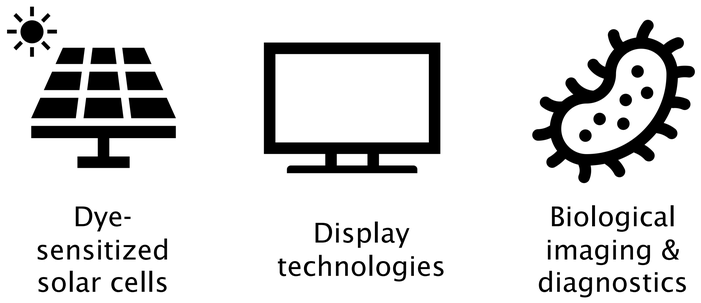
Artificial Intelligence Applications in the Design of Novel Dye Molecules with Targeted Optical Properties

Kevin P. Greenman[1], William H. Green[1], and Rafael Gómez-Bombarelli[2]
[1]Department of Chemical Engineering, MIT; [2]Department of Materials Science and Engineering, MIT

Massachusetts Institute of Technology

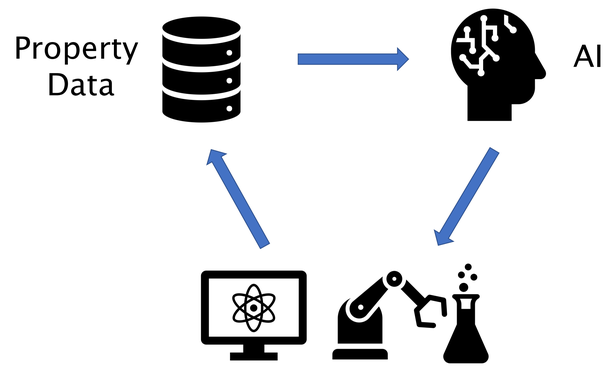
Background

Numerous applications of molecular dyes



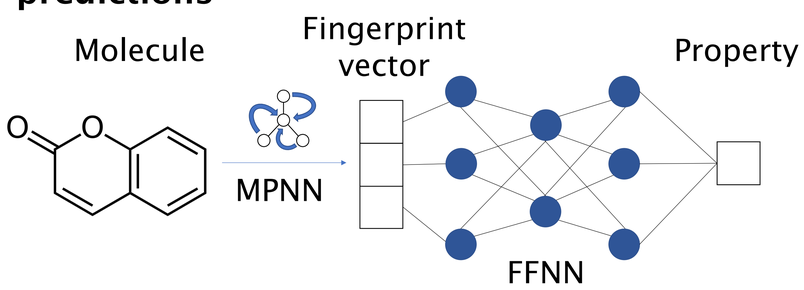
Dye-sensitized solar cells | Display technologies | Biological imaging & diagnostics

AI enables humans to navigate enormous molecular design space



Methods

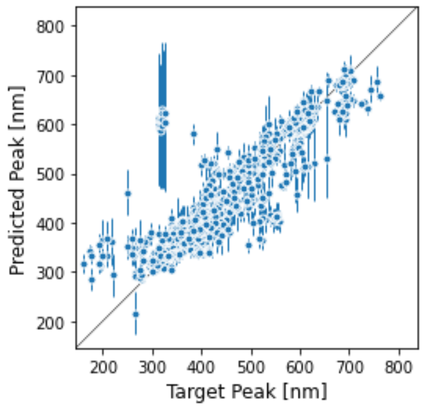
Message-passing neural network (MPNN) learns optimal representation for property predictions[1]



- Molecule represented as a graph (nodes & edges)
- "Messages" passed between neighboring atoms to learn chemical environment
- Aggregate atom vectors to get molecule vector
- Use traditional neural network on property regression problem

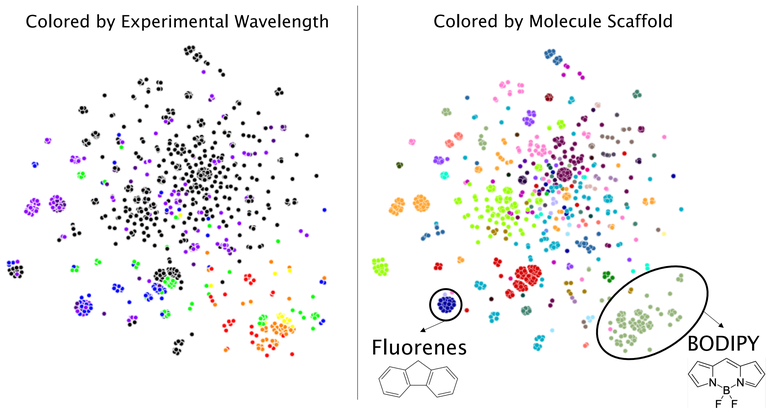
Results

MPNN predicts peak absorption wavelength accurately

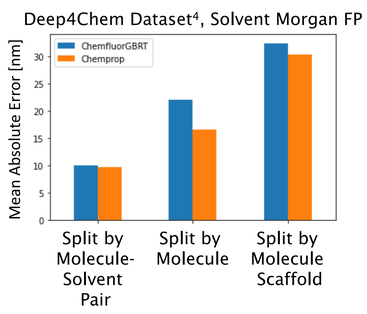


- MPNN outperforms physics-based methods (TD-DFT)
- Uncertainty from ensemble variance provides estimate of model's confidence in each prediction

MPNN embeddings are interpretable in 2-D



Predictions are generalizable to unseen chemistries



- Different train-validation-test split types assess generalizability
- MPNN method is superior to existing statistical methods[3]

Conclusions

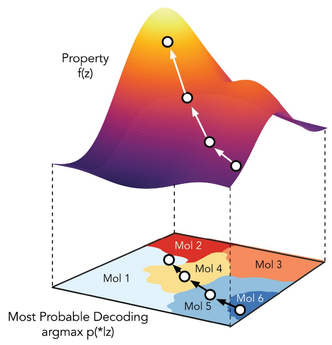
- MPNN method produces accurate, generalizable predictions of the peak absorption wavelength
- This method can be extended to predict other optical properties of interest in molecular dye design

Ongoing & Future Work

Active Learning

- Even with high-throughput techniques, gathering new data is expensive and time-intensive for large libraries
- Active learning uses model uncertainty to determine which new datapoints would be most useful to measure for model improvement

Inverse Design with Generative Models



- Generative models can be trained to propose molecular structures as outputs, given a desired property as an input[2]
- Gradient descent/ascent in continuous latent space that maps to discrete molecule space

Autonomous, Closed-Loop System

- Given a molecular design objective, a computer and a robotic experimentalist will make autonomous decisions to pursue that goal

References

1. Yang, K.; Swanson, K.; et al. *J. Chem. Inf. Model.* 2019, *59* (8), 3370–3388.
2. Gómez-Bombarelli, R.; et al. *ACS Cent. Sci.* 2018, *4* (2), 268–276.
3. Ju, C.-W. W.; et al. *J. Chem. Inf. Model.* 2021, *61* (3), 1053–1065.
4. Joung, J. F.; Han, M.; et al. *Sci. Data* 2020, *7* (1), 1–6.

Acknowledgements

This work is funded by the Defense Advanced Research Projects Agency (DARPA). Thank you to my team members Simon Axelrod, Camille Bilodeau, Brent Koscher, Charles McGill, Florence Vermeire, and Michael Forsuelo for their ongoing support and advice.