MESSAGE PASSING NEURAL NETWORKS TO PREDICT ADSORPTION PROPERTIES OF NANOPOROUS MATERIALS

An Application of Graph Neural Networks

DISCOVERING MATERIALS FOR GAS STORAGE AND SEPARATION

Covalent Organic Frameworks (COFs)

- Assembled modularly from organic molecules through strong covalent bonds.
- 560,000 discrete structures hypothesized computationally to date.
- Exceptional porosity affords exemplary storage and separation of gasses.
- Applications include carbon capture, low pressure fuel storage, cancer detection, etc.

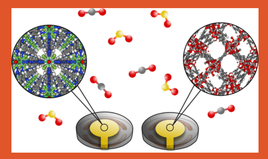


Fig 1. Gravimetric Gas Array Sensors using MOFs

TOOLS & PITFALLS IN COMPUTATIONAL CHEMISTRY

The adsorption properties of most COFs are unknown. **How do we know which are best suited for our application?**

- **Synthesis:** COFs can be synthesized and tested, but this is costly and inefficient.
- **Simulations:** Molecular dynamics are extremely accurate, but also computationally expensive.
- **Machine Learning:** Has the potential to be very accurate and efficient, but often relies on hand-crafted features as inputs, which are not readily available.

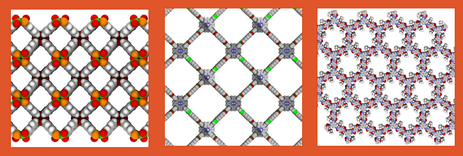


Fig 2. Various crystal structures of COFs

MATRIX FACTORIZATION & THE COLD START

An innovative approach to learning adsorption properties is the recommender system approach.

- Learns to predict adsorption by observing a partially complete target-set, overcoming the need for hand crafted features. However, each material requires 1 or more known targets to compute.

How do we approach a cold start problem, in which no targets, and only the molecular structure is known?

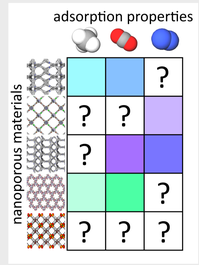


Fig 3. An incomplete material-property matrix with the cold-start problem

MESSAGE PASSING NEURAL NETWORKS (MPNNS)

We design and train an MPNN to predict uptakes of 16 different gasses. We train on a partially complete set of adsorption tasks (Θ) and the molecular graph. **Our model only requires the molecular graph as input during inference, thus overcoming the cold start problem.**

