Briefings in Bioinformatics
doi.10.1093/bib/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category
OXFORD

Data and text mining

Learning multi-scale heterogenous network topologies and various pairwise attributes for drug-disease association prediction

Hongda Zhang [1], Hui Cui [2], Tiangang Zhang [3], Yangkun Cao [4], Ping Xuan [1,*]

[1]School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China, [2]Department of Computer Science and Information Technology, La Trobe University, Melbourne 3083, Australia, [3]School of Mathematical Science, Heilongjiang University, Harbin 150080, China, [4]School of Artificial Intelligence, Jilin University, Changchun 130012, China.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Identifying new therapeutic effects for the approved drugs is beneficial for effectively reducing the drug development cost and time. Most of the recent computational methods concentrate on exploiting multiple kinds of information about drugs and disease to predict the candidate associations between drugs and diseases. However, the drug and disease nodes have neighboring topologies with multiple scales, and the previous methods did not fully exploit and deeply integrate these topologies.

Results: We present a prediction method, MTRD, to integrate and learn multi-scale neighboring topologies and the attributes of a pair of drug and disease nodes. First, for multiple kinds of drug similarities, multiple drug-disease heterogenous networks are constructed respectively to integrate the similarities and associations related to drugs and diseases. Moreover, each heterogenous network has its specific topology structure, which is helpful for learning the corresponding specific topology representation. We formulate the topology embeddings for each drug node and disease node by random walking on each heterogeneous network, and the embeddings cover the neighboring topologies with different scopes. Since the multi-scale topology embeddings have context relationships, we construct Bi-directional long short-term memory-based module to encode these embeddings and their relationships and learn the neighboring topology representation. We also design the attention mechanisms at feature level and at scale level to obtain the more informative pairwise features and topology embeddings. A module based on multi-layer convolutional networks is constructed to learn the representative attributes of the drug-disease node pair according to their related similarity and association information. Comprehensive experimental results indicate that MTRD achieves the superior performance than several state-of-the-art methods for predicting drug-disease associations. MTRD also retrieves more actual drug-disease associations in the top ranked candidates of the prediction result. Case studies on 5 drugs further demonstrate MTRD's ability in discovering the potential candidate diseases for the interested drugs.

Contact: xuanping@hlju.edu.cn

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

Hongda Zhang is studying for his master's degree in the School of Computer Science and Technology at Heilongjiang University, Harbin, China. His 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 data-driven and computerized models for biomedical and health informatics.

Tiangang Zhang, PhD (The University of Tokyo), is an associate professor of the Department of Mathematical Science, Heilongjiang University, Harbin, China. His current research interests include complex network analysis and computational fluid dynamics.

Yangkun Cao is studying for his PhD's degree in the School of Artificial Intelligence at Jilin University, Changchun, China. His research interests include Computational Biology and Complex Systems.

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.

1 Introduction

De novo drug discovery and development involves a lot of expences. It usually requires 10 to 15 years and costs 800 million to 1.5 billion US dollars [1–4]. Drug repositioning identifies new therapeutic effects of drugs approved by regulatory agencies [5–7]. Unlike de novo, repositioned drugs have well-characterized biological effectiveness, safety and pharmacological characteristics, which can shorten the drug development cycle to 6.5 years and reduce the cost to 300 million US dollars [8, 9].

Computational prediction of new therapeutic effects of approved drugs can identify true associations that can be confirmed by further experiments. [10–13]. Previously implemented methods can be broadly divided into three categories. Drugs usually obtain the therapeutic effects by interacting with their target genes. A drug and a disease is more likely to association with each other when they are associated or interacted with more common targets[14]. Therefore, several

methods [15–17] concerning the first strategy are based on relevant public targets to infer the propensity for drug-related diseases. Similarly, association propensity can also be evaluated using proteins that are associated with drugs and diseases [18–20]. However, these methods are very limited in discovering drug-disease associations without public genes or proteins.

The second strategy is to screen out drug-related candidate diseases by integrating a variety of data, including multiple attributes, similarities, associated targets, and known associations between drugs and diseases. A model based on the kernel function was established to integrate molecular and phenotypic information [21]. Several studies have constructed a heterogeneous network and proposed a method to predict associations by modeling the association information or using random walks [22–24]. Several methods are based on sparse subspace learning[25], matrix factorization with similarity constraints [26] or co-occurrence statistics in the literature [27] to infer candidate drug-disease pairs. However, these methods are rudimentary predictive models and cannot completely explore the deep and complex relationships between drugs and diseases.

Recently, several methods based on deep learning were proposed and the prediction performance were further improved. A model based on the Gaussian interaction profile and autoencoder was established to predict drug-related candidate diseases by fusing drug-related similarities and associations [28]. A framework composed of convolutional neural networks and a bidirectional long short-term memory network was established to integrate path information between drug diseases [29]. Additionally, the model of the graph auto-encoder has been applied to the association prediction of drugs and diseases [30]. The topological structure formed by a pair of neighbor nodes of drugs and diseases has many different scales and different information. In other words, a set of one-hop neighbor nodes only covers a small and partial topology, whereas a set of two-hop and three-hop nodes covers a larger range of topologies. However, the multi-scale neighbor topology of a node pair is neither extracted by the previous method nor considered.

In this article, we propose a new prediction method, MTRD, for learning, capturing, and integrating topological context representations and attribute representations from multiple heterogeneous networks. The contributions of our model can be summarized as follows.

