



Drug-Target Affinity Prediction Based on Multi-channel Graph Convolution

Hang Zhang¹, Jing Hu^{1,2,3}(✉), and Xiaolong Zhang^{1,2,3}

- ¹ School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China
{hujing, xiaolong.zhang}@wust.edu.cn
- ² Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan, China
- ³ Institute of Big Data Science and Engineering, Wuhan University of Science and Technology, Wuhan, Hubei, China

Abstract. Computer-aided drug design with high performance is a promising field, and the pre-diction of drug target affinity is an important part in computer-aided drug design. As a kind of deep learning model algorithm, graph neural network series model algorithm has been gradually applied in drug and protein research due to its excellent performance in structural feature learning. In the new field of drug target affinity prediction, graph neural network also has great potential. In this paper, a novel approach for drug target affinity prediction based on multi-channel graph convolution network is proposed. The method encodes drug and protein sequences into corresponding node adjacency matrix. The adjacency matrix together with the physical and chemical characteristics of drug and protein sequences are used as the inputs of the model to construct a multi-channel graph convolution network that aggregates the information of nodes at different distances. The drug feature and target feature vectors are concatenated, and then through the full connection layer, the concatenated vector is converted to the predicted value. The experiment results on Davis dataset and KIBA dataset show that the proposed method outperforms most relevant methods. While the experimental results is Slightly worse than GraphDTA, it shows that the proposed method can improve the prediction of drug-target affinity to a certain extent and aggregate more information from other nodes of higher order proximity in the graphs.

Keywords: Prediction of drug target affinity · Graph convolutional neural network · Nodes of high order proximity · Multichannel graph convolutional networks

1 Introduction

It takes a lot of money and development time to develop new drugs. According to statistics, FDA-approved drugs cost about \$2.6 billion and take 17 years to develop. Finding new uses for approved drugs can avoid the expensive and time-consuming drug development process [1–3]. To effectively change the use of approved drugs, it

is necessary for researchers to understand which proteins are targets for which drugs. High-throughput screening tests can detect drug target affinity, but these tests are costly and time-consuming [4, 5]. Moreover, the presence of a large number of drug-like compounds and potential protein targets makes thorough screening difficult [6–8]. However, the computational model based on the existing drug target experiments can effectively estimate the interaction intensity of new drug target pair, so this kind of method is gradually popular.

At present, many methods have been used to predict drug target interaction, which greatly promotes the development of drug target interaction research. Pahikkala et al. used the Kronecker Regularized Least Squares (KronRLS) algorithm to calculate the paired nuclear K from drug-drug and protein-protein Kronecker products [9]. He et al. Proposed SimBoost method to predict the affinity between the unknown drug and the target which used affinity similarities between drugs and targets to construct new features. All the methods mentioned above are traditional machine learning methods [10]. With the improvement of the accuracy of neural network and the continuous improvement of the high precision requirements of drug design, deep learning methods are also applied to the scoring and prediction of protein ligand interactions. Hakime Öztürk et al. showed DeepDTA methods to predict the affinity of the drug to the protein target [11]. They used SMILES [12], simplified molecular input line entry specification of drug molecules, and the protein sequence expression as the input of the model, respectively constructing two convolutional neural networks to extract the expressions of drugs and proteins, and finally combined the two expressions to predict the affinity between drugs and protein targets. Hakime Öztürk [13] proposed WideDTA method which was further improved based on DeepDTA. The model takes ligand SMILES (LS), ligand max common substructure [14] (LMCS), protein sequence [15] (PS), protein motifs and domains (PMD) as input, after convolution neural network training, then the representation vectors are concatenated and through full connection layers we can get the predicted values. Although the methods mentioned above are significantly better than traditional machine learning methods in predicting results, the representation of drug molecules and protein sequences as strings is not a natural way to express. Recently, graph neural network has been widely used in different fields. It has no restriction on the size of input graph and can express the structure of drug molecules and proteins more truly than the way of using string as input expression, so it can extract more deep molecular information in a more flexible form. The PADAME model designed by Q. Feng [16] utilizes molecular graph convolution in drug target interaction prediction, demonstrating the potential of graph convolutional neural networks in drug discovery. Like PADAME T. Nguyen et al. proposed GraphDTA method which took atoms as graph nodes and chemical bonds as graph edges to construct a drug molecule graph [17]. The drug molecule graphs, and protein sequences are inputs of the network, then through training and concatenation, the predicted affinity values of drugs and targets can be obtained. Compared with other methods, GraphDTA has an obvious improvement in the prediction performance of drug-target interaction. However, the model has only three-layer graph convolution, which is difficult to aggregate information of similar but distant node. Although increasing the number of graph convolution layers can realize the information aggregation of similar but distant nodes, the expression of nodes will gradually be projected to a stable state. Therefore, the number of graph

