Briefings in Bioinformatics, 22(5), 2021, 1–10

https://doi.org/10.1093/bib/bbab119
Method Review

Integrating multi-scale neighbouring topologies and cross-modal similarities for drug–protein interaction prediction

Ping Xuan, Yu Zhang, Hui Cui, Tiangang Zhang, Maozu Guo and Toshiya Nakaguchi

Corresponding author: Maozu Guo, School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture,Beijing 100044, China; E-mail: guomaozu@bucea.edu.cn

Abstract

Motivation: Identifying the proteins that interact with drugs can reduce the cost and time of drug development. Existing computerized methods focus on integrating drug-related and protein-related data from multiple sources to predict candidate drug–target interactions (DTIs). However, multi-scale neighboring node sequences and various kinds of drug and protein similarities are neither fully explored nor considered in decision making.
Results: We propose a drug-target interaction prediction method, DTIP, to encode and integrate multi-scale neighbouring topologies, multiple kinds of similarities, associations, interactions related to drugs and proteins. We firstly construct a three-layer heterogeneous network to represent interactions and associations across drug, protein, and disease nodes. Then a learning framework based on fully-connected autoencoder is proposed to learn the nodes' low-dimensional feature representations within the heterogeneous network. Secondly, multi-scale neighbouring sequences of drug and protein nodes are formulated by random walks. A module based on bidirectional gated recurrent unit is designed to learn the neighbouring sequential information and integrate the low-dimensional features of nodes. Finally, we propose attention mechanisms at feature level, neighbouring topological level and similarity level to learn more informative features, topologies and similarities. The prediction results are obtained by integrating neighbouring topologies, similarities and feature attributes using a multiple layer CNN. Comprehensive experimental results over public dataset demonstrated the effectiveness of our innovative features and modules. Comparison with other state-of-the-art methods and case studies of five drugs further validated DTIP's ability in discovering the potential candidate drug-related proteins.

Ping Xuan, PhD (Harbin Institute of Technology), is a professor at the School of Computer Science and Technology, Heilongjiang University, Harbin, China. Her current research interests include computational biology, complex network analysis and medical image analysis.
Yu Zhang is studying for her master's degree in the School of Computer Science and Technology at Heilongjiang University, Harbin, China. Her research interests include complex network analysis and deep learning.
Hui Cui, PhD (The University of Sydney), is a lecturer at Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia. Her research interests lie in machine learning, medical image analysis, and health informatics.
Tiangang Zhang, PhD (The University of Tokyo), is an associate professor of the School of Mathematical Science, Heilongjiang University, Harbin, China. His current research interests include complex network analysis and computational fluid dynamics.
Maozu Guo, PhD (Harbin Institute of Technology), is a professor at the School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China. His current research interests include bioinformatics, machine learning and image processing.
Toshiya Nakaguchi, PhD (Sophia University), is a professor at the Center for Frontier Medical Engineering, Chiba University, Chiba, Japan. His current research interests include complex network analysis, medical image processing and biometrics measurement.
Submitted: 9 January 2021; **Received (in revised form):** 15 February 2021

Introduction

Drugs usually exert their actions by targeting corresponding proteins. Therefore, revealing underlying drug–target relations plays an essential role in drug development [1–3]. Identifying drug-related proteins contributes to drug development, which in turn reduces the development period and costs [4, 5]. Computational methods have demonstrated the capacity to discover drug-protein interactions and screen reliable candidates for subsequent biological experiments [6–9].

Existing methods for DTIs prediction fall into two main categories: docking-based methods and machine learning-based methods. Docking-based methods are based on the 3D structure information of proteins. These methods, however, are limited by the availability of small part of structure information [10, 11]. Therefore, it is challenging to use docking-based methods for large-scale DTIs prediction [12].

The second category of prediction models is based on machine learning techniques to infer drug-target interactions. For instance, Bleakley *et al*., Keum *et al*. and Ma *et al*. construct a drug-protein bipartite graph and use a support vector machine model (SVM) [13, 14] to predict drug-related proteins. Graph-based methods such as random walk algorithm (RW) [15–17] are also explored to represent the relations between drugs and proteins. RW has been used over the drug similarity network, protein similarity network and drug-protein interaction network and other networks [18, 19]. RW, however, cannot learn the deep feature representations of drug nodes and protein nodes. Besides, the graph attribute information of drug and protein nodes, however, are not considered by existing methods. Ezzat *et al*. and Xuan *et al*. constructed drug–protein heterogeneous networks and established prediction models based on singular value matrix decomposition (SVD) [20] and gradient boosting decision tree (GBDT) [21, 22] respectively, to predict drug-related protein candidates. Nevertheless, those methods are shallow prediction models that cannot fully extract deep and complex associations between drug and protein nodes.

Recently, deep learning methods have been explored to integrate drug-related and protein-related information to identify potential protein candidates that interacted with drugs more accurately. For instance, a combined Long Short-Term Memory (LSTM) networks and convolutional neural network (CNN) [23] model is proposed for DTIs prediction. However, the method is not effective on multi-scale neighbor node sequences. Manoochehri *et al*. and Torng *et al*. proposed a graph convolutional network (GCN)-based prediction model [24, 25]. Sun *et al*. constructed a graph convolutional autoencoder and generative adversarial network-based method to predict interactions between drugs and proteins [26]. These three models, however, did not consider multiple kinds of drug similarities and protein similarities. The motivations behind those methods are that drug similarities and protein similarities can reflect intra-relations among the same types of nodes from different perspectives as drug nodes and protein nodes. However, diverse drug-related and protein-related similarities are neither fully extracted nor considered by existing methods. Besides, when using multi-layer heterogeneous networks to embed drug-protein-disease, the multi-scale neighboring node sequences formulated by random walking over the network are also important auxiliary information.

In this work, we propose a prediction model, DTIP, to learn and integrate pairwise representations of neighbour topology and attributes from multiple drug and protein related data sources. The unique contributions are summarized as below.

- We propose a new feature encoding module to associate various features of the drug, protein, and disease nodes, and model their low-dimensional feature representations. The module has a novel feature-level attention mechanism to extract and integrate three types of nodes' informative features.
- A drug-protein–disease heterogeneous network is constructed, which is composed of multiple kinds of edges to reflect and embed the interactions and associations among three types of nodes. The network benefits the generation of neighboring topological sequences by RW given multiple scales of drug and protein nodes. To learn the sequential information among the nodes within a neighbor node sequence, we propose a bidirectional gated recurrent unit (Bi-GRU) based strategy. The extracted multiple neighboring topological representations from drug and protein nodes are finally adaptively fused by representation-level attention mechanism.
- Due to various contributions from multiple kinds of drug similarities and protein similarities to DTIs prediction, we further define a similarity-level attention mechanism to discriminate the contributions and achieve pairwise attribute embedding.
- Finally, we design a multiple-layer convolution neural network to deeply integrate topological representations and attribute embeddings of drug–protein pairs. The improved prediction results are demonstrated by comprehensive evaluations of the innovations, comparison with recent prediction models, and case studies on five drugs.