- Three different drug-disease bi-layer networks were constructed for three kinds of drug similarities, which is beneficial for combining diverse connections among drugs and diseases and extracting the specific topology of each network.
- A novel topology embedding strategy based on random walks in each drug-disease network was designed to formulate various topology embeddings with multiple scales. Moreover, each drug-disease network has its specific multi-scale embeddings.
- A novel framework was constructed to learn two representations of the drug-disease node pair, where the neighboring topology representation integrates the context relationships among multi-scale topology embeddings, and attribute representation reveals the relationship among pairwise attributes.
- Two new attention mechanisms at the feature level and the scale level were established to discriminate the various contributions of pairwise features and multi-scale topology embeddings for drug-disease association prediction, respectively. The improved performance was confirmed by comparison with the state-of-the-art methods, the recall rates in top-ranked candidates, and case studies on five drugs.

2 Materials and methods

Multiple drug-disease heterogeneous networks with different drug similarity connections have been established to facilitate the integration of information on multiple drugs and the relationship between drugs and diseases. We propose a prediction model composed of two branches to predict potential disease candidates for drugs of interest. In the left branch, the topological embedding vectors of the drug and disease nodes are generated by random walks with restarts on multiple heterogeneous networks. A new feature-level attention mechanism was proposed to integrate the topological embeddings of drug-disease pairs from different heterogeneous networks. Furthermore, to distinguish the contributions of different topological scales, we designed a scale-level attention mechanism. Next, a module based on a bidirectional long short-term memory network (Bi-LSTM) was established that learns the topological representation of drug-disease pairs from the integrated vector. For the second branch, a convolutional neural network module was proposed to learn the attribute representation of drug-disease pairs by integrating similarity information and the known therapeutic relationships between drug and disease nodes. Finally, the topological representation containing multi-scale information and the attribute representation containing information about the similarities and associations were integrated through multilayer perceptron. The output score is regarded as the predicted probability of whether the drug and disease are related.

2.1 Dataset

Information on the three drug attributes selected in our study, including drug-related chemical substructures, drug-related target protein domains and gene ontology annotations, drug-disease associations, and disease semantic

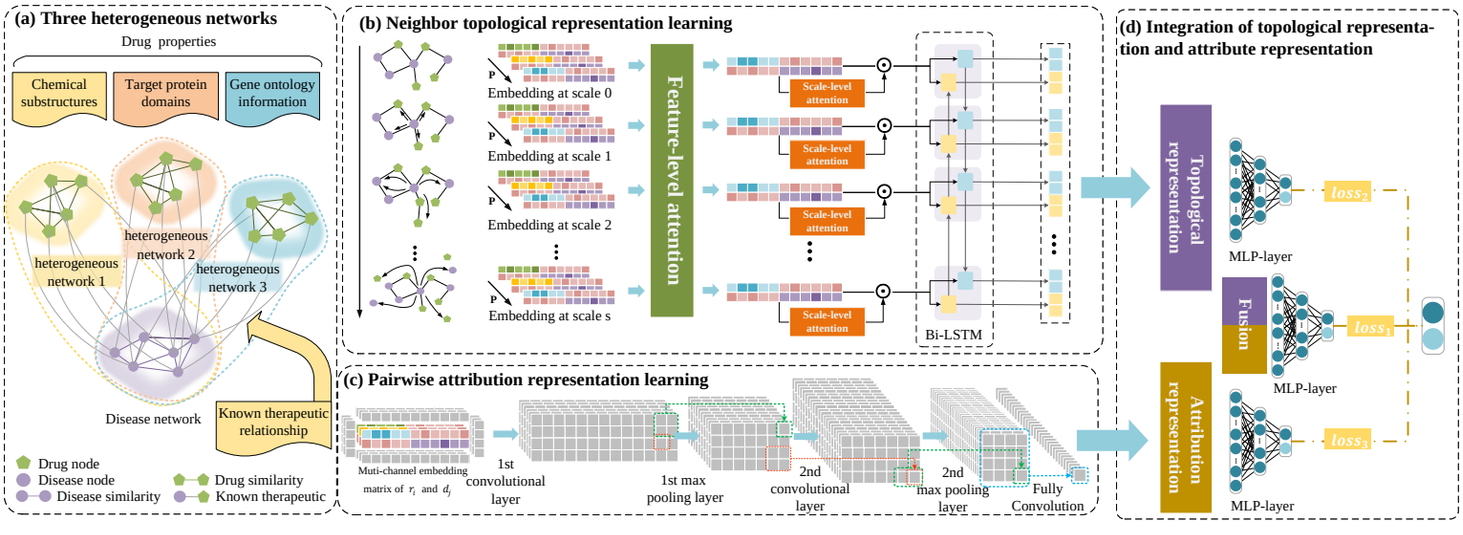

Fig. 1. Framework of the proposed MTRD model. (a) construct multiple drug-disease heterogeneous networks with different drug similarity connections (b) learn neighbor topological representation of r_i-d_j pair (c) learn attribute representation of r_i-d_j pair (d) deeply integrate multiple representations and estimate association score.

similarities, were obtained from previous studies. The dataset included 763 drugs and 681 diseases along with 3051 known associations between these drugs and diseases. The drug attribute information was acquired from DrugBank [31]; 623 drug-related chemical substructures and 1,426 protein domain items originated from the PubChem [32] and InterPro [33] database, and 4447 gene function annotations were obtained from the UniProt [34] database.

2.2 Construction of multiple drug-disease heterogeneous networks

We constructed three drug-disease heterogeneous networks using the three selected attributes, each of which uses a certain drug similarity, a disease similarity, and known treatment relationships to connect its nodes.

Drug-drug similarity networks

Multiple drug networks describe the similarity between drugs from multiple perspectives. The similarity between the drugs r_i and r_j increases with the number of common chemical substructures among the drugs. Additionally, the similarity among drugs increases when more common domains of target proteins or more gene ontology annotation items are present. For different attribute information, three different similarities were separately obtained using cosine similarity. These similarities describe the relevance of drugs from different drug attribute levels.

