

RESEARCH

Open Access



Multi-scaled self-attention for drug–target interaction prediction based on multi-granularity representation

Yuni Zeng¹, Xiangru Chen², Dezhong Peng^{2,3,4}, Lijun Zhang^{5,6} and Haixiao Huang^{7*}

*Correspondence:
huanghaixiao.cd@gmail.com

¹ School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China

² College of Computer Science, Sichuan University, Chengdu, China

³ Shenzhen Peng Cheng Laboratory, Shenzhen, China

⁴ Chengdu Sobey Digital Technology Co., Ltd, Chengdu, China

⁵ Sichuan Zhiqian Technology Co., Ltd, Chengdu, China

⁶ Chengdu RuiBei YingTe Information Technology Co., Ltd, Chengdu, China

⁷ Sichuan Provincial Commission of Politics and Law, Chengdu, China

Abstract

Background: Drug–target interaction (DTI) prediction plays a crucial role in drug discovery. Although the advanced deep learning has shown promising results in predicting DTIs, it still needs improvements in two aspects: (1) encoding method, in which the existing encoding method, character encoding, overlooks chemical textual information of atoms with multiple characters and chemical functional groups; as well as (2) the architecture of deep model, which should focus on multiple chemical patterns in drug and target representations.

Results: In this paper, we propose a multi-granularity multi-scaled self-attention (SAN) model by alleviating the above problems. Specifically, in process of encoding, we investigate a segmentation method for drug and protein sequences and then label the segmented groups as the multi-granularity representations. Moreover, in order to enhance the various local patterns in these multi-granularity representations, a multi-scaled SAN is built and exploited to generate deep representations of drugs and targets. Finally, our proposed model predicts DTIs based on the fusion of these deep representations. Our proposed model is evaluated on two benchmark datasets, KIBA and Davis. The experimental results reveal that our proposed model yields better prediction accuracy than strong baseline models.

Conclusion: Our proposed multi-granularity encoding method and multi-scaled SAN model improve DTI prediction by encoding the chemical textual information of drugs and targets and extracting their various local patterns, respectively.

Keywords: Drug–target interaction, Deep learning, Self-attention networks, Representations learning

Background

Drug–target interaction (DTI) indicates the binding of drug compounds to their targets. The targets refer to the proteins or some bio-molecules to which the drug directly binds, and which are responsible for the therapeutic efficacy of the drug in vivo [1]. The drugs exert their clinical effects in treating diseases by changing the structure of the targets or regulating their metabolism. Therefore, accurate identification of DTI is one crucial step of drug discovery and development [1–3]. For example, in process of drug repositioning



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

[4] task, DTI prediction is regarded as the foundation to find new targets of existing drugs. Nowadays, due to the high-cost and time-consuming traditional biological experiments, effective computational methods are urgently needed [5–7].

In response to this demand, many DTI prediction methods have been proposed in recent years. These methods mainly includes two parts: encoding methods and DTI prediction methods.

As for the encoding methods, most studies for DTI prediction label their inputs by a character-based dictionary. For example, in DeepDTA [6], with a dictionary like {‘C’:1,‘H’:2,‘N’:3,‘.’:63}, the drug simplified molecular input line entry system (SMILES) sequence ‘CN=C=O’ was labelled as [1 3 63 1 63 5]. It labelled each character of drug SMILES by its corresponding integer in the character-based dictionary. In addition, in other chemical compounds related fields, some works applied tokenization methods to extract substrings from drug sequences as their functional groups at the chemical level. Study [8] tokenized the names of chemical compounds by the open parser for systematic IUPAC nomenclature (OPSIN) tokenizer [9] and byte-pair-encoding (BPE) [10] in predicting chemical compounds task. Based on BPE, study [11] introduced a tokenization algorithm named SMILES pair encoding (SPE) to label the SMILES by the learned chemical groups. It has been applied to generative and predictive tasks and molecular tasks. Study [12] proposed a ChemBoost approach to predict protein-ligand binding affinity scores based on substrings extracted by Word2vec [13] and BPE. In these studies, tokenizer methods in the fields of natural language processing (NLP) were used for drug SMILES segmentation, and then the segmented SMILES were applied to compound-related tasks.

For DTI prediction methods, many efforts have been conducted to predict drug–target binding affinity scores in recent years. The traditional approach to DTI prediction mainly based on similarity [14, 15]. Study [16] used the 2D compound similarity of drugs and Smith-Waterman similarity of targets as the inputs. Then, the Kronecker regularized least squares (KronRLS) algorithm was employed to predict the binding affinity values of drug–target pairs. Study [17] also utilized a number of similarity-based information and features to predict DTI by a gradient boosting machine. DTINet [18] was based on the assumption that similar drugs may share similar targets. Taking a series of similar matrices as input, it was designed to find an optimal projection from drug space onto target space by the random walk with restart (RWR) algorithm.

