DoCEIS 2023
TECHNOLOGICAL INNOVATION FOR CONNECTED CYBER PHYSICAL SPACES
14th Advanced Doctoral Conference On Computing, Electrical And Industrial Systems
5 – 7 Jul 2023 Caparica, Portugal

Drug-Target Affinity Prediction Using Graph Attention Networks

Sonya Falahati[1], Fatemeh Zamani[1], Amin Khodamoradi[2]
(1) Babol Noshirvani University of Technology (NIT), Faculty of Electrical and Computer Engineering
(2) Universidade NOVA de Lisboa, NOVA School of Science and Technology (FCT NOVA) / Uninova, Center of Technology and Systems

motivation

Drug discovery (DD) is the process through which potential new medicines are identified. Drug-target affinity(DTA) prediction is a crucial step in the DD process, as it can help to identify and optimize potential drug candidates to target specific protein(s). DTA measures how strongly a drug binds to a protein. However, experimental methods are costly and slow, and existing computational methods have some limitations. For instance, some methods rely on 3D structural information of targets, which is not widely available, or use simple string representations of drugs, which may not capture their complex molecular properties.
Moreover, some methods use graph neural networks to represent the drugs as a graph but ignore the importance of different atoms and bonds. We propose graph attention networks, which can assign different weights to different nodes and edges in the graph based on their importance for the task.

Research Question

How to predict drug binding affinities to target proteins?
1. How to represent the protein sequences as an input model?
2. How to generate an informative graph of drugs?
a. How can exploit Graph ATtention Networks (GAT) to model the drugs as a molecular graph?

Hypothesis

We hypothesize that GAT can capture more information from the graph structure and node features.
- Using GAT can improve the accuracy and efficiency of Drug-Target Affinity.

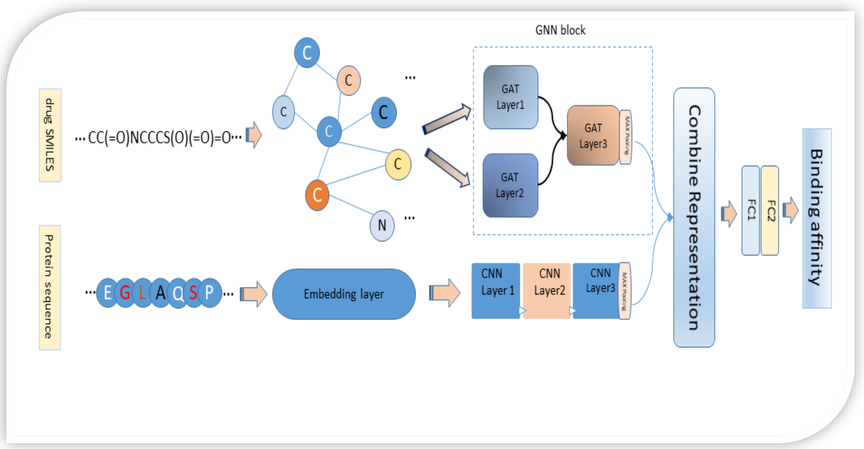
APPROACH

We propose an improved version of GraphDTA [1], a state-of-the-art model that predicts drug-target binding affinity using graph neural networks (GNNs) for drugs and convolutional neural networks (CNNs) for proteins. We use a stack of GAT layers and concatenate features from different layers to increase the expressiveness and complexity of our model. GATs calculate weights for different nodes and edges in the drug graphs based on their importance for the task.

Our model has two components:

- For drug graphs, we use three GAT layers to perform graph attention on the input features. We concatenate the outputs of the first two GAT layers before passing them to a linear layer. We then use ReLU activation and dropout after the linear layer. The linear layer takes the output of the third GAT layer as input and produces a scalar representation vector, which is a part of the input for the following Fully connected Layer (FC)s.
- For protein sequences, use an embedding layer followed by three 1D convolutional layers to learn different levels of features from the input. then apply the max pooling layer to get a representation vector. This approach is similar to the existing baseline models like DeepDTA [2] and GraphDTA [1].

We then combine the representation vectors of the drug and protein and treat them as features and pass them through two FCs to get an affinity output for each drug-target pair.



CPS Relation

CPS can improve the DD process and the healthcare outcomes by using DTA. DTA prediction can have applications in CPS scenarios, such as smart healthcare, personalized medicine, or drug repositioning.

- DTA prediction can use the latest and most advanced CPS technologies to improve its performance and efficiency. For example, smart healthcare can monitor, diagnose, treat, or prevent diseases with IoT devices. IoT devices can connect physical objects to the internet and enable data collection and transmission.
- Personalized medicine can use DTA prediction to avoid bad drug effects or combinations by selecting drugs that have high affinity with target proteins and low affinity with other proteins.
- CPS technologies can support the data and computation needs of DTA prediction by providing high-quality and diverse data sets, as well as fast and powerful computation models.

Result and Conclusion

We evaluated our model on the Davis dataset[3], which contains drug-target binding affinity data measured by Kd. We compared our model with five state-of-the-art methods: KronRLS[4], SimBoost [5], DeepDTA[2], WideDTA[6] and GraphDTA. We used two evaluation metrics: mean squared error (MSE) and concordance index (CI). Table 1 shows the expected performance of our model and the compared methods on the Davis dataset:

method	MSE	CI
KronRLS	0.379	0.871
SimBoost	0.282	0.871
DeepDTA	0.261	0.878
WideDTA	0.262	0.886
GraphDTA-GAT	0.232	0.892
GraphDTA-GCN	0.254	0.880
Our model	0.249	0.881

Our model achieves the second lowest MSE and the second highest CI among all the methods, GraphDTA-GAT achieves the lowest MSE and the highest CI among all the methods, showing that GATs can improve the representation learning of drugs. Our model also uses GATs for drugs, but differs from GraphDTA-GAT in the number of GAT layers and the concatenation of features from different layers.
Our model advantages are avoiding the limitation of 3D structural information of targets, capturing more information from the graph structure and node features of drugs, and enhancing the representation learning of our model. Our experimental results show our model achieves competitive performance on the Davis dataset compared to the state-of-the-art methods.

References

[1] Nguyen, Thin, et al. "GraphDTA: predicting drug–target binding affinity with graph neural
[2] Öztürk, Hakime, Arzucan Özgür, and Elif Ozkirimli. "DeepDTA: deep drug–target binding affinity
[3] Davis, Mindy I., et al. "Comprehensive analysis of kinase inhibitor selectivity." *Nature biotechnology* 29.11 (2011): 1046-1051.
[4]Nascimento, André CA, Ricardo BC Prudêncio, and Ivan G. Costa. "A multiple kernel learning algorithm for drug-target interaction prediction." *BMC bioinformatics* 17 (2016): 1-16.
[5]He, Tong, et al. "SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines." Journal of cheminformatics 9.1 (2017): 1-14
[6] Öztürk, Hakime, Elif Ozkirimli, and Arzucan Özgür. "WideDTA: prediction of drug-target binding affinity." arXiv preprint arXiv:1902.04166 (2019).