When constructing the drug–drug network, the corresponding connecting edges are added between drug nodes with similarity greater than 0, and the weight of the edge is set to its similarity value. Based on the similarity values of the three different drug attributes, three drug networks were constructed. The matrix $R^p = [R^p_{i,j}] \in \Re^{N_r \times N_r}$ is used to represent the drug similarity network under the attribute $p \in \mathbf{P}$, where $\mathbf{P} = \{\text{PubChem}, \text{TarProDom}, \text{GeneOntoInfoTerm}\}$, PubChem represents the public chemical substructure, TarProDom represents the target protein domain, and GeneOntoInfoTerm represents the gene ontology term. N_r represents the number of drugs, $\Re^{N_r \times N_r}$ represents the real number space of $N_r \times N_r$. The similarity $R^p_{i,j}$ of r_i and r_j was obtained using the following formula:

$$R^p_{i,j} = Cosine(r_i, r_j), p \in \mathbf{P} \tag{1}$$

$Cosine(\cdot)$ represents the cosine similarity, and the similarity values are in the interval $[0, 1]$, r_i and r_j represent the one-hot vectors under attribute p.

Disease-disease network

Measuring the similarities between diseases is essential for constructing disease networks. Semantic similarity measures the relationship between diseases. Wang's method [35], which considers the topological information of two diseases in the ontology, was used to find the semantic similarity of diseases in our work.

When constructing a disease network, two diseases with a similarity greater than zero are connected by an edge. Its matrix is expressed as $D = [D_{i,j}] \in \Re^{N_d \times N_d}$, N_d is the number of diseases, $\Re^{N_d \times N_d}$ represents the real number space of $N_d \times N_d$, $D_{i,j}$ is the similarity between the drug d_i and the drug d_j, and the value of disease similarity is in the interval $[0, 1]$.

Multiple drug-disease heterogeneous networks

According to the known drug-disease-treatment relationship, a bipartite graph connecting the drug and disease nodes was constructed. The edges in the constructed bipartite graph were used to connect N_r drugs and N_d diseases from different networks. The set of edges is represented as $T = [T_{i,j}] \in \Re^{N_r \times N_d}$, u and v are drug and disease nodes, respectively, and $T_{u,v}$ is used to represent the weight between them. When r_i has a known therapeutic relationship with the d_j, $T_{i,j} = 1$, when it is unknown, $T_{i,j} = 0$. The drug network R^p and the disease network D are connected through T, and they form a drug-disease heterogeneous network $H^p = [H^p_{u,v}] \in \Re^{(N_r+N_d)\times(N_r+N_d)}$,

$$H^p_{u,v} = \begin{cases} R^p_{u,v} & u, v \text{ are drugs} \\ D_{u,v} & u, v \text{ are diseases} \\ T_{u,v} & \text{otherwise} \end{cases} \tag{2}$$

each element $H^p_{u,v}$ in H^p represents the connection weight between node u and node v. When both u and v are drugs, the weight between them is represented by the drug similarity $R^p_{r_u,r_v}$. When both u and v are diseases, their connection

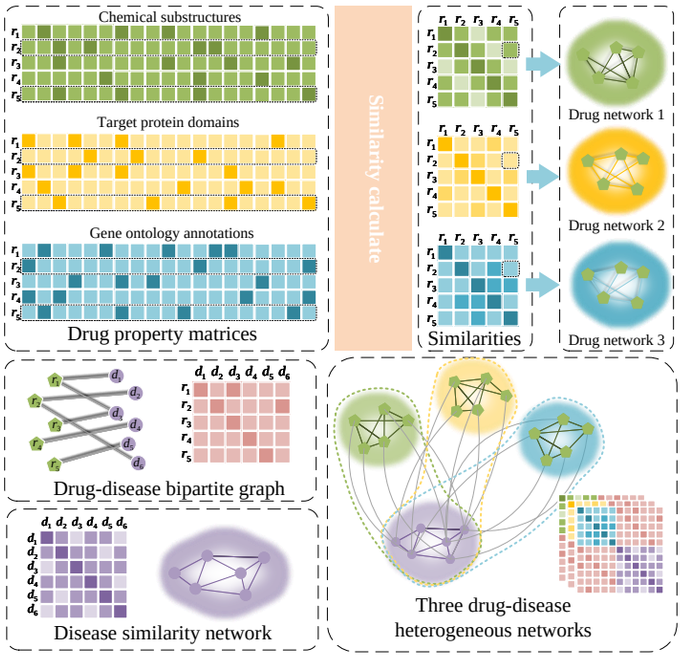

Fig. 2. Construction of multiple drug-disease heterogeneous networks based on different kinds of drug similarities.

weights are equal to the disease similarity D_{d_u,d_v}. When u and v are drug and disease nodes, respectively, the known drug-disease association $T_{u,v}$ is used to represent the weight between them. The heterogeneous network H^p with different drug properties can be regarded as different views, which together constitute the multi-attribute view of the heterogeneous network $\{H^p | p \in \mathbf{P}\}$, where $\mathbf{P} = \{\text{PubChem}, \text{TarProDom}, \text{GeneOntoInfoTerm}\}$ is a collection of drug properties.

2.3 Learning pairwise neighboring topological representation

The model of the left branch consists of three parts: a feature-level attention mechanism, a scale-level attention mechanism, and an encoding module based on bidirectional LSTM modules. Different drug attributes have different effects on the prediction of drug-disease association. The attention mechanism at the feature level integrates the drug-disease pair embedding vectors under different attribute views. The scale-level attention mechanism evaluates the importance of each scale view and assigns a score. The bidirectional LSTM module deeply integrates the contextual information between multiscale topologies and learns the topological representation of r_i-d_j pairs.