With the significant success of deep learning in computer vision, speech recognition and NLP, deep learning models are widely used in DTI prediction. DeepDTA [6] employed two convolutional neural network (CNN) models to extract features for deep representations of drugs and targets. Then, an fully connected network was utilized to predict the interaction of drug and protein representations. OnionNet [19] also utilized CNNs for drug and protein representations and so as to predict the binding affinity values. GANsDTA [20] used the generative adversarial networks (GANs) to learn deep representations for drugs and targets, and then predicted the binding affinity scores of drug–target pairs. DeepCDA [21] also was proposed for binding affinity score prediction. It employed two CNNs to extract feature of drug and target. Then, long-short-term memory (LSTM) layers and a two-side attention mechanism were used in interaction learning to predict DTIs. Moreover, self-attention networks (SANs) also were applied to

generate deep representations of drugs and targets [22–24]. Especially, study [23] proved that SANs have the ability to capture the long-distance relation between atoms in drug and target sequences.

Despite these efforts, the existing methods have several areas for improvement:

- The existing encoding method labels molecular input character by character and it cannot encode fundamental chemical groups: (1) atoms with multiple characters in compounds, like 'Br', 'Cl', and (2) chemical functional groups, like 'CC', 'OH'. These chemical groups are the determining part of chemical compounds and protein sequences. Therefore, the existing encoding method leads to the loss of essential chemical information.
- The existing deep models do not fully model different chemical correlations between atoms and atoms, atoms and chemical groups, chemical groups and chemical groups. Although CNNs can capture local features of these correlations, they failed to model long-distant atoms [23]. Besides, SANS focus on the overall input sentence, but they may overlook fine-grained information in drug and target sequences [25]. Thus, the existing deep model for DTI prediction need to improve.

In order to address the above problems, we introduce a new multi-scaled SAN model for drug–target binding affinity prediction based on multi-granularity representations in this work. Taking protein sequences and drug SMILES sequences as inputs, we first introduce a multi-granularity encoding method for them. The multi-granularity encoding is built upon the BPE algorithm which is a widely used tokenization algorithm in field of NLP. BPE calculates the frequency of occurrence of each consecutive byte pair, and then forms a vocabulary from high-frequency byte pairs. The multi-granularity representations are labelled by the vocabulary and then transmitted as inputs to our proposed multi-scale SAN model. By assigning different window sizes to heads in SAN, the multi-scaled SAN is exploited to learn the multi-scaled local patterns and generate deep representations of drugs and targets. Finally, the prediction is made on fused deep representations.

To the end, we evaluate the effectiveness of our proposed model on benchmark datasets (Davis [26] and KIBA [27]). Experimental results demonstrate that our multi-granularity multi-scaled model yields better accuracy over baselines and existing DTI deep models. Moreover, the experiment analyses reveal that both the multi-granularity encoding and multi-scaled features extracted by our multi-scaled SANs are beneficial to DTI prediction.

Methods

In this work, we propose a multi-granularity multi-scaled method for DTI prediction, as shown in Fig. 1. The proposed method includes four components: multi-granularity encoding, drug representation learning, protein representation learning, and the interaction learning part. Firstly, we introduce a multi-granularity encoding method for drug and protein input sequences. In this process, the input sequences are encoded by a multi-granularity vocabulary, which are generated by a segmentation method. Then, taken the multi-granularity representations as inputs, a multi-scaled SAN is proposed to

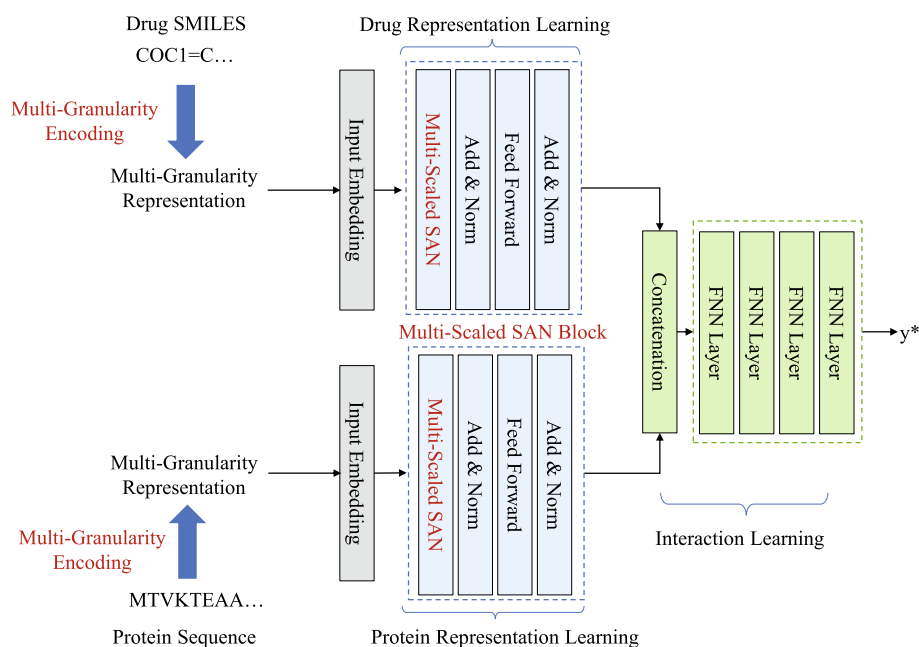


Fig. 1 Our proposed multi-granularity multi-scaled SAN model for DTI prediction

extract and fuse multi-scaled local features. Finally, the prediction is made on fused deep drug representations and deep protein representations by fully connected feed-forward networks.

