Predicting Battery Degradation: A Simple Linear Regression Approach

Amey Kulkarni
Department of Big Data Analytics
San Diego State University

Email: ameykulkarni888@gmail.com / akulkarni7388@sdsu.edu

***Abstract*—Batteries play a vital role in powering various devices, from smartphones to electric vehicles, but over time, their performance degrades, resulting in shorter usage periods and reduced efficiency. This project focuses on developing a predictive model to estimate how a battery's capacity (how much energy it can store) and voltage (the electrical pressure) change over time as the battery undergoes repeated charging and discharging. Using data from real-world battery tests, the project applies a mathematical technique called linear regression to analyze the relationship between different factors like the number of cycles (charges and discharges), temperature, and voltage, to predict the future degradation of a battery. The results, presented in graphs, show how well the model can forecast a battery's remaining life and help manage battery usage more effectively. This approach can lead to better battery management in devices, helping to maximize battery lifespan and improve performance prediction, even for complex systems like electric vehicles and renewable energy storage.**

I. Introduction

A. Why Do Batteries Matter?

Batteries are integral to many modern technologies, from consumer electronics to electric vehicles (EVs). Their performance, however, degrades over time, which impacts their efficiency and overall lifespan. As the demand for high-performance, long-lasting batteries increases, particularly in electric vehicles and renewable energy storage systems, the need for accurate battery degradation prediction becomes more critical. This paper explores the prediction of battery degradation, specifically focusing on capacity and voltage loss, through a predictive model based on real-world data.

B. Importance of Battery Degradation Prediction

Battery degradation prediction plays a crucial role in optimizing battery life and improving system performance. For instance, in electric vehicles, accurate degradation forecasts can enable better battery management, extending the vehicle's operational life and improving its energy efficiency. Similarly, in renewable energy applications, understanding how batteries degrade is vital for ensuring reliable power storage. Capacity and voltage are the primary performance indicators affected by degradation, and their accurate prediction can help users maximize battery usage while avoiding premature failure

C. What Is Battery Degradation?

Battery degradation refers to the gradual loss of performance due to chemical and physical changes inside the battery. Common causes include:

- Repeated charge and discharge cycles.
- High operating temperatures.
- Aging, even when the battery is not in use.

Understanding these processes is essential to predict when a battery will no longer meet its requirements.

D. Battery Basics

To better understand degradation, it is essential to define a few key battery properties:

- **Capacity** (C): Measures the total energy a battery can store, typically in ampere-hours (Ah). Mathematically:
$$C = I \cdot t \tag{1}$$
where I is current in amperes and t is time in hours.
- **Voltage** (V): Represents the electric potential difference across the battery terminals, measured in volts (V). Voltage decreases over time due to internal resistance and chemical reactions.
- **Energy Stored** (E): Calculated using:
$$E = V \cdot Q \tag{2}$$
where Q is the charge in coulombs.

A battery's performance is defined by two main parameters: **capacity** and **voltage**. Over time, as a battery undergoes charge and discharge cycles, both its capacity and voltage degrade.

The degradation in capacity Dc (t) and voltage Dv (t) in cycle t can be quantified using the following equations:

Capacity Degradation: Dc (t) = C (0) C (t) where C (0) is the initial capacity at cycle 0, and C(t) is the capacity at cycle t.

Voltage Degradation: Dv (t) = V (0) V (t) where V (0) is the initial voltage in cycle 0, and V(t) is the voltage at cycle t.

These formulas help quantify how much a battery's performance has deteriorated at any given time.

E. Objectives

The goal of this project is to:

1) Analyze real-world data from batteries.
2) Develop a simple model to predict:
   - Capacity degradation: The loss of energy storage capability.
   - Voltage degradation: The reduction in energy delivery efficiency.
3) Provide insights into extending battery life.

