Title: Predicting Crystal Systems in Lithium-ion Battery Cathode Materials

Group Number: Group 12

Course: CHEE 6397

Date: 11/21/2023

Group Members: Hao Wang (2344601), Udom Ifeanyi (2220031), Kevin Tieu (1902070)

Abstract

This project explores the prediction of crystal systems in lithium-ion battery cathode materials using machine learning techniques. Utilizing a dataset from Kaggle, the study involves data analysis, including data distribution and correlation, and a gradient boosting classification model. The gradient boosting model demonstrated promising results with an accuracy of 0.663. To further analyze the final model, a 5-fold cross validation was used, and the model achieved an average accuracy of 0.651, indicating its performance and potential for crystal structure prediction application.

Introduction

Lithium-ion batteries are pivotal in modern technology, with their performance significantly influenced by the properties of cathode materials. This study aims to predict the crystal systems of these materials, enhancing understanding and potentially contributing to improved battery designs.

Background

Cathode materials in lithium-ion batteries play a crucial role in their overall performance. The crystal structure of these materials can significantly affect their functionality. Predicting these structures is thus vital for advancing battery technology.

Literature Review

In lithium-ion battery (LIB) research, significant focus has been on cathode materials, given their critical role in battery performance. Recent studies have highlighted the importance of these materials' energy density, thermal stability, and charge rates [1], [2]. Research advancements have been made in exploring new cathode materials, showing promising results in battery capacity and life. The crystal structure of cathode materials is a key determinant in LIBs' performance. Studies indicate that the crystallography of these materials influences lithium-ion diffusion rates, directly impacting the efficiency of charging and discharging [3], [4]. Machine learning (ML) has emerged as a powerful tool for predicting the crystal structure of cathode materials [5], [6]. These methodologies, leveraging algorithms based on material composition and properties, have shown accuracy in structure prediction, marking a significant step forward in computational material science. However, the field continues to face challenges, particularly in the generalization of these predictive models across different types of cathode materials. Future research is geared towards developing more sophisticated ML models that can incorporate a broader array of variables for enhanced prediction accuracy [7].

Dataset Description

The dataset includes several data for each of the 339 cathode materials for lithium-ion batteries. The features of each data point include formation energy, energy above hull, band gap, number of sites, density, volume, band structure, and the material’s chemical composition.

Methodology

Prior to data analysis, the dataset is processed to eliminate invalid data points and convert the raw data into manageable values. String values of the target (crystal system) are converted into distinct integers using one-hot encoding to digest easier. Also, chemical formulas for all cathode materials are broken down into their element compositions for better data processing.

Data analysis is conducted using several approaches. First, the correlations between features are investigated using a heat map. It will indicate the pairs of features showing strong correlation to one another. In addition, data distribution is used to analyze the feature and target data. This information will indicate whether the dataset is balanced and biased toward a label or feature.

After the data analysis steps are completed, a gradient boosting classification model is used to predict crystal system. The model is fine-tuned by optimizing its three hypermeters (number of learners, learning rate and max tree depth). The model’s performance is evaluated using the F1 scoring metrics. In addition, to prevent the model from being overfitting, a five-fold cross-validation is used to evaluate the model on different training and test sets.

Results

Based on the analysis performed in the Jupyter notebook, the following key findings were obtained:

1. Data Distribution and analysis:

The histograms generated for key features like space group, crystal structure, number of sites, and chemical formula indicated varied distributions. These visualizations were critical in understanding the dataset's properties and the prevalence of different crystal systems.

A graph of different colored bars

Description automatically generated

A graph with a bar and a few squares

Description automatically generated with medium confidence

1. Classification Model Performance:

Gradient Boost Model: The Gradient Boost classification model was employed to predict crystal structures. The model achieved an average 5-fold cross-validation accuracy of 0.651, indicating a strong predictive performance.

A green squares with white text

Description automatically generated

A screen shot of a graph

Description automatically generated

A graph of different types of energy

Description automatically generated

1. Correlation Analysis:

The extended heat map provided additional insights into the correlations between different material properties. These correlations are vital for understanding the crystal structures in lithium-ion battery cathode materials.

A green squares with black and white numbers

Description automatically generated

5. Visualizations and Comparative Analysis:

Scatter plots were used to compare the number of sites and space group, colored by crystal system. These visualizations played a crucial role in illustrating the relationships between these variables and their influence on the crystal system.

[Insert Scatter Plots from Notebook here showing the comparison between the number of sites and space group. ]

6. Key Takeaways:

The analysis highlights the potential of machine learning techniques, especially gradient boost classification, in predicting the crystal systems of lithium-ion battery cathode materials. The regression models and visualizations further deepen the understanding of the relationships between various material properties.

These results demonstrate the effectiveness of the chosen analytical methods and contribute significantly to the field of lithium-ion battery research. For the final report, you should insert each mentioned element (graphs, plots, histograms) directly from the Jupyter notebook outputs to provide visual and empirical evidence supporting these findings.

Conclusion and Future Work

The study provides insights into the relationships between various properties of lithium-ion battery cathode materials. Future research could focus on refining models for more accurate predictions and exploring additional classification algorithms.

References

[1] M. S. Whittingham, "Lithium batteries and cathode materials," Chemical Reviews, vol. 104, no. 10, pp. 4271-4302, 2004.

[2] J. B. Goodenough and K. S. Park, "The Li-ion rechargeable battery: a perspective," Journal of the American Chemical Society, vol. 135, no. 4, pp. 1167-1176, 2013.

[3] M. Armand and J. M. Tarascon, "Building better batteries," Nature, vol. 451, no. 7179, pp. 652-657, 2008.

[4] Q. Yan et al., "Machine learning design of polymer dielectrics: machine learning of polymer dielectrics," Chemical Science, vol. 8, no. 3, pp. 1759-1765, 2017.

[5] B. Kang and G. Ceder, "Battery materials for ultrafast charging and discharging," Nature, vol. 458, no. 7235, pp. 190-193, 2009.

[6] Y. Sun, N. Liu, and Y. Cui, "Promises and challenges of nanomaterials for lithium-based rechargeable batteries," Nature Energy, vol. 1, no. 7, pp. 1-12, 2016.

[7] A. Jain et al., "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," APL Materials, vol. 1, no. 1, 011002, 2013.