Multi-granularity encoding

The current labeling method is not sufficient to encode chemical sequences since it ignores the chemical textual information from chemical groups in drugs and proteins, for example, chemical functional groups '[C@@H]', 'Br'. Thus, the intuitive way for representing a chemical sequence is to find out the substrings in the sequence by a computational method. Here, the substring is the chemical functional groups or atoms with multiple characters.

BPE [10] is a data compression method that can obtain high-frequency substrings to segment the sequence. In the field of NLP, BPE is widely used in different text tasks and as the first step to understand text sentences. BPE initializes the symbol vocabulary with the character vocabulary, and then it iteratively counts the frequency of adjacent character pairs in the corpus and merges the pair with the highest frequency to a new symbol. Finally, the vocabulary update is stopped when the number of merge operations reaches a threshold.

In this work, we utilize the BPE algorithm to generate vocabularies for encoding molecular inputs (SMILES or proteins). First, the segmentation datasets of drugs and targets are built and used to train BPE. Then, the BPE model trained by drug data would generate a vocabulary V_d with a threshold T_d for drugs, and V_p and T_p for targets. T determines the size of the generated vocabulary which consists of the segmented inputs by BPE. For example, taken the 'COC1=C(C=C2C(=C1)N=CN=C2NC3=C(C(=CC=C3)Cl)F)CN4CCCC[C@@H]4C(=O)N' as the input, the segmented outputs of BPE is shown in Table 1 with different T .

Table 1 Examples of segmented outputs of 'COC1=C(C=C2C(=C1)N=CN=C2NC3=C(C(=CC=C3)C)F)CN4CCCC[C@@H]4C(=O)N' with different T

T	Segmented SMILES Sequence (Vocabulary)
1k	COC1=C(C=C, 2C(=C1), N=C, N=C2, N, C3=C, (C(=CC=C, 3)C, I), F)C, N, 4, CCCC, [C@@H]4, C(=O)N
5k	CO, C1=C, (C=C, 2, C(=C1), N=C, N=C, 2, N, C3=C, (, C(=CC=C, 3)C, I), F,)C, N, 4, CCCC, [C@@H], 4, C(=O), N
25k	C, O, C1=C, (C=C, 2, C(=C, 1), N=C, N=C, 2, N, C3, =C, (, C(, =CC=C, 3,)C, I,), F,)C, N, 4, CCCC, [C@@H], 4, C(=O), N

Finally, a multi-granularity dictionary is constructed by assigning each group in the vocabulary a corresponding integer like the character-level dictionary in study [6]. Thus, an input sequence is labelled as multi-granularity representation $X = \{x_1, x_2, \dots, x_i, \dots\}$ where $x_i \in N^*$ and the length of X is varied, which depends on the length of the input sequence.

Multi-scaled self-attention model for drug–target binding affinity prediction

Our multi-scaled SAN is built upon Transformer block [28] which has shown excellent capability on sequence processing tasks. Given a drug multi-granularity representation X_d and protein multi-granularity representation X_p , we first adopt an input embedding module to integrate multiple embeddings. Then, for drug embedding E_d and protein embedding E_p , two multi-scaled SAN blocks are exploited to capture the local patterns features of drugs and proteins, respectively. Finally, an interaction block is proposed to fuse and extract interaction features from deep drug representations R_d and deep protein representations R_p . The final prediction y^* is the output of the interaction block.

Input embedding

Given a multi-granularity drug input as

$$X_d = \{d_1, d_2, \dots, d_{l_d}\}, \quad (1)$$

and a multi-granularity protein input as

$$X_p = \{p_1, p_2, \dots, p_{l_p}\}, \quad (2)$$

we define a hyper-parameter l to restrict the max input length. Specially, l_d restricts drug input X_d and l_p restricts target input X_p . If the length of X is shorter than l , the lack value is setting as 0. According to Transformer [28] and MT-DTI [23], the input of multi-scaled SAN is the sum of token embedding E_t of the input sequence and position embedding E_p of the input sequence, that is calculated as:

