

Protein-Protein Binding Affinity

Alen Aliev

MIPT

Protein-Protein Binding Affinity

We analyze protein complexes interactions.

Our goal is to provide a model for predicting binding affinity of complexes.

We suggest using deep-learning methods, i.e. Graph Neural Networks and CNN.

Protein Interaction Problem

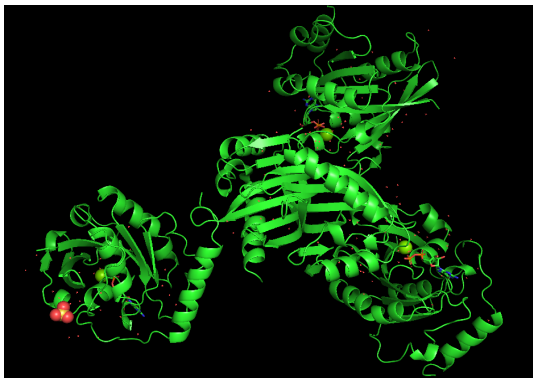


Figure 1: Protein Complex Example

$$\mathcal{A} = \{Ala, Arg, Asn, \dots, Val\}$$

$$a_i \in \mathbf{R}^3 \text{ - atom}$$

$$PR = \{a_i\} \in \mathbf{R}^{3*} \text{ - single protein}$$

$$PC = (PR_0, PR_1) \text{ - protein complex}$$

Literature

- ▶ K Yugandhar and M Michael Gromiha. Computational approaches for predicting binding partners, interface residues, and binding affinity of protein–protein complexes. Springer, 2017
- ▶ Anna Vangone and Alexandre MJJ Bonvin. Contacts-based prediction of binding affinity in protein–protein complexes, 2015.
- ▶ Iain H Moal, Rudi Agius, and Paul A Bates. Protein–protein binding affinity prediction on a diverse set of structures. Bioinformatics, 2011.

Protein Dataset Description

Dataset consists of protein chains, taken from pdbind.org

Our hypothesis is that the interacting atoms are the ones selected with KNN method.

Two chains are considered interacting if they maximize number of atoms-neighbours.

Affinity Prediction Analysis

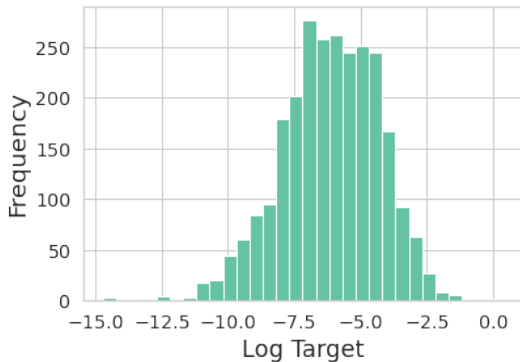


Figure 2: Log of Target Distribution

We predict logarithm of target.

We benchmark models against existing state-of-the-art solutions using MSE metric and batch accuracy.

Problem Solution: 3D grid

We define C_{PR_0, PR_1} as center of mass of all interacting atoms in chains.

Each complex is now a 3D grid centered around C_{PR_0, PR_1} .

This grid is used with 3D CNN.

Problem Solution: Graph

Let $V = PR_0 \cup PR_1$

We say that $(a_i, a_j) \in E \leftrightarrow a_j$ is in K Nearest Neighbours of a_i .

This graph $G = (V, E)$ representation is used with Graph Neural Network.

Computational Experiment

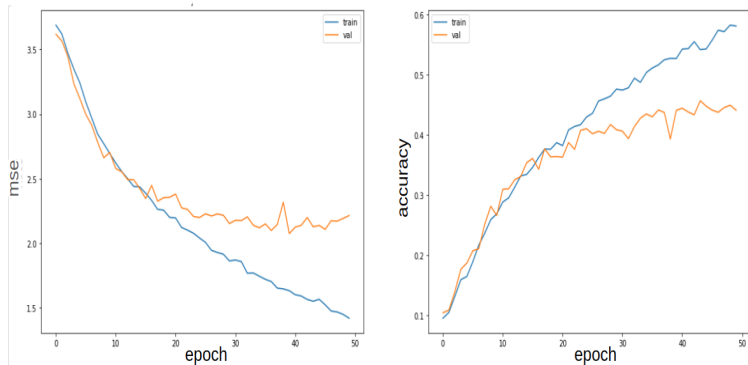


Figure 3: CNN performance

Around 50 epochs is sufficient to fit the model according to stoppage criterion.

Error Analysis

Test metrics		
Model	MSE	Accuracy
3D-CNN, VGG based	0.8715	0.72
Graph NN	0.9	0.71
CatBoost	1.2	0.53
Basic Model	4.8	0.172

Conclusion

- ▶ We provide two different interpretations of protein complex
- ▶ Implemented the model for predicting affinity of PPB.

Next: hyperparameters fitting.