Machine Learning Interatomic Potentials: Reference Training Data on the Hands of Workflows

Alireza Ghasemi

Erlangen National High Performance Computing Center (NHR@FAU), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Martensstraße 1, 91058 Erlangen, Germany

Machine learning interatomic potentials (MLIPs) have drawn great attention in recent years. A variety of satisfactory methods have been developed thus far. Nevertheless, a key success of every MLIP heavily relies on the quality of the training dataset whose generation typically requires considerable treatment by a qualified scientist. To generate diverse and comprehensive training reference data points, we employ workflows that execute in a cyclic process way the combination of crystal structure prediction (CSP) and training of MLIPs. The generated conformations in CSP runs are post-processed to exclude similar structures and append those which either fill the holes of the current feature space or expand it. All steps are performed automatically with minimal human intervention.