Evaluating Contrastive Learning methods performance on Many-body UMLIPs architecture

Thai-Bach Le,[1] Huu T. Do,[2] and Truong Son Hy[3]
[1] *Fulbright University Vietnam, Ho Chi Minh City, Vietnam*
[2] *University of Illinois, Chicago, Illinois, USA*
[3] *Indiana State University, Indiana, USA*
Email: TruongSon.Hy@indstate.edu

ID: PST05
Poster
Aug. 24, 16.00 - 17.30

Pre-training techniques leverage unlabeled data to address the challenges of scarce and resource-intensive labeled datasets, while enhancing out-of-distribution prediction performance. Recent advancements in Universal Machine Learning Interatomic Potential (MLIP) architectures which incorporated many-body interaction techniques, have achieved a balance between ab initio accuracy and the computational efficiency of force field methods. Despite extensive research on pre-training techniques, most studies have focused on simpler architectures like GemNet and GIN on organic domain, leaving a gap in evaluating MLIP architectures and their application to inorganic materials. In this study, we assess node-level pre-training techniques applied to the M3GNet architecture. We utilized the OQMD Dataset with The Materials Project for unsupervised learning, and employ The Materials Project with Formation Energy for the downstream task. Our findings are consistent with prior works, demonstrating the limitations of naive contrastive learning methods as disrupt causation in dataset molecular structures. Additionally, our results indicate promising potential for generative self-supervised training in future MLIPs models toward transfer learning into specific domains.

Keywords: Pretraining techniques, Universal Machine Learning Interatomic Potential (MLIP), Many-Body Interaction, Contrastive Learning, Generative Self-Supervised Training, GNN, M3GNet Architecture, OQMD Dataset, The Materials Project