Systems biology

Drug–target interaction predictions with multi-view similarity network fusion strategy and deep interactive attention mechanism

Wei Song [ID][1], **Lewen Xu**[1], **Chenguang Han**[1], **Zhen Tian** [ID][1,2,*], **Quan Zou** [ID][2,*]

[1]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450000, China
[2]Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China

*Corresponding authors. E-mail: ieztian@zzu.edu.cn (Z.T.); E-mail: zouquan@nclab.net (Q.Z.)

Associate Editor: Xin Gao

Abstract

Motivation: Accurately identifying the drug–target interactions (DTIs) is one of the crucial steps in the drug discovery and drug repositioning process. Currently, many computational-based models have already been proposed for DTI prediction and achieved some significant improvement. However, these approaches pay little attention to fuse the multi-view similarity networks related to drugs and targets in an appropriate way. Besides, how to fully incorporate the known interaction relationships to accurately represent drugs and targets is not well investigated. Therefore, there is still a need to improve the accuracy of DTI prediction models.

Results: In this study, we propose a novel approach that employs Multi-view similarity network fusion strategy and deep Interactive attention mechanism to predict Drug–Target Interactions (MIDTI). First, MIDTI constructs multi-view similarity networks of drugs and targets with their diverse information and integrates these similarity networks effectively in an unsupervised manner. Then, MIDTI obtains the embeddings of drugs and targets from multi-type networks simultaneously. After that, MIDTI adopts the deep interactive attention mechanism to further learn their discriminative embeddings comprehensively with the known DTI relationships. Finally, we feed the learned representations of drugs and targets to the multilayer perceptron model and predict the underlying interactions. Extensive results indicate that MIDTI significantly outperforms other baseline methods on the DTI prediction task. The results of the ablation experiments also confirm the effectiveness of the attention mechanism in the multi-view similarity network fusion strategy and the deep interactive attention mechanism.

Availability and implementation: https://github.com/XuLew/MIDTI.

1 Introduction

The prediction of drug–target interactions (DTIs) holds a critical position in the process of drug development and repurposing (Yıldırım *et al.* 2007). Researchers estimate that the development of one new drug to be approved by the traditional wet experimental approach for clinical use will cost over $1 billion and take 10–15 years (Hinkson *et al.* 2020). Therefore, in pursuit of accelerating the drug and target screening process, researchers have resorted to computational-based approaches to assist the rapid design and development of drugs (Lin *et al.* 2020). The essential step for these computational models is the prediction of underlying DTIs, aiming to identify novel targets for existing drugs.

Computational-based DTI prediction approaches (DTIs), could be divided into three categories: structure-based approaches, ligand-based approaches, and machine learning-based approaches (Ding *et al.* 2022). Structure-based and ligand-based methods are two types of traditional computation-based prediction approaches. Specifically, structure-based methods usually consider the structures of both drug molecules and their targets, as well as the binding sites (Shaikh *et al.* 2016). However, the known structures of some drug targets such as membrane proteins are still limited

(Bagherian *et al.* 2021). Ligand-based approaches always utilize existing active small molecule structures to establish pharmacophore models or quantitative structure-activity relationships (Keiser *et al.* 2007). These methods usually need a large number of known binding ligands for interested targets and they are always ineffective when only a few ligands are known to bind with their targets (Tian *et al.* 2022).

Currently, many machine learning-based approaches have been widely proposed, which usually exploit the chemical structure of drugs and the genomic sequences of targets to extract their significant features efficiently. These methods always treat the DTI prediction problem as a binary classification task where they extract potential representations of drugs and targets, and take the concatenated embeddings of drug–target pairs as inputs for classification separately (Lee *et al.* 2019, Nguyen *et al.* 2021, Quan *et al.* 2019, Shin *et al.* 2019). For example, DeepConv-DTI learned the target protein features based on the convolution neural networks (CNN) model and the drug features from Extended Connectivity Fingerprint (ECFP) through the fully connected layer to predict DTIs. Nevertheless, the fully connected layer cannot capture potential relationships among distant atoms in raw molecule sequences (Lee *et al.* 2019). Meanwhile, the

rapid development of graph neural networks (GNNs) has extended the application of machine learning to the graph domain, and related methods have also been applied for feature extraction. For example, GraphCPI (Quan *et al.* 2019) and GraphDTA (Nguyen *et al.* 2021) employed GNNs to capture the structural information of drugs and improve the predictive ability of DTIs. To comprehensively capture the relationship among atoms in the sequences, Shin *et al.* proposed a Transformer-based DTI model, which utilized multi-layered bidirectional transformer encoders to learn the high-dimensional structure of molecules from the simplified molecular input line entry system (SMILES) string (Shin *et al.* 2019). These methods above learned molecular representation only based on the molecular structure of drugs and targets themselves but ignored the interaction contributions between each DTI pair.

In addition to the chemical and genomic features, the relationships between biological entities (e.g. drugs, targets, diseases, and side effects) usually contain rich semantic information, which could offer a system-level understanding DTIs (Li *et al.* 2023). Thus, establishing meaningful networks that incorporate this heterogeneous biological information could contribute to the prediction of potential DTIs. For example, DITNet integrated diverse heterogeneous information and obtained the representations of nodes through the random walk with a restart (RWR) and diffusion component analysis (DCA) models (Luo *et al.* 2017). Based on DTINet, Peng *et al.* added a denoising-auto encoder-based feature selector and a CNN-based interaction predictor to improve the accuracy in DTI prediction (Peng *et al.* 2020). However, these approaches above usually extracted the features based on each network separately, which may be unable to leverage complex relationships from the heterogeneous networks consistently. Afterward, MVGCN employed a neighborhood information aggregation (NIA) layer designed for iteratively updating the embeddings of nodes from different views (Fu *et al.* 2022). EEG-DTI performed a message-passing strategy based on different types of edges in the heterogeneous networks (Peng *et al.* 2021). However, MVGCN and EEG-DTI considered different types of edges separately, which made it difficult to apply on large-scale networks with multiple types of data sources. It is also a challenge to integrate the representations of entities from multiple views in an appropriate manner.

Nowadays, various types of biological data related to drugs and targets are easily accessible, laying the foundation for multi-view similarity network construction. Since these networks may have varying rates of false positives and -negatives, a multi-view network fusion strategy should be raised for establishing a more robust biological network, which could accurately capture the underlying complex relationships. Meanwhile, previous studies only simply concatenate the representations of drugs and targets, which neglects the interactive contributions between the embeddings of drugs and targets. The attention mechanism in deep neural networks has shown an outstanding role in representation learning. Therefore, by means of the effective attention mechanism coupled with the known DTI information, we can obtain their discriminative representations in a feasible manner.

Inspired by multi-view similarity network fusion strategy and deep interactive attention mechanism, here we propose a novel method called MIDTI to predict DTIs. The overall framework of MIDTI (see Fig. 1) mainly contains four steps.