2.3.1 Multi-scale topology embedding by random walk with restart

In the drug-disease heterogeneous network H^p, if two nodes are connected by only one edge, we refer to this adjacency relationship as the adjacency of scale 1. If two nodes are connected by two (or more) end-to-end edges, we refer to this adjacency relationship as the adjacency of scale 2 (or multi-scale adjacency). In the network H^p, there are three adjacencies with a scale of 1 that have different meanings. d_1-d_2, r_1-r_2 represent the similarity between diseases and the similarity relationship between drugs, respectively, r_1-d_1 represents the known drug-disease relationship. The adjacency of scale 2 or more in the network can reflect more connections between drugs and diseases. For example, r_1-r_2-d_1 means "r_1 and r_2 have similar chemical composition or function, and r_2 and d_1 have a known association, therefore r_1 may also be associated with d_1". To identify and mine the potential association information contained in the adjacency topologies of different scales, we used the random walk with restart (RWR) to construct the topological structure embedding vector of the drug-disease pair. For random walks in a heterogeneous network of drug diseases, the

transition matrix A^p is defined as follows:

$$A^p = [A^p_{u,v}] = \frac{H^p_{u,v}}{\sum_v H^p_{u,v}} \tag{3}$$

where each $A^p_{u,v}$ represents the probability of the walker going from node u to v under the attribute p. When the walker starts from node u, after passing the s edges (i.e., crossing the s scale), the probability vectors of reaching the other nodes are $[r^p(s); d^p(s)]$. $r^p(s)$ and $d^p(s)$ are the drug and disease parts of the probability vector, respectively, and $sum(r^p(s)) + sum(d^p(s)) = 1$. At this time, the sth scale embedding vector $x_u^{(s,p)}$ of the node u can be composed of $r^p(s)$ and $d^p(s)$.

$$x_u^{(s,p)} = \begin{bmatrix} r^p(s) \\ d^p(s) \end{bmatrix} \tag{4}$$

For the initial walking state $x_u^{0,p}$ of the walker under attribute p, if u is a drug

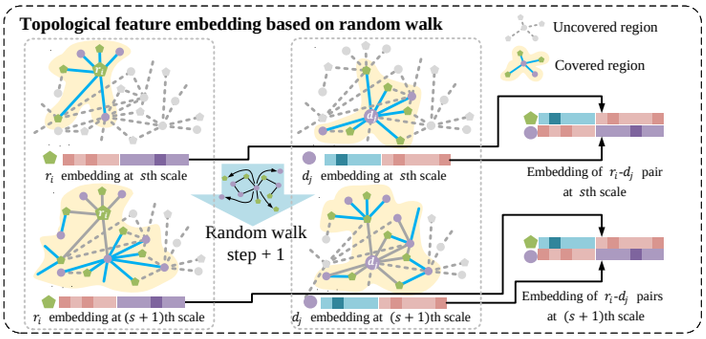

Fig. 3. Illustration of the proposed multi-scale topological embedding of drug-disease node pair.

node, only the i element in $r^p(0)$ is set to 1, and the other elements are set to 0. Similarly, if u is a disease node, only the value of the j element in $d^p(0)$ is 1, and the rest are 0. Based on the transition matrix A^p and the initial state x_u^0, The probability that the walker that starts from u will reach each node at the $(s+1)$th scale is defined as follows:

$$x_u^{(s+1,p)} = (1-\alpha)A^p x_u^{s,p} + \alpha x_u^{(0,p)} \tag{5}$$

where $\alpha \in [0,1]$ is the restart probability, which represents the probability that the walker jumps back to the initial node u to walk again during the walk. The obtained $x_{r_i}^{(s,p)}$, $x_{d_j}^{(s,p)}$ are spliced end-to-end as the topological structure of the embedding vector $x_{r_i,d_j}^{(s,p)}$, s is the corresponding scale, and p is the corresponding attribute.

$$x_{r_i,d_j}^{(s,p)} = x_{r_i}^{(s,p)} \| x_{d_j}^{(s,p)} \tag{6}$$

$\|$ represents the splicing of the head and tail, $x_{r_i,d_j}^{(s,p)} \in \Re^{1\times(N_r+N_d)}$. First, the topological structure embedding vectors of the same scale and different attributes are stacked up and down:

$$x_{r_i,d_j}^{(s,\cdot)} = \underset{p\in P}{\textbf{a-stack}}\, x_{r_i,d_j}^{(s,p)} \tag{7}$$

a-stack is stacked up and down, and $x_{r_i,d_j}^{(s,\cdot)} \in \Re^{N_p\times(N_r+N_d)}$. Next the different groups are stacked up and down according to the scale:

$$x_{r_i,d_j}^{(\cdot,\cdot)} = \underset{s=1\ldots S}{\textbf{s-stack}}\, x_{r_i,d_j}^{(s,\cdot)} \tag{8}$$

where the symbol **s-stack** means stacking up and down according to the scale, and $x_{r_i,d_j}^{(\cdot,\cdot)} \in \Re^{N_s\times N_p\times(N_r+N_d)}$ is used as the input of the left branch module.

2.3.2 Attention at feature level to integrate topology embeddings

Under different networks, the edges of the same disease pair present different similarities from a unique perspective. Therefore, an attention mechanism at the feature level was constructed. The attention score of pth attribute vector $x_{r_i,d_j}^{(\cdot,p)}$ is e_p,

$$e_p = W^P\, tanh(W_p x_{r_i,d_j}^{(\cdot,p)} + b_p) \tag{9}$$

W_p is the parameter matrix and b_p is the bias vector. W^P is a parameter matrix used to capture the interrelationships between different views. Softmax was used to normalize the attention score e_p to the attention weight, α_p.

