Machine learned Potential Accuracy and Cost: The Role of Training Set Size, Potential Complexity, and Cut-Off Radius

Jason Gibson, Jan Janssen, Laura Lopes, Richard Hennig, Danny Perez

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

The promise of obtaining accuracy on par with first-principle calculations at the computational cost of empirical potentials has made machine-learned interatomic potentials (MLP) attractive alternatives for studying and characterizing materials. However, while many MLPs have reported accuracy within several meV/atom of first principle approaches, the limited diversity of available data has resulted in poor transferability of MLPs and a hindered understanding of how the MLPs parameters affect their accuracy. This study investigates the relationship between the size of the training set, MLP complexity, and cut-off radius using three well-established MLPs on a diverse tungsten dataset. Our results show that there is a strong interdependence between these factors. We also examine the balance between MLP accuracy, cost of curating the training data and the computational cost of using the MLP in simulation. We conclude by providing a practical approach to designing an MLP efficiently.