Firstly, MIDTI constructs different drug similarity networks based on drug-related association information and obtains an integrated drug similarity network with a multi-view similarity network fusion strategy. MIDTI also establishes an integrated target similarity network similarly. Secondly, MIDTI adopts the GCNs as the encoders to learn drug and target embeddings from the integrated drug similarity network, the integrated target similarity network, the drug–target bipartite network as well as the drug–target heterogeneous network respectively. Thirdly, MIDTI learns the discriminative embeddings based on the known DTI relationships with the deep interactive attention mechanism. Lastly, we feed the learned representations of drug–target pairs into the multilayer perceptron (MLP) to predict DTIs. Our main contributions can be summarized as follows:

- We put forward a novel multi-view similarity network fusion strategy, which could integrate different similarity networks in an unsupervised manner with the multi-view attention mechanism, as long as the nodes and sizes of these networks are consistent.
- MIDTI employs the deep interactive attention mechanism to learn the discriminative embeddings of drugs and targets with known DTI information.
- Extensive experimental results fully indicate that MIDTI is superior to other SOTA approaches in DTI prediction tasks.

2 Materials and methods

2.1 Data collection

In this study, DTIs were initially downloaded from Luo's dataset (Luo *et al.* 2017). After processing, the experiment data mainly contains 12 015 nodes and 1 895 445 edges, which includes four types of nodes (drugs, targets, diseases, and side effects) and six types of edges (drug–protein interactions, drug–drug interactions, drug–disease associations, drug–side-effect associations, protein–disease associations, and protein–protein interactions). There are 1923 DTIs, related to 708 drugs and 1512 targets. Meanwhile, to comprehensively evaluate the performance of MIDTI, we also perform the comparison experiment on Yamanishi's (Yamanishi *et al.* 2008) and Zheng's dataset (Zheng *et al.* 2018). The description for these two datasets could be referred to the Supplementary Section S4.

2.2 Multitype network construction

As is shown in Step 1 in Fig. 1, we first construct different drug similarity networks and target similarity networks. Then MIDTI employs the similarity network fusion strategy to establish the integrated drug similarity network and the integrated target similarity network. Lastly, we construct the drug–target heterogeneous network.

2.2.1 Similarity network construction for drugs and targets

MIDTI firstly establishes five similarity networks for drugs, based on (i) drug–drug interactions, (ii) drug–disease associations, (iii) drug–side-effect associations, (iv) drug–protein associations, and (v) drug–chemical structures. Meanwhile, MIDTI establishes four similarity networks for targets based on (i) protein–protein interactions, (ii) protein–disease associations, (iii) drug–protein associations, and (iv) genome sequences. The construction process for these similarity
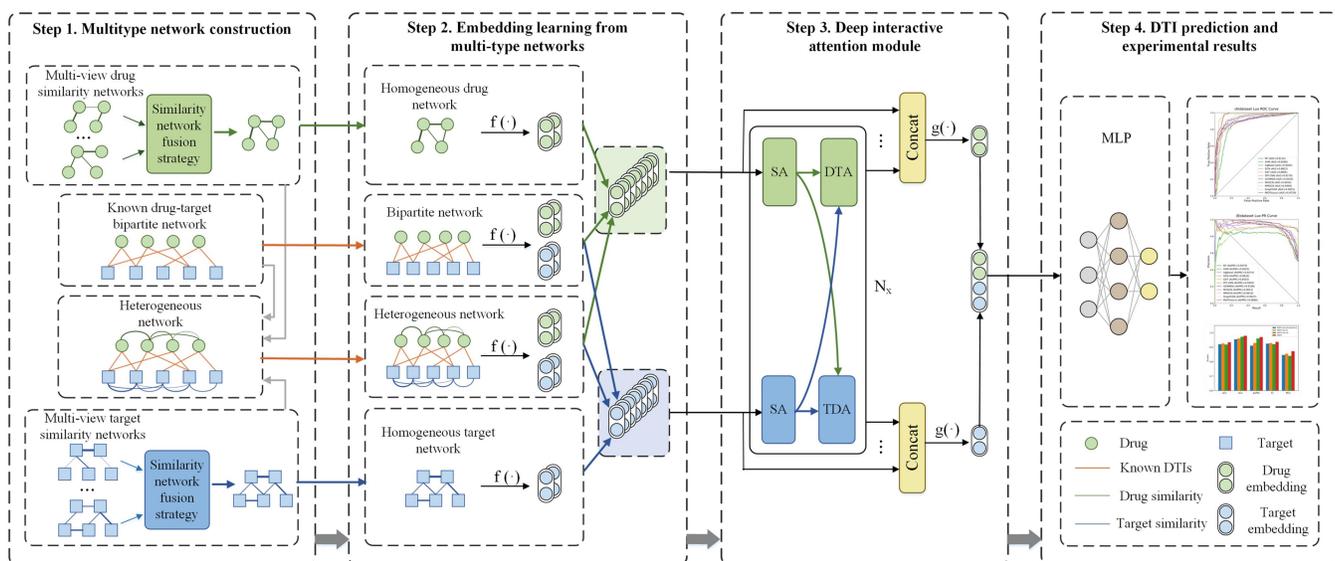
Figure 1. The overall framework of MIDTI. In Step 1, MIDTI constructs the integrated similarity networks of drugs and targets with their multisource information, as well as the drug–target bipartite network and the drug–target heterogeneous network. In Step 2, MIDTI learns the embeddings of drugs and targets from multiple networks respectively. In Step 3, MIDTI adopts the deep interactive attention mechanism to learn discriminative representations of drugs and targets. In Step 4, MIDTI predicts the potential DTIs with the MLP classifier.

networks of drugs and targets has been displayed in Supplementary Section S1.

2.2.2 Similarity network fusion strategy
Inspired by BIONIC (Forster *et al.* 2022), MIDTI will integrate different similarity networks of drugs and targets with the similarity network fusion strategy (see Fig. 2). The integrated network could accurately reflect the topologies of the underlying original networks and capture functional information. Different from BIONIC, MIDTI adds a multi-view attention mechanism that adaptively learns the importance of features from different similarity networks. A detailed description of the similarity network fusion strategy has been presented in Supplementary Section S2.

In this way, MIDTI could establish the integrated drug similarity matrix A_{homo_d} and the learned drug feature representation X_d. Besides, MIDTI could also establish the integrated target similarity matrix and target feature matrix represented as A_{homo_t} and X_t.

2.2.3 Drug–target heterogeneous network construction
MIDTI establishes a drug–target heterogeneous network named N_{hete} based on the integrated drug similarity network, integrated target similarity network and drug–target bipartite network.

