CS 273B Literature Review: AtomNet

Joe Paggi
jpaggi@stanford.edu

Andrew Lamb
andrewl3@stanford.edu

Kevin Tian
kjtian@stanford.edu

Irving Hsu
irvhsu@stanford.edu

Pierre-Louis Cedoz
plcedoz@stanford.edu

Prasad Kawthekar
pkawthek@stanford.edu

1 Introduction

The problem of predicting molecular binding affinity is an interesting one, particularly in the choice of feature selection in the model. AtomNet is a novel approach to the problem of this bioactivity prediction, which uses structural information in addition to ligand-based features in a deep neural net architecture. We think the method shows promise against many benchmarks in the prediction task, but also raise some questions regarding their descripion of features, and propose ideas for extension in incorporating other relevant features.

2 Key Contributions

AtomNet is novel in that it is the first deep convolutional neural network (DCNN) for molecular binding affinity prediction, and the first deep learning system that incorporates *structural* information about the target to make predictions. Previous methods for predicting the bioactivity of molecules primarily include ligand-based virtual screening. Random forests and SVMs have also been used for QSAR. Ligand-based techniques are limited in several respects, primarily in 1) the prediction of novel targets and 2) feature discovery. For the former, ligand-based methods are restricted to targets in which substantial amounts of prior data are already available. For the latter, existing deep neural networks for ligand-based models take as input molecular fingerprints such as ECFP. As such, they are unable to discover *arbitrary* features, and are instead restricted to compositions of pre-specified molecular structures defined during the fingerprinting process. Lastly, ligand-based models are blind to the target, and as a result are unable to elucidate which potential interactions are left unfulfilled by a molecule.

AtomNet addresses these weaknesses by combining information about the ligand with structural information about the target. By requiring the locations of each atom in the binding site of the target, AtomNet is able to uncover arbitrary molecular features that describe both favorable and unfavorable interactions between ligands and targets. These features can also be applied to targets for which no binders are known by the model.

Datasets

AtomNet is trained and validated on three datasets: the DUDE benchmark, a dataset the researchers constructed from ChEMBL-20 that is similar to DUDE, and dataset of molecules that are experimentally verified to be inactive.

The DUDE benchmark is a standard for predicting bioactivity. The positive and negative examples are constructed from a diverse set of active molecules and property matched decoys (PMD), respectively. A PMD is a molecule that shares characteristics of the active molecules (e.g. molecular weight), but is structurally different - the key assumption here is that the molecule is inactive because it is chemically different from the active molecule. The researchers use both the DUDE benchmark and a similar dataset constructed from ChEMBL-20.

The AtomNet researchers also use a dataset of experimentally verified inactive molecules (i.e. not PMDs). This allows negative cases that are structurally similar to active molecules, a challenging case for activity prediction.

3 Results and Validation

Comparison to Smina

The performance of the DCNN was evaluated in comparison to the performance of Smina, the baseline algorithm for the SB evaluation. The metrics were the area under the receiver operating characteristic (AUC) and logAUC over the three benchmarks. LogAUC is a measurement similar to AUC that emphasizes early enrichment performance by putting more weight at the beginning of the curve.

On each of our four evaluation data sets, AtomNet achieves an order-of-magnitude improvement over Smina at a level of accuracy useful for drug discovery. For instance on the ChEMBL-20 inactives benchmark, AtomNet had a mean AUC of 0.745 whereas Smina mean AUC was 0.607 (Figure 1). The performance of AtomNet is even more impressive if we consider the AUC and logAUC results with respect to different performance thresholds (Figure 2). For example on the full DUDE set (DUDE-102), AtomNet achieves or exceeds 0.9 AUC on 59 targets whereas Smina only achieves 0.9 AUC for a single target. AtomNet achieves 0.8 or better AUC for 88 targets while Smina achieves it for 17 targets.

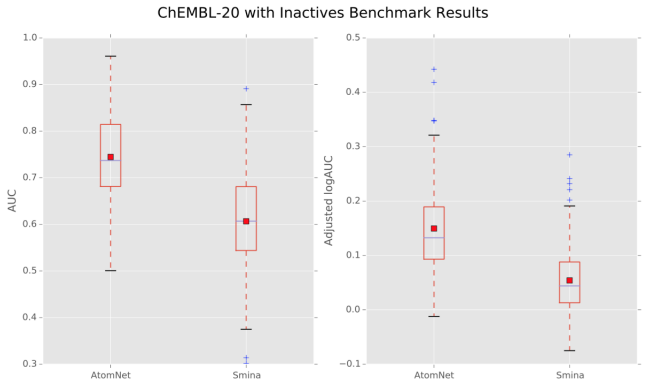

Figure 1: Distribution of AUC and logAUC values of 149 ChEMBL-20-inactives targets for Atom-Net and Smina