$$E_d = E_t^d + E_p^d. \quad (3)$$

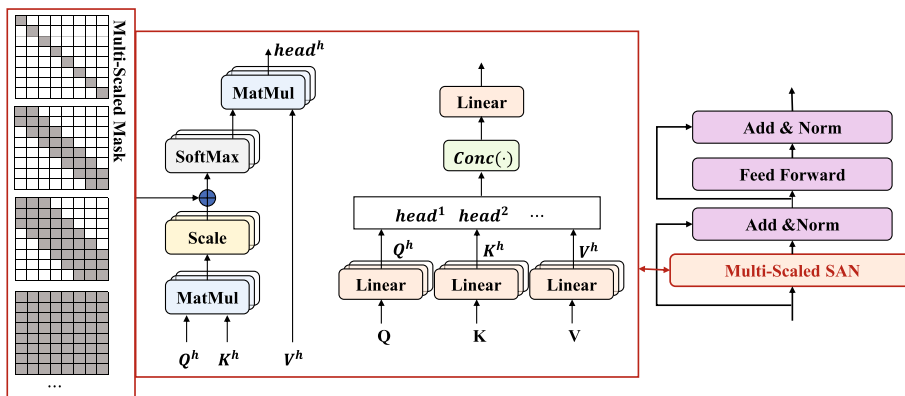


Fig. 2 Our proposed multi-scaled SAN block

Here, the token embedding $E_t^d \in \mathbb{R}^{l_d \times e_d}$ has a trainable weight $W_t^d \in \mathbb{R}^{v_d \times e_d}$. The v_d is the vocabulary size of drugs and e_d is the embedding length of drugs. The position embedding $E_p^d \in \mathbb{R}^{l_d \times e_d}$ has a trainable weight $W_p^d \in \mathbb{R}^{l_d \times e_d}$. As for protein embedding,

$$E_p = E_t^p + E_p^p. \tag{4}$$

where $E_t^p \in \mathbb{R}^{l_p \times e_p}$ is the token embedding of X_p , $E_p^p \in \mathbb{R}^{l_p \times e_p}$ is the position embedding of X_p and e_p is the embedding size of protein sequence.

Multi-scaled self-attention block

Multi-head SAN is the main component of Transformer [28]. It performs multiple self-attention modules on input expressions, then jointly pay attention to the information of different expression at different position. In this work, in order to generate a more informative deep representations of drugs and proteins, we adopt multi-scaled SAN to their embeddings, which assign different window size to heads in multi-head SAN, that is formulated as,

$$R_d = MSSAN(E_d, L_d), \tag{5}$$

$$R_p = MSSAN(E_p, L_p). \tag{6}$$

where $MSSAN(\cdot)$ denotes a multi-scaled self-attention block, as shown in Fig. 2. L_d and L_p are the hyper-parameters notating the number of multi-scaled SAN blocks.

Especially, suppose the input to multi-scaled SAN blocks is E . Our model first transforms input sequence into N subspace with different linear projections,

$$Q^h, K^h, V^h = EW_Q^h, EW_K^h, EW_V^h, \tag{7}$$

where $1 \leq h \in \mathbb{N}^+ \leq N$ is the index and $W_*^h \in \mathbb{R}^{e_d \times d_*^h}$, the d^h denotes the dimensionality of the h^{th} head subspace. Then, we utilize a mask matrix $M^h \in \mathbb{R}^{l \times l}$ for the h^{th} head to achieve multi-scaled SAN. The output of h^{th} head on multi-scaled SAN is calculated as,

$$head^h = softmax\left(\frac{Q^h K^{hT}}{\sqrt{d_h}} + M^h\right) V^h, \tag{8}$$

where M^h is determined by a hyper-parameter named window size m^h ,

$$M_{i,j}^h = \begin{cases} -\infty, & j < (i - m^h), \\ 0, & (i - m^h) \leq j \leq (i + m^h), \\ -\infty, & j > (i + m^h). \end{cases} \quad (9)$$

Then, the h heads are concatenated,

$$heads = conc[head^1, \dots, head^h, \dots, head^N] \quad (10)$$

where $conc(\cdot)$ is a concatenation function. Next, a residual connection [29] and the layer normalization (LN(\cdot)) [30] are employed,

$$Z = LN(heads + E). \quad (11)$$

Thus, the output of a multi-scaled SAN block is formulated,

$$MSSAN(E, 1) = LN(Z + FFN(Z, 1)), \quad (12)$$

where $FFN(Z, 1)$ denotes one fully connected feed-forward layer (FCN) with ReLU activation [31] and Z as input. The hidden size of the FCN is e_d .

Interaction block

The interaction block in this work is to combine deep drug and protein representations and predicts the binding affinity scores of drug–target pairs. Mathematically, firstly,

$$R = conc[R_d, R_p]. \quad (13)$$

Next, 4 layers of FCN are employed to capture the interaction information from R .

$$y^* = FFN(R, 4) \quad (14)$$

where y^* is the predicted binding affinity value of the drug–target pair.

Data and experiments

Datasets

Benchmark datasets for DIT prediction

We evaluated our proposed model on Davis [26] and KIBA [27] datasets because they are widely used in existing drug–target interaction studies. Specially, in order to ensure the uniqueness of drug input sequence, we only use Isomeric SMILES strings in this paper. The number of proteins, compounds and interactions of the Davis and KIBA dataset are summarised in Table 2. In particular, the Davis dataset contains the 442 kinase proteins, their relevant inhibitors (68 ligands) and their respective dissociation constant (K_d) value. The binding affinity scores of drug–target pairs were transformed K_d into log space pK_d , as [6, 17], as follows,

$$pK_d = -\log_{10}\left(\frac{K_d}{1e9}\right). \quad (15)$$

Table 2 The summary of the benchmark datasets

	Proteins	Compounds	Interactions	Training Data	Test Data
Davis	442	68	30056	25046	5010
KIBA	229	2111	118254	98545	19709

Table 3 Summary of parameter settings

Parameter	KIBA	Davis
l_d	80	36
l_p	800	900
m^h	0,1,2,3	0,1,2,3
e_d	128	64
L_d	2	2
L_p	2	1
Hidden size in FFN	1024,1024,512,1	
Epoch	300	300
Dropout	0.1	0.1
Optimizer	Adam	Adam
Learning rate	0.0001	0.0001

The used KIBA dataset comprised 229 proteins, 2111 drugs and their KIBA scores. Here, the KIBA scores measure the kinase inhibitor bioactivities as the binding affinity values in following experiments.