2.3 Embedding learning from multitype networks
In this section, MIDTI learns the embeddings of drugs and targets with GCNs from multi-type networks, which are integrated drug similarity network N_{homo_d}, integrated target similarity network N_{homo_t}, drug–target bipartite network N_{bi} and drug–target heterogeneous network N_{hete}, respectively. Meanwhile, their corresponding adjacency matrices are denoted as A_{homo_d}, A_{homo_t}, A_{bi} and A_{hete}.

Taking the N_{homo_d} as an example, MIDTI adopts GCNs to learn embeddings of drugs and the output of $(l+1)$ layer is denoted as:

$$X_{homo\_d}^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}_{homo\_d}\tilde{D}^{-\frac{1}{2}}X_{homo\_d}^{(l)}W_{homo\_d}^{(l)}\right) \quad (1)$$

where $\tilde{A}_{homo\_d} = A_{homo\_d} + I$ and I is the identity matrix with the same shape as A_{homo_d}, $\tilde{D}$ is the degree matrix of A_{homo_d}. Besides, feature matrix $X_{homo\_d}^{(0)} = X_d$.

Similarly, MIDTI learns the embeddings of drugs from N_{bi} and N_{hete} networks. The output at $(l+1)$ layer is denoted as $X_{bi\_d}^{(l+1)}$ and $X_{hete\_d}^{(l+1)}$, respectively.

Besides, MIDTI learns the embeddings of targets from N_{homo_t}, N_{bi}, and N_{hete} networks in a similar way. The output at $(l+1)$ layer is denoted as $X_{homo\_t}^{(l+1)}$, $X_{bi\_t}^{(l+1)}$, and $X_{hete\_t}^{(l+1)}$, respectively.

Given one drug, MIDTI learns its embeddings from each layer based on each network and stacks them together in a concatenated manner, which is formulated as:

$$x_d = \{x_{homo\_d}^{(1)}, x_{homo\_d}^{(2)}, \ldots, x_{homo\_d}^{(l)}, x_{bi\_d}^{(1)}, x_{bi\_d}^{(2)} \\ , \ldots, x_{bi\_d}^{(l)}, x_{hete\_d}^{(1)}, x_{hete\_d}^{(2)}, \ldots, x_{hete\_d}^{(l)}\} \quad (2)$$

where $x_d \in \mathbb{R}^{3l \times F_m}$. Likewise, the embedding of one target can be represented as:

$$x_t = \{x_{homo\_t}^{(1)}, x_{homo\_t}^{(2)}, \ldots, x_{homo\_t}^{(l)}, x_{bi\_t}^{(1)}, x_{bi\_t}^{(2)} \\ , \ldots, x_{bi\_t}^{(l)}, x_{hete\_t}^{(1)}, x_{hete\_t}^{(2)}, \ldots, x_{hete\_t}^{(l)}\} \quad (3)$$

2.4 Deep interactive attention module
In the deep interactive attention module, MIDTI employs the attention mechanism to further learn the discriminative
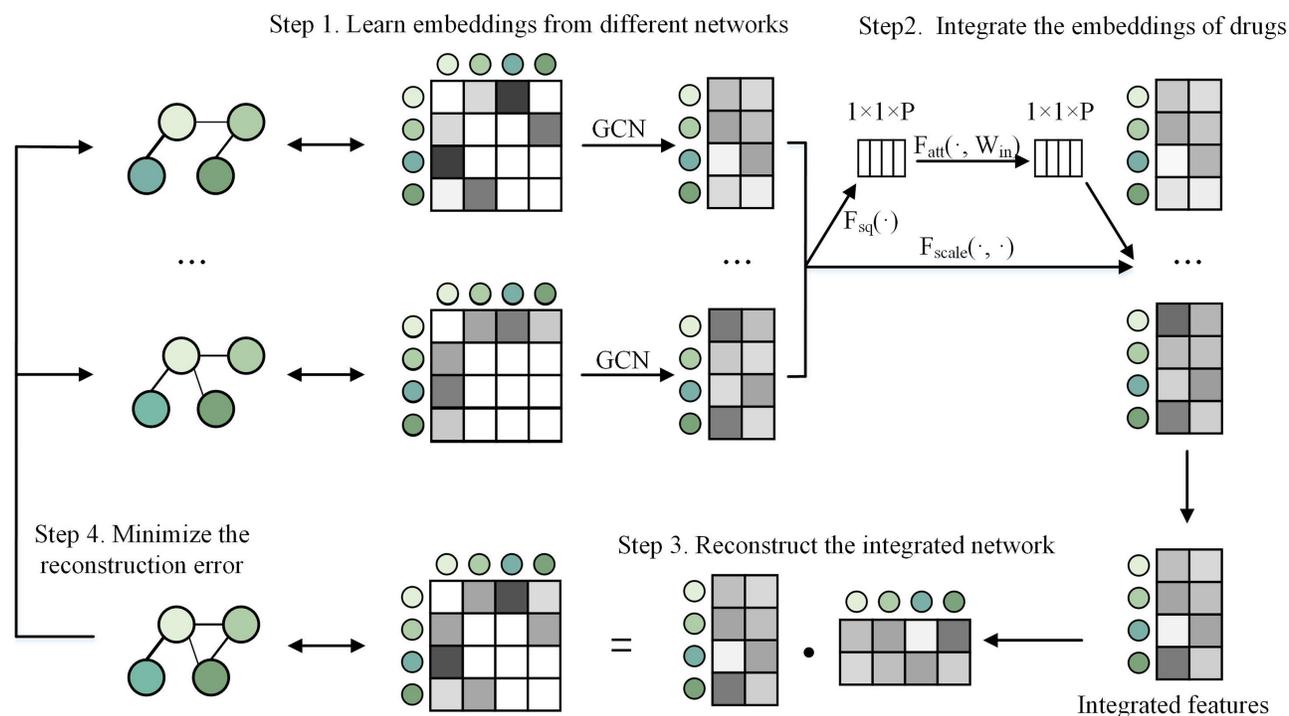
Figure 2. The four steps of multi-view drug similarity network fusion strategy. Step 1: Take different similarity networks of drugs as input and learn the embeddings of drugs from different networks. Step 2: Integrate the embeddings of drugs with the multi-view attention mechanism. Step 3: Reconstruct the integrated drug network through dot product operation on the integrated drug features. Step 4: Train MIDTI by minimizing reconstruction error between the reconstructed network and each original drug similarity network.

representations of drugs and targets. Figure 3 demonstrates the three types of mechanisms in the deep interactive attention module, which are the self-attention (SA) mechanism, drug–target attention (DTA) mechanism, and target-drug attention (TDA) mechanism.

2.4.1 Embedding learning with SA mechanism

Specifically, the SA mechanism takes x_d (x_t) as input and learns the embedding of drugs (targets), which is shown in Fig. 3A. The input of the scaled dot-product attention consists of three matrices: queries, keys and values, which are formulated as follows:

$$Q = Linear_Q(x_{emb}) \quad (4)$$
$$K = Linear_K(x_{emb}) \quad (5)$$
$$V = Linear_V(x_{emb}) \quad (6)$$

where $x_{emb} = x_d$ when feeding drug embeddings and $x_{emb} = x_t$ when feeding target embeddings.