$$\alpha_{p,i} = \frac{exp(e_{p,i})}{\sum_{p=1}^{N_p} exp(e_{p,i})} \tag{10}$$

The representation vector under the scale s is the element-wise sum of the weights of N_p attribute vectors $y_{r_i,d_j}^s \in \Re^{1\times(N_r+N_d)}$ and is calculated as follows

$$y_{r_i,d_j}^s = \sum_{p\in P} \alpha_p \otimes x_{r_i,d_j}^{(s,p)} \tag{11}$$

2.3.3 Attention at scale level

Connections of different scales have different contributions to the prediction of drug-disease associations. Therefore, at the scale level, we established an attention mechanism. The attention score e_s and weight α_s of the scale view s are calculated as follows,

$$e_s = h^S\, tanh(W^s y_{r_i,d_j}^s + b^s) \tag{12}$$

$$\alpha_s = \frac{exp(e_s)}{\sum_{s=1}^{N_s} exp(e_s)} \tag{13}$$

W^s and b^s are the attention weight and bias vector, respectively, and h^S is a parameter vector shared by topologies of various scales. The vector enhanced by the attention mechanism is z_{r_i,d_j}^s,

$$z_{r_i,d_j}^s = \alpha_s y_{r_i,d_j}^s \tag{14}$$

2.3.4 Learning contextual relationship of multi-scale topologies by Bi-LSTM

$z_{r_i,d_j}^1, \ldots, z_{r_i,d_j}^s, \ldots, z_{r_i,d_j}^{N_s}$ together form a topological vector sequence of r_i-d_j. To better capture the contextual relationship between multi-scale sequences, a network module based on bidirectional LSTM was built to learn the topological representation of drug-disease pairs. The forward $\overrightarrow{\text{LSTM}}$ and the reversed $\overleftarrow{\text{LSTM}}$ modules read each scale topology in the order of $1, 2, 3, \ldots, N_s$ and $N_s, N_s-1, \ldots, 1$ individually. The hidden layer outputs of the forward and the backward modules are $\overrightarrow{h_s}$ and $\overleftarrow{h_s}$ at the scale s.

$$\overrightarrow{h_s} = \overrightarrow{\text{LSTM}}(z_{r_i,d_j}^s, \overrightarrow{h_{s-1}}) \tag{15}$$

$$\overleftarrow{h_s} = \overleftarrow{\text{LSTM}}(z_{r_i,d_j}^s, \overleftarrow{h_{s+1}}) \tag{16}$$

$\overrightarrow{h_s}$ and $\overleftarrow{h_s}$ are spliced to get the output of bidirectional LSTM,

$$h_s = Concatenate(\overrightarrow{h_s}, \overleftarrow{h_s}) \tag{17}$$

$Concatenate(\cdot)$ denotes the concatenation. The BiLSTM output sequences at different scales are stitched together to obtain the final feature output $o_{\text{topo}} \in \Re^{(N_s\times N_{\text{hid}})\times 1}$:

$$o_{\text{topo}} = h_1 \| h_2 \| \cdots \| h_{N_s} \tag{18}$$

where $\|$ represents head-to-tail splicing, and N_{hid} is the hidden layer of the bidirectional LSTM.

2.4 Learning pairwise multiple attribute representations

2.4.1 Construction of attribute embedding matrix

When predicting the association between r_i and disease d_j, we hypothesized that if r_i and d_j are related (or similar) to more public drugs (or diseases), they are more to be related. Therefore, as the picture shows, for the three drug-disease

heterogeneous networks, three embedding matrices are established to capture the nodes that both r_i and d_j are connected in each network.

We used the vector x_r^l in the lth matrix to represent the connection between r_i and other drugs and disease nodes, and we obtain the disease vector x_d^l. x_r^l and x_d^l are spliced, and the embedding matrix X^l of r_i-d_j pair on the lth network was obtained. The embedding matrices of each l are used as a different channel, which together constitute the input $X_{r_i,d_j}^{\text{attr}} \in \Re^{3\times 2\times(N_r+N_d)}$ of the right branch module.

2.4.2 Attribute representation learning based on multi-layer CNN

The multi-channel embedding matrix $X_{r_i,d_j}^{\text{attr}}$ of r_i and d_j was fed into the right CNN module to learn its attribute representation. This convolution module contained two convolution pooling layers. First, to prevent loss of information when processing boundaries, a circle of zeros was added around the input feature matrix $X_{r_i,d_j}^{\text{attr}}$ and they were recorded together as $\widetilde{X}_{r_i,d_j}^{\text{attr}}$. We considered the first convolutional layer and the first pooling layer as examples to introduce our right convolution module. For the first convolutional layer, the number of input

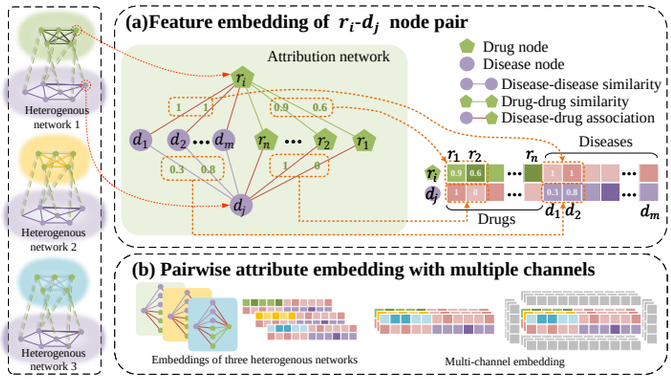

Fig. 4. Constructing attribute embedding of drug-disease node pair with multiple channels.

channels is n_{in} (that is, the number of attribute views N_p); we set up n_{c_1} filters with a receptive field of $w_c \times h_c$, and let them slide on the feature matrix in steps of s_{c_1}. Each filter learns a connection mode at r_i-d_j pair and obtains a feature map $Y_{c_1} \in \Re^{n_{c_1}\times(4-w_c+1)\times(2+N_r+N_d-h_c+1)}$. $X(a,b)$ is the receptive field when the kth filter slides to the position of the ath row and the bth column of the element.

