Geometric Deep Learning for Protein-Protein Binding Affinity Prediction

- Alen Aliev, aliev.ae@phystech.edu
- Ilya Igashov, dummy@phystech.edu
- Arne Schneuing, dummy@phystech.edu

2022

Содержание

Abstract

Proteins are involved in several biological reactions by means of interactions with other proteins or with other molecules such as nucleic acids, carbohydrates, and ligands. Among these interaction types, protein–protein interactions (PPIs) are considered to be one of the key factors as they are involved in most of the cellular processes.
In this work we aim to compile a novel benchmark of PPIs with known binding affinity values from refined data and benchmark the resulting deep learning geometry method against existing state-of-the-art approaches.

Introduction

The binding of two proteins can be viewed as a reversible and rapid process in an equilibrium that is governed by the law of mass action. Binding affinity is the strength of the interaction between two (or more than two) molecules that bind reversibly (interact). It is translated into physico-chemical terms in the dissociation constant Kd, the latter being the concentration of free protein at which half of all binding sites of the second protein type are occupied [2].
Predicting the affinity of protein–protein complexes has been a topic of active research for more than two decades. The availability of experimental data on binding affinity prompted researchers to explore the principles and develop methods for prediction [1]. However, the amount of experimentally observed three-dimensional protein-protein complexes with known binding constants still remains extremely limited [3, 4], which complicates the application of modern deep learning methods in this task. The most recent computational research on PPI binding affinity prediction is mainly built around the idea of utilizing standard statistical [5, 6] and machine learning methods [4, 7, 8, 9] trained on various handcrafted descriptors such as QSAR features [8], inter-residue contacts and buried surface area [5], surface tension area and hydrophobicity [4], and sequence-based descriptors [6, 9].
Recent advances in adjacent computational problems such as protein structure prediction [10] or protein-protein interaction prediction [11] demonstrated how powerful deep learning methods can be as long as enough training data is provided and correct inductive biases are set up.
In this work, we will apply geometric deep learning methods for predicting protein-protein binding affinity. We believe that geometry is a clue for understanding protein-protein interactions, and aim to notably move forward the state of the art in binding affinity prediction with the aid of graph neural networks. To the best of our knowledge, geometric deep learning methods have never been applied to the protein-protein binding affinity prediction problem so far.

Objectives

Three main objectives of this work can be formulated as follows:

- Refine PDBbind [12] data and a standard binding affinity dataset [3], and compile a novel benchmark of PPIs with known binding affinity values
- Employ graph-learning toolset to predict binding affinities of PPIs from the new dataset
- Benchmark the resulting method against existing state-of-the-art approaches

Problem Statement

Let $\mathcal{D} = \{(X, y)\}$ be the given dataset, where X is referred to as design matrix and y as target vector. Let X be a vector of PDB codes, each representing an atomic coordinate file of a molecular structure. Thus, $X \in \mathsf{X}^n, y \in \mathsf{R}^{\mathsf{n}}_{+}$. The problem is now defined as optimizing model w^* with respect to following functional: $w^* = \underset{w}{\operatorname{argmin}} = S(w(\mathcal{D}_T), y_T)$, where S is the error function. The set $\mathcal{D}_T \subset \mathcal{D}$ is a training design subset with corresponding target y_T. For the validating purposes we use the quadratic error function: $S(a, b) = ||a - b||_2^2$.