convolution layers is limited, and it is difficult to aggregate high-order similar nodes. Therefore, it is necessary to develop a convolutional architecture with high computational efficiency to utilize the information of high-order neighboring nodes through appropriate aggregators while maintaining the heterogeneity of nodes. Zhou et al. proposed a multi-channel graph convolutional network (MCGCN) model to achieve the aggregation of high-order information by enriching the number of input channels [18]. In this paper, the multi-channel graph convolutional neural network was applied to the prediction of drug target affinity to further optimize the experimental results. Compared with other methods, the proposed method solves the problem caused by too many convolution layers by aggregating the node information of different distances in each channel and learns more comprehensive graph data. Secondly, the proportion of aggregated node information of each channel is adjusted by parameters to make the model more rational.

2 Methods

In this experiment, we referred to a variety of methods for predicting drug target affinity based on deep learning and obtained the prediction results by extracting the expressions of drug molecules and proteins respectively and then splicing them together. Compared with other methods, the innovation of this experiment lies in the introduction of multi-channel graph convolution into the training of drug molecular graph data. Compared with traditional graph convolution, multi-channel graph convolution can better obtain the structural information with different distances in drug molecular graph.

2.1 Molecular Representation

In the experimental dataset, we obtained the model input expression using the Simplified Molecular Linear Input Specification (SMILES). Smiles enables molecular data to be read by computers for efficient applications such as fast retrieval and substructure search. The compound SMILES strings of the Davis dataset are extracted from the PubChem compound database according to their PubChem CIDs. The KIBA dataset needs to first convert the ChEMBL ID to PubChem CID, and then extract the SMILES string through the corresponding CID. Expressed by SMILES, molecular graphs can be constructed with atoms as nodes and chemical bonds as edges. In the experiment, the atomic number of the drug molecule, the set of atomic pairs at both ends of the chemical bond and the physical and chemical characteristics of each atom were taken as the input expression of the drug molecule. To ensure that the node features are fully considered in the graph convolution process, self-loop is added into the graph convolution structure to improve the performance of drug molecules. The graph construction for molecular features is shown in Fig. 1. The molecular features are illustrated in Table 1, which are the same as those in DGraphDTA.

Contact map is one of the outputs of structure prediction method, usually in matrix form. Assume that the length of the protein sequence is L , then the predicted contact map M is a matrix of L rows and L columns, where each element m_{ij} of M represents whether the corresponding residue pairs, namely residues I and J , are contacted. In general, two

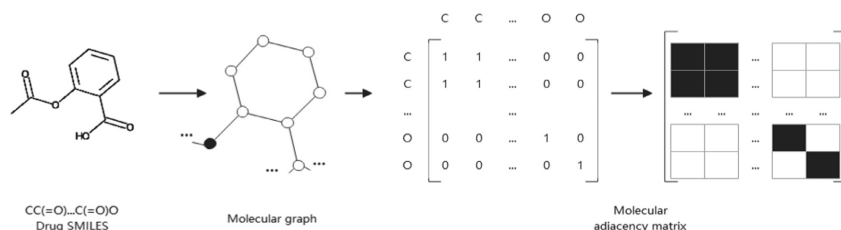


Fig. 1. Graph construction for molecular graph

Table 1. Node features (atom)

Feature	Dimension
One-hot encoding of the atom element	44
One-hot encoding of the degree of the atom in the molecule, which is the number of directly-bonded neighbors (atoms)	11
One-hot encoding of the total number of H bound to the atom	11
One-hot encoding of the number of implicit H bound to the atom	11
Whether the atom is aromatic	1

residues are in contact if the Euclidean distance between the C_{β} atoms (in glycine's case, the C_{α} atoms) is less than a specified threshold.

In this study, Pconsc4 open source method was used to predict contact maps efficiently and quickly. Pconsc4 uses the U-NET [19] architecture, which operates on 72 features calculated from each position in a multi-sequence alignment. Pconsc4 takes the probability of the attachment of residue pairs as the output, and then takes 0.5 as the threshold to obtain the contact map of size (L, L) , which also corresponds to the adjacency matrix of protein sequences.

