Capstone 3 Proposal

By Zachary Brown

Title: Breathing easier: designing metal-organic frameworks for carbon dioxide capture from power plant exhaust

Problem Statement

The goal of this project is to use two theoretical metal-organic framework (MOF) databases to develop a machine learning model that will correlate MOF properties with CO_2 capacity at atmospheric temperature and pressure. A range of regression models will be trained on a combination of the QMOF database which has calculated electronic properties of 14,000 MOFs and the MOFXDB which has simulated isotherms for over 160,000 MOFs. The models will correlate MOF properties and the CO_2 capacity at atmospheric pressure and 25°C (298 K). Once a model has been optimized I'll test it against the NIST database of experimental MOF isotherms to determine whether the theoretical isotherms are a good prediction of experimental adsorption results.

Context

As the world works to reduce greenhouse gas emissions the need to carbon capture and sequestration increases. Scientists have been working for decades to develop porous materials capable of capturing greenhouse gases from power plant exhaust. One class of material that seems ideal for this task is metal-organic frameworks (MOFs). These materials are composed of organic linkers connected by metal oxide nodes, both of which have a wide range of options making them extremely flexible for tailoring to a given need. There are many reports on MOFs that have been tailored for exactly this purpose, and fortunately, given the repeating nature of their structure, chemists have been able to model large libraries of theoretical MOFs to help guide experimentation. While there are many other important factors to consider when designing a MOF for CO_2 capture, base CO_2 capacity sets a ceiling that gives an idea of the best a material can possibly perform. By identifying materials with the highest CO_2 capacity, we create a good starting group of materials worth probing further.

Criteria for Success

For this project to be successful I need to develop a regression model that can accurately predict the CO_2 capacity of MOFs that it hasn't been exposed to yet.

The Data

Electronic and structural properties of the training MOFs will be provided by the QMOF database. Isotherms of the training MOFs will be retrieved from the MOFXDB database. Experimental isotherms of synthesized MOFs will be used from the NIST/ARPA-E database of adsorbent materials.

Method

I'll download the CO_2 isotherm files from the MOFXDB as well as the full QMOF dataset. I'll join the two on the MOF ID, and then reduce the isotherm data just to atmospheric pressure (1 bar) and temperature (298 K). I'll train and test a wide range of regression models on the joined dataset until I've narrowed down the most accurate model possible. I'll then download the NIST/ARPA-E dataset and use it as a test set for the final model.

Deliverables

This project will result in a written report summarizing the data manipulation, model testing, and final model performance. It will also produce a presentation detailing these same features of the project. I'll also save the algorithm so that it can be used for current and future projects to extend this analysis in the future.