ISyE 6740 – Fall 2023
Project Proposal

Team Member Names: Patrick Gardocki

Project Title: Predicting Critical Temperature of Superconductors from Chemical Composition

Introduction

Superconductors are a material that under a certain critical temperature, the material is able to conduct current with near zero resistance. This makes superconductors very favorable in certain applications. The most common application is the use of them in fMRI machines used by the healthcare industry, as well as various research facilities, such as the Large Hadron Collider. Since the discovery of superconductors in 1911 by Heike Kamerlingh Onnes, there have been two main limitations that prevent more abundant use of superconductors [1]. One being that superconductors favorable properties emerge only at very low temperatures making it challenging to keep the material under its critical temperature. The other issue is that there is no current model to predict the critical temperature of superconductors.

The goal of this project is to attempt to apply machine learning methods in order to predict the critical temperature of a superconductor given its chemical formula. Materials with known critical temperatures will be used to evaluate the applied models and determine their accuracy. The models developed could later be used to find the theoretical critical temperature. Since there is no way to validate such result, that will not be tested in this project.

The data will be accessed from the Superconducting Material Database maintained by Japan's National Institute for Materials Science (NIMS) [2]. There are 21,263 superconductor entries with 81 variables per entry. These variables will be features of chemical properties of the material such as atomic mass, thermal conductivity, or number of valance electrons.

Data Source

The dataset used for this project was borrowed from 2018 paper titled, "A Data-Driven Statistical Model for Predicting the Critical Temperature of a Superconductor" [3]. It contains 21,263 rows and 81 input columns. The 81 input columns are made up of various feature extractions of each semiconductor material. Eight element properties were used to create features. The eight properties are atomic mass, first ionization energy, atomic radius, density, electron affinity, fusion heat, thermal conductivity, and number of valence. For each of the eight properties, ten methods were applied to extract features.
The following are the ten ways that the element data is manipulated.

Mean: $\mu = \frac{t_1+t_2}{2}$ *(1)*

Weighted Mean: $\nu = (p_1 t_1) + (p_2 t_2)$ *(2)*

Geometric Mean: $(t_1 t_2)^{\frac{1}{2}}$ *(3)*

Weighted Geometric Mean: $t_1^{p_1} t_2^{p_2}$ *(4)*

Entropy: $-\omega_1 \, ln(\omega_1) - \omega_2 \, ln(\omega_2)$ *(5)*

Weighted Entropy: $-Aln(A) - Bln(B)$ *(6)*

Range: $t_1 - t_2$ *(7)*

Weighted Range: $p_1 t_1 - p_2 t_2$ *(8)*

Standard Deviation: $(\frac{1}{2}((t_1 - \mu)^2 + (t_2 - \mu)^2))^{\frac{1}{2}}$ *(9)*

Weighted Standard Deviation: $(p_1(t_1 - \nu)^2 + p_2(t_2 - \nu)^2))^{\frac{1}{2}}$ *(10)*

Where,

t_i is an element's given property,

p_i is the ratio of an element in a material,

ω_i is the fraction of a given property of a material,

A = $\frac{p_1 \omega_1}{p_1 \omega_1 + p_2 \omega_2}$

B = $\frac{p_2 \omega_2}{p_1 \omega_1 + p_2 \omega_2}$

This is how 80 input columns were generated with the last one being the number of elements in the material. The element data was gathered using ElementData from Mathematica, and the superconductors material information was obtained from the Superconducting Material Database supported by NIMS.

Methodology

Initially, dimensionality reduction techniques will be used to determine if all 81 inputs are required to accurately predict the critical temperature, and if not, how many will be sufficient. Principle Component Analysis will be applied to the data to see if a few principal components will be able to capture most of the data's variance.

Neural Networks and Regression will be utilized to attempt to predict the critical temperatures of superconductors, and the two models will be compared in accuracy, efficiency, and reliability. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 will be used to evaluate the model's performance.

The neural network model will be tuned by adjusting its hyperparameters, such as activation types, learning rate, alpha and, number of layers. Most importantly, the number of neurons and total number of parameters will be based on the following equations. The first equation (11) suggests a number of neurons and the next (12) suggests the total number of parameters to use [4].

Number of Neurons: $N_h = \frac{N_s}{\alpha * (N_i + N_o)}$ (11)

Number of Parameters: $Ns * (Ni + No)$ (12)

Extreme Gradient Boosting, XGBoost, is another type of machine learning model that is a tree learning model. It adds new models iteratively to correct mistakes of previous models. XGBoost's framework contains custom loss functions and regularization methods to prevent overfitting. XGBoost is used due to its efficiency in speed and memory use. It utilizes parallel processing and can handle very large datasets very well.

Both models will be implemented and compared to determine if one performs better when applied to predicting critical temperatures of superconductors.

References

[1] ‘Superconductivity’ CERN. Available: https://home.cern/science/engineering/superconductivity

[2] MDR SuperCon Datasheet. National Institute for Materials Science. Available: https://mdr.nims.go.jp/collections/5712mb227

[3] Hamidieh, Kim. (2018). ‘A Data-Driven Statistical Model for Predicting the Critical Temperature of a Superconductor’. Available: https://arxiv.org/abs/1803.10260

[4] ‘Training an artificial Network – Intro’. Frontline Solvers. Available: https://www.solver.com/training-artificial-neural-network-intro