PSSM [20] (position-specific scoring matrix) is a common protein expression pattern in proteomics. In PSSM, each residue position can be scored according to the sequence alignment results and used to represent the residue node features. In this experiment, PSSM and the physicochemical properties of each residue node were taken as the features of protein sequences. The specific features of these nodes are shown in Table 2 (Fig. 2).



Fig. 2. Graph construction for protein graph

Table 2. Node features (residue)

Feature	Dimension
One-hot encoding of the residue symbol	21
Position-specific scoring matrix (PSSM)	21
Whether the residue is aliphatic	1
Whether the residue is aromatic	1
Whether the residue is polar neutral	1
Whether the residue is acidic charged	1
Whether the residue is basic charged	1
Residue weight	1
The negative of the logarithm of the dissociation constant for the -COOH group	1
The negative of the logarithm of the dissociation constant for the -NH3 group	1
The negative of the logarithm of the dissociation constant for any other group in	1
The pH at the isoelectric point	1
Hydrophobicity of residue (pH = 2)	1

2.2 Multichannel Graph Convolution Structure

In recent years, the success of convolutional neural networks in computer vision, speech recognition and natural language processing has stimulated researchers to study the field of graph neural networks. Graph neural network solves two main problems when convolutional neural network is extended to graphs: (1) forming receptive fields in graphs where data points are not arranged according to Euclidean grids; (2) Pool the graph under sampling. After years of rapid development, Graph Neural Network has derived many powerful variants, such as Graph Convolution Network (GCN) Graph Attention Network (GAT) Graph Isomorphism Network (GIN), these models are very effective for graph feature extraction.

For GCN, each layer will perform the convolution operation through (1):

$$H^{l+1} = f(H^l, A) = \sigma\left(\widehat{D}^{-\frac{1}{2}} \widehat{A} \widehat{D}^{-\frac{1}{2}} H^l W^{l+1}\right) \quad (1)$$

In the equation, A is the adjacency matrix of the protein graph of shape (n, n) , n is the number of nodes in the graph, $\widehat{A} = A + I$, where I is the identity matrix, \widehat{D} is the diagonal node degree matrix calculated by A and its shape is the same as matrix A , W^{l+1} is the weight matrix of $l + 1$ layer, H^l is the output of the last layer of shape (n, F^l) , F^l is the number of output channels in layer l , $H^0 = X$, where X is the input eigenvector of the node.

In essence, graph convolutional networks treat the network structure as a computational graph and train the entire neural network model in an end-to-end manner. By adopting an appropriate message passing mechanism in each convolution layer of the

graph convolutional network, each node can aggregate attribute information from adjacent nodes in the network. However, as the depth of the graph convolutional network increases, the nodes will aggregate information from other nodes of higher order proximity. During this process, the node representation is projected to a steady state after several aggregation steps. Therefore, the number of existing graph convolutional network layers should not be too large. In practical applications, nodes with the same/similar structural roles may be far away from each other in the network, and graph convolutional networks with limited depth cannot aggregate the information of nodes with similar roles but far away from each other. Therefore, this paper does not increase the depth of graph neural network, but chooses rich information channel, that is, uses multi-channel graph convolutional network to support any order of information aggregation through the network.

Like graph convolution, multi-channel graph convolutional network uses (2) to implement message delivery:

$$H_k = \begin{cases} X & k = 0 \\ \sigma(\hat{A}H_{k-1}W_{k-1}) & k = [1, l] \end{cases} \quad (2)$$

Specifically, the number of layers of multi-channel graph convolutional neural network is l . In the current k layer, $H_0 = X$ represents the eigenmatrix X as the input of the model. In addition, $H_k \in R^{N \times d_k}$ is the output node expression of layer K and the input node expression of layer $K + 1$, so the node information will be aggregated through the message passing model. σ represents the message propagation function that aggregates information through the network, \hat{A} represents the renormalized adjacency matrix, and W_{k-1} is the weight matrix of the k^{th} layer.

