Message Passing Neural Networks to Predict Adsorption Properties of Nanoporous Materials

[1]Faaiq Waqar
[1]Tristan Gavin
[1]Nikolas Achatz
[1]Xiaoli Fern
[2]Cory M. Simon

[1]School of Electrical Engineering and Computer Science. *Oregon State University. Corvallis, OR.*
[2]School of Chemical, Biological and Ecological Engineering. *Oregon State University. Corvallis, OR.*

What are Covalent Organic Frameworks (COFs)?

Nanoporous materials (NPMs) leverage physical interactions in nanometer-scale pores to selectively adsorb and concentrate gasses and find use in applications such as sensing, storage, and catalysis.

COFs are a diverse subclass of modular, lightweight NPMs, that are assembled from organic molecules through strong covalent bonds. There exist **560,000 discrete structures hypothesized in silico to date**.

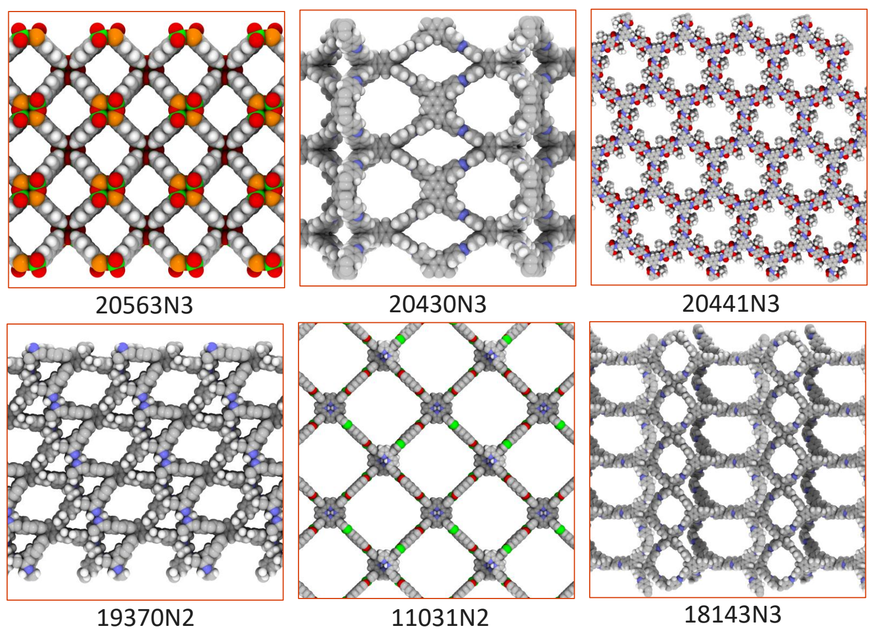

Fig 1. Various crystal structures of COFs observed in our dataset

Why Machine Learning?

With so many potential structures to choose from, **how can we determine which COFs are best suited for our application?**

- **Synthesis:** Manufacturing and testing every COF would be expensive and burdensome.
- **Simulation:** Molecular dynamics are extremely accurate, at the expense of computational complexity.
- **Machine learning:** Potentially very accurate and efficient, but often rely a vast input feature space.

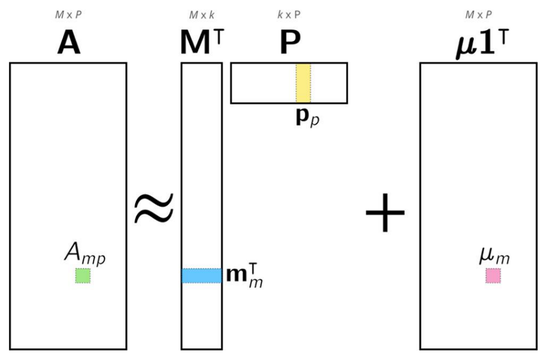

Fig 2. Low-rank model of the material–property matrix where $A \approx M^T P + \mu 1^T$

Matrix Factorization

Matrix factorization utilizes partially observed targets as inputs to impute missing values. An incomplete NPM-property matrix (A) can be completed by factoring the learned material (M^T) and property space (P).