Materials and methods

To predict candidate proteins associated with a given drug, we propose a drug-target interaction prediction model, which is composed of two branches as shown in Figure 1. In the first branch, we learn low-dimensional representations of drug, protein and disease nodes based on fully connected autoencoders. Afterwards, feature representations of various nodes and heterogenous drug–protein–disease topology are integrated by graph convolutional neural network to learn neighboring topological representations of drug and protein nodes. For the second branch, a novel similarity-level attention mechanism is proposed to integrate the attribute embedding of pairs of drug and target nodes. Finally, neighboring topology representation and attribute embedding of a drug-protein pair are incorporated by convolutional and fully connected layers. The output probabilities are considered as the drug-protein interaction scores. The higher the score, the more likely they will interact with each other [27, 28].

Dataset

We extract drug similarities, drug–disease associations, drug–protein interactions, protein–disease associations and protein similarities from public databases and published articles for DTIs prediction. 1923 known DTIs are from published papers [19], which include 708 drugs from DrugBank database 3.0 [29] and 1512 proteins from HPRD database 9.0 [30]. We extract 199 214 pairs of drug-disease associations and 15 96 745 pairs of target-disease associations from the Comparative Toxicogenomics Database [31], which contains 5603 diseases except drugs and proteins. Drug similarities and protein similarities are calculated based on chemical substructure and sequence, respectively.

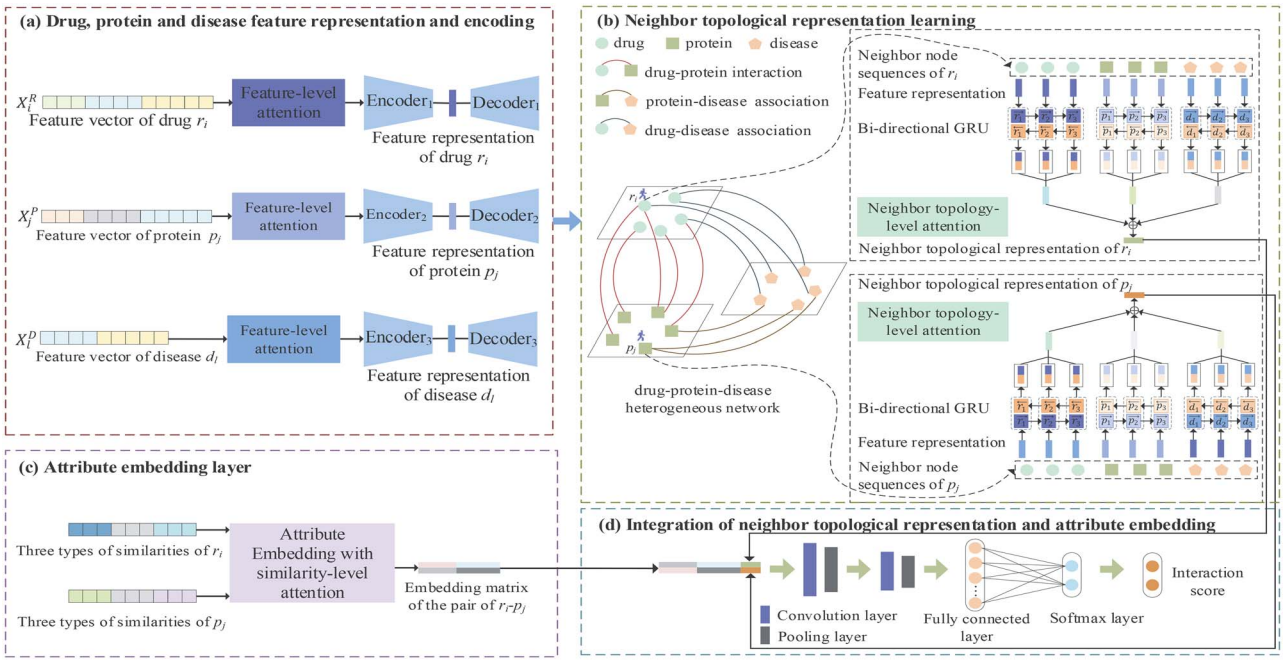

Figure 1. Framework of the proposed DTI model. **(A)** feature-level attention for drug-related, protein-related and disease-related more informative features, and encode their low-dimensional feature representation, respectively **(B)** neighbor representation of drug and protein node by random walk and bidirectional GRU to integrate r_i and p_j corresponding neighbor low-dimensional feature representation **(C)** embedding matrix of r_i-p_j by an attribute embedding mechanism **(D)** integration of neighbor topological representation and attribute embedding for final drug–protein interaction prediction.

Multi-source data representation and calculation

Representation of drug similarities, drug-protein interactions and drug-disease associations

Two drugs may have similar chemical substructures which are calculated by Tanimoto coefficient [32]. Given matrix $A^{drug-similarity} \in \mathbb{R}^{N_r \times N_r}$ representing similarities of chemical substructure among N_r drugs, where the i-th row of $A^{drug-similarity}$ represent the coefficient between r_i and N_r drugs, and i-th row also represents the feature vector of r_i at the chemical substructure level. Let matrix $A^{drug-protein} \in \mathbb{R}^{N_r \times N_p}$ represent interactions between N_r drugs and N_p proteins, each row of $A^{drug-protein}$ corresponds to a drug and each column corresponds to a protein, where $A_{ij}^{drug-protein}$ is 1 if drug r_i has been observed with interaction with protein p_j, and or 0 otherwise. And the entire row where r_i locates represents the feature vector of r_i at the drug–protein interaction level. The associations between N_r drugs and N_d diseases are represented by matrix $A^{drug-disease} \in \mathbb{R}^{N_r \times N_d}$, where $A_{ij}^{drug-disease}$ is the corresponding associated value where the row of drug r_i and the column of disease d_j locates. $A_{ij}^{drug-disease}$ is set as 1 if r_i is associated with d_j according to Figure 2, and 0 otherwise. The feature vector of r_i located in $A^{drug-disease}$ is represented by the association value between r_i and N_d diseases at the level of drug-disease associations. In summary, $A^{drug-similarity}$, $A^{drug-protein}$ and $A^{drug-disease}$ are feature matrices of N_r drugs from three different perspectives, where the feature vectors at i-th rows of three matrices represent different levels of feature representation. Thus, they are concatenated as a drug feature matrix $X^R \in \mathbb{R}^{N_r \times (N_r+N_p+N_d)}$, where the data in the i-th row of represents the feature vector of r_i.

Representation of protein-drug interactions, protein similarities and protein-disease associations

The interactions between N_p proteins and N_r drugs are represented by matrix $B^{protein-drug} \in \mathbb{R}^{N_p \times N_r}$, where each row of $B^{protein-drug}$ represents a protein, and each column represents a drug. If protein p_i interacts with drug r_j, $B_{ij}^{protein-drug}$ is 1, otherwise 0. The feature vector of p_i is represented by the i-th row in matrix $B^{protein-drug}$. $B^{protein-similarity} \in \mathbb{R}^{N_p \times N_p}$ records the sequence similarities among N_p proteins, where the similarity is calculated by Smith–Waterman score [33]. $B_{ij}^{protein-similarity}$ represents the similarity between i-th protein and j-th protein. The vector in the i-th row of $B^{protein-similarity}$ represents the feature of protein p_i at the similarity level. There are N_p proteins and N_d diseases in matrix $B^{protein-disease} \in \mathbb{R}^{N_p \times N_d}$. If protein p_i and disease d_j are associated with each other, the association value is 1, otherwise 0. And the i-th row in $B^{protein-disease}$ represents the feature of p_i. $B^{protein-drug}$, $B^{protein-similarity}$ and $B^{protein-disease}$ are concatenated together to get a new protein matrix $X^P \in \mathbb{R}^{N_p \times (N_r+N_p+N_d)}$.