II. Methods: How Do We Predict Battery Health?

A. Introduction to Linear Regression

Linear regression is a technique that models relationships between variables. Imagine plotting how many times you've used a battery (cycles) against how much energy it can store (capacity). Over time, capacity decreases, forming a downward trend. Linear regression fits a straight line to this trend, allowing us to:

- Estimate how much capacity is left after a certain number of cycles.
- Predict the total lifespan of the battery.

The general form of a linear regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{3}$$

where:

- y represents the predicted degradation (capacity or voltage),
- $x_1, x_2, \ldots, x_n$ are the input features (e.g., cycle number, temperature),
- $\beta_0, \beta_1, \ldots, \beta_n$ are the model coefficients that quantify the effect of each feature on the degradation.

By fitting this model to historical battery data, we can predict how a battery will degrade under various conditions.

B. The Dataset

The dataset used in this project comes from NASA's publicly available Li-ion Battery Aging Dataset, which provides real-world battery data collected from a variety of battery cycles. The dataset is part of NASA's research into improving battery life predictions and understanding the degradation mechanisms of lithium-ion batteries. The data includes measurements taken during charge-discharge cycles of batteries, including capacity, voltage, temperature, and other relevant variables.

The data is stored in .mat (MATLAB) file format and consists of several variables recorded during battery tests. These variables include:

- **Cycle Number**: The number of charge-discharge cycles the battery has undergone.
- **Capacity**: The amount of charge the battery can store, typically measured in ampere-hours (Ah).
- **Voltage**: The voltage across the battery terminals at different points during charging and discharging.
- **Temperature**: The temperature of the battery at different stages of the charge-discharge cycles.
- **Discharge Data**: Information on battery discharge behavior, including current and voltage measurements during discharge.

The data is provided for several different batteries and is intended for research into the modeling and prediction of battery degradation. Each .mat file corresponds to a specific battery, with each battery undergoing several cycles, providing detailed records of its performance over time. The dataset's purpose is to offer insights into the degradation behavior of lithium-ion batteries and is widely used in the development of machine learning models for battery life prediction.

The specific batteries used in this project include:

- **Battery B0005**
- **Battery B0006**
- **Battery B0007**
- **Battery B0018**

By analyzing these datasets, we can develop predictive models for capacity and voltage degradation, which are key performance indicators in battery health and longevity.

C. Modeling Process

The modeling process for predicting battery degradation involves several steps, including data preprocessing, feature extraction, model selection, and evaluation. The primary goal is to develop a model that can accurately predict the degradation in battery capacity and voltage over time, based on historical cycle data. The following steps outline the modeling process:

1) 1. Data Preprocessing: Before applying any machine learning models, it is essential to preprocess the raw data. This includes:

- **Data Extraction**: Extract relevant variables (e.g., cycle number, temperature, voltage, capacity) from the .mat files provided by NASA.
- **Handling Missing Data**: Handle any missing or inconsistent data through interpolation, imputation, or removal, depending on the nature of the dataset.
- **Feature Engineering**: Combine and transform features, if necessary, to ensure the model can interpret them effectively.
- **Normalization**: Scale the features to a common range to ensure that all features contribute equally to the model. This is done using StandardScaler from `scikit-learn` to normalize the features like cycle number, temperature, and voltage.

2) 2. Model Selection: For this project, we have chosen **linear regression** as the predictive model due to its simplicity and effectiveness in capturing linear relationships between features and degradation outcomes. Linear regression models the relationship between input features ($x_1, x_2, \ldots, x_n$) and the output variable (y) as a linear equation.

As we saw earlier, the general equation for linear regression is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (4)$$

The goal of linear regression is to find the values of the coefficients $(\beta_0, \beta_1, \ldots, \beta_n)$ that minimize the residual sum of squares between the observed actual outcomes and the predicted outcomes.