Segmentation dataset

We collect drug SMILES sequences from the National Center for Biotechnology Information (NCBI)¹ and protein sequences from The Universal Protein Resource². Finally, 147546 SMILES sequences and 114500 protein sequences are collected as segmentation data to train the segmentation methods.

Experiment setup and metric

Table 3 summaries other hyper-parameter settings. We use five-time leave-one-out cross-validation to train our model and list the average results on test data. All models were trained on 1 NVIDIA 3080 GPU.

To measure the performance of our model, three metrics are included: mean squared error (MSE), Concordance Index (CI) and the r_m^2 metric. MSE is the loss of the optimizer in the deep model.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i^* - y_i)^2 \quad (16)$$

¹ <https://pubchem.ncbi.nlm.nih.gov/>.

² https://ftp.expasy.org/databases/uniprot/current_release/uniparc/

where the y^* is the predicted binding affinity value, y is the ground-truth and n is the number of drug–target pairs.

CI is the probability of the predicted scores of two randomly chosen drug–target pairs in the correct order, as

$$CI = \frac{1}{N} \sum_{\delta_i > \delta_j} f(t_i - t_j) \quad (17)$$

where t_i is the predicted value with larger affinity δ_i , t_j is the prediction score for smaller affinity δ_j and N is a normalization constant. Moreover, the $f(x)$ is a step function [16],

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ 0.5, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases} \quad (18)$$

Then r_m^2 metric [32, 33] is another widely used metric in this field. Mathematically,

$$r_m^2 = r^2 * (1 - \sqrt{r^2 - r_0^2}), \quad (19)$$

where r^2 and r_0^2 are the squared correlation coefficient values between the observed and predicted values with and without intercept, respectively. Especially, the r_m^2 value of an acceptable model should be larger than 0.5.

Experiments 1: Effects of the segmentation method

In this paper, the BPE algorithm is utilized as the segmentation method to learn the substrings in drug SMILES and protein sequences. As seen in Table 1, the threshold T determines the degree of segmentation. The larger T in BPE indicates the more fine-grained and longer segmentation outputs. We first investigated the effects of T to DTI prediction on KIBA and Davis dataset. We extract various multi-granularity representations by setting different T , and then build DeepDTA [6] models with these representations as inputs. As plotted in Figs. 3 and 4, the prediction results on KIBA and Davis dataset are demonstrated, respectively.

Discussion: For both KIBA and Davis dataset, the $T_d = 20k$ and $T_g = 36k$ is superior to other settings. It is clear that when $T_d < 20k$ and $T_g < 36k$, the prediction quality goes up as T increases. Conversely when $T_d > 20k$ and $T_g > 36k$, the increase of T seems to cause performance degradation. One possible reason is that the segmented SMILE with $T_d = 20k$ and the segmented protein sequences with $T_g = 36k$ include more chemical textual information for predicting DTI. As the result, $T_d = 20k$ and $T_g = 36k$ in following experiments.

Experiments 2: Encoding methods for DTI prediction

The starting point of our approach is an observation in encoding methods. Considering the improvements of existing character-based encoding methods, we adopt segmentation method to learn the chemical groups in drug and target sequences. Thus, in this subsection, we evaluate whether deep representations learned from multi-granularity representations contains more drug–target interaction information than deep representations learned from character encoded representations, We also implemented

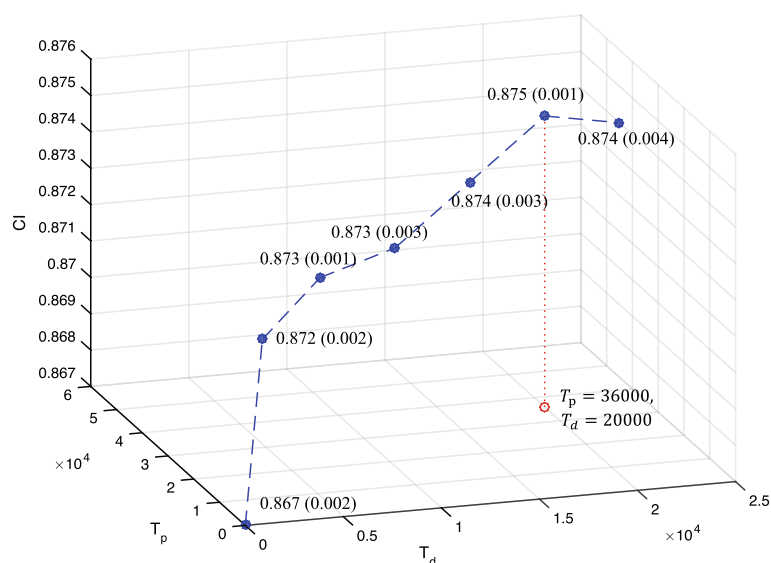


Fig. 3 Results of DeepDTA [6] model on the KIBA dataset with different multi-granularity representations as inputs. These multi-granularity representations are encoded by BPE algorithm with different threshold T . Here, T_d is the threshold T for drug segmentation and T_p is the threshold T for protein segmentation