Representation of disease–drug associations and disease–protein associations

The matrix $C^{disease-drug} \in \mathbb{R}^{N_d \times N_r}$ denotes associations between diseases and drugs, which recodes associations between N_d disease and N_r drugs. The confirmed disease-drug associations and the unconfirmed are represented by 1 and 0, respectively. The feature vector of the disease d_i are represented by the i-th row in $C^{disease-drug}$. $C^{disease-protein} \in \mathbb{R}^{N_d \times N_p}$ is an association matrix which covers N_d disease nodes and N_p protein nodes. If d_i has confirmed association with p_j, $C_{ij}^{disease-protein}$ is 1, otherwise, 0. A new disease feature matrix $X^D \in \mathbb{R}^{N_d \times (N_r+N_p)}$ is obtained by joining $C^{disease-drug}$ and $C^{disease-protein}$ head to tail.

Attention extraction at feature level

Given feature matrices of drugs, proteins and diseases X^R, X^P and X^D, taking drug r_i as an example, X_i^R contains similarities between r_i and drugs $r_1 \ldots r_{N_r}$, interactions between r_i and protein $p_1 \ldots p_{N_p}$, and associations between r_i and disease $d_1 \ldots d_{N_d}$. However, each

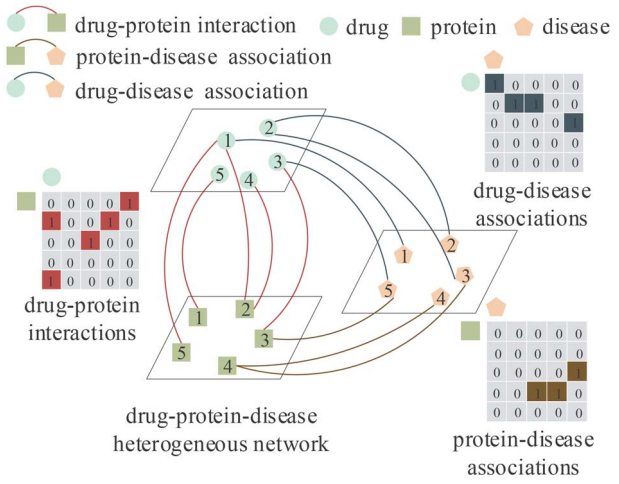

Figure 2. Three-layer heterogeneous network composed of drug–protein interactions, drug–disease associations and protein–disease associations.

feature in X_i^R may contribute with various weights to DTIs prediction. Thus, we propose an attention mechanism at feature level to capture informative features. The attention score of each feature belonging to r_i is defined as:

$$s_i^{att} = H^{att} tanh\left(W_X^{att} X_i^R + b^{att}\right), \tag{1}$$

where $s_i^{att} = \{s_{i,1}, s_{i,2}, ..., s_{i,(N_r+N_p+N_d)}\}$ is a score vector of each feature in X_i^R. H^{att} and W_X^{att} are weight matrices to be learnt during training process, and b^{att} is a bias vector. The normalized attention score α_{ij}^{att} for each feature in X_i^R is represented by

$$\alpha_{ij}^{att} = \frac{exp\left(s_{ij}^{att}\right)}{\sum_k exp\left(s_{ik}^{att}\right)}. \tag{2}$$

Thus, each potential feature of y_i can be denoted as follows,

$$y_i = \alpha_i^{att} \circ X_i^R, \tag{3}$$

where '$\circ$' is an element-wise product operator. We also perform feature-level attention extraction on each drug feature vector in X^R to generate attention enhanced drug feature matrix $Y^R \in \mathbb{R}^{N_r \times (N_r+N_p+N_d)}$. Each feature vector of a protein and a disease node takes the same operations to generate enhanced protein matrix $Y^P \in \mathbb{R}^{N_p \times (N_r+N_p+N_d)}$ and disease matrix $Y^D \in \mathbb{R}^{N_d \times (N_r+N_p)}$, respectively.

Drug, protein and disease feature representation and encoding based on fully connected autoencoder

Due to high dimensions of Y^R, Y^P and Y^D, we perform dimension reduction based on fully connected autoencoder to obtain dense and low-dimensional features representations of drug, protein and disease: $F^R \in \mathbb{R}^{N_r \times n_f}$, $F^P \in \mathbb{R}^{N_p \times n_f}$, $F^D \in \mathbb{R}^{N_d \times n_f}$, where $n_f < N_r < N_p < N_d$, and n_f is the dimension of feature vector after dimensionality reduction. The fully connected autoencoder consists of encoding and decoding parts, where each of them is composed of two neural network layers. Since the encoding and decoding processes with drugs, proteins and diseases are similar, we use drugs as an example to explain the process.

Encoder. The encoder is composed of two hidden layers, where each of them contains a linear layer and a nonlinear activation function. The drug feature matrix Y_R is fed into the first hidden layer as input. The representations of two hidden layers with the encoding module $H_{encoder}^{(1)}$, $H_{encoder}^{(2)}$ are defined as follows,

$$H_{encoder}^{(1)} = \sigma\left(W_{encoder}^{(1)} Y^R + B_{encoder}^{(1)}\right), \tag{4}$$

$$H_{encoder}^{(2)} = \sigma\left(W_{encoder}^{(2)} H_{encoder}^{(1)} + B_{encoder}^{(2)}\right), \tag{5}$$

where $W_{encoder}^{(1)}$ and $W_{encoder}^{(2)}$ are weight matrices of the first and second layer respectively, and $B_{encoder}^{(1)}$ and $B_{encoder}^{(2)}$ are corresponding bias matrices. σ is the activation function Relu [34].

Decoder. The essence of decoder is to upgrade the dimensions of learnt features. The decoder is composed of two neural networks. Taking $H_{encoder}^{(2)}$ as the input of the decoding module, its corresponding hidden layer representation is defined as follows,

$$H_{decoder}^{(1)} = \sigma\left(W_{decoder}^{(1)} H_{encoder}^{(2)} + B_{decoder}^{(1)}\right), \tag{6}$$

$$H_{decoder}^{(2)} = \sigma\left(W_{decoder}^{(2)} H_{decoder}^{(1)} + B_{decoder}^{(2)}\right), \tag{7}$$

where $W_{decoder}^{(1)}$ and $W_{decoder}^{(2)}$ are weight matrices of decoder, and $B_{decoder}^{(1)}$ and $B_{decoder}^{(2)}$ are corresponding bias matrices.

