



Contribution ID: 49

Type: Talk

Interpretable Affinity Prediction and Generative Design for Protein-Ligand Interactions

Thursday, 15 February 2024 16:15 (30 minutes)

In this talk I will introduce our recent works on machine learning for protein-ligand interactions, including explainable prediction of protein-ligand binding affinity and structure-based *de novo* ligand design.

First, as deep learning methods for modeling protein-ligand interactions are increasingly improving their accuracy, their interpretability is often under-explored. We had previously developed DeepAffinity that, starting with protein sequences and chemical identities, predicts both protein-ligand affinities and underlying intermolecular contacts. DeepAffinity unified recurrent and convolutional neural networks, exploited both labeled and unlabeled data, and adopted attention mechanisms for interpreting affinity predictions.

Our recent advances cover two aspects. (1) As attention mechanisms alone are inadequate, we regularize attentions with predicted 3D structural contexts and supervise attentions with non-bonded intermolecular contacts, which leads to DeepAffinity+. We further design DeepRelations with a physics-inspired, intrinsically explainable architecture. (2) Thanks to the recent breakthroughs in protein structure prediction, we consider protein data as available in both modalities of 1D amino-acid sequences and predicted 2D contact maps; and we introduce cross-modality protein embedding schemes. Moreover, we pre-train the protein embeddings through self-supervision using unlabeled data. Our results indicate that attention supervision, cross-modality, and self-supervision further improve the accuracy, the interpretability, and the generalizability of affinity prediction especially for unseen proteins.

Second, for simultaneous *de novo* design and structure prediction, we formulate the problem as 3D graph generation conditioned on target protein structures; and we solve the problem through diffusion models, a recent class of generative AI methods. By embedding 3D ligand graphs' identity and geometry jointly into a latent space, we diffuse the 3D graph iteratively in the latent space and maintain roto-translational equivariance. Compared to other generative models, our latent diffusion model generates tight and diverse molecules while being an order of magnitude faster to train.

Submitting to:

8th CAPRI assessment meeting

Primary author: SHEN, Yang (Texas A&M University)

Presenter: SHEN, Yang (Texas A&M University)

Session Classification: CAPRI