Does Inter-Protein Contact Prediction Benefit from Multi-Modal Data and Auxiliary Tasks?

Arghamitra Talukder, Rujie Yin, Yuanfei Sun, Yang Shen, Yuning You

Hypotheses

Proteins are essential biomolecules of life. Proteins are represented by 1D sequences folding into 3D structures (multi-modality) and interact to form assemblies to function (multi-scale functionality). Unlike protein structure prediction, protein complex/assembly structure prediction has yet seen as successful technology breakthroughs. We hypothesize improvements might root from two perspectives:

• incorporating multi-modal features for inputs, and
• Synergizing contact predictions with auxiliary predictive tasks.

Model Architectures in Experiments

A general framework of model architecture using encoders, common and task-specific MLP layers.

Results: Multi-Modal Features

Multi-modal incorporating model (NHBG) improved 18.75% compared to single modality (H).

Best performing model (NHBG) improved 34.38% compared to single modality (H).

Best test AUPRC also surpasses SOTA score by 26.47%.

References