The receptive fields $X(a,b) \in \Re^{n_{in}\times w_c\times h_c}$ and the element $Y_{c_1}(k,a,b)$ at ath row, bth column and kth feature map are defined as follows:

$$X(a,b) = \widetilde{X}_{r_i,d_j}^{\text{attr}}[1:N_{in}, a:a+w_c, b:b+h_c] \tag{19}$$

$$Y_{c_1}^k(a,b) = f(W_{c_1}^k * X(a,b) + b_{c_1}^k) \tag{20}$$

$$1 \le a \le 4-w_c+1, 1 \le b \le 2+N_r+N_d-h_c+1, 1 \le k \le n_{c_1}$$

$W_{c_1}^k$ and $b_{c_1}^k$ represent the learnable parameters at the kth filter, respectively. $f(\cdot)$ represents the ReLU activation function, and $*$ is sum of the product of position elements in two matrices.

Maximum pooling was used to extract representative features from the multichannel feature map, Y_{c_1}. Similarly, Y_{c_1} was added to a circle of zero to obtain $\widetilde{Y}_{c_1}$. w_{p_1} and h_{p_1} as the length and height of the pooling window, respectively. The output after the pooling layer is Y_{p_1}:

$$Y_{p_1}^k(a,b) = \max(\widetilde{Y}_{c_1}[k, a:a+w_{p_1}, b:b+h_{p_1}]) \tag{21}$$

$$1 \le a \le 4-w_{c_1}+1, 1 \le b \le 2+N_r+N_d-h_{c_1}+1, 1 \le k \le n_{c_1}$$

where $Y_{p_1}^k(a,b)$ represents the element in the ath row and the bth column of the feature map of the kth channel in Y_{p_1}.

After the feature matrix X was pooled using two convolutions, the feature map Y is obtained. Y is tiled into r_i-d_j's attribute representation vector $o_{\text{attr}} \in \Re^{w'h'n_{out}\times 1}$:

$$o_{\text{attr}} = flatten(Y), Y \in \Re^{n_{out}\times w'\times h'}$$

$$w' = \lceil \frac{4}{w_{c_1} w_{c_2}} \rceil, h' = \lceil \frac{N_r+N_d+2}{h_{c_1} h_{c_2}} \rceil$$

$\lceil\cdot\rceil$ represents rounding up, w' and h' represent the width and height of the feature map, respectively, and $flatten(\cdot)$ means to expand into a vector.

2.5 Loss of association prediction

To completely utilize the representation vectors o_{topo} and o_{attr}, a combined strategy for training and final prediction was designed. o_{topo} and o_{attr} are in-depth representations obtained from the perspective of node pair attributes and network topology, and we concatenate them as the final representation vector of r_i-d_j to predict the probability of their association.

$$o = [o_{\text{topo}}; o_{\text{attr}}] \tag{22}$$

A multilayer perceptron activated with $sigmoid$ accepts o and outputs its associated predicted probability p:

$$p = sigmoid(Wo + b) \tag{23}$$

The $loss_1$ between the predicted probability p and true probability g is defined as follows:

$$loss_1 = -\sum_p g\log(p) + (1-g)\log(1-p) \tag{24}$$

When there is a known association between r_i and d_j, g is 1; otherwise, g is 0.

To ensure the reliability of the learned topology representation o_{attr} and attribute representation o_{topo} and their ability to predict independently, both representations have their own classifier. The loss of the probability distribution is calculated as follows:

$$p^{topo} = sigmoid(W^{topo} o_{\text{topo}} + b^{topo}) \tag{25}$$

$$loss_2 = -\sum_{p^{topo}} g\log(p^{topo}) + (1-g)\log(1-p^{topo}) \tag{26}$$

$$p^{attr} = sigmoid(W^{attr} o_{\text{attr}} + b^{attr}) \tag{27}$$

$$loss_3 = -\sum_{p^{attr}} g\log(p^{attr}) + (1-g)\log(1-p^{attr}) \tag{28}$$

p^{topo} and p^{attr} are the associated prediction probability distributions obtained from topological information and attribute information, respectively, The introduction of supervision information in $loss_2$ and $loss_3$ improves the attribute representation and topological representation capabilities in the model and makes the correlation prediction of (r_i, d_j) more comprehensive. The final loss L of the proposed model is defined as follows:

$$L = \lambda_1 loss_1 + \lambda_2 loss_2 + \lambda_3 loss_3 \tag{29}$$

$\lambda_1, \lambda_2, \lambda_3$ are hyperparameters used to weigh the proportion of each loss item in training.

3 Experimental Evaluations and Discussions

3.1 Evaluation metrics

Five-fold cross-validation was used to systematically evaluate MTRD's performance. All known drug-disease associations were treated as positive samples and randomly divided into five equal parts. Four of them were used for training, and the remaining one was set as unknown for testing. All drug-disease associations that have not been observed were considered negative cases. We randomly selected negative samples with an equal number of positive samples for training, and the remaining negative samples were used for testing.

Evaluation metrics included the area under the ROC curve (AUC) and the area under the precision-recall curve (AUPR). The samples whose predicted score is

greater than a threshold θ are considered positive examples; otherwise, they are considered negative examples. True Positive Rate (TPR) and False Positive Rate (FPR) under θ are defined as follows:

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP} \tag{30}$$

where TP (TN) is the number of samples correctly identified as positive samples (negative samples), and FP (FN) is the number of false positive samples (negative samples).

The ROC curve shows the performance of the model at all classification thresholds, and the AUC is considered an evaluation criterion because of its statistical consistency[36]. The number of negative samples is far greater than that of positive samples, and there is an obvious category imbalance. When evaluating the prediction results of category imbalance, the precision-recall curve is more informative than the ROC [37]. Therefore, we also used AUPR for evaluation. Precision and Recall are defined as follows:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \tag{31}$$

Precision represents the proportion of positive identifications that was actually correct. Recall and TPR are equivalent. We first calculated the TP, FP, FN, and TN in each fold experiment, and used the mean of the five-fold crossover to obtain the final score.

