Materials Science With Machine Learning

Regression Models for Lattice Thermal Conductivity Prediction

Rahul Ghosh
NISER - MLC3291

Nabendu Mondal
NISER - MLC3294

Introduction

Machine learning is a branch of artificial intelligence where computer algorithms learn patterns from data and improve their performance on tasks without being explicitly programmed.

ML can accelerate the materials discovery by understanding the structure-property relation in feature space, bypassing traditional costly calculations like DFT.

Apply different machine learning regression algorithms to a database of lattice thermal conductivity in order to evaluate their performance and draw inferences.

Database contains thousands of unique materials with lattice thermal conductivity (AFLOW)

Download and curate the data

Featurization

Train Models and Validate

Select The best model for Out of The Dataset Prediction

Introduction

Lattice thermal Conductivity(κ_l) → Performance of thermoelectric materials

Thermoelectric performance, ZT = $\dfrac{S^2\sigma}{\kappa_e + \kappa_l}$

Electrical Conductivity

Seebeck Coefficient

Electronic thermal conductivity

Lattice thermal conductivity(very expensive calculations needed eg. DFPT)

For thermoelectric applications we need the materials with low lattice thermal conductivity (*300K).

Base Paper

Machine learning assisted hierarchical filtering: a strategy for designing magnets with large moment and anisotropy energy

By Arijit Dutta and Prasenjit Sen
Submitted in **Journal of Materials Chemistry C, 2022**

They employed both **classification** and **regression** approaches: classification to distinguish between stable and unstable materials, to determine magnetic states, and to categorize materials based on magnetic anisotropy energy; and regression to predict properties such as heat of formation, magnetic moment, and anisotropy energy. For these tasks, they used algorithms including SVM (with linear and RBF kernels) and Random Forests etc.

They have used chemical, compositional configurational and stoichiometric features to train the models.

As a conclusion of this paper, it shows using physics based features to predict the magnetic properties can reduce the cost upto some orders while maintaining DFT level accuracy.

Flow of Our Project

1. **Download and curate the available data(~ 5.5k unique compounds) with lattice thermal conductivity available on Aflow.**

1. **Perform Featurization**(*as mentioned in the base paper)
    a. Compositional average quantities
    b. Stoichiometric features
    c. Chemical Features
    d. We will be skipping configurational features due to computational and time constraint.

1. **Perform Feature Selection methods like RFE(Recursive Feature Selection) and PCA to make the correlated features with target and remove the correlation between the features.**

1. **Train, validate and test the following models**
    a. Linear and Polynomial Regression b. Ridge c. KNN d. Decision Tree
    d. Random forest e. Boosting etc.

1. **Checking the best models performance on the out of the data material (for which case we have chosen** Bi_2Te3 **)**

Next Deadline Goals

Retrieved lattice thermal conductivity data from AFLOW (restricted to 300 K) and curated it for machine learning training.

Perform Feature Engineering.

Train and validate with some of the base models like Linear Regression, KNN, Decision Tree.

***If time permits we’ll also be using Neural Networks for out project.**

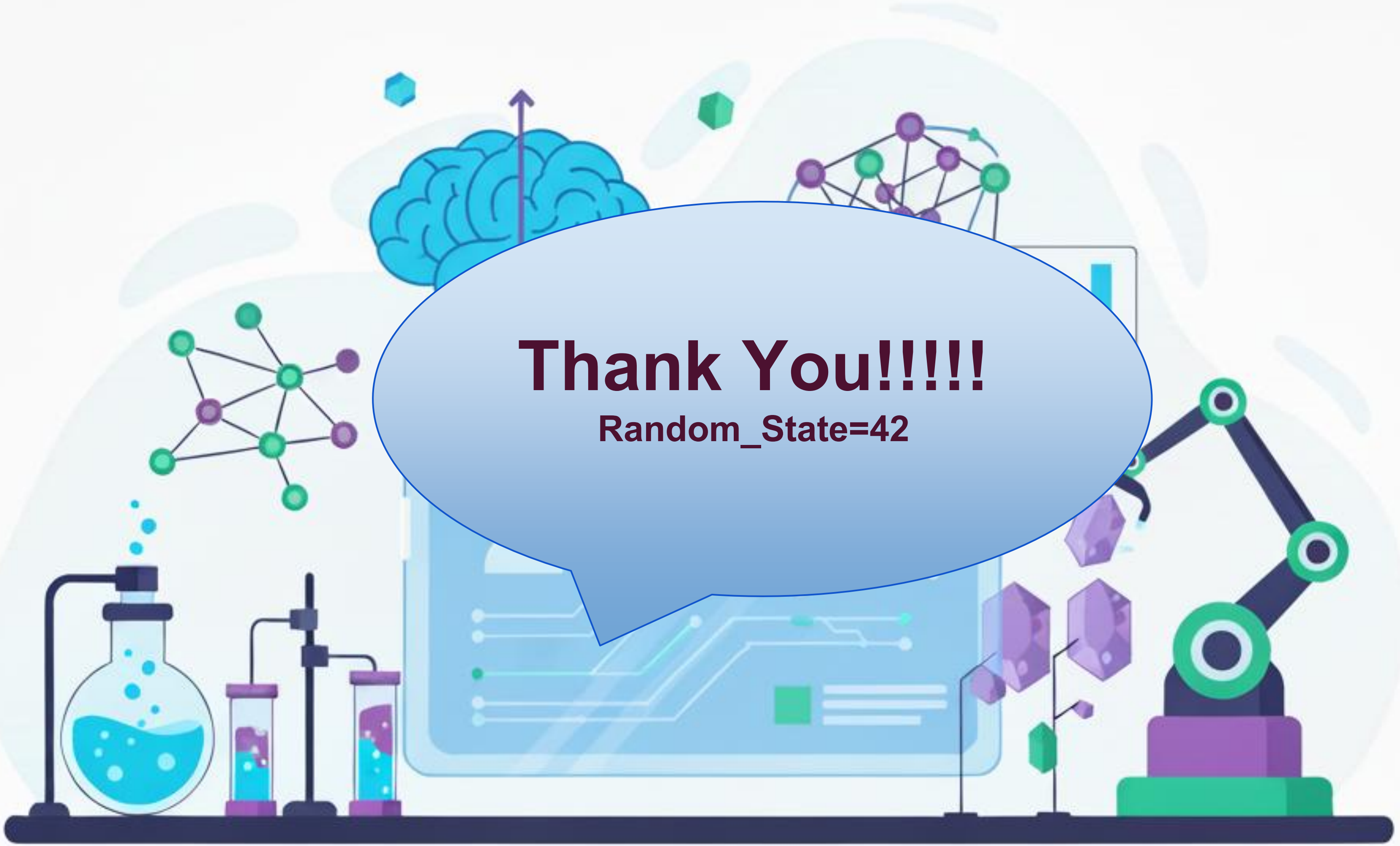
Conclusions

Machine Learning enables us to bypass expensive DFT/DFPT calculations by learning structure–property relationships directly from data.

By using curated AFLOW datasets and physics-informed features, we can predict lattice thermal conductivity with high accuracy.

Comparing models like Random Forest, Boosting, and Neural Networks will allow us to identify the most reliable approach for out-of-dataset predictions.

Ultimately, ML will help us rapidly screen materials and accelerate the finding of thermoelectric compounds with low lattice thermal conductivity, achieving our project's objectives efficiently.


Thank You!!!!!
Random_State=42