PREDICTION OF PROTEIN-LIGAND BINDING AFFINITY

BOLTZMANN LABS PRIVATE LIMITED

BY – YASHI KESARWANI

OBJECTIVE

- The objective of this project is Protein−Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks.

- The reference review paper for this project is K-deep which proposes an end-to-end framework based on the 3D Convolution Neural Networks for predicting protein-ligand binding affinity.

BINDING AFFINITY

- Binding affinity is the strength of the binding interaction between a single biomolecule (e.g. protein or DNA) to its ligand/binding partner (e.g. drug or inhibitor).
- Binding affinity is typically measured and reported by the equilibrium dissociation constant (Kd), inhibition constant (Ki) and half-concentration constant (IC). These properties are used to evaluate and rank order strengths of bimolecular interactions.
- The smaller the Kd value, the greater the binding affinity of the ligand for its target. The larger the Kd value, the more weakly the target molecule and ligand are attracted to and bind to one another.

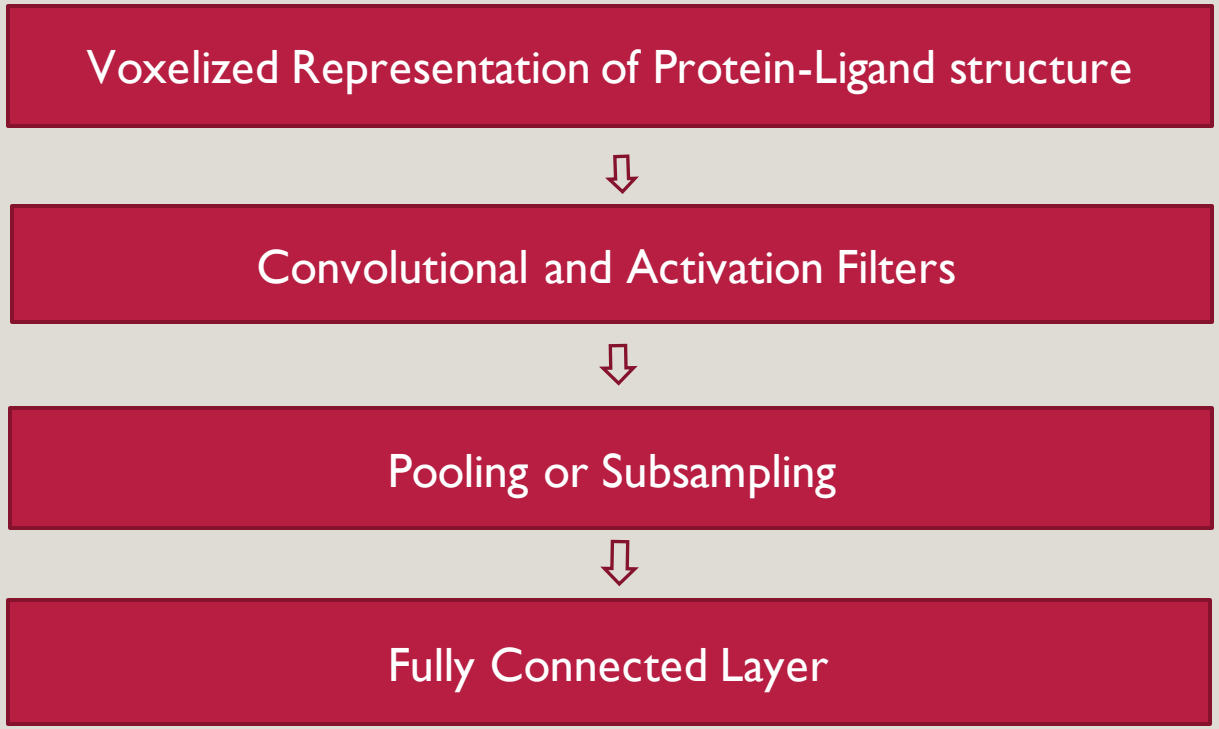
DATASET DESCRIPTION

- PDBbind (v.2016) database, containing 13,308 protein−ligand complexes and their corresponding experimentally determined binding affinities collected from literature and the Protein Data Bank (PDB), in terms of a dissociation (Kd), inhibition (Ki) or half-concentration (IC50) constant.
- A smaller refined subset (4057 Protein-ligand complexes) is extracted from it following quality protocols addressing structural resolution and experimental precision of the binding measurement.

WORKFLOW OF ALGORITHM

- Both protein and ligand are featurized via a voxelized 24 A˚ representation of the binding site considering pharmacophoric-like properties.
- These descriptors are used by a three-dimensional convolutional neural network model, which in turn learns the binding affinity of the complex given enough training examples. Once trained, the network is used to predict unseen examples.

WORKFLOW OF ALGORITHM



VOXELIZATION OF PROTEINS AND LIGANDS

- Atom typing for the protein requires the protein to be protonated and to include the atom bond information.
- The protonation will a) move atoms to optimize hydrogen networks b) add missing sidechains and c) the bond guessing can go wrong if atoms are very close to each other which can happen when adding sidechains.
- Now we can calculate the voxel information for the protein. By using getVoxelDescriptors, we will calculate the bounding box of the molecule and grid it into voxels. As we don't use point properties but smooth them out over space, we will add 1 A buffer space around the protein for the voxelization grid so that the properties don't cut off at the edge of the grid.

VOXELIZATION OF PROTEINS AND LIGANDS

- prot.view and viewVoxelFeatures will visualize each voxel channel as a separate VMD molecule so that you can inspect each individually.
- To perform 3D convolutions we need to reshape the data to (nsamples, nchannels, d, h, w) with the last three dimensions corresponding to the 3D elements. Our data is in (d*h*w, nchannels) format so, we first transpose and then reshape it.

MODEL USED - SQUEEZENET

Architectural Design Strategies

Strategy 1. Replace 3×3 filters with 1×1 filters

Strategy 2. Decrease the number of input channels to 3×3 filters

Strategy 3. Down-sample late in the network so that convolution layers have large activation maps

Strategies 1 and 2 are about judiciously decreasing the quantity of parameters in a CNN while attempting to preserve accuracy. Strategy 3 is about maximizing accuracy on a limited budget of parameters.


THANK YOU!