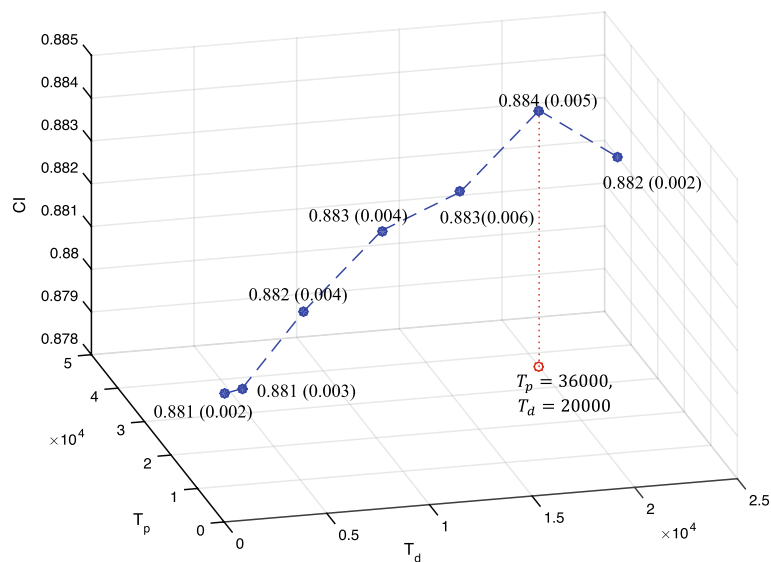


Fig. 4 Results of DeepDTA [6] model on the Davis dataset with different multi-granularity representations as inputs. These multi-granularity representations are encoded by BPE algorithm with different threshold T . Here, T_d is the threshold T for drug segmentation and T_p is the threshold T for protein segmentation

DeepDTA [6], as baseline, with multi-granularity representations and character encoded representations as inputs. Table 4 lists the average results of the drug–target binding affinity prediction on KIBA and Davis dataset.

Discussion: As seen, the multi-granularity encoding method improves the prediction quality in both two datasets, reconfirming the necessity of encoding the chemical groups in drug and protein sequences.

Table 4 Results of DeepDTA (CNN model) on KIBA and Davis dataset with character-based and multi-granularity encoding. Especially, the character-based encoding methods is original labelling method in DeepDTA [6]

	Encoding Method	CI	MSE	r_m^2
KIBA	Character Encoding	0.863 (0.002)	0.194	0.673 (0.009)
	Multi-Granularity	0.875 (0.001)	0.185	0.696 (0.017)
Davis	Character Encoding	0.878 (0.004)	0.261	0.630 (0.017)
	Multi-Granularity	0.884 (0.005)	0.250	0.655 (0.019)

Bold values indicate the best results on the datasets

Table 5 Results of deep models on KIBA and Davis dataset with multi-granularity representations as inputs

	Deep models	CI	MSE	r_m^2
KIBA	CNNs	0.863 (0.002)	0.194	0.673 (0.009)
	SANs	0.875 (0.003)	0.179	0.691 (0.019)
	Multi-scaled SANs	0.890 (0.002)	0.155	0.742 (0.010)
Davis	CNNs	0.878 (0.004)	0.261	0.630 (0.017)
	SANs	0.888 (0.004)	0.232	0.689 (0.007)
	Multi-scaled SANs	0.890 (0.005)	0.233	0.681(0.014)

Bold values indicate the best results on the datasets

Experiments 3: Multi-scaled SAN for DTI prediction

In this section, we conducted experiments about deep models based on multi-granularity encoding. Table 5 gives the average test results on the drug–target binding affinity prediction tasks. One intuition of our work is to capture the local patterns in multi-granularity representations by multi-scaled SANs. To evaluate it, we implemented models with CNNs from DeepDTA [6], SANs from Transformer [28] which also employed in MT-DTI [23] and our multi-scaled SAN.

Discussion: As shown in Table 5, the multi-scaled SAN outperforms the SANs model, indicating that the local pattern information can raise the ability of SANs to capture the drug–target interaction information. Moreover, as all known, CNNs have the ability to capture the local features. According to Table 5, the multi-scaled model achieved higher results than CNNs model, revealing extracting local features by the dynamic weights of multi-scaled SANs is superior to fixed weight from CNNs.

Experiments 4: Comparison to existing approaches

In the end, we compare our multi-granularity multi-scaled SANs model to traditional methods, like KronRLS [16], SimBoost [17], and other recent deep sequence representation methods, like DeepDTA [6], MT-DTI [23], GANsDTA [20], CrossAttentionDTI [24]. Table 6 lists the results of these models on drug–target binding affinity prediction task.