Based on the three matrices, the attention score can be calculated by

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)$$

where $\sqrt{d}$ turns the attention matrix into the standard normal distribution.

To learn the embeddings of drugs and targets from different representation subspaces, the multi-head attention (MHA) mechanism is incorporated into the SA mechanism. The embedding of x_d obtained with the MHA mechanism is

represented as $x_{(d,mha)} \in \mathbb{R}^{3l \times F_m}$, which would be fed into the feed-forward layer and dropout layer.

The residual connection and layer normalization are incorporated to further improve the robustness of MIDTI. Finally, the embedding of drug x_d with SA mechanism is presented as:

$$x_d = Layernorm\left(x_{(d,mha)} + Dropout\left(FL\left(x_{(d,mha)}\right)\right)\right) \quad (8)$$

where $FL(\cdot)$ denotes the feed-forward layer, $Dropout(\cdot)$ denotes the dropout layer, and $Layernorm(\cdot)$ denotes the operation of layer normalization.

2.4.2 Embedding learning with DTA and TDA mechanism

DTA mechanism is designed to estimate the contributions of different parts of a target to the drug (Fig. 3B). To be specific, DTA receives two inputs, which are x_d and x_t. Queries are computed by x_d, and keys and values are obtained by x_t. Hence the embeddings of targets guide the drug representations learning in DTA according to the attention scores.

Analogously, TDA also receives x_d and x_t but measures the effect of drug embedding on learning the target embedding (Fig. 3C). Keys and values are obtained by x_d and queries are computed by x_t.

2.4.3 Embedding learning with deep interactive attention mechanism

We put forward the interactive attention layer which contains SA, DTA, and TDA mechanisms (see Fig. 4). The embeddings of drugs and targets are initially separately fed into the SA layer, and then the TDA mechanism updates the target features with the help of the contributions of drug embeddings,
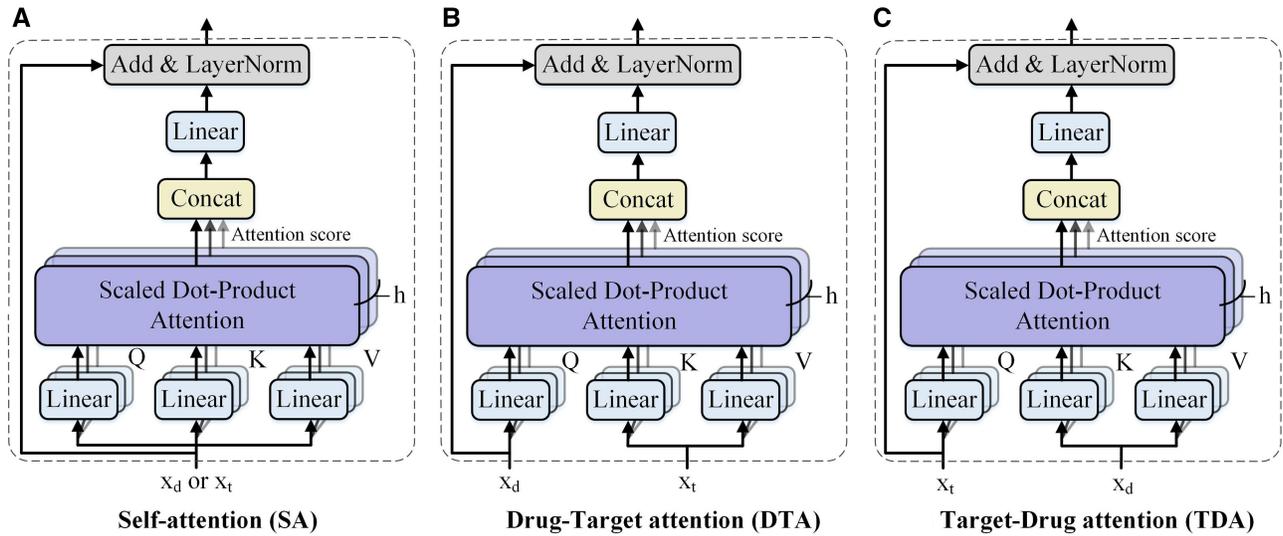
Figure 3. Three types of mechanism in deep interactive attention module. (A) SA mechanism, (B) DTA mechanism, (C) TDA mechanism.
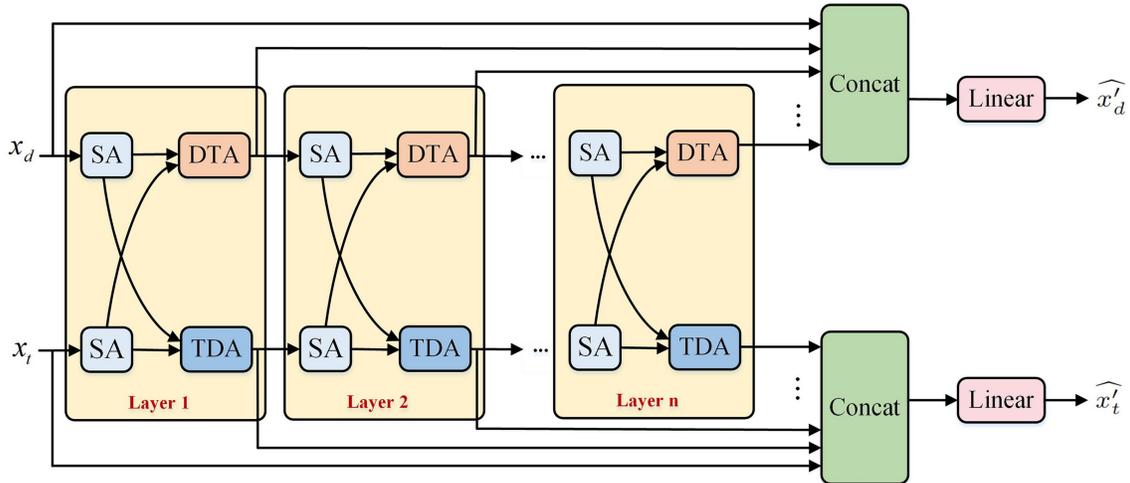

Figure 4. The deep interactive attention mechanism based on a cascade of interactive attention layers. Each interactive attention layer contains the corresponding SA, DTA, and TDA mechanisms respectively.

while the DTA mechanism updates the embedding of drugs with the help of target embeddings.