Our learning framework's optimisation goal based on fully connected autoencoder is to make $H_{decoder}^{(2)}$ consistent with input Y^R as much as possible. The loss is defined as,

$$loss_1 = \frac{1}{T_r}\sum_1^{T_r}\left(Y_i^R - H_{decoder,i}^{(2)}\right)^2, \tag{8}$$

where $H_{decoder}^{(2)}$ is the output of decoder, and Y^R is the input of encoder, Y_i^R is the corresponding training sample (feature vector of drug r_i), T_r is the number of training samples (number of drugs). The loss function (8) is optimized by Adam algorithm [35]. The autoencoder is trained by the standard backpropagation algorithm [36]. With iteration algorithm, $H_{decoder,i}^{(2)}$ is recorded as the low-dimensional feature representation of the drug r_i. Let $H_{decoder}^{(2)} = F^R$, $F^R \in \mathbb{R}^{N_r \times n_f}$ is the low-dimensional feature matrix of N_r drugs. Similarly, we can get the low-dimensional feature matrix of protein as $F^P \in \mathbb{R}^{N_p \times n_f}$ and that of disease as $F^D \in \mathbb{R}^{N_d \times n_f}$.

Neighboring topological representation

Three-layer heterogeneous network constructions

We constructed a three-layer heterogeneous network $G = (V, E)$ as shown in Figure 2. Node set $V = \{V^R \cup V^P \cup V^D\}$ consists of three types of nodes: drug nodes V^R, protein nodes V^P and disease nodes V^D. The edge set $E = \{E^{R-P} \cup E^{R-D} \cup E^{P-D}\}$ contains drug–protein interactions, drug–disease associations and protein–disease associations. Edge $e_{ij} \in E$ exists if there is a connection between node $v_i \in V$ and node $v_j \in V$. The i-th row of F^R represents the feature vector of drug node r_i in G. Similarly, the feature vectors of protein node p_j and disease node d_l are represented by the j-th row and the l-th row in F^P

and F^D, respectively. For an unknown drug–protein pair in G, if the drug and protein are connected to a common drug node, protein node or disease node, there are high probabilities that the drug–protein pair interacts with each other. To extract such relations between drug and protein nodes in the heterogeneous network, we formulate neighbour node sequences of each drug and protein node by random walk with restart (RWR).

Multi-scale neighbor node sequence extraction by RWR

The random walker can start from any one of the drug and protein node in the heterogeneous network. For instance, taking r_i as the starting point, a restart probability p is randomly generated. If p is greater than the threshold 0.5, a node, which is directly connected to r_i in the network is selected as r_i's neighbor node, and r_{neigh} is set as a new starting node. Otherwise, the walker returns to the starting node r_i. The above process is repeated until converge, resulting in N_h drug neighbor nodes, N_h target neighbor nodes and N_h disease neighbor nodes of r_i. It is noted that neighbor nodes can appear repeatedly. Afterwards, the obtained neighbor nodes of r_i are sorted according to the occurence frequency. The top $N_s(N_s<N_h)$ ranked nodes are preserved as the final neighbor node sequences of r_i, thus forming it's multi-scale neighbor node sequences. r_i's drug multi-scale neighbor node sequences are denoted by T_R, protein multi-scale neighbor node sequences are presented by T_P, and disease multi-scale neighbor node sequences are denoted by T_D.

Representation of the same type neighbor nodes based on Bi-GRU

To effectively combine the node features and network topology in G, we propose a Bi-GRU based approach to represent neighbor node sequences. Since neighbour node sequences of drug, protein and disease nodes have similar ways to represent information, we take the multi-scale neighbor node sequences of r_i as an example to describe the process. For r_i, connections exist in its multi-scale neighbor sequences of drug nodes, as well as protein nodes and disease nodes. Therefore, the neighbor nodes set T_R of r_i is fed into the Bi-GRU, which consists of forward GRU ($\overrightarrow{GRU}$) and backward GRU ($\overleftarrow{GRU}$). $\overrightarrow{GRU}$ deals with the information of r_i from the first to the N_s-th neighbor sequences of drug nodes, and $\overleftarrow{GRU}$ deals with neighbor sequences from the N_s-th to the first one, where each neighbor of drug node is represented as $v \in T_R$. $\overrightarrow{GRU}$ and $\overleftarrow{GRU}$ are combined in the hidden layer together to form the representation of neighbor set of drug nodes as $Z_1^r(r_i)$,

$$Z_1^r(r_i) = \frac{\sum_{v \in T_R} \left[\overrightarrow{GRU}\left\{F_v^R\right\} \oplus \overleftarrow{GRU}\left\{F_v^R\right\}\right]}{N_s}, \tag{9}$$

where $Z_1^r(r_i) \in \mathbb{R}^{N_F \times 1}$, $\overrightarrow{GRU}\left\{F_v^R\right\} \in \mathbb{R}^{N_F/2 \times 1}$, $\overleftarrow{GRU}\left\{F_v^R\right\} \in \mathbb{R}^{N_F/2 \times 1}$, N_F is the dimension of the neighbor vector of drug. The operator $\oplus$ denotes concatenation. Similarly, the representation of r_i's protein neighbor and disease neighbor can be denoted as $Z_1^p(r_i)$ and $Z_1^d(r_i)$.

Integration of the representation among different neighbor nodes based on neighbor topological attention

Since $Z_1^r(r_i)$, $Z_1^p(r_i)$ and $Z_1^d(r_i)$ of r_i have different contributions to the drug–target interaction prediction, we propose attention mechanism at neighbor topology level to get the enhanced representation of neighbor topology of r_i. The attention score at the level of neighbor topology is s_t^{nei},

$$s_t^{nei} = h^{nei} tanh\left(W_t^{nei} Z_1^t(r_i) + b^{nei}\right), \tag{10}$$

where $t \in \{r, p, d\}$, W_t^{nei} is the weight matrix, b^{nei} and h^{nei} are the bias vector and weight vector, respectively. $\alpha_{r_i,t}^{nei}$ is the normalized attention weight,

$$\alpha_{r_i,t}^{nei} = \frac{exp\left(s_t^{nei}\right)}{\sum_{j \in 3} exp\left(s_j^{nei}\right)}, \tag{11}$$

where *exp* is an exponential function. The representation of neighbor topology of r_i which is enhanced by the attention mechanism is denoted as $Z_2(r_i)$,

$$Z_2(r_i) = \sum\nolimits_{t \in \{r,p,d\}} \alpha_{r_i,t}^{nei} Z_1^r(r_i), \tag{12}$$

where $Z_2^r(r_i) \in \mathbb{R}^{N_F \times 1}$. Pairwise representation of neighbor topology $M_{ij} \in \mathbb{R}^{2 \times N_F}$ is finally obtained by concatenating $Z_2(r_i)$ and $Z_2(p_j)$ learned in the graph neural network.

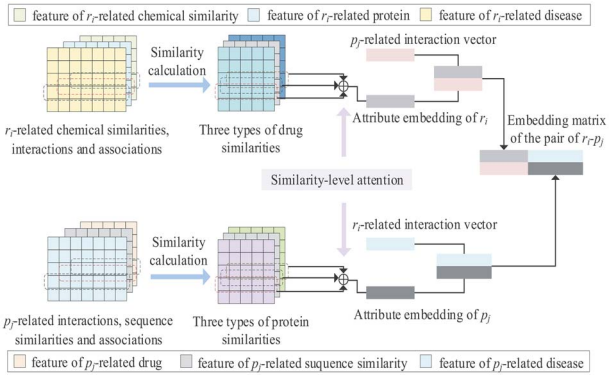

Figure 3. An illustration of the proposed attribute embedding between a pair of drug and protein node.