3) 3. Model Training and Prediction: Once the features are normalized and the model is defined, the next step is to train the model using the historical data:

- **Training the Model**: The training data is used to fit the linear regression model. This is done by minimizing the error between the predicted values and the actual observed values using the normal equation:

$$\beta = (X^T X)^{-1} X^T y \quad (5)$$

where X is the feature matrix (including the input features), and y is the vector of observed degradation values.

- **Predicting Degradation**: After training, the model can predict the degradation (capacity or voltage) for new data points (unseen cycles). The prediction is made by plugging the new feature values into the linear regression equation:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (6)$$

where $\hat{y}$ is the predicted degradation.

4) 4. Evaluation of Model Performance: To evaluate the performance of the model, several metrics are used:

- **Mean Absolute Error (MAE)**: Measures the average of the absolute errors between predicted and actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

where y_i is the actual value, and $\hat{y}_i$ is the predicted value.

- **Root Mean Squared Error (RMSE)**: Measures the square root of the average squared differences between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

- **R-squared (R^2)**: Indicates how well the model explains the variance in the data. An R^2 value closer to 1 indicates a better fit.

The model is assessed using these metrics to ensure its accuracy and reliability in predicting battery degradation.

5) 5. Code Implementation Logic: The code implements the following general logic:

- **Data Extraction**: The code extracts discharge data and capacities from the .mat files using a custom function that handles the parsing of MATLAB file format.
- **Feature Preparation**: Features such as cycle number, temperature, and voltage are extracted and stacked into a feature matrix.
- **Normalization**: Features are normalized using `StandardScaler` to ensure that all variables contribute equally.
- **Model Training**: The linear regression model is trained using the normal equation to find the best-fitting coefficients.
- **Prediction**: The trained model predicts battery degradation based on the input features.
- **Evaluation and Plotting**: The code also evaluates the model's performance using MAE, RMSE, and R^2, and visualizes the results by plotting actual vs predicted degradation, voltage, and capacity over time.

III. Results

A. Capacity Degradation

Figure 1 shows the actual and predicted capacity degradation for Battery B0005. The red dashed line closely follows the blue points, indicating accurate predictions.

As we can see, at cycle 0, the degradation is 0 and as the number of cycles increase, so does the degradation.

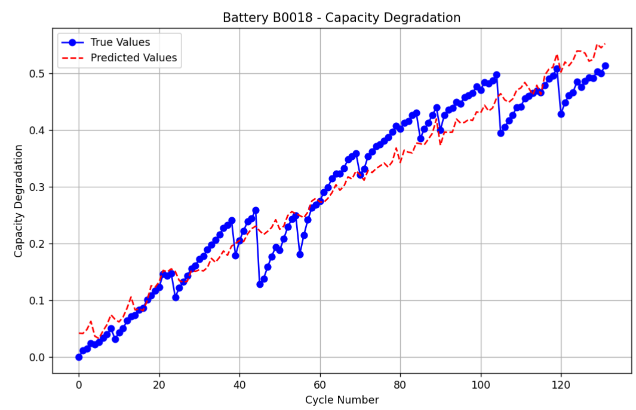

Fig. 1. Predicted vs. Actual Capacity Degradation for Battery B0018

B. Voltage Degradation

Voltage predictions are equally reliable, as shown in Figure 5. Although the voltage fluctuates, the model captures how the it degrades over the period of time.

C. Combined Insights

Figure 4 summarizes the predictions for all four batteries, demonstrating the model's consistency across different datasets.

In previous plot, we saw the degradation consistance of capacity. Here we are plotting the capacity of the battery against number of cycles, and Voltages against cycles.