Discussion: As seen, these sequence-based deep models improve prediction quality than transitional methods, reconfirming the effectiveness of modeling sequence information. Besides, our proposed model improves CI to 0.890 on both KIBA and Davis

Table 6 Results on KIBA and Davis of our proposed multi-granularity multi-scaled SANs model, transitional methods and existing deep sequence representation methods

	Method	Drug	Protein	Interaction	CI	MSE	r_m^2
KIBA	KronRLS [16]	Pubchem Sim	S-W	–	0.782 (0.001)	0.411	0.342 (0.001)
	SimBoost [17]	Pubchem Sim	S-W	–	0.836 (0.001)	0.222	0.629 (0.007)
	DeepDTA [6]	CNNs	CNNs	Concatenation	0.863 (0.002)	0.194	0.673 (0.009)
	MT-DTI [23]	SANs	SANs	Concatenation	0.882 (0.002)	0.152	0.738 (0.006)
	GANsDTA [20]	GANs	GANs	Concatenation	0.866 (–)	0.224	0.675 (–)
	CrossAttentionDTI [24]	Cross SANs	Cross SANs	Concatenation	0.874 (0.001)	0.175	–
	Ours	MSSAN	MSSAN	Concatenation	0.890 (0.002)	0.155	0.742(0.010)
Davis	KronRLS [16]	Pubchem Sim	S-W	–	0.871 (0.001)	0.379	0.407 (0.005)
	SimBoost [17]	Pubchem Sim	S-W	–	0.872 (0.001)	0.282	0.644 (0.006)
	DeepDTA [6]	CNNs	CNNs	Concatenation	0.878 (0.004)	0.261	0.630 (0.017)
	MT-DTI [23]	SANs	SANs	Concatenation	0.887 (0.003)	0.245	0.665 (0.014)
	GANsDTA [20]	GANs	GANs	Concatenation	0.881 (–)	0.276	0.653 (–)
	CrossAttentionDTI [24]	Cross SANs	Cross SANs	Concatenation	0.876 (0.006)	0.244	–
	Ours	MSSAN	MSSAN	Concatenation	0.890 (0.005)	0.233	0.681 (0.014)

Bold values indicate the best results on the datasets

dataset, and improve r^2m to 0.742 and 0.681 on KIBA and Davis dataset, respectively. Thus, our model outperforms the recent sequence-based works, indicating the superiority of the proposed approaches.

Discussion

DTI prediction is to identify the interactions between drugs and targets, which is a substantial task in the drug discovery field. Many studies proposed computation methods to reduce dependence on time, cost and traditional biological experiments. Based on these related works, we proposed a deep model for DTI prediction based on the multi-granularity encoding and the multi-scaled SAN model in this work. The main contribution of this paper can be summarized as follows.

- In order to encode fundamental chemical groups, a multi-granularity encoding method is introduced to label the molecular inputs of drugs and targets as the corresponding multi-granularity representations (Section Method).
- In order to model the multiple kinds of chemical correlations, a multi-scaled SAN model is proposed to learn the local patterns in drugs and targets by the dynamic weights (Section Method).
- Our proposed method achieves higher results on KIBA and DAVIS datasets, compared to traditional methods and recent deep sequence representation methods (Section Experiments).

Via in-depth analyses, our work may contribute to subsequent researches on this topic: (1) the multiple encoding methods of SMILES sequence and protein sequence in DTI prediction as well as other bioinformatics tasks, (2) the learning method for

local patterns in sequence, and (3) the representation learning of drug and target sequences.

Conclusion

In this paper, we investigate and propose effective approaches to improve drug–target binding affinity prediction from both encoding method and model architecture perspectives. As for the encoding method, we employ the BPE algorithm and segmentation dataset to train a multi-granularity encoding method for drug SMILES and protein sequences. It contributes to encode atoms with multiple characters and chemical functional groups. Secondly, we build a multi-scaled SAN model for their multi-granularity representations by assigning various window size to heads in original SANs. Experimental results demonstrate that the proposed approach not only is of benefit to predict DTI but also makes DTIs prediction surpass baselines on various metrics.

Our proposed method achieves the improvements by benefiting from the encoding method for chemical groups and the local patterns modeled by the representation learning model. In the encoding process, we collected a large of unlabeled data of drugs and targets to train the encoding method. Meanwhile, we found the lack of labeled data limits the improvements of deep models to predict new DTIs. Thus, our future work may focus on the utilization of these unlabeled data, like the unsupervised learning method for DTI learning.

Abbreviations

DTI	Drug–target interaction
SMILES	Simplified molecular input line entry system
OPSIN	open parser for systematic IUPAC nomenclature
GAN	Generative adversarial network
CNN	Convolutional neural network
SAN	Self-attention network
BPE	Byte-pair-encoding
NLP	Natural language processing
KronRLS	Kronecker regularized least squares
RWR	Random walk with restart
LSTM	Long-short-term memory
NCBI	National Center for Biotechnology Information
MSE	Mean squared error
CI	Concordance Index

Acknowledgements

Not applicable.

Author Contributions

YZ and HH conceived the research work. YZ and XC implemented the proposed model and conducted experiments. DP and LZ supervised the experiments and analysed the experimental results. YZ and HH drafted and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 61971296, U19A2078, 61836011]; and the Sichuan Science and Technology Planning Project [grant number 2021YFG0317, 2021YFG0301].

