NJIT | INSTITUTE FOR DATA SCIENCE

Data Science Seminar Series

Fast and Accurate Prediction of Potential Energy Functions for 3D Molecular structures

Sutanay Choudhury, Ph.D.

Chief Data Scientist

Pacific Northwest National Laboratory

Date: Wednesday, February 15, 2023
Time: 4:00 PM - 5:00 PM
Location: Zoom Virtual Room
Web Link: RSVP Here

Whether it is designing lightweight materials or discovering energy storage solutions for tomorrow, accurate modeling of atomistic interactions is critical. Nature is fond of geometry, and different molecular structures show unique geometric traits. However, training neural networks to predict properties of such geometric structures comes with a unique set of challenges. As molecular structures grow larger, the number of possible input states for 3D-GNNs increases exponentially. This presents a challenge for generalizability of the models and requires innovation to answer questions, such as can a model trained over a region of the chemical space be applied to another? How do we push the limits of today's ML frameworks and hardware architectures to train these models in hours? In this talk I'll present our contributions toward solving those distinct problems, including: 1) introducing one of the largest machine learning (ML) benchmark datasets for chemistry; 2) presenting graph-theoretic approaches for both development and interpretation of neural network models; 3) providing reference implementations via message-passing frameworks, such as PyTorch Geometric; and 4) exploring co-design of novel software-hardware architectures for rapid training of these models.

Dr. Sutanay Choudhury is a Chief Data Scientist at PNNL with 10+ years of experience in Artificial Intelligence and Data Science. His current research focuses on development and application of graph representation learning, prediction, and reasoning methods towards solving problems in chemistry, medicine and power grid. Dr. Choudhury serves in multiple leadership roles at PNNL, such as the Co-Director of PNNL's Center for Theoretical and Computational Chemistry and thrust lead for PNNL's internal initiative on AI research. He has served as PI, Co-PI and contributor on multiple projects funded by the DoD, DHS, DOE and PNNL's internal research programs developing multiple systems focusing on graph analytics and graph neural networks. He developed StreamWorks, a streaming graph analytics system that received R&D100 award for novel applications in cyber-security.