IV. Discussion

A. Implications

The results show that even simple models like linear regression can provide meaningful insights into battery health. This has several practical applications:

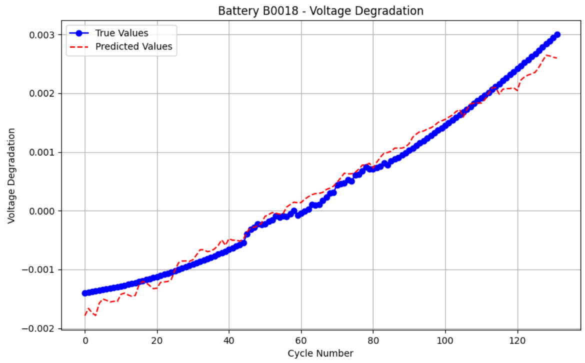

Fig. 2. Predicted vs. Actual Voltage Degradation for Battery B0018

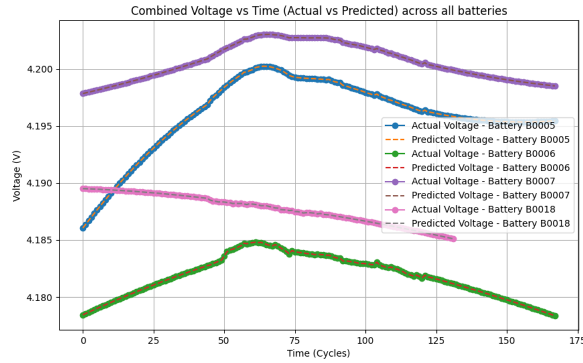

Fig. 4. Combined Voltage Degradation Across All Batteries

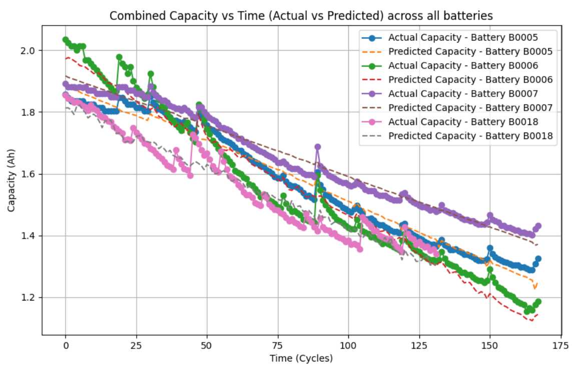

Fig. 3. Combined Capacity Degradation Across All Batteries

- **For Consumers**: Know when to replace batteries to avoid unexpected failures.
- **For Manufacturers**: Design batteries that last longer by studying degradation trends.
- **For Researchers**: Use predictive models to test new materials and technologies.

B. Limitations and Future Work

While effective, the model has limitations:

- It assumes degradation follows a straight line, which might not capture real-world complexities.
- It uses limited features; adding more data (e.g., current, state of charge) could improve predictions.

Future work could explore advanced models like neural networks to handle non-linear trends and make more accurate predictions.

V. Conclusion

This project demonstrates a practical and data-driven approach to modeling the degradation of lithium-ion batteries, leveraging real-world datasets provided by NASA. By utilizing linear regression, the study establishes a predictive framework to assess capacity and voltage degradation over time.

The analysis highlights the following key achievements:

- **Predictive Insights**: The model successfully captures the relationship between operational parameters, such as cycle number and temperature, and degradation metrics like capacity and voltage. These predictions offer valuable insights into battery health over its lifecycle.
- **Simplicity and Effectiveness**: Linear regression, despite its simplicity, proves effective in understanding the underlying trends in battery degradation, making it a robust choice for first-pass analysis.
- **Visualization and Trends**: The generated plots and metrics, such as R-squared values and RMSE, validate the reliability of the predictions and provide intuitive visual representations of degradation over time.

```
Results for Battery B0005:
  Capacity Degradation - MAE: 0.0240, RMSE: 0.0284
Results for Battery B0006:
  Capacity Degradation - MAE: 0.0243, RMSE: 0.0311
Results for Battery B0007:
  Capacity Degradation - MAE: 0.0196, RMSE: 0.0234
Results for Battery B0018:
  Capacity Degradation - MAE: 0.0218, RMSE: 0.0284
```

Fig. 5. MAE and RMSE for each battery

The findings have significant implications for battery design, maintenance, and usage. Predicting degradation trends can assist manufacturers in optimizing battery materials and designs, while enabling end-users to make informed decisions about battery maintenance and replacement. Moreover, this modeling approach could serve as a foundation for integrating more complex machine learning algorithms in future studies, potentially improving prediction accuracy and adaptability to varying datasets.