After the operation on multiple interactive attention layers, we can fully learn the embeddings of drugs and targets. The result of $(n+1)$th interactive attention layer is formulated as:

$$x\prime_d^{(n+1)} = DTA\left(SA\left(x\prime_d^{(n)}\right), SA\left(x\prime_t^{(n)}\right)\right) \quad (9)$$

$$x\prime_t^{(n+1)} = TDA\left(SA\left(x\prime_t^{(n)}\right), SA\left(x\prime_d^{(n)}\right)\right) \quad (10)$$

where $x\prime_d^{(n)}, x\prime_t^{(n)}$ are the representations of drug and target through nth interactive attention layers, $x\prime_d^{(0)} = x_d, x\prime_t^{(0)} = x_t$, $SA(\cdot)$ denotes the operations of the SA mechanism, $DTA(\cdot)$ and $TDA(\cdot)$ denotes the operations of DTA and TDA mechanism respectively.

MIDTI is designed to concatenate the outputs of the 0-st to nth interactive attention layers. Then we transform them

through a linear layer to restore the same dimension as the inputs, which is formulated as:

$$x\prime_d = \left(x\prime_d^{(0)} \| x\prime_d^{(1)} \| x\prime_d^{(2)} \ldots \| x\prime_d^{(n)}\right) W_r \quad (11)$$

$$x\prime_t = \left(x\prime_t^{(1)} \| x\prime_t^{(1)} \| x\prime_t^{(2)} \ldots \| x\prime_t^{(n)}\right) W_s, \quad (12)$$

where $\|$ denotes the concatenation operation, and $W_r, W_s \in \mathbb{R}^{(n+1) \cdot F_m \times F_m}$, and in this study $x\prime_d, x\prime_t \in \mathbb{R}^{3l \times F_m}$.

Finally, we average the above embeddings in the first dimension and obtain the final discriminative representations of drugs and targets, which are denoted as:

$$\hat{x\prime}_d = mean(x\prime_d) \quad (13)$$

$$\hat{x\prime}_t = mean(x\prime_t) \quad (14)$$

where $\hat{x\prime}_d, \hat{x\prime}_t \in \mathbb{R}^{1 \times F_m}$.

2.5 DTI prediction

For the representation $\hat{xl}_d$ and $\hat{xl}_t \in \mathbb{R}^{1 \times F_m}$, we concatenate them to represent the embedding of the drug–target pair and feed it into the MLP decoder, which is formulated as:

$$f = \hat{xl}_d \| \hat{xl}_t \tag{15}$$

$$\hat{y} = \text{Tanh}(W \cdot f + b) \tag{16}$$

where $\|$ denotes the concatenation operation, $W \in \mathbb{R}^{2F_m}$ is the weight matrix, $b \in \mathbb{R}$ is the bias, and $Tanh(\cdot)$ is the activation function. Besides, $\hat{y}$ is the predicted value of DTI, which denotes the interaction probability between the drug and the target.

MIDTI utilizes the cross-entropy loss as the objective function to train MIDTI, and the loss L is minimized as:

$$L = -\frac{1}{N} \sum_i^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{17}$$

where $y_i \in \{0, 1\}$ is the ground truth label, and $\hat{y}_i$ is the predicted label, and N is the number of training samples.

3 Results

3.1 Experimental setup and evaluation metrics

In this study, we conduct the experiments on three datasets, which are Luo's (Luo *et al.* 2017), Yamanishi's (Yamanishi *et al.* 2008), and Zheng's dataset (Zheng *et al.* 2018). For each dataset, we consider all the known DTI pairs as positive samples, and the remained drug–target pairs as negative samples. We initially select positive samples at a ratio of 1:1, 1:5, and 1:10 with negative samples to form three experimental datasets, respectively. Besides, MIDTI adopts the 5-fold cross-validation strategy (Tian *et al.* 2022) to evaluate its performance.

In this study, we adopt Accuracy (ACC), Area Under the receiver operating Characteristic curve (AUC), Area Under the Precision-Recall curve (AUPR), F1 score and Matthews Correlation Coefficient (MCC) as the evaluation metrics. All comparison methods adopt the same 5-fold cross-validation (5-CV) as MIDTI, and the results shown here are the average values of the five-time experiments. The implementation details of the experiments are presented in Supplementary Section S3. The time and space complexity analysis for MIDTI is presented at Supplementary Section S10.

3.2 Comparison with other baseline methods

Ten competitive approaches are selected for comparison with MIDTI and we evaluate them with ACC, AUC and AUPR metrics. They are Random Forests (RF) (Pedregosa *et al.* 2011), Support-Vector Machine (SVM) (Chang and Lin 2011), eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin 2016), GCN (Kipf and Welling 2016), Graph Attention Networks (GAT) (Veličković *et al.* 2017), DTI-CNN (Peng *et al.* 2020), GCNMDA (Long *et al.* 2020), MVGCN (Fu *et al.* 2022), MMGCN (Tang *et al.* 2021), GraphCDA (Dai *et al.* 2022), and DTINet (Luo *et al.* 2017). The description for these comparison approaches is presented in Supplementary Section S5.

The comparison results are shown in Fig. 5 and Table 1, and Fig.5(A) and Fig.5(B) denote the ROC and PR curves respectively. MIDTI wins the best performance among all SOTA methods. Specifically, MIDTI gets the scores on ACC, AUC, and AUPR metrics are 0.9340, 0.9787, and 0.9701. MIDTI is 2.55%, 2.31%, and 2.30%, higher than the ACC of MMGCN, AUC of MMGCN and AUPR of GraphCDA, respectively. Preliminary experiments suggest that MIDTI is the most competitive drug–target association prediction method on this dataset. The comparison results under the 1:1 ratio on Yamanishi's dataset are presented at Supplementary Section S6 in Supplementary Table S1. The comparison
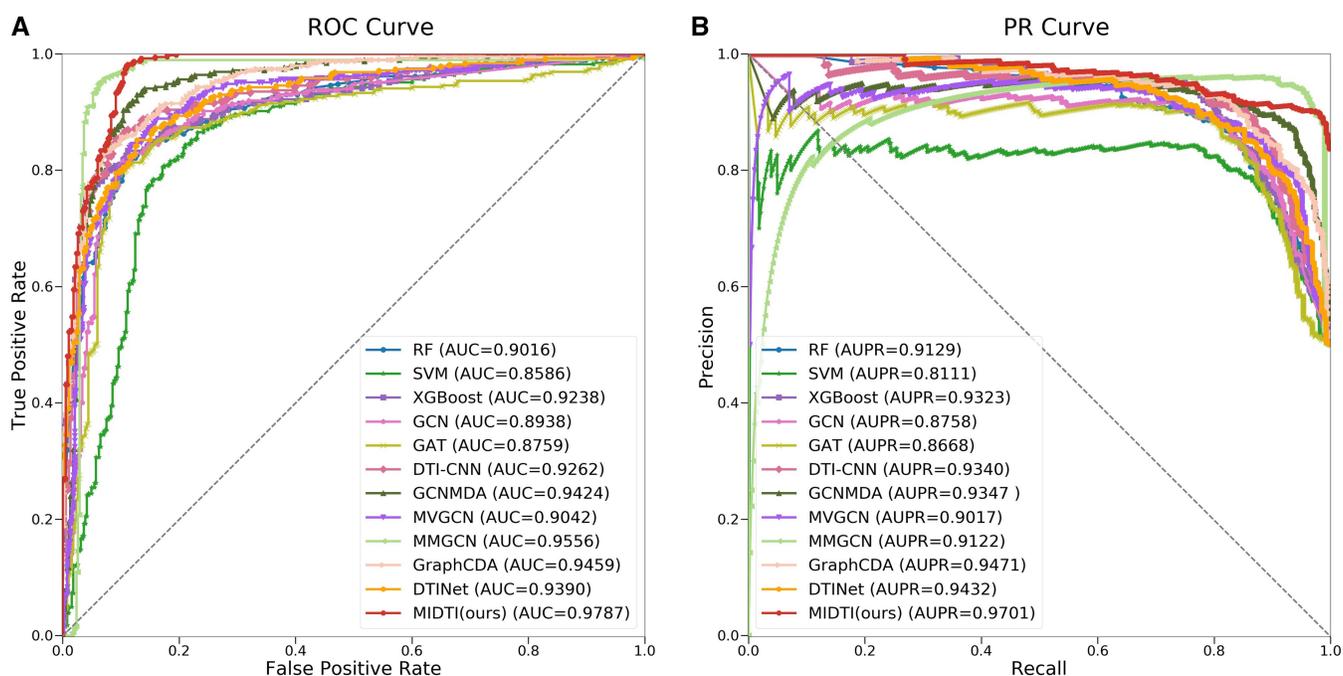
Figure 5. The results of MIDTI as well as other baseline approaches on AUC and AUPR metrics.