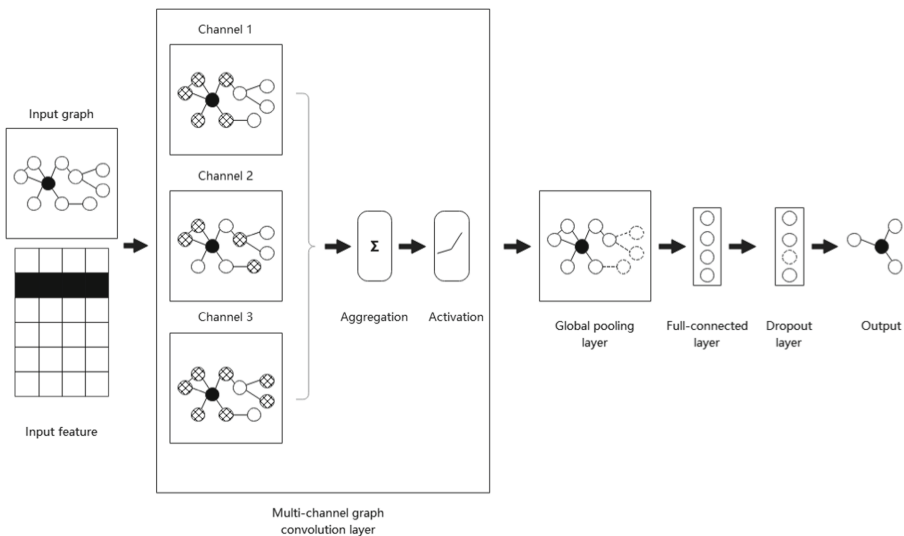


Fig. 3. Multi-channel convolution architecture

The multi-channel convolution architecture is shown in Fig. 3. The model takes the feature matrix $X \in R^{N \times d}$ as input, each row of which represents the features of a node, and the node information can be aggregated in different channels respectively. Specifically, the propagation network in channel k corresponds to a specific matrix \hat{A}^k , which is the k power of the normalized adjacency matrix \hat{A} . The forward propagation expression of the model is shown in (3):

$$H = AGG(\hat{A}XW_1, \hat{A}^2XW_2, \hat{A}^3XW_3, \dots) \quad (3)$$

In the equation, $\hat{A}^i XW_i$ represents a high-order GCN channel that gets information from the i^{th} order neighbor, and AGG is used to aggregate node information from all channels. In this paper, we considered using the summation operator as aggregation function from two aspects: first of all, the general schemes of GCN polymerization can be viewed as functions on a set of domain nodes, and in the different aggregation functions only the summation operator can get the complete set, so the summation operator more than other operators will be able to distinguish between different network structure [21]; Secondly, the implementation of the summation operator can obtain the weighted summation over different convolution channels, which can amplify the relatively important information. Therefore, the forward propagation model can be calculated as (4):

$$H = \sum_{i=1}^k (\hat{A}^i XW_i) \quad (4)$$

In the equation, k represents the total number of channels of the model, and W_i represents the learnable weight. W_i can be also regarded as a pre-processing operation on node characteristics in each channel. To reduce model parameters and avoid overfitting, the experiment uses shared weight W_s for different channels of the model. At the same time, the nonlinear function σ was used in the experiment to improve the expression ability of the model, and the parameter α was used for appropriate adjustment between different channels. Finally, the equation of the forward propagation process of the model is rewritten as following:

$$H = \sigma \left[\sum_{i=1}^k (\alpha \hat{A})^i XW_s \right] \quad (5)$$

2.3 Model Structure

Figure 4 shows the complete structure of the model. Through experiments, we found that when the drug data was 3-channel graph convolution and the protein data was 3-layer graph convolution, the experimental results were the best. The drug molecule graph data and protein graph data are input into the convolutional layer of the model. After that, the characterization vectors of drug molecules and protein sequences are respectively obtained through a pooling layer and two layers of full-connection layers. Finally, vector concatenated is conducted to obtain the predicted value of the model through the two full-connection layers.