While the project achieves its core objectives, it also identifies potential avenues for improvement. Future work could explore nonlinear regression models, incorporate additional features such as charge/discharge rates, and evaluate the model's performance across diverse datasets. These enhancements would further refine the predictive capability and extend the applicability of the analysis.

In conclusion, this study emphasizes the importance of data-driven modeling in understanding battery degradation and showcases how simple, interpretable models can provide actionable insights, bridging the gap between academic research and real-world applications.

Appendix: Experimentation and Explorations

In this appendix, I detail a few additional experiments conducted during the course of this project, which reflect the exploration of various numerical methods, feature engineering, and data transformations. These experiments were pursued to understand how different parameters or techniques could influence the prediction of battery degradation.

Experiment 1: Including the Current Feature

One of the first experiments I conducted involved adding the *current* feature to the dataset. I hypothesized that the current, being an essential parameter in battery charging and discharging cycles, might play a significant role in degradation prediction. However, upon inspection of the data, I realized that the current values were either -2A or 0A for many of the data points, which suggested that the data might be either incorrect or unreliable for modeling purposes.

Thus, I decided to remove the *current* feature, as its inclusion did not improve the model and could potentially introduce noise. This decision led to a more stable and interpretable model. Removing unreliable features is important in ensuring that the model's predictions are based on valid and useful data.

Experiment 2: Data Normalization

To improve the model's performance and ensure that all features contributed equally to the predictions, I applied data normalization. Normalization rescales the data to a standard range, preventing features with larger numerical values (e.g., cycle number) from dominating the model. The normalized data allows the linear regression algorithm to treat each feature equally, improving prediction accuracy.

The formula used for normalization is:

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{9}$$

where x is the feature value, μ is the mean of the feature, and σ is the standard deviation.

Experiment 3: Smoothing the Voltage Data Using the Savitzky-Golay Filter

Another key experiment involved smoothing the voltage data to reduce noise. Voltage readings are often subject to fluctuations due to measurement errors or temporary changes in battery conditions. By applying the Savitzky-Golay filter, I was able to reduce high-frequency noise while preserving the important trends in the voltage data.

The Savitzky-Golay filter is a popular smoothing technique that fits successive polynomials to a moving window of data points. It can smooth the data while also preserving its higher-order derivatives, which is beneficial for capturing the underlying trends in battery degradation.

In this project, I applied the Savitzky-Golay filter using the Python `scipy.signal.savgol_filter` function with a window size of 5 and a polynomial order of 2:

$$V_{\text{smoothed}}(t) = \text{savgol\_filter}(V(t), \text{window\_length} = 5, \text{polyorder} = 2) \tag{10}$$

where $V(t)$ is the voltage value at time t, and the filter smooths the data based on the specified window length and polynomial order. This technique helped create a cleaner input for the regression model, leading to more consistent and reliable predictions.

Other Experiments and Future Work

Other techniques, such as exploring different feature scaling methods (e.g., min-max scaling), investigating polynomial regression models, or experimenting with advanced models like monte carlo methods, were considered but not explored in depth due to time constraints.

Future work could involve further experiments using these advanced models, which may be better suited to capturing non-linear degradation trends. Additionally, more granular data features, such as the state of charge, could be incorporated to see if they improve prediction accuracy.

Acknowledgments

The author thanks San Diego State University for providing access to resources and supporting this project.

References

1) NASA Ames Prognostics Center of Excellence, "Li-ion Battery Aging Datasets," 2017. [Online]. Available: https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb/about$_data$.