When a material has no known observations, the algorithm fails. **How can we approach this cold-start problem, in which only the molecular structure of a material is known?**

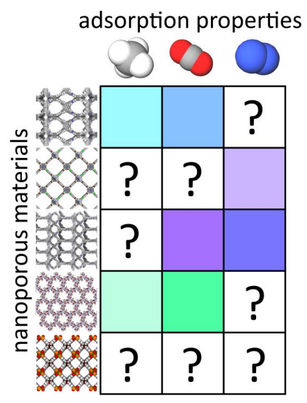

Fig 3. An incomplete material-property matrix with the cold-start problem

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 bonding structure of the molecule as a graph. **Our model only requires the bonding graph as input during inference, thus overcoming the cold-start problem.**

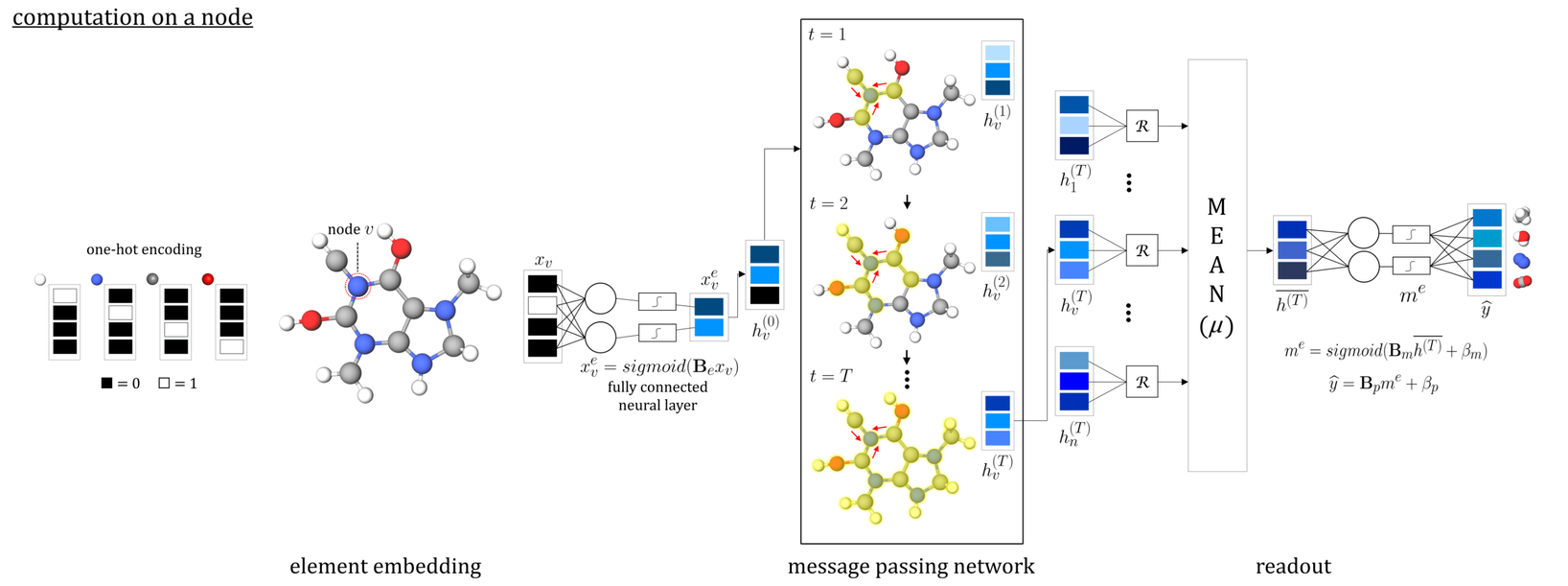

Fig 4. Our MPNN Architecture, which learns a material and property embedding, analogous to the matrix factorization approach: $m_m^T \leftrightarrow m^e,\ P \leftrightarrow B_p,\ \mu 1^T \leftrightarrow \beta_p,\ A_m \leftrightarrow \hat{y}$