Table 1. The performance of MIDTI as well as other baseline approaches for predicting DTI under different ratios on Luo's dataset.[a]

Models	1:1			1:5			1:10		
	ACC	AUC	AUPR	ACC	AUC	AUPR	ACC	AUC	AUPR
RF (Pedregosa *et al.* 2011)	0.8409	0.9016	0.9129	0.9103	0.9093	0.7836	0.9438	0.9176	0.7156
SVM (Chang and Lin 2011)	0.7993	0.8586	0.8111	0.9074	0.8917	0.6962	0.9380	0.8871	0.6078
XGBoost (Chen and Guestrin 2016)	0.8573	0.9238	0.9323	0.7982	0.8586	0.8111	0.9550	0.9311	0.7864
GCN (Kipf and Welling 2016)	0.8393	0.8938	0.8758	0.9068	0.8895	0.7100	0.9299	0.8617	0.5817
GAT (Veličković *et al.* 2017)	0.8219	0.8759	0.8668	0.8710	0.8558	0.6339	0.9268	0.8525	0.5340
DTI-CNN (Peng *et al.* 2020)	0.8523	0.9262	0.9340	0.9269	0.9281	0.8286	<u>0.9558</u>	0.9319	<u>0.7957</u>
GCNMDA (Long *et al.* 2020)	0.8850	0.9424	0.9347	0.9044	0.9354	0.7520	0.9302	0.9423	0.6573
MVGCN (Fu *et al.* 2022)	0.8489	0.9042	0.9017	0.9132	0.9209	0.7777	0.9445	0.9163	0.6959
MMGCN (Tang *et al.* 2021)	<u>0.9085</u>	<u>0.9556</u>	0.9122	<u>0.9403</u>	<u>0.9671</u>	0.8038	**0.9582**	<u>0.9715</u>	0.7684
GraphCDA (Dai *et al.* 2022)	0.8796	0.9459	<u>0.9471</u>	0.9221	0.9484	0.8353	0.9377	0.9133	0.6435
DTINet (Luo *et al.* 2017)	0.8672	0.9390	0.9432	0.8983	0.9017	<u>0.8511</u>	0.9029	0.9003	0.7883
MIDTI (ours)	**0.9340**	**0.9787**	**0.9701**	**0.9413**	**0.9813**	**0.9075**	0.9539	**0.9794**	**0.8431**

[a] The best results are marked in bold and the second best is underlined.

results under the 1:1 ratio on Zheng's dataset are presented at Supplementary Section S6 in Supplementary Table S4.

3.3 Experimental results of MIDTI with different ratios between positive and negative samples

The different ratios of positive samples to negative samples can affect the performance of MIDTI and baseline methods. Hence, we conduct evaluation experiments under positive samples to negative sample ratios of 1:5 and 1:10 on Luo's dataset.

For the results with the 1:5 ratio, MIDTI gets the first rank on ACC, AUC and AUPR, and their scores are 0.9413, 0.9813, and 0.9075, respectively. The scores of MMGCN (rank second) on ACC and AUC are 0.1% and 1.42% lower than those of MIDTI. DTINet is lower than MIDTI by 6.2% on AUPR metric. For the results with the 1:10 ratio, results show that MIDTI also performs best on AUC and AUPR metrics, which are 0.9794 and 0.8431. MMGCN achieves the highest score on ACC metric, which is 0.9582. The other results have already been presented in Table 1.

Besides, the comparison results under the 1:5 and 1:10 ratio on Yamanishi's dataset are presented at Supplementary Section S6 in Supplementary Tables S2 and S3. The comparison results under 1:5 and 1:10 ratio on Zheng's dataset are presented at Supplementary Section S6 in Supplementary Table S4.

3.4 Model ablation study

In order to verify the effectiveness of the essential modules in MIDTI, we conduct three sets of ablation experiments.

The first ablation experiment is to verify the effectiveness of each attention mechanism. MIDTI applies the attention mechanism at two stages: one is the multi-view attention (VA) mechanism to fuse multiple drug-similarity networks and target-similarity networks respectively, and the other is the deep interactive attention (IA) mechanism to learn the embedding of drugs and targets. Here, MIDTI splits VA and IA into four combinations, which are displayed in Table 2.

The results in Table 2 demonstrate that MIDTI achieves the best performance in all five metrics. The values for ACC, AUC, AUPR, F1 and MCC are 0.9340, 0.9787, 0.9701, 0.9370, and 0.8726, respectively. And the performance of MIDTI w/o all attention is worst on AUC and AUPR since it does not employ any attention. Results indicate the deep interactive attention mechanism and the view-attention

Table 2. The ablation experimental results on the view-attention mechanism and interactive attention mechanism for MIDTI.

VA	IA	ACC	AUC	AUPR	F1	MCC
✗	✗	0.9202	0.9547	0.9108	0.9247	0.8464
✓	✗	<u>0.9259</u>	0.9603	0.9291	<u>0.9289</u>	<u>0.8550</u>
✗	✓	0.9184	<u>0.9723</u>	<u>0.9603</u>	0.9212	0.8402
✓	✓	**0.9340**	**0.9787**	**0.9701**	**0.9370**	**0.8726**

The best results are in bold and the second best results are underlined.

