Discovery of Superionic Solid-state Electrolytes using Unsupervised Machine Learning

Data driven machine learning approaches to solid-state batteries

Shengyi Hu, Gabriel Krenzer, Anthony Onwuli, Aron Walsh

Abstract

Several known **lithium superionic conductor (SIC) materials** show high ionic conductivities at room temperature, making them promising solid electrolyte candidates for next-generation solid-state Li-ion batteries [1].

- **Ionic transport** is the key to understanding the conductivities of lithium SICs and the crystal structural framework determines the energy landscape of the ion migration [2]. Vibrational properties are chosen as the representation of Li-ion crystals in the databases from The Materials Project [3] and the Phonon Database at Kyoto University (PhononDB) [4][5][6].
- To decrease the issues of lacking good quality labelled data, unsupervised machine learning was employed to **find new crystals** for solid-state electrolytes from vibrational properties data due to the advantages of no labelled training data required and the ability to identify new species similar to known good solid electrolytes candidates [7].
- Various calculations were performed using phonon density of state (DOS) and the results were employed as descriptors for multiple methods of clustering. The outcomes were obtained, and further research is suggested.

Motivation

- The project aimed to **discover solid-state electrolytes** with desired properties used in solid-state batteries which are potentially less flammable and more electrochemical stable.
- Identify **new trends** in superionic behaviour by using different data visualisation approaches on calculation results and clustering outcomes.

Results

- The codes were written in different Jupyter Notebook files based on their functionalities. Data extraction from both databases and visualisation from clustering results could be easily produced by running corresponding files.
- Inputs for clustering were either a 1D feature from the calculations or a multi-dimensional feature array of calculation result combinations.
- Dendrograms were produced when performing Hierarchical Clustering. While 2D plots could be generated when two-dimensional feature arrays were used as an input for K-Means Clustering and Spectral Clustering.
- Variance and Silhouette Analysis could display values or scores for different numbers of clusters. The highest values or scores represent the optimal number of clusters for the chosen input and clustering method. The clusters could also be visualised from the Silhouette Analysis results.
- For the current selection of inputs, the optimal numbers of clusters were shown as either 2 or 3.

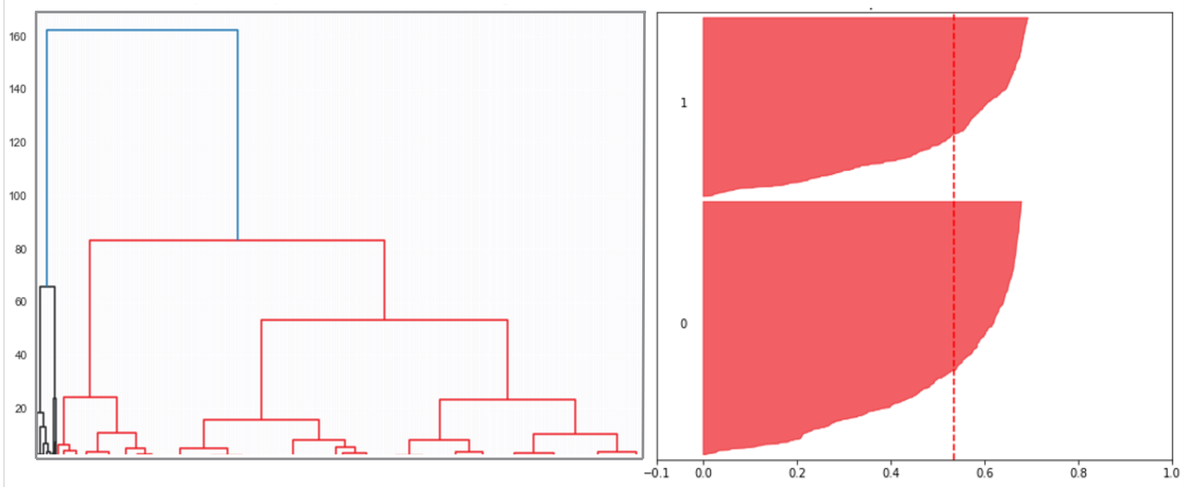
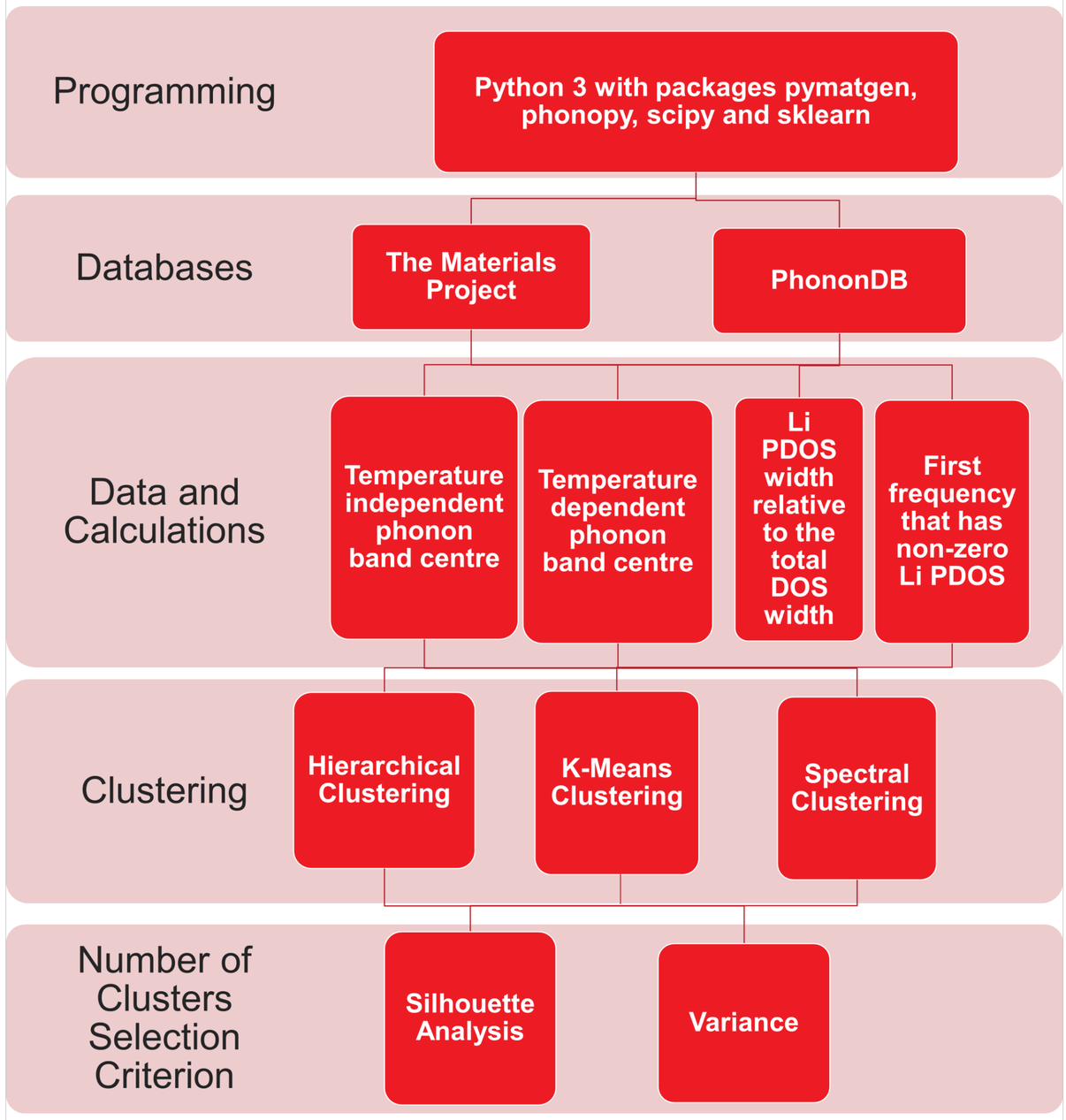


