A Material Informatics Approach for Graph Neural Network Learning of the Relaxed Energy of Unrelaxed Structures

Jason Gibson

Department of Material Science and Engineering, University od Florida, Gainesville FL

Computational materials discovery has continually grown in utility over the past decade due to advances in computing power and crystal structure prediction algorithms (CSPA). However, the computational cost of the ab initio calculations limits its utility to small unit cells, reducing the compositional and structural space the algorithms can explore. Past studies have bypassed many unneeded ab initio calculations by utilizing machine learning methods to predict formation energy and determine the stability of a material. Specifically, graph neural networks (GNN) display high fidelity in predicting formation energy. Traditionally GNN are trained on large datasets of relaxed structures. Unfortunately, the unrelaxed candidate structures produced by CSPA are often far from the minima of the relaxed structures. This leads to poor model performances when trying to filter energetically unfavorable structures produced by CSPA prior to ab initio evaluation. In this work, we show that as training progresses, the prediction error on relaxed structures reduces, while the prediction error on unrelaxed structures actually increases, suggesting an inverse correlation between relaxed and unrelaxed structure prediction accuracy. To remedy this behavior, we propose a simple, physically motivated, computationally cheap perturbation technique that augments our data to dramatically improve predictions on unrelaxed structures. On our test set consisting of 716 Nb-Sr-H hydrides, we found that training a crystal graph convolutional neural networks, utilizing our augmentation method, reduced formation energy prediction mean absolute error by 70.5% compared to training with only relaxed structures. This error reduction accelerates CSPA by improving the model's ability to accurately filter out energetically unfavorable structures.