Table 3. The ablation experimental results on the homogeneous similarity network, the bipartite network and the drug–target heterogeneous network for MIDTI.

N_{homo}	N_{bi}	N_{hete}	ACC	AUC	AUPR	F1	MCC
✓	✗	✗	0.5637	0.7142	0.6681	0.6137	0.1870
✗	✓	✗	0.9192	0.9767	0.9583	0.9308	0.8701
✗	✗	✓	0.8648	0.9514	0.9380	0.8656	0.7412
✓	✓	✗	0.9229	0.9760	0.9615	<u>0.9313</u>	0.8714
✓	✗	✓	0.8692	0.9479	0.9332	0.8688	0.7399
✗	✓	✓	<u>0.9273</u>	<u>0.9783</u>	<u>0.9688</u>	0.9300	<u>0.8715</u>
✓	✓	✓	**0.9340**	**0.9787**	**0.9701**	**0.9370**	**0.8726**

The best results are in bold and the second best results are underlined.

mechanism both play an essential role in improving the performance of MIDTI.

The second ablation experiment is to verify the effectiveness of the homogeneous similarity network, the bipartite network and the heterogeneous network of drugs and targets, which are denoted as N_{homo}, N_{bi}, N_{hete}, respectively. Specifically, three networks are divided into seven different combinations. The results for each combination are displayed in Table 3.

The results demonstrate that MIDTI achieves the best performance in all five metrics, and the corresponding values are 0.9340, 0.9787, 0.9701, 0.9370, and 0.8726. The results in this set of experiments illustrate that all three similarity networks are essential in learning drug and target embeddings.

The third ablation experiment is to verify the effectiveness of the proposed similarity network fusion strategy. Here, we also select other two similar network fusion strategies, which are called MIDTI_ave and MIDTI_pro. For MIDTI_ave strategy, we measure the arithmetic average values from different networks as the integrated similarity values. For MIDTI_pro

strategy, the integrated similarity value is formulated as $S = 1 - \prod_{i=1}^{n} S_i$, where S_i denotes the similarity values from the ith similarity network. The corresponding results for each combination are displayed in Table 4. The results

Table 4. The evaluation results of MIDTI with different similarity network fusion strategy on Luo's, and Zheng's datasets.

Datasets	Strategy	ACC	AUC	AUPR
	MIDTI_ave	0.9078	0.9547	0.9336
Luo	MIDTI_pro	0.8961	0.9611	0.9501
	MIDTI	**0.9340**	**0.9787**	**0.9701**
	MIDTI_ave	0.8162	0.8789	0.8665
Zheng	MIDTI_pro	0.8048	0.8803	0.8714
	MIDTI	**0.8836**	**0.9546**	**0.9497**

The best results are in bold and the second best results are underlined.

demonstrate that MIDTI achieved the best performance on Luo's dataset and Zheng's dataset, which could confirm the effectiveness of the proposed similarity network fusion strategy. The results on Yamanishi's dataset are displayed in Supplementary Section S8.

3.5 Parameter analysis experiments

In this section, we discuss the sensitivity of several parameters of MIDTI. These parameters mainly include the embedding size, the learning rate of the optimizer, the number of interactive attention heads, the number of GCN layers, the number of interactive attention layers and the number of MLP layers. The corresponding experiment results are all evaluated with ACC, AUC, AUPR, F1, and MCC, respectively.

From the results shown in Supplementary Figs S1 and S2 in Supplementary Section S7, we can see that MIDTI adopts the
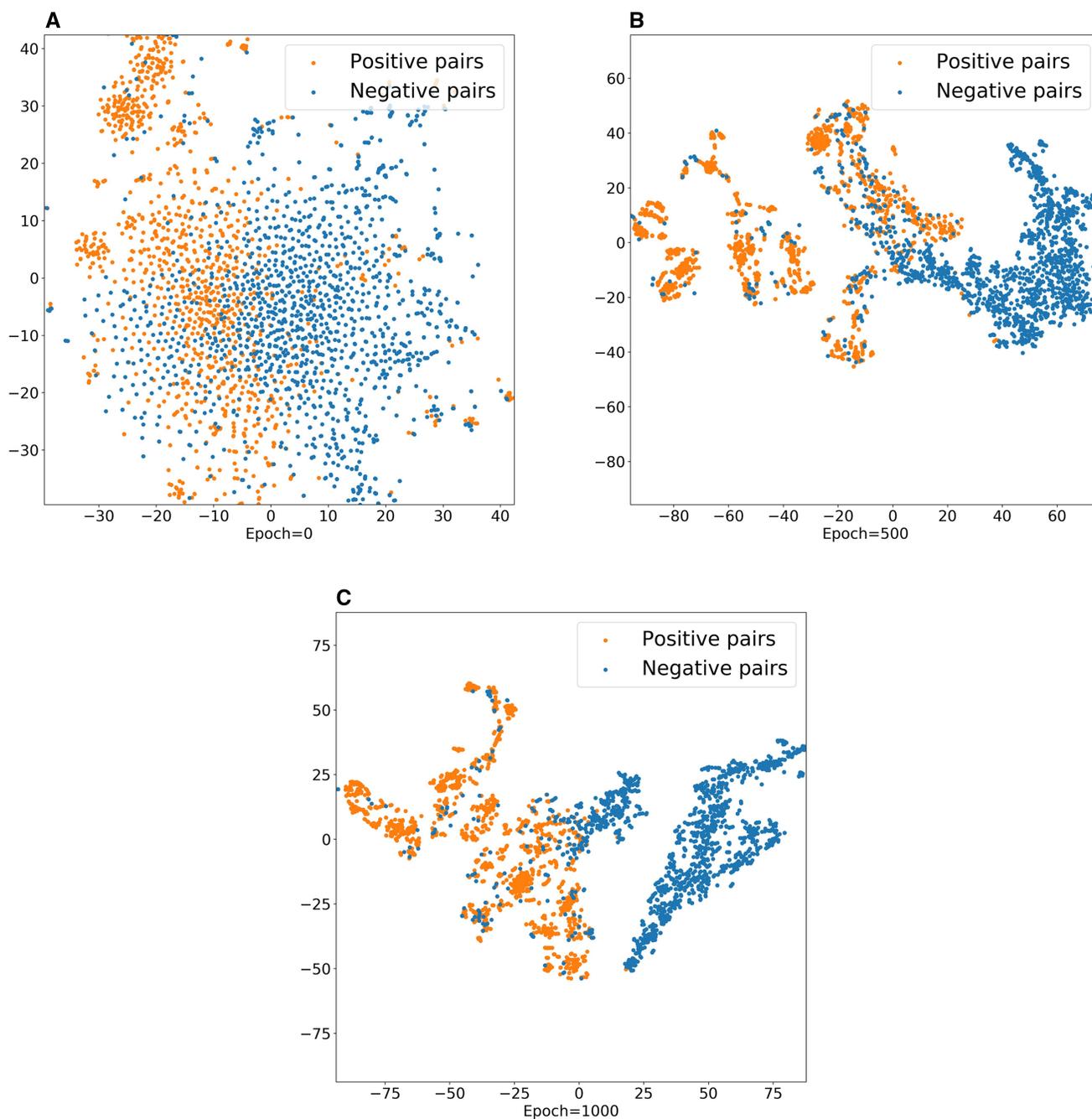

Figure 6. Visualization of the learned drug–target embeddings by MIDTI under different epochs.

embedding size, learning rate, the number of interactive heads, the number of GCN layers, the number of interactive attention layers, the number of MLP layers as 512, 0.1, 8, 3, 3, 3, respectively. A more detailed description of this group experiment is displayed in Supplementary Section S7.