Figure 1. The left plot is an example Hierarchical Clustering dendrogram for temperature-independent phonon band centres input. The clusters merge successively from individual samples at the bottom. The right plot is an example visualisation for the Silhouette Analysis of two clusters. The x-axis is the scores. The red line represents the average score for all clusters.

Methods



Conclusions

- The outputs from the optimal number of clusters selection were either 2 or 3, showing the clustering methods were likely to separate the samples into a high value of calculation results cluster and a low-value cluster with sometimes a medium value cluster.
- The representation for phonon DOS might be too brief to conclude the phonon DOS patterns for different species at various frequency points.
- Further exploration of vibrational properties representation is required and more trials on multi-dimensional array input for clustering are necessary.

Impact / Next steps

- Using machine learning for materials screening could potentially reduce the cost of expensive first-principles screenings, resource usage and environmental impact.
- This project explored the possibilities to **discover new solid-state electrolyte materials** using unsupervised machine learning. The codes are available as a base for future work.
- **Exploration of different properties** will be carried out and the **featurisation of phonon DOS** will also be a meaningful focus.
- Potential new solid-state electrolyte materials could be discovered following a similar workflow from this project.

References

[1] T. Famprikis, P. Canepa, J. Dawson, M. Islam and C. Masquelier, "Fundamentals of inorganic solid-state electrolytes for batteries", *Nature Materials*, vol. 18, no. 12, pp. 1278-1291, 2019.

[2] X. He, Y. Zhu and Y. Mo, "Origin of fast ion diffusion in super-ionic conductors", *Nature Communications*, vol. 8, no. 1, 2017.

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

[4] S. Ong et al., "Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis", *Computational Materials Science*, vol. 68, pp. 314-319, 2013.

[5] S. Ong et al., "The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles", *Computational Materials Science*, vol. 97, pp. 209-215, 2015.

[6]"Phonon database at Kyoto university — phonondb documentation", Phonondb.mtl.kyoto-u.ac.jp, 2022. [Online].

[7] Y. Zhang et al., "Unsupervised discovery of solid-state lithium ion conductors", *Nature Communications*, vol. 10, no. 1, 2019.

Intern bio

- Shengyi (Amelia) Hu is studying MEng Materials Science and Engineering at Imperial College London as an upcoming third-year student.
- Interested in Theory and Simulation of Materials and using Machine Learning for Materials Discovery.
- Aspiring to learn more about different methods and software for Materials Modelling.
- Aspiring to undertake a PhD in Materials Science and Engineering.

THE FARADAY INSTITUTION Modelling
MULTI-SCALE MODELLING

Imperial College London

UK Research and Innovation