Uncertainty estimation for exploring chemical space with neural networks

Jon Paul Janet[a] and Heather J. Kulik[a]

[a] *Department of Chemical Engineering, Massachusetts Institute of Technology*
hjkulik@mit.edu

Machine learning has enormous potential to accelerate material discovery by supplementing first-principles density functional theory (DFT) simulation with low-cost approximations from artificial neural networks (ANNs). This allows many more configurations to be explored than would be possible using DFT alone, and is particularly valuable where existing knowledge is limited and simulations are expensive and challenging, such as in the design of transition metal (TM) complexes, which have important applications as selective homogeneous catalysts and molecular sensors.

However, even models that show good performance on test data can fail to generalize when used to explore diverse chemical space because training data invariably reflects the biases and limitations of the data generation process, often being restrictive in chemical variety or focusing on already well-explored chemical space, e.g. small organic molecules. These problems are exacerbated for TM complexes where large databases are not available and understanding of similarity is less established. If ML models are to realize their potential for discovery, they must be equipped with reliable measures of system-specific uncertainty. We previously introduced a simple heuristic based on distance to training data in feature space that correlated well with model uncertainty, and we used it to design spin crossover materials and complexes with targeted frontier orbital properties from diverse pools of candidates using a genetic algorithm. Because we could control model uncertainty, a majority of our leads were validated using DFT.

Unfortunately, such geometric estimates are dependent on good feature engineering and are not straightforwardly applicable in cases with large feature vectors or when using deep networks to learn from simple representations, e.g. graph convolutions. Instead, we propose measuring similarity in the ANN latent space, which provides substantial improvements over raw feature distances. We convert distances into quantitative uncertainty metrics using a simple probabilistic model, and compare to commonly-utilized uncertainty metrics for ANNs: ensemble models and Monte Carlo dropout, in two applications 1) generalization of spin splitting energy predictions trained on simple ligands to experimental structures from the Cambridge Structural Database and 2) atomization energy predictions on a benchmark organic data set. We observe that the proposed model provides the best performance, without requiring additional investment in training multiple models. This makes the proposed model an attractive candidate for active learning and chemical discovery.