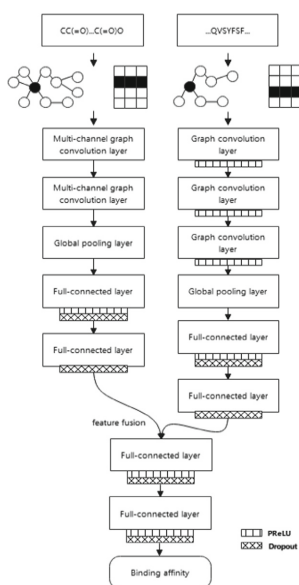


Fig. 4. The complete model structure

3 Results and Discussion

3.1 Dataset

To compare with other drug target affinity prediction models such as GraphDTA, WideDTA and DeepDTA, Davis [22] and KIBA [22] were selected for training and testing. The Davis dataset contains selected entities from the kinase protein family and related inhibitors, in addition to their respective dissociation constants. The Davis dataset contains 442 kinase proteins and 68 related inhibitors, as well as the dissociation constants for 30056 interactions. The KIBA dataset differs from the Davis dataset in that it contains bioactivity of kinase inhibitors from different sources, including K_i , K_d , and IC_{50} , which are processed as scores for the model to train and predict in the KIBA dataset. The KIBA data set initially contained 467 targets and 52,498 drugs, which was filtered by He et al. to contain only drugs and at least 10 interacting targets, resulting in 299 unique proteins and 2111 unique drugs. Table 3 shows two datasets of protein and drug molecules and their interactions.

Table 3. Dataset

Dataset	Proteins	Compounds	Binding entities
Davis	442	68	30056
KIBA	229	2111	118254

For the Davis dataset, the dissociation constant was converted to the exponential space to obtain pK_d as the affinity prediction, with the specific expression shown in (6):

$$pK_d = -\log\left(\frac{K_d}{10^9}\right) \quad (6)$$

He et al. took the negative value of each KIBA score and then selected the minimum value among the negative values and added the absolute value of the minimum value to all the negative values to construct the final form of KIBA score.

3.2 Metrics

Concordance index [23] (CI) and mean square error [24] (MSE) are both applied in the experiment which are also used in other state of the art methods.

Concordance index (CI) is obtained through (7), which is mainly used to calculate the difference between the predicted value and the actual value. The greater the value, the more consistent the predicted value is with the actual value.

$$CI = \frac{1}{Z} \sum_{d_x > d_y} h(b_x - b_y) \quad (7)$$

In the equation, b_x is the predictor of the larger affinity d_x , b_y is the predictor of the smaller affinity d_y , Z is a normalized constant, $h(x)$ is the step function, and the equation is shown in (8).

$$h(x) = \begin{cases} 1 & x > 0 \\ 0.5 & x = 0 \\ 0 & x < 0 \end{cases} \quad (8)$$

The mean square error (MSE) is also a common measure of the difference between the predicted value and the actual value. The smaller the value, the closer the predicted value is to the true value. For n samples, the mean square error is the average of the sum of squares of the difference between the predicted values of P_i ($i = 1, 2, \dots, n$) and the true values of y_i , as shown in (9).

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - y_i)^2 \quad (9)$$

In WideDTA, another new evaluation index, namely Pearson correlation coefficient [25], was introduced into the standard set, which was also used to compare the performance of experimental results. The higher the value, the better the performance of the model. Its expression is shown in (10).

$$Pearson = \frac{cov(p, y)}{\sigma(p)\sigma(y)} \quad (10)$$

In the equation, cov is the covariance between the predicted value and p and the actual value y , where σ represents the standard deviation.

3.3 Performance of Various Channels

Multi-channel graph convolution is an important step in drug feature extraction. Selecting the appropriate number of graph convolution channels can effectively improve drug target interaction prediction results. In this experiment, 1, 2 and 3 channels were selected for testing, and the evaluation indexes obtained in different situations were put together for comparison. The Davis dataset was selected as the model data in the experiment, and the results are shown in Table 4.

Table 4. Performance of various channels

Number of channels	CI	MSE	Pearson
1	0.760	0.577	0.540
2	0.776	0.575	0.554
3	0.763	0.620	0.526

As can be seen from the results in the table, when the number of channels is 2, the result obtained by multi-channel graph convolution is the best, and the performance of the three evaluation indexes is better than that of the number of channels is 1 and 3. If the number of channels is too small, it may be difficult to aggregate distant nodes, and if the number of channels is too large, the model calculation may be too large and the model effect may be reduced due to over-fitting.

3.4 Performance of Various Activation Functions

The activation function adds nonlinear factors to the model and solves problems that can't be solved by the linear model. Selecting the appropriate activation function can significantly improve the expression ability of the model. In this experiment, ReLU, PReLU and LeakyReLU activation functions were selected for testing, and Davis data set was used as the model data in the experimen. The results are shown in Table 5.