3.6 Visualization and interpretation for the embeddings of drug–target pairs learned by MIDTI

To further demonstrate the ability of MIDTI in learning embeddings of drugs and targets, we conduct the visualization experiment. Specifically, with the learned embeddings of drugs and targets, we can generate embeddings for positive and negative drug–target pairs. All the embeddings of drug–target pairs are plotted into a 2D space using t-SNE tool (Van der Maaten and Hinton 2008). The visualization results are displayed in Fig. 6.

It can be seen that the positive pairs and the negative pairs are gradually distinguished with the epochs increasing (see Fig. 6(A, B, C)). The embeddings of positive pairs and the negative pairs are in chaos when the epoch number is 0. The embedding distribution is gradually clear with the epochs increase. Finally, the positive pairs (yellow points) and the negative pairs (blue points) are almost separated when the number of epochs equals 1000. This observation further confirms that the learned embeddings of drug–target pairs are discriminative and interpretable, which improves the accuracy of MIDTI in predicting DTIs.

3.7 Case study

In practice, discovering the interactions accurately for some common drugs and targets is another effective manner to verify the effectiveness of DTI prediction models (Tian *et al.* 2022). In this section, we mainly conduct two types of experiments, and the corresponding results have been presented in Supplementary Section S9. The predicted targets for selected drugs almost could be verified. The analysis indicates that MIDTI has the powerful ability to discover potential DTIs, which has essential implications for drug screening and drug repositioning.

4 Conclusion

In this study, we propose a novel method called MIDTI to predict DTIs. We conducted extensive experiments to evaluate the performance of MIDTI. The comparison results demonstrate that MIDTI achieves the best performance with different ratios. Besides, the ablation experiments and parameter sensitivity experiments are also performed to further confirm the effectiveness of MIDTI. Finally, the results of the case study are supported by different published databases.

Next, we could work on the following two aspects. Firstly, we can utilize other related data sources of drugs and targets for their embedding learning. Secondly, MIDTI can be applied to other link prediction problems, such as miRNA–disease association prediction.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

References

Bagherian M, Sabeti E, Wang K *et al.* Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief Bioinform* 2021;**22**:247–69.

Chang C-C, Lin C-J. Libsvm: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011;**2**:1–27.

Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785–94.

Dai Q, Liu Z, Wang Z *et al.* Graphcda: a hybrid graph representation learning framework based on gcn and gat for predicting disease-associated circrnas. *Brief Bioinform* 2022;**23**:bbac379.

Ding Y, Tang J, Guo F *et al.* Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization. *Brief Bioinform* 2022;**23**:bbab582.

Forster DT, Li SC, Yashiroda Y *et al.* Bionic: biological network integration using convolutions. *Nat Methods* 2022;**19**:1250–61.

Fu H, Huang F, Liu X *et al.* MVGCN: data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks. *Bioinformatics* 2022;**38**:426–34.

Hinkson IV, Madej B, Stahlberg EA *et al.* Accelerating therapeutics for opportunities in medicine: a paradigm shift in drug discovery. *Front Pharmacol* 2020;**11**:770. page

Keiser MJ, Roth BL, Armbruster BN *et al.* Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;**25**:197–206.

Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv, arXiv:1609.02907, 2016, preprint.

Lee I, Keum J, Nam H *et al.* Deepconv-dti: prediction of drug–target interactions via deep learning with convolution on protein sequences. *PLoS Comput Biol* 2019;**15**:e1007129.

Li M, Cai X, Xu S *et al.* Metapath-aggregated heterogeneous graph neural network for drug–target interaction prediction. *Brief Bioinform* 2023;**24**:bbac578.

Lin X, Li X, Lin X *et al.* A review on applications of computational methods in drug screening and design. *Molecules* 2020;**25**:1375.

Long Y, Wu M, Kwoh CK *et al.* Predicting human microbe–drug associations via graph convolutional network with conditional random field. *Bioinformatics* 2020;**36**:4918–27.

Luo Y, Zhao X, Zhou J *et al.* A network integration approach for drug–target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**:573.

Nguyen T, Le H, Quinn TP *et al.* Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 2021;**37**:1140–7.

Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;**12**:2825–2830.

Peng J, Li J, Shang X *et al.* A learning-based method for drug–target interaction prediction based on feature representation learning and deep neural network. *BMC Bioinformatics* 2020;**21**:394.

Peng J, Wang Y, Guan J *et al.* An end-to-end heterogeneous graph representation learning-based framework for drug–target interaction prediction. *Brief Bioinform* 2021;**22**:bbaa430.

Quan Z, Guo Y, Lin X *et al.* Graphcpi: Graph neural representation learning for compound-protein interaction. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2019, 717–722.

Shaikh N, Sharma M, Garg P *et al*. An improved approach for predicting drug–target interaction: proteochemometrics to molecular docking. *Mol Biosyst* 2016;**12**:1006–14.

Shin B, Park S, Kang K *et al*. Self-attention based molecule representation for predicting drug–target interaction. In: *Machine Learning for Healthcare Conference*. PMLR. 2019, 230–248.

Tang X, Luo J, Shen C *et al*. Multi-view multichannel attention graph convolutional network for mirna–disease association prediction. *Brief Bioinform* 2021;**22**:bbab174.

Tian Z, Peng X, Fang H *et al*. Mhadti: predicting drug–target interactions via multiview heterogeneous information network embedding with hierarchical attention mechanisms. *Brief Bioinform* 2022; **23**:bbac434.

Van der Maaten L, Hinton G. Visualizing data using t-sne. *J Mach Learn Res* 2008;**9**(11).

Veličković P, Cucurull G, Casanova A *et al*. Graph attention networks. *J. stat*, 2017;**1050**(20):10–48550.

Yamanishi Y, Araki M, Gutteridge A *et al*. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**:i232–40.

Yıldırım MA, Goh K I, Cusick M E *et al*. Drug–target network. *Nat Biotechnol* 2007;**25**(10):1119–1127.

Zheng Y, Peng H, Zhang X *et al*. Predicting drug targets from heterogeneous spaces using anchor graph hashing and ensemble learning. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018. 1–7.