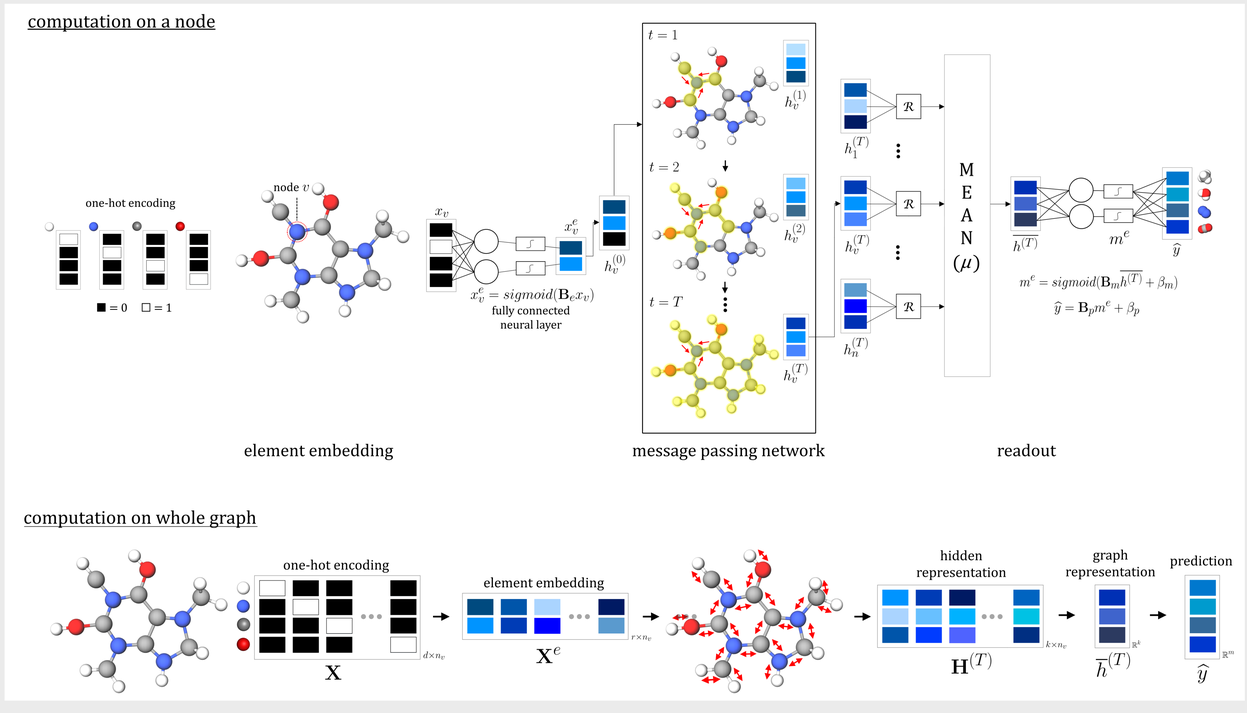


Fig 4. Our MPNN Architecture, which learns a vector representation of the COF m^e that is then used to generate prediction $\hat{y}$

Learn More!

GRAPH SIMILARITY

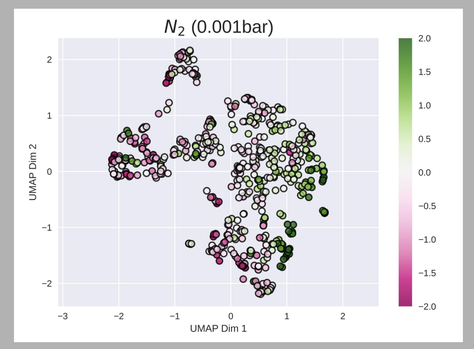


Fig 5. UMAP where each point represents a COF colored by its N_2 adsorption.

We find that the learned vector representation m^e correlates well to adsorption tasks, seen by the clusters of COFs with similar uptake in *Fig 5*.

MODEL PERFORMANCE

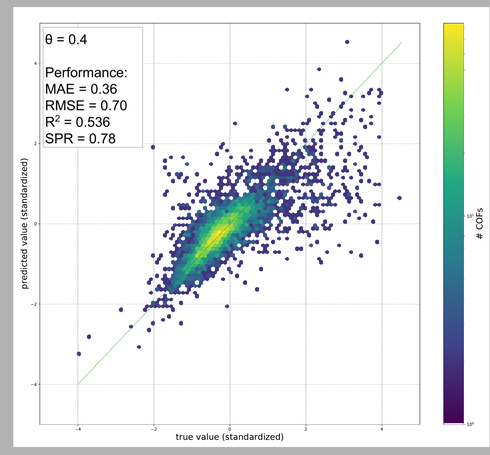
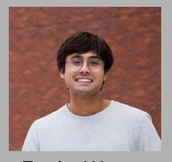
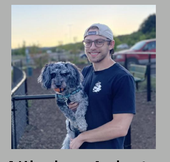
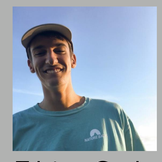


Fig 6. Parity plot mapping all predicted vs actual adsorption tasks

ACKNOWLEDGEMENTS


Faaiq Waqar
waqarf@orst.edu


Nikolas Achatz
achatzn@orst.edu


Tristan Gavin
gavint@orst.edu

Thank you to Dr. Xiaoli Fern (Associate Professor, OSU) and Dr. Cory M. Simon (Assistant Professor, OSU) for their support and mentorship. Thank you to the National Science Foundation (NSF) for the necessary funding and support (**Award # 1920945**).

Oregon State University