FOSILS: FrOm SmILes to Stars - an end-to-end workflow for estimating the capacitance of new battery materials

Michał Wiliński, Łukasz Sztukiewicz, Ignacy Stępka, Wojciech Mączka, Dariusz Brzeziński

Poznan University of Technology, Poznań, Poland

Group of Horribly Optimistic STatisticians

Motivation

- In modern world ability to rapidly iterate through the scientific discovery process is important
- The biggest bottleneck in material science is the time-consuming process of creating and combining new substances
- We need to strategically prioritize candidate molecules for synthesis

Objectives

- Simplify the entire discovery process, spanning from proposing molecules to ranking them effectively.
- Integrate established heuristics utilized by scientists into the prediction framework.
- Create a user-friendly web application tailored for material scientists, designed to boost research efficiency.

Solution

- Build comprehensive end-to-end solution that generates new molecules and predicts the target capacity of CTFs based solely on the encoded representation of the molecular substrates in Simplified Molecular Input Line Entry System (SMILES)

Results

- No signal in the data we received from our experts --> random ranking
- With LightGBM variant and the combined dataset (our data + data from literature with already extracted features) , **74% of pairwise comparisons resulted in a correct ranking**

Future Work

- Obtain more ground truth data containing SMILES
- Create more flexible molecule generation method using GNNs

