Invited Online Talk

HydraGNN: Scalable Machine Learning and Generative AI for Accelerating Materials Design

Dr. Jong Youl Choi

Time: Thursday, 05/30/2024, 1:00 PM — 2:00 PM (CST)

Zoom Link: https://unt.zoom.us/j/86512602380

Email: choij@ornl.gov

Abstract: The rapid progress in CPU technologies, GPU accelerators, and high-performance computing (HPC) systems has significantly enhanced AI and deep learning (DL) capabilities across various fields. We explore the challenges in developing large-scale training for generative AI models focused on material design. Employing HydraGNN, a scalable graph neural network (GNN) framework, in conjunction with DDStore, a distributed in-memory data store, we enable extensive data distribution across the supercomputing resources provided by the US Department of Energy (DOE). We will discuss insights into our implementation and the notable reduction in I/O overhead within HPC environments. The effectiveness of HydraGNN and DDStore is demonstrated through their application in molecular design, where a GNN model predicts the ultraviolet-visible spectrum using a dataset of over 10 million molecules. Enabling efficient scaling of training to thousands of GPUs on the Summit and Perlmutter supercomputers, DDStore accelerates DL training speed significantly, achieving up to a 6.15-fold increase in performance compared to our initial methods. We will explore the performance improvements on the new Frontier supercomputer at the Oak Ridge National Laboratory (ORNL), shedding light on the evolving landscape of supercomputing in AI research.

Bio: Dr. Jong Youl Choi is a researcher working in Discrete Algorithms Group, Computer Science and Mathematics Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee, USA. He earned his Ph.D. degree in Computer Science at Indiana University Bloomington in 2012 and his MS degree in Computer Science from New York University in 2004. His areas of research interest span data mining and machine learning algorithms, high-performance data-intensive computing, parallel and distributed systems. More specifically, he is focusing on researching and developing data-centric machine learning algorithms for large scale data management, in situ/in-transit data processing, and data management for code coupling.