Biologists are more willing to choose the top-ranked drug-disease association prediction results for further experimental verification. Therefore, we also calculated the average recall value on the top k prediction results in the five-fold crossover as another performance indicator standard of the model. The higher the recall value, the higher is the score given by the model contains more real associations, and the more reliable is the prediction result.

3.2 Comparison with other methods

MTRD was compared with six state-of-the-art methods for drug-disease association prediction, including GFPred [30], CBPred[38], SCMFDD[26], LRSSL [25], MBiRW[22], and HGBI[39].

In MSTD, the random walk step size S is set to 3, the restart probability is α=0.3, and both the forward and reverse hidden layer output dimensions of the BiLSTM module are 80. The number of hidden layer neurons in the attention mechanism is half of the number of input neurons, kernal size=2×2, padding=1 in each layer of CNN, the convolution operation stride=1, the channels are 16, 32, and the pooling operation stride=2. $\lambda_1, \lambda_2, \lambda_3 = [0.6, 0.2, 0.2]$, $lr = 0.001$ and the early stopping strategy (the loss value on the training set does not fluctuate more than 5% for 10 consecutive rounds) was used to train 150 epochs. For a fair comparison, MTRD and the other six methods use the same data set. Furthermore, the other six methods use their own optimal parameter settings for reporting: drug attri-encoder with four neuron layers (4000,2000,1000,700), disease attri-encoder with two neuron layers (1000,700), combination module with two convolutional layers (kernels 2×4 and 3×3, poolings 2×2 and 1×3, and channels 16 and 32) for GFPred; lr = 0.001 and λ = 0.12 for CBPred; k=45%,λ = 4 and u = 1 for SCMFDD; $u = \lambda = 0.01$, k = 10 and γ = 2 for LRSSL; $d = log(9999)$, c = -11,α = 0.3 and $r = l$ =2 for MBiRW; and α= 0.4 for HGBI.

In each fold, we obtained the prediction score for each drug-disease pair and calculated the average of five crossover experiments as the final prediction score. The ROC curves and P-R curves of all the drugs evaluated using the seven methods are shown in Figure 5.

On comparing all methods for 763 diseases, we observefd that MTRD reached the highest 0.973 with an average AUC, which was 2.8% higher than GFPred, 4.7% higher than CBPred, 25% higher than SCMFDD, 14.2% higher than LRSSL, 14.5% higher than MBiRW, and 27.1% higher than HGBI.

HGBI and SCMFDD only consider some similarity information between nodes, and have the lowest performance among the seven methods. The performance of MBiRW using the comprehensive similarity metric was improved by 12.6% compared to that of HGBI and 10.5% higher than that of SCMFDD. LRSSL integrates multiple attribute information of drugs and diseases based on sparse subspace learning, which is better than the above-mentioned similarity-based learning. The method performance has been further improved (0.3% for MBiRW,12.9% for HGBI, and 10.8% for SCMFDD). Two deep

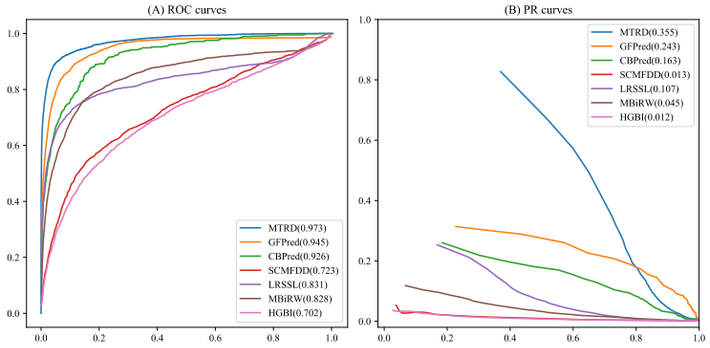

Fig. 5. ROC curves and PR curves of different methods for drug-disease association prediction.

learning-based methods, GFPred and CBPred, not only integrate the multi-source attribute information but also consider the neighbor node attributes of the drug node and the route between the drug disease nodes; this significantly improved the prediction performance (0.945 and 0.926, respectively). However, for AUPR, CBPred's performance was 1.9% lower than that of LRSSL's 0.107. One possible reason is that the known drug-disease nodes have very few associations and the network is sparse, which makes it difficult for the model to find enough paths to learn the expression of the drug-disease pairs. Compared with all the above methods, MTRD formally takes the multi-scale topology and attribute information of the network as the main learning objectives of the model; on the one hand, it integrates the attributes of neighbor nodes, and on the other hand, it avoids the loss of prediction accuracy caused by the lack of path information. MTRD also achieved the highest score of 0.355 on AUPR, which was 11%, 26.7%, 33.2%, 24.8%, 31%, and 34.3% higher than GFPred, CBPred, SCMFDD, LRSSL, MBiRW, and HGBI, respectively.

Table 1. The statistical results of the paired Wilcoxon test on the AUCs over all the 763 drugs by comparing MTRD and all other six methods.

	GFPred	CBPred	SCMFDD	LRSSL	MBiRW	HGBI
p-value of AUC	1.95384e-28	1.01639e-62	1.35768e-25	1.16259e-05	6.80685e-03	2.45055e-44
p-value of AUPR	1.31464e-56	3.73263e-31	5.01886e-50	8.31601e-65	4.99916e-65	4.80922e-64

We used the Wilcoxon test to verify whether MTRD was better. We obtained the average AUC and AUPR of seven groups implementing the seven methods in the five-fold crossover experiment, and each group contained the AUC and AUPR scores of 763 drugs. We calculated the results of six pairs of Wilcoxon tests between our method and other methods. The statistical results shown in the table 1 revealed that our method is significantly better than the other six methods (p-value< 0.05).

