COS 472 Project Proposal

The Materials Project

Ben Franklin, Lucas Matheson, Colby Wirth

October 18, 2024

Overview

1. Project Introduction
2. Motivation
3. Goals
4. Methods
5. Evaluation

Project Introduction

What is the Materials Project?

- Multi-national
- Multi-institutional
- Collaborative project aimed at computing the properties of new materials
- Enables expedited scientific research by eliminating costs associated with conducting experiments on compounds

Motivation

Why contribute to the Materials Project?

- Test and Implement our new skills in a meaningful manner
- Contribute to a project much greater than ourselves
- Learn more about Machine Learning (and Chemistry!)
- Establish inter-department relationships!

Goals

- Goal 1:
  - Initially implement a ML model that can predict a handful of already known attributes with a high level of precision.
- Goal 2:
  - Predict unknown attributes for a large subset of compounds from the original dataset

Methods

- Choose which attributes are predictable and which are computable
  - We will not predict values that are easily computed.
  - Not trivial given our collective knowledge gap in Chemistry.
- Implement a Knowledge Discovery in Database (KDD) pipeline
  - Preprocess data (lots to do here)
  - Choose and train ML models such as Linear, L1 or L2 Regression
  - Evaluate and tune hyper-parameters
- Repeat for multiple attributes
- Analyze results for the final paper

Evaluation Criteria

- Goal 1:
  - Achieve 95% accuracy rate with predicting continuous attributes like 'band-gap' (and potentially other) attributes across $\sim$ 50,000 compounds
- Goal 2:
  - Predict these attributes for the remaining $\sim$ 100,000 compounds
- If time permits, repeat Goals 1 and 2 for many attributes.

End

End of Presentation

Questions?