Attribute embedding of multiple similarities

In the second branch of the prediction model as shown in Figure 3, we firstly calculate multiple drug similarities and protein similarities. The calculated drug chemical structure similarities, drug–protein interactions-based similarity, drug–disease associations-based similarity, protein–drug interactions-based similarity, protein sequence and protein-disease associations-based similarity are provided in supplementary files SF1-SF6. Then a CNN based module is constructed to learn the attribute embedding of a pair of drug-protein.

Based on the biological premise that similar drugs are more likely to interact with similar proteins, the similarity between two drugs is measured via the interacted proteins. For example, if r_i interacts with p_2, p_4, p_5 and r_j interacts with p_3, p_5, p_6, the similarity between $M_1 = \{p_2, p_4, p_5\}$ and $M_2 = \{p_3, p_5, p_6\}$ is considered as the similarity between r_i and r_j. The similarity between r_i and r_j, $R_{ij}^{(2)}$ is calculated by cosine similarity [21] as

$$R_{ij}^{(2)} = \frac{A_i^{drug-protein} \cdot A_j^{drug-protein}}{\| A_i^{drug-protein} \| \times \| A_j^{drug-protein} \|}, \tag{13}$$

where $R^{(2)} \in \mathbb{R}^{N_r \times N_r}$ is the drug matrix, values in $R_{ij}^{(2)}$ vary from 0 to 1. Similarly, the other two drug similarity matrices $R^{(1)} \in \mathbb{R}^{N_r \times N_r}$ and $R^{(3)} \in \mathbb{R}^{N_r \times N_r}$ are calculated based on the drug chemical substructure and drug-related disease, respectively. For protein nodes, the protein similarity matrices $P^{(1)} \in \mathbb{R}^{N_p \times N_p}$, $P^{(2)} \in \mathbb{R}^{N_p \times N_p}$ and $P^{(3)} \in \mathbb{R}^{N_p \times N_p}$ are calculated based on the protein–drug interaction matrix $B^{protein-drug}$, protein sequence similarity matrix $B^{protein-similarity}$ and protein–disease association matrix $B^{protein-disease}$. Multiple types of drug similarities (or protein similarities) may be regarded as the multi-modal similarities of drugs (or protein).

$R_i^{(1)}$, $R_i^{(2)}$ and $R_i^{(3)}$ reflect the similarities among drug nodes from different levels, which contribute differently to attribute embedding of r_i. Therefore, we perform attention mechanism at the similarity level to obtain the attribute embedding of r_i. The attention score s_i^{attr} and the corresponding normalized attention weight β_i^{attr} are calculated as follows,

$$s_i^{attr} = h^{attr} tanh\left(W_{mat}^{attr} R_i^{(mat)} + b^{attr}\right), \tag{14}$$

$$\beta_i^{attr} = \frac{exp\left(s_i^{attr}\right)}{\sum_{j \in 3} exp\left(s_j^{attr}\right)}, \tag{15}$$

where $mat \in \{1, 2, 3\}$, W_{mat}^{attr} and h^{attr} are weight matrix and weight vector, respectively, b^{attr} is a bias vector. The attribute embedding vector of r_i is g_i,

$$g_i = \sum\nolimits_{mat \in \{1,2,3\}} \beta_i^{attr} R_i^{(mat)}. \tag{16}$$

Similarly, we can obtain p_j attribute embedding vector as v_j. If r_i and p_j have the similarity and interaction connections with more common drugs, then they are more likely to interact with each other. Therefore, we stack g_i and $B_j^{protein-drug}$ up and down to get the matrix $X^{(1)} \in \mathbb{R}^{2 \times N_r}$. In the same way, there is a possible interaction between r_i and p_j when they have the interaction and similarity connections with common proteins, thus $A_i^{drug-protein}$ and v_j are stacked up and down to get $X^{(2)} \in \mathbb{R}^{2 \times N_p}$. We concatenate $X^{(1)}$ and $X^{(2)}$ left and right to obtain the pairwise r_i-p_j embedding matrix $X_{ij} \in \mathbb{R}^{2 \times (N_r + N_p)}$.

Integration of neighbor topological representation and attribute embedding based on CNN

The representation of neighbor topology of r_i-p_j learned from the first branch and the attribute embedding from the second branch are combined to get the final r_i-p_j pairwise representation $f \in \mathbb{R}^{2 \times (N_F + N_r + N_p)}$. Then the interaction score of a pair of r_i-p_j can be obtained by CNN.

The CNN module consists of two convolution layers, two pooling layers and a fully connected layer. For convolution layer, we set the filter's length and width as l_c and w_c, respectively, which means that a filter will act on the region composed of $l_c \times w_c$ features. To learn the marginal information of f, we perform zero padding. The number of filters is n_{conv}. $f_{m,n}$ is the element in the m-th row and the n-th column of f, and $f_{q,m,n}$ means that the q-th filter slides on $f_{m,n}$,

$$f_{q,m,n} = f\left(m : m + w_c, n : n + l_c\right), \tag{17}$$

where $m \in [1, 4 - w_c + 1]$, $n \in [1, 2 + (N_F + N_r + N_p) - l_c + 1]$, $q \in [1, n_{conv}]$. The filter $W_{q,m,n}$ slides on $f_{m,n}$ to get the element value $Z_q(m, n)$ at the q-th feature map Z_q as

$$Z_q(m, n) = \sigma\left(W_{q,m,n} f_{q,m,n} + b(q)\right), \tag{18}$$

where σ is the Relu activation function, and b is the bias item. $Z_q(m, n)$ is an element in the m-th row and the n-th column of the q-th feature map.

The max-pooling layer selects more important feature in the q-th feature map. The length and width of the filter in max-pooling layer are l_p and w_p, respectively. Value $P_{q,m,n}$ is in the m-th row and n-th column of the q-th feature, which is calculated as:

$$P_{q,m,n} = \max\left(Z_q\left(m : m + w_p, n : n + l_p\right)\right) \tag{19}$$

The kernel size and number of filters are set as 2x2 and 16 in the first convolutional layer in CNN module. The second convolutional layer has 32 filters, and kernel size=2x2, stride=1 in both first and second convolutional layers. The kernel size, stride and zero padding are set as 2×2, 1 and 0 in both pooling layers.

After two convolution layers and two max-pooling layers, we obtain the output vector z_f. z_f is then fed into a fully connected layer and a softmax layer [37] to get the r_i-p_j interaction score,

$$score = softmax\left(W_f z_f + b_f\right), \tag{20}$$

where W_f and b_f are weight matrix and bias vector, respectively. *score* is the probability distribution of interaction classified by C ($C = 2$), which includes the probability of the interaction between the drug and the protein and the possibility of no interaction.

Loss of interaction prediction

In our model, loss function is defined as the cross-entropy between the ground truth distribution and the predicted distribution of the drug-target interaction score as

$$loss = -\sum\nolimits_{i \in T} \sum\nolimits_{j=1}^{C} z_j \log p_j, \tag{21}$$

where $i \in T$, T is the set of training samples. z_j is the probability of the drug-protein interaction. z_j is 1 if the drug indeed interacts with the protein, and 0 otherwise.

Experimental evaluations and discussions

Evaluation metrics

We use five-fold cross-validation to evaluate the performance of DTIP. All known drug-protein interactions are regarded as positive samples and randomly divided into five subsets, where four of them are for training and the remaining for testing. All unobserved drug–protein interactions are taken as negative samples. We randomly select negative samples whose number is equal to positive samples for training, and the remaining negative samples for testing.

