of the peaks, peak heights, and peak widths.

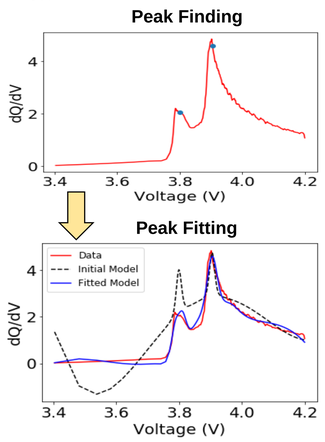
Pseudo-Voight Distribution

$$f_v(x, A, \mu, \sigma, \alpha) = \frac{(1-\alpha)A}{\sigma_g\sqrt{2\pi}}\exp\left[-(x-\mu)_2/2\sigma_g{}^2\right] + \frac{\alpha A}{\pi}\left[\frac{\sigma}{(x-\mu)^2+\sigma^2}\right]$$

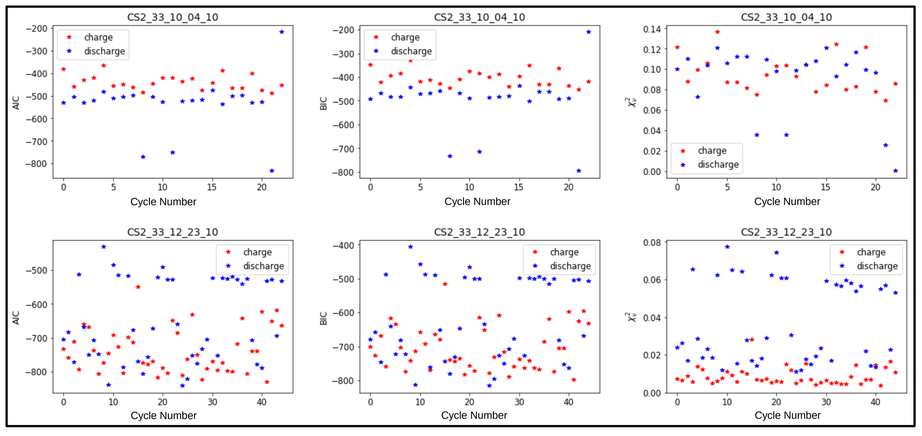
$$\sigma_g = \sigma/\sqrt{2\ln 2}$$

Polynomial Fit

$$f(x) = \sum_{i=0}^{3} c_i x^i + \sum_{j=1}^{n} f_{vj}(x, A_j, \mu_j, \sigma_j, \alpha_j)$$



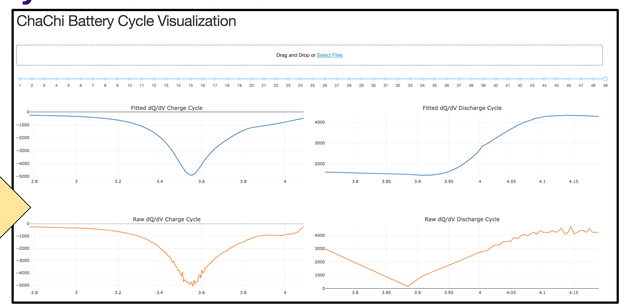
Fitting Error Analysis of Two Batteries



Results

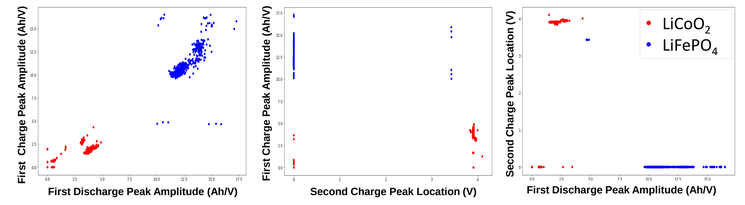
Plotly & Dash Visualization

Data Visualization



Reducing the Feature Set

The Feature Set was reduced by running least absolute shrinkage and selection operator (LASSO), a Linear Model trained with L1 prior as the regularizer. With an alpha of 0.005, LASSO scaled the original feature vector containing 32 unique descriptors down to 3.



Machine Learning

The support vector machine (SVM) algorithm was found to be the best algorithm to classify this data. The linear fit gave the best training errors over a polynomial fit and using a radial basis function. Data was separated into an 20/80 test/train split. The model was trained on training data, and validated using test data.

Conclusions

The final model included 3 features which include the first charge peak amplitude, the first discharge peak amplitude, and the location of the second charge peak. The SVM linear model had a **training error of 20% and a test set error of 24%.**

Future Work

This work is an important step in the understanding of materials design for Clean Energy applications. It provides a basis for extending our classification algorithm so that we can build database to compare battery data across labs and examine more exotic chemistries. We propose using this package as tool at “hack weeks” to further extend the datasets available to train the ML model on.

Acknowledgements

Dave Beck and Kelly Thornton

Datasets were taken from publically available battery cycling data from the University of Maryland’s Center for Advanced Life Cycle Engineering (CALCE)

Only open source packages were used in this work,
documentation of all packages used can be found at our github at https://github.com/tacohen125/chachies