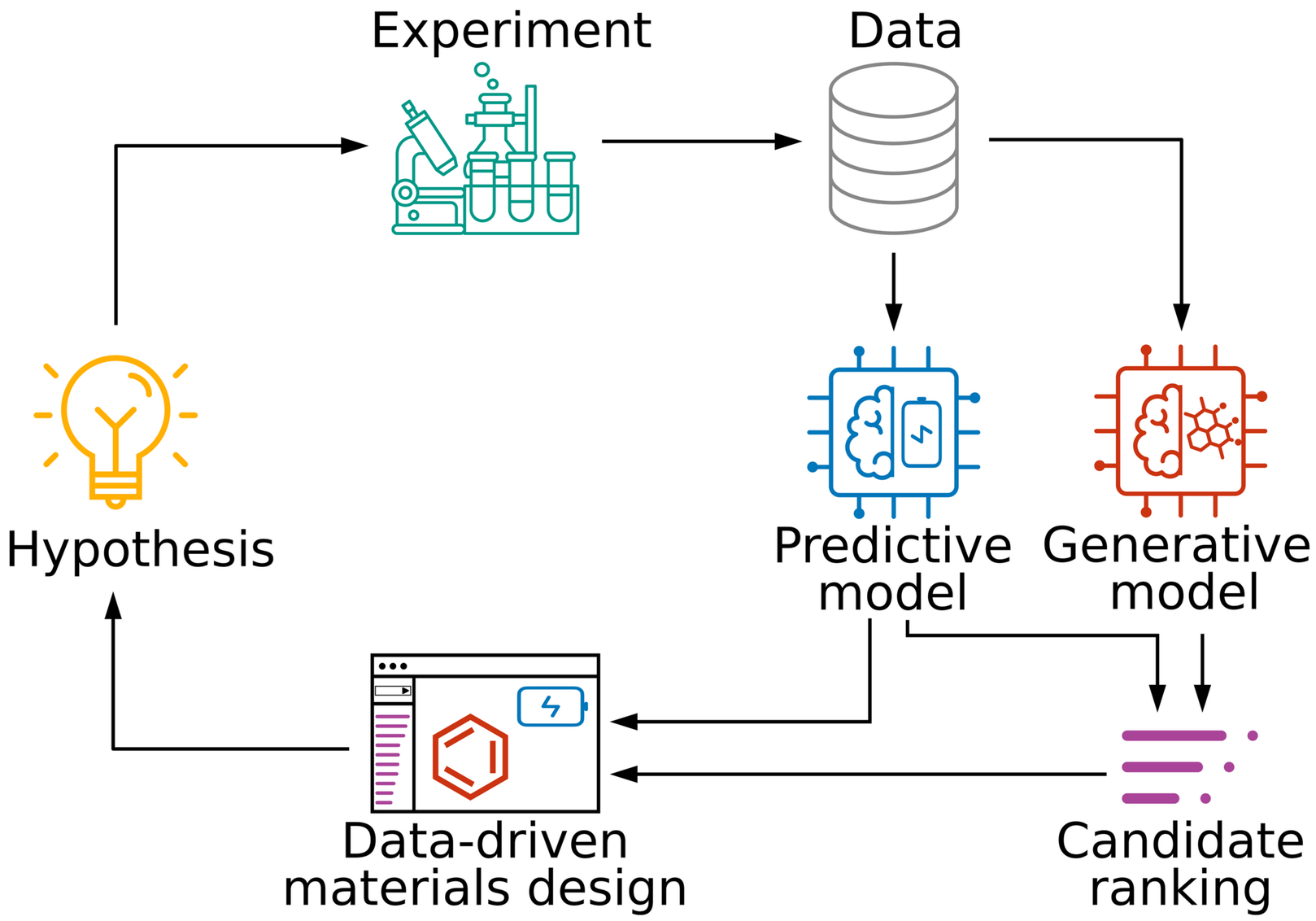
Acknowledgments

- We thank Prof. Płońska-Brzezińska and Dr. Hryniewicka for providing theoretical chemistry background
- We also acknowledge the involvement of Paweł Śnioszek and Antoni Klorek in the development process

References

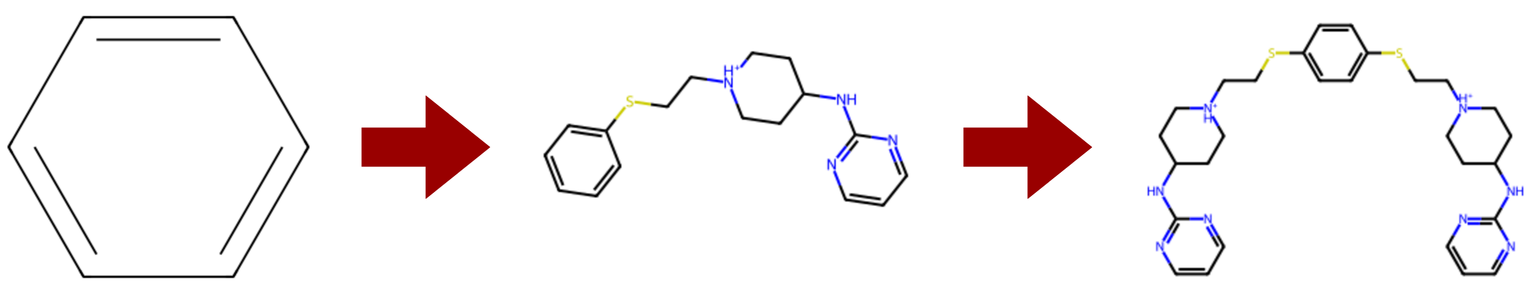
1. Saad et al., “Data-Driven Machine Learning Approach for Predicting the Capacitance of Graphene-Based Supercapacitor Electrodes.”
2. Ishida et. al, “Chemtsv2: Functional molecular design using de novo molecule generator”
3. Liao et al., “Advances in the Synthesis of Covalent Triazine Frameworks.”
4. Zhu et al., “Artificial Neural Network Enabled Capacitance Prediction for Carbon-Based Supercapacitors.”

Framework



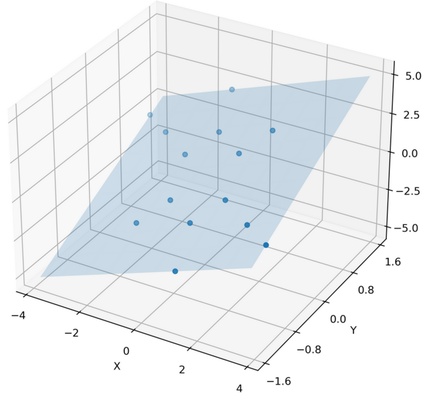
Methods

Generation

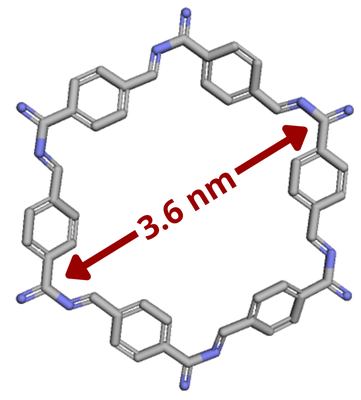


Feature engineering

Spatial flatness



Pore size



CTF capacity prediction