Evaluation metrics include positive rates (TPRs), false-positive rates (FPRs), area under the receiver operating characteristic (ROC) curve (AUC) [38], the area under the precision-recall curve (AUPR) [39]. When the predicted interaction score of

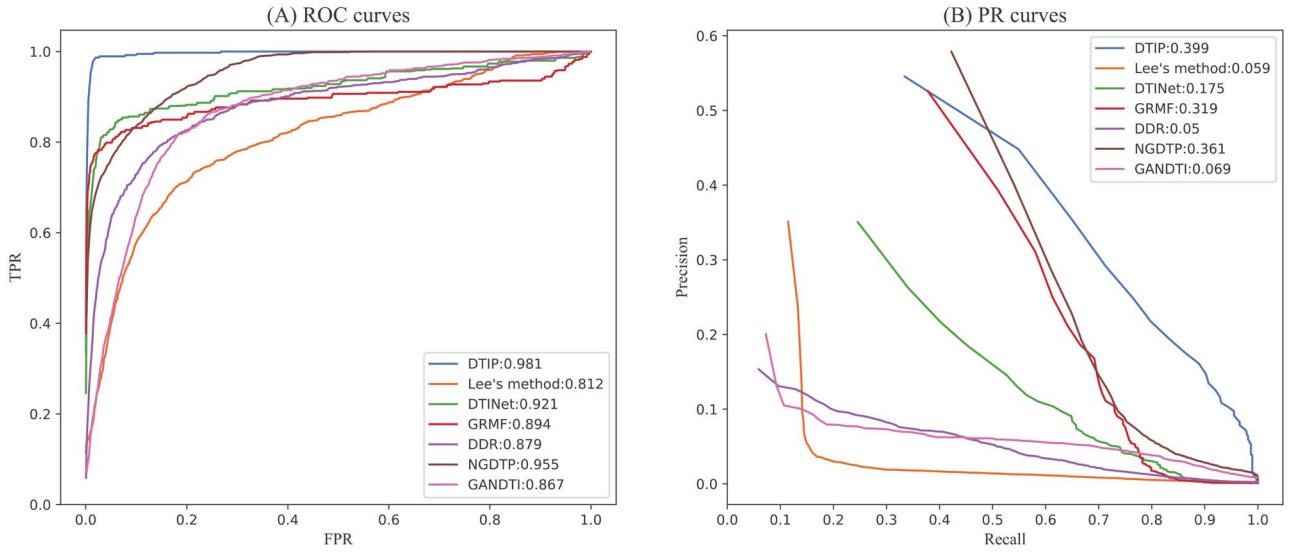

Figure 4. ROC curves and PR curves of all the methods in comparison of all the 708 drugs.

Table 1. The statistical results of the paired Wilcoxon test on the AUCs and AUPRs over all the 708 drugs by comparing DTIP and all other six methods

	Lee's Method	DTINet	GRMF	DDR	NGDTP	GANDTI
p-value of AUC	2.67609e-98	1.87391e-41	3.49947e-40	2.58338e-208	1.01899e-220	5.04374e-218
p-value of AUPR	9.53462e-30	6.67507e-20	4.60184e-24	2.12481e-36	4.68283e-60	2.05711e-41

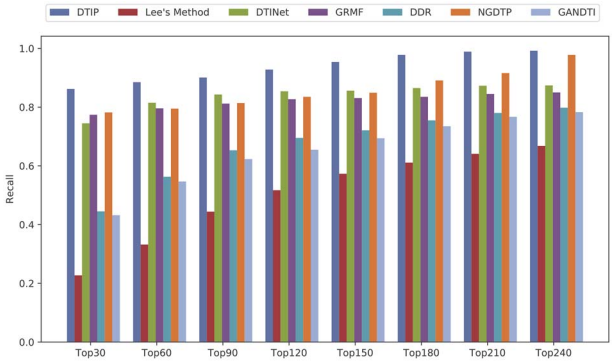

Figure 5. The average recalls over all the drugs at different top *k* values.

a drug–protein pair is greater than the threshold θ, the sample is considered as a positive sample, otherwise a negative sample. TPR (FPR) is the proportion correctly (incorrectly) identified as positive (negative) samples among all positive (negative) samples. *TPR* and *FPR* are defined as,

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP}, \tag{22}$$

where *TP*(*TN*) is the number of correctly identified positive (negative) samples, *FN*(*FP*) is the number of incorrectly identified positive (negative) samples.

AUC is widely used to evaluate the performance of the prediction model with probability estimations [41]. Since the number of positive samples is much smaller than negative samples, there are imbalanced distributions. AUPR is more informative than AUC to evaluate prediction performance [39]. *Precision* and *Recall* are defined as:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, \tag{23}$$

where *Precision* is the percentage of correctly identified positive samples among all the predicted samples, and *Recall* is the same as *TPR*. Averaging cross validation is performed to calculate AUC and AUPR [40]. We firstly calculate average AUC and AUPR from each fold, and the mean value under 5-folds is taken as the final score.

Since biologists often choose the top-ranked prediction results for further validation, it's crucial to discover the actual drug–target interactions. Therefore, we also calculate the drug-protein pair candidate's recall rate in the top *k* of the prediction results.

Comparison with other methods

DTIP is compared with six state-of-the-art methods for drug-protein interaction prediction including Lee's method [18], DTINet [19], GRMF [20], DDR [42], NGDTP [21] and GANDTI [26]. A brief description of each comparison method is added to supplementary file SF7. For fair comparison, the best performance of DTIP and the other six methods are reported by using their own best parameter settings. Specifically, $r = 0.8$ for Lee's method; $r = 0.8$, $\lambda_d = 1$ for DTINet; $\eta = 0.5$, $d = 0.1$, $t = 0.1$, $l = 0.2$ for GRMF; $n = 600$, $k = 5$ for DDR; $a_1 = 0.1$, $a_2 = 0.1$, $a_3 = 0.1$ for NGDTP; $l = 500$, $k = 200$, $epoch = 2000$, learning rate=0.05 for GANDTI.

We firstly calculate AUC and AUPR from each fold for each drug, and then calculate the mean value for each of 708 drugs over 5-folds. All the AUCs (or AUPRs) of 708 drugs are averaged as