Latent Space

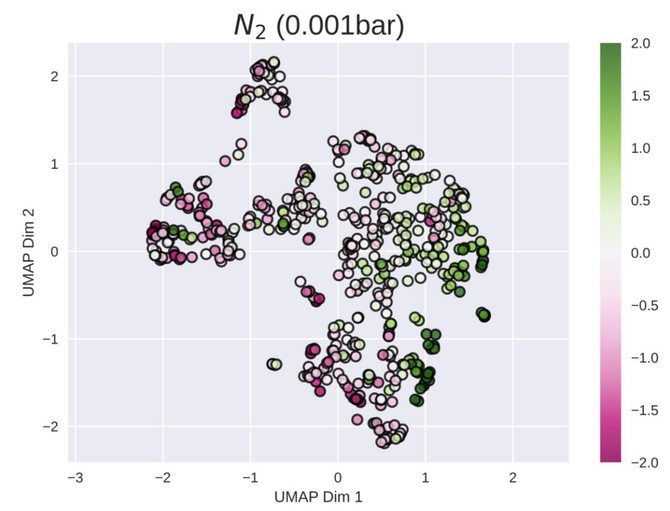

Fig 5. UMAP of Latent Representation vs. Low Pressure N_2 Adsorption

We learn a rich latent representation of the input graph in the form of m^e. This maps to the predicted uptakes ($\hat{y}$) through a feed forward neural network.

Performance

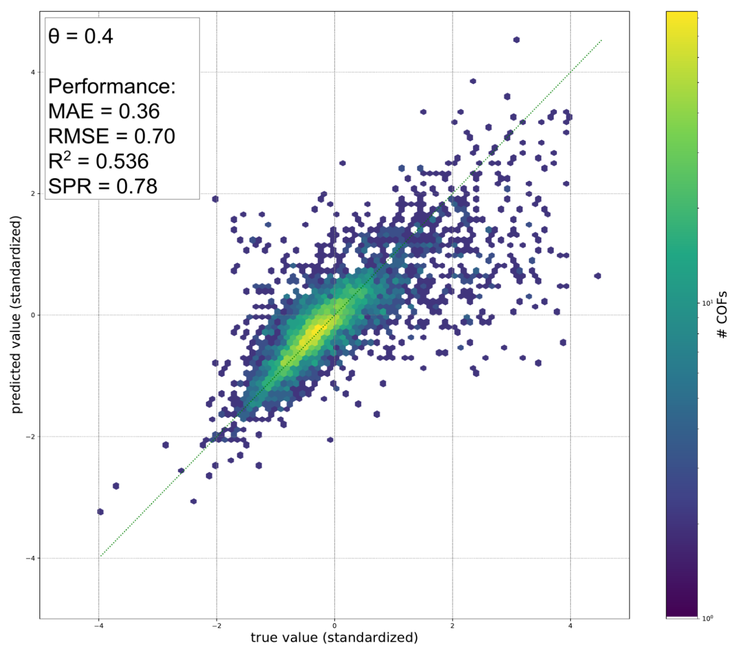

Fig 6. Parity plot mapping all predicted vs ground truth adsorption tasks

Looking forward…

> Changes to Training and Architecture

> In a multitask setting, shared parameters may cause sacrificial performance in some tasks. We look to improve performance by using our architecture as a pre-training step.

> Bayesian Optimization (BO)

> Using the latent space of materials generated from the MPNN, we may be able to optimize screening using BO.

Acknowledgements

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

Learn More!

Scan this QR code to see our group website to learn more about our projects and code.