AUC		> 0.5	> 0.6	> 0.7	> 0.8	> 0.9
ChEMBL-20 PMD	AtomNet	49	44	36	24	10
	Smina	38	10	4	1	0
DUDE-30	AtomNet	30	29	27	22	14
	Smina	29	25	14	5	1
DUDE-102	AtomNet	102	101	99	88	59
	Smina	96	84	53	17	1
ChEMBL-20 inactives	AtomNet	149	136	105	45	10
	Smina	129	81	31	4	0

Figure 2: The number of targets on which AtomNet and Smina exceed given AUC thresholds

To sum up, Atomnet outperforms Smina on every benchmark with respect to overall and early enrichment performances (AUC and log(AUC)). The most challenging benchmark was ChEMBL-20 inactives because it includes challenging classification cases (inactives) of structurally similar molecules with different labels but Atomnet reaches good performance even on this benchmark (Figure 1). They also compared there results to other commercial docking algorithms (Surflex-Dock, Dock3.7 and Dock6.7) and they reported improvements of one order of magnitude on the DUDE benchmark.

Model Interpretation

As the filters in the trained DCNN are 3-d, the authors take an indirect approach to visualize them by examining the location where they fire most on input data. This allows them to identify chemical functions evaluated by the filter. For example, by using this technique they are able to observe complex chemical features such as sulfonyl/sulfonamide detection (Figure 3).

4 Vagueness of Input Features

The primary weakness of this paper is that it is vague, perhaps intentionally to protect their business interests. In this section, we will review their procedure for generating input features for their model, highlighting where they are unclear and inferring what they are perhaps doing.

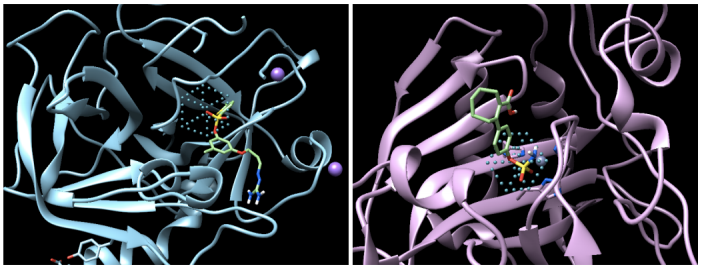

Figure 3: Sulfonyl/sulfonamid detector filter

For each target, they define a ligand binding site by identifying empty space surrounding the site of a bound ligand in a co-crystallized complex. They then place a $20 \times 20 \times 20$ 1Å grid over the ligand binding site they identify. The input features of their model are then functions of the atoms occupying each grid square. However, for the vast majority of the ligands in their datasets, there are not known ligand binding poses (how the ligand is oriented in the binding site) and it is not clear how they chose poses for these ligands. Furthermore, it is not clear what information they are in fact recording for each cube. Both of these tasks are in fact active research areas.

Our guess is that they are sampling ligand binding poses by using a physics-based docking code such as Glide [1]. Physics-based docking codes sample many ligand orientations and report back those with low energy based on a physics-inspired scoring function. They claim to include several binding poses, so we infer that they take the top n ranked poses outputted by the docking software. This is problematic because they are presumably unable to obtain good predictions for ligands for which they cannot obtain sound binding poses. This is precisely the domain in which docking codes are likely performing poorly for virtual screening purposes.

They state that their "basic structural features can vary from a simple enumeration of atom types to more complex protein-ligand descriptors such as SPLIF, SIFt, APIF". There is no obvious way to implement any of these options. For atom types, they could mean anything from the element alone to more complex descriptions such as an extended connectivity fingerprint. The latter descriptors represent interactions between the receptor and ligand, such as hydrogen bonding or lipophilic interactions. It is unclear where precisely features of this type should be placed as they are defined at the level of functional groups, not atoms.

5 Extensions

While the most important "extension" we believe the authors should make would be to fill in some of the missing details enumerated above, here we describe an additional possible direction. They appear to currently be ignoring information from known crystal structures, except for the task of identifying binding sites. They could train their model using databases of known crystal structures, such as scPDB, in addition to their current datasets. This would allow them to train on poses known to be correct, perhaps dissipating the affects of training on likely inaccurate poses. Furthermore, in the case of targets for which there are known crystal structures, they could use this information to inform the poses of other ligands binding to the same target [2].

6 Conclusion

Ultimately, AtomNet proves to be successful against existing prediction benchmarks, and other existing bioactivity prediction methods. Its key contribution is proving the effectiveness of incorporating structure-based information in a deep learning model for prediction. We think their specific means for choosing features are interesting, and think it would lend to a better understanding of the overall method. Furthermore, it seems reasonable that exploring other molecular affinity influencing factors, such as known crystal structures, would help the prediction model as well.

References

[1] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, and D.T. Mainz. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, 49(21):6177–6196, 2006.

[2] D. Fourches, R. Politi, and A. Tropsha. Target-specific native/decoy pose classifier improves the accuracy of ligand ranking in the csar 2013 benchmark. *Journal of Chemical Information and Modeling*, 55(1):63–71, 2014.