Table 2. The top 10 candidate targets of five drugs

Drug name	Rank	Target	Evidence	Rank	Target	Evidence
	1	ADRA1A	DrugBank/DrugCentral	6	HTR6	DrugBank/DrugCentral
	2	ADRA2A	DrugBank/DrugCentral	7	GSTP1	DrugBank
Clozapine	3	ADRB1	DrugCentral	8	HTR1F	DrugCentral/Uniprot
	4	HTR2C	DrugBank/DrugCentral	9	CALY	DrugBank
	5	CHRM5	DrugBank/DrugCentral	10	ADRA1D	DrugCentral
	1	PPARG	Literature [45]	6	DRD1	DrugBank/DrugCentral
	2	CHRM1	DrugBank/DrugCentral	7	ADRA2C	DrugBank/DrugCentral
Aripiprazole	3	HTR2C	DrugBank/DrugCentral	8	OPRD1	DrugBank
	4	CHRM2	DrugBank/DrugCentral	9	HTR5	DrugBank/DrugCentral
	5	DRD3	DrugBank/DrugCentral	10	GRIN1	DrugBank
	1	HTR2A	DrugCentral	6	KCNH2	DrugCentral
	2	HTR2B	DrugCentral	7	NR3C1	Literature [45]
Ziprasidone	3	ADRA2A	DrugBank/DrugCentral	8	ESR1	Literature [45]
	4	ADRB1	DrugCentral	9	DRD1	DrugCentral
	5	HRH2	DrugCentral	10	CYP3A4	DrugBank
	1	DRD3	DrugCentral	6	HRH2	DrugBank/DrugCentral
	2	HTR1D	DrugBank	7	KCNH2	DrugBank/DrugCentral
Amitriptyline	3	HTR6	DrugBank/DrugCentral	8	RET	DrugCentral
	4	HTR2B	DrugCentral	9	SCN2A	DrugCentral
	5	DRD5	DrugCentral	10	DRD1	DrugCentral
	1	HTR2C	DrugBank/DrugCentral	6	OPRM1	Literature [45]
	2	CHRM1	DrugCentral	7	PPARG	Literature [45]
Asenapine	3	CHRM4	Literature [45]	8	ESR1	Literature [45]
	4	HTR7	DrugCentral	9	CACNA1C	Literature [45]
	5	SLC6A4	Literature [45]	10	HTR1A	DrugCentral

the final value. As shown in Figure 4, DTIP achieved the best average AUC of 0.981 among all the methods in comparison over 708 drugs, which outperforms Lee's method by 16.9%, DTINet by 6%, GRMF with 8.7%, DDR with 10.2%, NGDTP with 2.6% and GANDTI by 10.5%. The second best model NGDTP integrates drug-related disease annotations, side effects, interacted drugs and proteins, and protein-related disease annotations and interacted proteins. The results show that incorporating various kinds of drug and protein attributes information improved the prediction accuracy. The AUC of DTINet is 2.7% higher than GRMF. However, the AUPR of DITNet is 14.4% lower than the latter. The performance of DDR and Lee's method is worse than the previous four methods. A possible reason is that DDR neglects the attribute information of drug node and protein node. Lee's method only considered drug-related and protein-related information but ignored drug similarity and protein similarity information. GANDTI integrated drug–protein interactions, drug chemical structure similarities and protein sequence similarities without taking advantage of multiple similarities information over drug and protein nodes. In summary, our DTIP achieved the best results due to the integration of various drug and protein similarities, and the deeply learned representation of drug-protein neighbor topology and attribute embeddings. When it comes to AUPR over 708 drugs, DTIP achieved 34.0%, 22.4%, 8.0%, 34.9%, 3.8% and 37.3% better results than Lee's method, DTINet, GRMF, DDR, NGDTP and GANDTI, respectively.

For each of the prediction methods, we can obtain 708 average 5-fold AUC (AUPR) results for 708 drugs. When comparing two methods, we calculate Paired Wilcoxon test using 708 paired results. The statistical results in Table 1 demonstrated that the improvement was statistically significant (p-value < 0.05).

The recall rates of top k ranked target candidates are given in Figure 5. The higher the rate, the more drug-related target candidates are correctly identified. Our model's average recall rates over 708 drugs are 86.2% in top 30, 88.5% in top 60, and 92.8% in top 120, which are consistently higher than the other methods. The recall rates of NGDTP are very close to those of GRMF in the top 30, 60 and 120. The former model obtained rates of 78.2%, 79.5% and 83.5%, while the latter obtained rages of 77.4%, 79.6% and 81.2%. When k is 30, 60 and 120, the recall rate of DTINet is 74.5%, 81.5% and 85.4%, respectively. DDR ranked 44.5% in the top 30, 56.3% in the top 60, and 69.5% in the top 120, which is slightly lower than DTINet. GANDTI has a lower recall rate with 43.2%, 54.7%, 65.5% when k is 30, 60 and 120. Lee's method ranked 22.7%, 33.2% and 51.7% in top30, 60 and 120, respectively, which is the worst.

Case studies on five drugs

To further demonstrate the capability of DTIP in drug–target prediction, we conducted case studies on Clozapine, Aripiprazole, Ziprasidone, Amitriptyline and Asenapine. We ranked interaction scores of protein candidate for each drug in descending order, collected and analyzed the top 10 protein candidates. The results are given in Table 2.

DrugBank is a database containing comprehensive molecular information, mechanism of action of drugs, interactions and proteins [29]. DrugCentral records information about the interactions between drugs and proteins, as well as drug actions and indications [43]. Uniprot mines drug-target interactions from relevant external literature [44]. Among the 50 candidate proteins, 23 candidates are included by DrugBank, 35 candidates and one candidate are recorded by DrugCentral and Uniprot respectively. It indicates that candidate targets indeed interact with corresponding drugs. Nine candidate proteins labelled as 'Literature' were supported by the published literature [45], which demonstrates that they are more likely to interact with these drugs. In summary, case studies among five drugs further

demonstrated the capability of DTIP in discovering potential drug-target interactions.

Prediction of novel proteins related to drugs

In the end, we apply the trained model to predict protein candidates which are related to each of the interested drugs. The top 10 ranked protein candidates predicted by our model are provided in the supplementary table ST1 to assist the biologists in discovering real novel drug-related proteins with further wet-lab experiments. Apart from multiple kinds of similarities, interactions and associations related to drugs and proteins in our work, different types of disease-disease similarities have been measured in Comparative Toxicogenomics Database, DisGeNET and iDrug [46]. Our future work includes the investigation of integrating multiple disease similarities to improve DTI prediction performance.

Conclusion

We presented a method to formulate and fuse inter- and intra- similarities, interactions and associations across multi-sourced data, and multi-scale neighbour sequences for drug–protein interaction prediction. The proposed three-layer heterogeneous network benefited the formulation of neighbor node sequences based on RWR. A framework based on fully connected autoencoder, bidirectional GRU, and multiple-layer convolution neural networks was constructed for encoding and integrating the neighbor topological representations and the attribute representations. Three attention mechanisms were proposed to assign the higher weights for the more important features, topologies and drug similarity types. DTIP should be more attractive for the biologists as its top-k candidates often contain more actual drug–protein interactions during the cross-validation process. Both cross-validation and case studies confirm DTIP's powerful prediction ability. Our prediction model can be served as a prioritization tool to screen potential DTI candidates for subsequent discovering the true DTIs by wet laboratory experiments.

Key Points

- A three-layer heterogeneous network is constructed to assist the extraction of similarities, interactions and associations related to drugs and proteins from the multiple-sourced data.
- A novel framework to learn two levels of drug-protein representations where neighbor topological representation reveals the sequential information among multiple neighbors, and attribute representation reveals the underlying relationship among common related drugs, proteins and diseases.
- New attention mechanisms at feature level, neighbouring topological level, and similarity level to discriminate the different contributions of node features, topologies and various similarities for the final DTI prediction, respectively.
- Improved prediction performance validated by comprehensive evaluations including comparison with six state-of-the-art models over public dataset, top k candidates retrieval, and case studies of five drugs.

Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

Funding

The Natural Science Foundation of China (61972135, 62031003); Natural Science Foundation of Heilongjiang Province (LH2019F049 and LH2019A029); China Postdoctoral Science Foundation (2019M650069, 2020M670939); Heilongjiang Postdoctoral Scientific Research Staring Foundation (BHLQ18104); Fundamental Research Foundation of Universi-ties in Heilongjiang Province for Technology Innovation (KJCX201805); Innovation Talents Project of Harbin Science and Technology Bureau (2017RAQXJ094); Fundamental Research Foundation of Universities in Heilongjiang Province for Youth Innovation Team (RCYJTD201805); the Foundation of Graduate Innovative Research (YJSCX2020-073HLJU).

References

1. Chen X, Yan CC, Zhang X, *et al*. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2015;**17**(4):696–712.
2. Yu L, Huang J, Ma Z, *et al*. Inferring drug-disease associations based on known protein complexes. *BMC Med Genomics* 2015;**8**Suppl 2(Suppl 2):1–13.
3. Yu L, Zhao J, Gao L. Drug repositioning based on triangularly balanced structure for tissue-specific diseases in incomplete interactome. *Artif Intell Med* 2017;**77**(C):53–63.
4. Bagherian M, Sabeti E, Wang K, *et al*. Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. *Brief Bioinform* 2020;**22**(1):1467–5463.
5. Chen Z-H, You Z-H, Guo Z-H, *et al*. Prediction of drug-target interactions from multi-molecular network based on deep walk embedding model. *Front Bioeng Biotechnol* 2020;**8**:338.
6. Ding Y, Tang J, Guo F. Identification of drug-target interactions via multiple information integration. *Inform Sci* 2017;**418-419**:546–60.
7. Ding Y, Tang J, Guo F. Identification of protein-ligand binding sites by sequence information and ensemble classifier. *J Chem Inf Model* 2017;**57**(12):3149–61.
8. Ding Y, Tang J, Guo F. Identification of drug-side effect association via multiple information integration with centered kernel alignment. *Neurocomputing* 2019;**325**:211–24.
9. Shen C, Ding Y, Tang J, *et al*. An ameliorated prediction of drug-target interactions based on multi-scale discrete wavelet transform and network features. *Int J Mol Sci* 2017;**18**:781.
10. Donald BR. *Algorithms in structural molecular biology*. Mass.: MIT Press: Cambridge, 2011.
11. Morris GM, Huey R, Lindstrom W, *et al*. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;**30**(16):2785–91.
12. Li Y, Huang Y-A, You Z-H, *et al*. Drug-target interaction prediction based on drug fingerprint information and protein sequence. *Molecules* 2019, 2019;**24**(16):2999.
13. Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 2009;**25**(18):2397–403.

14. Keum J, Nam H. SELF-BLM: prediction of drug-target interactions via self-training SVM. *PLoS One* 2017;**12**(2):e0171839.
15. Chen X, Liu M-X, Yan G-Y. Drug-target interaction prediction by random walk on the heterogeneous network. *Mol Biosyst* 2012;**8**(7):1970–8.
16. Lee I, Nam H. Identification of drug-target interaction by a random walk with restart method on an interactome network. *BMC Bioinformatics* 2018;**19**:1–10.
17. Yan X-Y, Zhang S-W, He C-R. Prediction of drug-target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods. *Comput Biol Chem* 2019;**78**:460–7.
18. Li Z-C, Huang M-H, Zhong W-Q, *et al*. Identification of drug-target interaction from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics* 2015;**32**(7):1057–64.
19. 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**(1):573.
20. Ezzat A, Zhao P, Wu M, *et al*. Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans Comput Biol Bioinform* 2016;**14**:1–1.
21. Xuan P, Chen B, Zhang T, *et al*. Prediction of drug-target interactions based on network representation learning and ensemble learning. *IEEE/ACM Trans Comput Biol Bioinform* 2020;**4**:1–12.
22. Xuan P, Sun C, Zhang T, *et al*. Gradient boosting decision tree-based method for predicting interactions between target genes and drugs. *Front Genet* 2019;**10**:459.
23. Zheng X, He S, Song X, *et al*. DTI-RCNN: new efficient hybrid neural network model to predict drug-target interactions. *Artif Neural Networks Machine Learning – ICANN* 2018;**2018**:104–14.
24. Manoochehri HE, Pillai A, Nourani M. Graph convolutional networks for predicting drug-protein interactions. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 2019;**2019**:1223–5.
25. Torng W, Altman RB. Graph convolutional neural networks for predicting drug-target interactions. *J Chem Inf Model* 2019;**59**(10):4131–49.
26. Sun C, Xuan P, Zhang T, *et al*. Graph convolutional autoencoder and generative adversarial network-based method for predicting drug-target interactions. *IEEE/ACM Trans Comput Biol Bioinform* 2020;**1**:1–11.
27. Zeng X, Ding N, Rodríguez-Patón A, *et al*. Probability-based collaborative filtering model for predicting gene-disease associations. *BMC Med Genomics* 2017;**10**(5):76.
28. Zou Q, Li J, Song L, *et al*. Similarity computation strategies in the microRNA-disease network: a survey. *Brief Funct Genomics* 2016;**15**:55–64.
29. Wishart D, Djoumbou Y, Guo AC, *et al*. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2017;**46**:D1074–D1082.
30. Keshava Prasad TS, Goel R, Kandasamy K, *et al*. Human protein reference database–2009 update. *Nucleic Acids Res* 2009;**37**(Database issue):D767–72.
31. Davis AP, Murphy CG, Johnson R, *et al*. The comparative toxicogenomics database: update 2013. *Nucleic Acids Res* 2013;**41**(Database issue):D1104–14.
32. Iorio F, Bosotti R, Scacheri E, *et al*. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci* 2010;**107**(33):14621.
33. Wang W, Yang S, Zhang X, *et al*. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics 2014* 2014;**30**(20):2923–30.
34. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, Israel: Omnipress. 2010, 807–14.
35. Kingma D, Ba J. Adam: a method for stochastic optimization. *Int Learn Represent* 2014;**1412**:1–15.
36. Mircea Petrini. Improvements to the Backpropagation Algorithm. *Annals of the University of Petrosani, Economics*, University of Petrosani, Romania 2012;**12**(4):185-92.
37. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. In: *International Conference on Learning Representations*, San Diego, United States: ICLR. 2015, abs/1409.0473.
38. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med* 2013;**4**(2):627–35.
39. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;**10**(3):e0118432.
40. Pahikkala T, Airola A, Pietila S, *et al*. Toward more realistic drug-target interaction predictions. *Brief Bioinform* 2015;**16**(2):325–37.
41. Ling CX, Huang J, Zhang H. AUC: a better measure thanaccuracy in comparing learning algorithms. In: *Conferenceof the Canadian Society for Computational Studies of Inteligence*. Kingston, Canada: Springer, 2003, 329–41.
42. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics* 2017;**34**(7):1164–73.
43. Ursu O, Holmes J, Knockel J, *et al*. DrugCentral: online drug compendium. *Nucleic Acids Res* 2017;**45**(D1):D932–9.
44. Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2018;**47**(D1):D506–15.
45. Bosc N, Atkinson F, Felix E, *et al*. Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *J Chem* 2019;**11**(1):4.
46. Chen H, Cheng F, Li J. iDrug: integration of drug repositioning and drug-target prediction via cross-network embedding. *PLoS Comput Biol* 2020;**16**(7):e1008040.