UNC GREENSBORO | Department of Computer Science

Department of Computer Science

Colloquium Series

Supercharging Molecule Design, Protein Folding Path and Allosteric Regulation Studies with Machine Learning

Yong Wei, Ph.D.

Professor, Department of Computer Science
High Point University

Date: Tuesday, March 28th,2023
Time: 10:00-11:00am
Location: Eberhart Building 161

Biography:

Yong Wei is a Computer Science professor at High Point University with a primary research focus on applied machine learning and its applications in various fields, such as computational chemistry, materials science, computational physics, and computer vision. Despite a demanding teaching schedule, he has authored over twenty papers in both journals and international conference proceedings, including the Journal of Chemical Theory and Computation, ACS LANGMUIR, ACS Journal of Physical Chemistry (b), ACM Multimedia Conference, IEEE Trans. on Multimedia, ACM Trans. on Multimedia Computing, Communications and Applications, and IEEE International Conference on Machine Learning and Applications.

Abstract:

Chemical Environment Adaptive Learning

The understanding material structure-property relationship is crucial in material discovery. Graphitic carbon nitride (g-C_3N_4) and its doped variants recently gained much attention due to their wide applications as optical materials, demanding accurate prediction of their band gaps. However, quantum simulation methods are computationally intractable when treating many doped molecular structures. The recent success of graph neural networks (GNNs) in chemical structure learning offers an opportunity to predict material properties and accelerate material discovery despite their inefficient exploitation of local chemical environment information underlying distinct molecular structures. We propose a novel machine learning (ML) algorithm called Chemical Environment Adaptive Learning (CEAL), which can satisfactorily capture the characteristics of atoms' local chemical environments. Our benchmark results demonstrate more than 100% improvement in band gap prediction accuracy for CEAL compared with all established GNNs on doped g-C_3N_4. Most importantly, a single unified CEAL model was found to be able to precisely foresee the band gaps of various doped g-C_3N_4 structures, making it a valuable tool for performing high-throughput prediction in material design.

Self-supervised Learning

Computational chemistry simulations have been used to study the protein folding trajectory. Most data produced in these simulations are unlabeled, demanding self-supervised ML algorithms to extract in-depth information from them for downstream tasks.

Case 1: Understanding the interfacial behaviors of biomolecules is crucial to applications in biomaterials and nanoparticle-based biosensing technologies. Discontinuous molecular dynamics (DMD) simulations have been utilized to study the lysozyme folding path in the adsorption process on a hydrophobic graphene surface. Using DMD and autoencoder (a self-supervised ML algorithm), the stages of the lysozyme structure trajectory in the adsorption process were identified, revealing the roles of residue-surface hydrophobic interaction and the $\pi - \pi$ stacking interactions in the adsorption process.

Case 2: Allosteric regulation is common in protein-protein interactions and is thus promising in drug design. Nevertheless, the mechanism of allosteric regulation remains elusive for most proteins, including SARS-CoV-2 spike protein, despite extensive experimental endeavors over the past years. All-atom explicit solvent molecular dynamics (MD) simulations have been used to study the allosteric regulation of SARS-CoV-2 protein when peptide binds to the polybasic cleavage sites. Contrastive learning is the state-of-the-art self-supervised ML algorithm that can capture good feature representations from unlabeled data. Utilizing the contrastive learning algorithm in conjunction with all-atom explicit solvent MD simulations, we were able to reveal the propagation route of the spike protein's backbone fluctuation, paving the way for the rational design of allosteric antibody inhibitors.