As can be seen from the table, the PReLU activation function is the best, followed by the LeakyReLU function and the ReLU function. LeakyReLU and PReLU functions have improved on the negative area compared to ReLU function, solving the Dead ReLU problem, so the results of the former two are close to and better than ReLU function.

Table 5. Performance of various activation functions

Activation function	CI	MSE	Pearson
ReLU	0.774	0.575	0.547
PReLU	0.778	0.574	0.549
LeakyReLU	0.769	0.587	0.531

3.5 Performance of Various Pooling Methods

Pooling function plays the role of down sampling, which can reduce the number of data and feature parameters and ensure the consistency of experimental data length. In this experiment, the max pooling function, the mean pooling function and the sum pooling function are selected to test, and the experimental results are shown in Table 6.

Table 6. Performance of various pooling functions

Pooling function	CI	MSE	Pearson
Global max pooling	0.770	0.576	0.538
Global mean pooling	0.757	0.592	0.514
Global sum pooling	0.765	0.587	0.520

As can be seen from the table, the max pooling function results are the best among the three, slightly better than the mean pooling function and significantly better than the sum pooling function. It may be that there are great differences between the input of drug and protein target data. The global max pool can better preserve main features.

3.6 Performance of Various Methods

To verify the performance of multi-channel graph convolution network in predicting drug target affinity, the prediction results were compared with the current common drug target affinity prediction methods, including KronRLS, SimBoost, DeepDTA, WideDTA and GraphDTA. The results testing in Davis dataset are shown in Table 7.

Table 7. Performances of various methods on Davis dataset

Method	CI	MSE	Pearson
KronRLS Error! Reference source not found.	0.871	0.379	
SimBoost Error! Reference source not found.	0.872	0.282	
DeepDTA Error! Reference source not found.	0.878	0.261	
WideDTA Error! Reference source not found.	0.886	0.262	0.820
GraphDTA Error! Reference source not found.	0.893	0.229	
MCGraphDTA	0.890	0.250	0.832

As can be seen from the table, MCGraphDTA, our experiment model performs better than KronRLS, SimBoost, DeepDTA and WideDTA, but worse than GraphDTA. Compared with KronRLS, SimBoost and DeepDTA models, the mean square error value of this experiment model decreased by 34%, 11% and 4% respectively, and the Concordance index value increased by 2%, 2% and 1% respectively. Compared with

WideDTA, the MSE value decreased by 5% and the CI value was 0.5% higher. However, compared with GraphDTA, the MSE value of MCGraphDTA is 9% higher and the CI value of MCGraphDTA is 0.3% lower.

The results testing in KIBA dataset are shown in Table 8. Compared with KronRLS, SimBoost and DeepDTA models, the mean square error value of this experiment model decreased by 62%, 29% and 19% respectively, and the Concordance index value increased by 13%, 6% and 2% respectively. Compared with WideDTA, the MSE value decreased by 12% and the CI value was 0.8% higher. However, compared with GraphDTA, the MSE value of MCGraphDTA is 11% higher and the CI value of MCGraphDTA is 1% lower.

Table 8. Performances of various methods on KIBA dataset

Method	CI	MSE	Pearson
KronRLS Error! Reference source not found.	0.782	0.411	
SimBoost Error! Reference source not found.	0.836	0.222	
DeepDTA Error! Reference source not found.	0.863	0.194	
WideDTA Error! Reference source not found.	0.875	0.179	0.856
GraphDTA Error! Reference source not found.	0.891	0.139	
MCGraphDTA	0.882	0.157	0.876

There is a gap between the test results of the experiment model and the expectation. Although the parameters of the model have been adjusted for many times according to the type and value range in the training stage, the model test results obtained by the “best” parameters are still worse than GraphDTA. There are two reasons for the failure of the affinity prediction: firstly, there are correlations between some parameters in the model. For example, when the value or type of parameter a is adjusted, the optimal value or type of parameter B will also change. In this case, if the original value or type is still used for the experiment, it is difficult to obtain the optimal parameter or type. Secondly, there are discrete and continuous parameters, so it is unreasonable to uniformly sample and test the continuous parameters, and the best value of these parameters may be between the sampling points. To solve these two problems, we intend to try to use Algorithms for Hyper-Parameter Optimization, such as random search, Bayesian Optimization and hyperband, to adjust the parameters of the model to improve the effect of the model.