Availability of data and materials

The segmentation datasets are freely available at <https://pubchem.ncbi.nlm.nih.gov/> and https://ftp.expasy.org/databases/uniprot/current_release/uniparc/. The training and testing datasets for this paper are freely available at [6].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 May 2022 Accepted: 22 July 2022

Published online: 03 August 2022

References

1. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. 2017;16(1):19–34.
2. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief Bioinform*. 2021;22(1):247–69.
3. Ye Q, Zhang X, Lin X. Drug–target interaction prediction via multiple classification strategies. *BMC Bioinform*. 2022;22S(12):461.
4. Jarada TN, Rokne JG, Alhadj R. SNF-CVAE: computational method to predict drug–disease interactions using similarity network fusion and collective variational autoencoder. *Knowl Based Syst*. 2021;212: 106585.
5. Agyemang B, Wu W, Kpiebaareh MY, Lei Z, Nanor E, Chen L. Multi-view self-attention for interpretable drug–target interaction prediction. *J Biomed Inform*. 2020;110: 103547.
6. Öztürk H, Özgür A, Olmez EO. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*. 2018;34(17):821–9.
7. Monteiro NR, Ribeiro B, Arrais J. Drug–target interaction prediction: end-to-end deep learning approach. *IEEE/ACM Trans Comput Biol Bioinform*. 2020.
8. Omote Y, Matsushita K, Iwakura T, Tamura A, Ninomiya T. Transformer-based approach for predicting chemical compound structures. In: Proceedings of the 1st conference of the Asia-Pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing, ACL/IJCNLP, Suzhou, China; 2020. pp. 154–162.
9. Lowe DM, Corbett PT, Murray-Rust P, Glen RC. Chemical name to structure: OPSIN, an open source solution. *J Chem Inf Model*. 2011;51(3):739–53.
10. Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL, August 7–12, Berlin, Germany (2016).
11. Li X, Fourches D. SMILES pair encoding: a data-driven substructure tokenization algorithm for deep learning. *J Chem Inf Model*. 2021;61(4):1560–9.
12. Özçelik R, Öztürk H, Özgür A, Ozkirimli E. ChemBoost: a chemical language based approach for protein–ligand binding affinity prediction. *Mol Inf*. 2020.
13. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: 27th Annual conference on neural information processing systems, December 5–8, 2013, Lake Tahoe, Nevada, United States, pp. 3111–3119.
14. Buza K, Peska L. Drug–target interaction prediction with bipartite local models and hubness-aware regression. *Neurocomputing*. 2017;260:284–93.
15. Mei J, Kwok CK, Yang P, Li X, Zheng J. Drug–target interaction prediction by learning from local information and neighbors. *Bioinformatics*. 2013;29(2):238–45.
16. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, Aittokallio T. Toward more realistic drug–target interaction predictions. *Brief Bioinform*. 2015;16(2):325–37.
17. 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*. 2017;9(1):24–12414.
18. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, Peng J, Chen L, Zeng J. A network integration approach for drug–target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun*. 2017;8(1):1–13.
19. Zheng L, Fan J, Mu Y. OnionNet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. *ACS Omega*. 2019;4(14):15956–65.
20. Zhao L, Wang J, Pang L, Liu Y, Zhang J. GANsDTA: predicting drug–target binding affinity using gans. *Front Genet*. 2020;10:1243.
21. Karim A, Parvin R, Antti P, Massoud A, Ghasemi JB, Ali MN. DeepCDA: deep cross-domain compound–protein affinity prediction through lstm and convolutional neural networks. *Bioinformatics*. 2020;36(17):4633–42.
22. Huang K, Xiao C, Glass LM, Sun J. MolTrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*. 2021;37(6):830–6.
23. Shin B, Park S, Kang K, Ho JC. Self-attention based molecule representation for predicting drug–target interaction. In: Proceedings of the machine learning for healthcare conference, MLHC, Ann Arbor, Michigan, USA, vol. 106; 2019. pp. 230–248.
24. Koyama K, Kamiya K, Shimada K. Cross attention DTI: drug–target interaction prediction with cross attention module in the blind evaluation setup. In: 19th International workshop on data mining in bioinformatics, BIOKDD, Aug 24, San Diego, USA; 2020.
25. Guo M, Zhang Y, Liu T. Gaussian transformer: a lightweight approach for natural language inference. In: The thirty-AAA conference on artificial intelligence, Honolulu, Hawaii, USA; 2019. pp. 6489–6496. (2019).
26. Davis DJ, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2011;29:1046–51.

27. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, Aittokallio T. Anaking sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model*. 2014;54(3):735–43.
28. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Lu, Polosukhin I. Attention is all you need. In: *Advances in neural information processing systems, NIPS*; 2017. pp. 5998–6008.
29. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *IEEE conference on computer vision and pattern recognition, CVPR, Las Vegas, NV, USA, June 27–30*; IEEE Computer Society; 2016. p. 770–778.
30. Ba JL, Kiros JR, Hinton GE. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).
31. Agarap AF. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375* (2018).
32. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN. Some case studies on application of “ r_m^2 ” metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *J Comput Chem*. 2013;34(12):1071–82.
33. Roy PP, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive qsar models. *Molecules*. 2009;14(5):1660–701.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