4 Conclusions

Predicting drug-target affinity is of great importance to drug development. Applying computational models to predict the drug-target affinity can not only save cost of drug development, but also accelerate the drug development cycle. In this paper, we have proposed a multi-channel graph convolution network to predict drug target affinity. To solve the problem that the number of existing graph convolutional network layers should not be

too large, multi-channel graph convolution is introduced to aggregate the information of nodes with similar roles but far away from each other. Evaluation of the model on Davis and KIBA datasets demonstrates that the proposed method outperforms most relevant methods, suggesting the effectiveness of the proposed approach in predicting the affinity of drug and protein pairs. Though the performance of the proposed method is worse than that of GraphDTA, in the future work, we will further optimize the model parameters and apply the attention mechanism to get a more rational aggregation proportion for each channels.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (No. 61972299).

References

1. Asher, M.: New drugs cost US \$2.6 billion to develop. *Nat. Rev. Drug Disc.* **13**(12), 877 (2014)
2. Ashburn, T.T., Thor, K.B.: Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Disc.* **3**(8), 673 (2004)
3. Roses, A.D.: Pharmacogenetics in drug discovery and development: a translational perspective. *Nat. Rev. Drug Disc.* **7**(10), 807 (2008)
4. Cohen, P.: Protein kinases—the major drug targets of the twenty-first century? *Nat. Rev. Drug Disc.* **1**(4), 309 (2002)
5. Noble, M.E.M., Endicott, J.A., Johnson, L.N.: Protein kinase inhibitors: insights into drug design from structure. *Science* **303**(5665), 1800–1805 (2004)
6. Deshpande, M., Kuramochi, M., Wale, N., Karypis, G.: Frequent substructure-based approaches for classifying chemical compounds. *IEEE Trans. Knowl. Data Eng.* **17**(8), 1036–1050 (2005)
7. Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S.: The protein kinase complement of the human genome. *Science* **298**(5600), 1912–1934 (2002)
8. Stachel, S.J., et al.: Maximizing diversity from a kinase screen: identification of novel and selective pan-Trk inhibitors for chronic pain. *J. Med. Chem.* **57**(13), 5800–5816 (2014)
9. Pahikkala, T., et al.: Toward more realistic drug–target interaction predictions. *Brief. Bioinform.* **16**, 325–327 (2014)
10. He, T., Heidemeyer, M., Ban, F., Cherkasov, A., Ester, M.: SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *J. Cheminform.* **9**(1), 24 (2017)
11. Öztürk, H., Özgür, A., Ozkirimli, E.: DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* **34**(17), i821–i829 (2018)
12. Weininger, D.: SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**(1), 31–36 (1988)
13. Öztürk, H., Ozkirimli, E., Özgür, A.: arXiv Preprint. [arXiv:1902.04166](https://arxiv.org/abs/1902.04166) (2019)
14. Woźniak, M., et al.: Linguistic measures of chemical diversity and the “keywords” of molecular collections. *Sci. Rep.* **8**, 1–10 (2018)
15. Sigrist, C.J.A., et al.: Prosite, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* **38**(suppl_1), D161–D166 (2009)
16. Feng, Q., Dueva, E., Cherkasov, A., Ester, M.: arXiv Preprint. [arXiv:1807.09741](https://arxiv.org/abs/1807.09741) (2018)
17. Nguyen, T., Le, H., Venkatesh, S.: bioRxiv, p. 684662 (2019)
18. Zhou, C., Shi, B., Qiu, H., Liu, J.: *IEEE Access* **8**, 92802–92813 (2020)

19. Ronneberger, O., Fischer, P., Brox, T.: International Conference on Medical image Computing and Computer Assisted Intervention, pp. 234–241 (2015)
20. Cheol, J.J., Lin, X., Chen, X.-W.: *IEEE/ACM Trans. Comput. Biol. Bioinf.* **8**(2), 308–315 (2010)
21. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? *arXiv: 1810.00826* (2018). <http://arxiv.org/abs/1810.00826>
22. He, T., Heidemeyer, M., Ban, F., Cherkasov, A., Ester, M.: *J. Cheminf.* **9**, 24 (2017)
23. Gönen, M., Heller, G.: *Biometrika* **92**, 965–970 (2005)
24. Allen, D.M.: *Technometrics* **13**, 469–475 (1971)
25. Benesty, J., Chen, J., Huang, Y., Cohen, I.: *Noise Reduction in Speech Processing*, pp. 1–4. Springer, Heidelberg (2009). <https://doi.org/10.1007/978-3-642-00296-0>