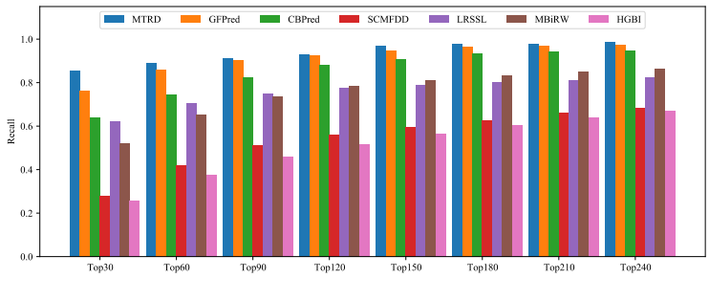

Fig. 6. The average recalls over all the drugs at different top k values.

The candidate association recall rate of the top k is shown in 6. The higher the recall rate, the more accurate is identification of drug-disease associations by the model. The average recall rates of MTRD for 763 drugs were 85.50% in the top 30, 89.13% in the top 60, and 91.10% in the top 90, which is significantly higher than those of the other six methods. MTRD was higher than GFPred by 11.21%, 8.34%, and 5.90% for the top 30, top 60, and top 90, respectively, and 16.70%, 12.21%, and 9.63% higher than CBPred. The recall of MTRD, GFPred, and

CBPred in the top k are significantly higher than those of the other four methods. The recall rates of LRSSL and MBiRW were very similar. When the values of k are 30, 60, and 90, the former are 62.12%, 70.61%, and 74.88%, which is slightly higher than the latter's 53.06%, 66.30%, and 74.71%, respectively. When the value of k is 150, 180, and 210, MBiRW performs slightly better than LRSSL. The recall rates of SCMFDD were 27.97%, 50.68%, 59.47%, and 65.97% in the top 30, 90, 150, and 210, respectively, and those of HGBI were 25.70%, 45.61%, 55.58%, and 63.47%, respectively, which were the worst among the seven methods.

3.3 Case studies on five drugs

We conducted case studies on five drugs: ampicillin, ciprofloxacin, doxorubicin, etoposide, and hydrocortisone. This study further proved the ability of MTRD to predict drug-disease associations. We used descending order to sort the candidate disease associations of each drug, and collected and analyzed the top 10 disease candidate results in the table 2.

The Comparative Toxicogenomics Database (CTD) is a literature-based toxicological information database that is reviewed and manually sorted by professional biocurators. It encodes and correlates important information about human diseases, such as drugs, genes, and phenotypes [40]. The DrugBank database contains detailed information about FDA-approved drugs and experimental drugs that have passed the FDA approval process [31]. In the table 2, there are 38 and 29 drug-disease associations included in CTD and DrugBank, respectively, indicating that these drugs can indeed be used to treat the predicted diseases.

ClinicalTrials.gov (http://clinicaltrials.gov) is the world's largest online clinical trial registry, which includes various trials funded by state grants or private funds around the world. In the case analysis, we used the experimental records included in the case, and the status was marked as "completed" as the supporting material for the candidate disease. PubChem[32] is currently the largest collection of chemical information supported by the US National Institutes of Health (NIH), which also includes associated disorders and diseases of small-molecule drugs. Forty candidate associations were found in clinical trials, and 25 were found in PubChem. They indicated that these drug-disease associations have been verified in clinical trials.

In addition to manually curated associations, CTD also includes drug and disease associations inferred from the literature. In the case study, the relationship between Pseudomonas infections and ampicillin was inferred from five studies in the CTD. This indicates that Pseudomonas infections are very likely to be treated with ampicillin. Among all 50 candidate drug-disease associations, there is one association that has not been confirmed by observational evidence and is marked as "unconfirmed." In conclusion, the case analysis further demonstrates the ability of MTRD to determine potential drug-disease associations.

3.4 Prediction of novel drug-disease associations

Finally, we used all known drug-disease associations to train the model and predict disease candidates for the drug of interest. The top 15 disease candidates predicted by MTRD are listed in Supplementary Table ST1. It can be used to help biologists discover new drug-disease associations through further wet-lab experiments.

4 Conclusion

We proposed a method to encode and fuse the similarity and association connections with various scales within multiple bi-layer networks for prediction of candidate drug-related diseases. The constructed multiple drug-disease networks are benefit for formulating their specific neighboring topology embeddings based on random walks. The multi-scale pairwise neighboring topology embeddings are formed by adjusting the ranges that the random walker may walk in the networks. A Bi-LSTM based module is constructed to capture the context relationships among multi-scale topology embeddings, and to learn the pairwise topology representation. A multi-layer convolutional networks based module is also constructed to encode and integrate the various pairwise attributes from multiple heterogenous networks. Two attention mechanisms are established to assign greater weights for the more informative features and topology embeddings. The cross-validation results indicated that MTRD outperforms than other six advanced methods in terms of both AUC and AUPR. The higher recall rates in the top ranked candidates and the case studies demonstrated MTRD's ability in discovering the potential drug-disease associations.

Funding

The work was supported by the Natural Science Foundation of China (61972135, 62172143); Natural Science Foundation of Heilongjiang Province (LH2019F049 and LH2019A029); China Postdoctoral Science Foundation (2019M650069, 2020M670939); Hei-longjiang 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 (YJSCX2021-199HLJU).

Table 2. The top 10 candidate diseases of five drugs.

Drug name	Rank	Disease name	Evidence	Rank	Disease name	Evidence
ampicillin	1	Streptococcal Infections	CTD,DrugBank, ClinicalTrials, PubChem	6	Bone Diseases Infectious	ClinicalTrials
	2	Pneumonia Bacterial	CTD,DrugBank, ClinicalTrials, PubChem	7	Bacterial Infections	CTD, DrugBank, ClinicalTrials
	3	Pseudomonas Infections	inferer by 5 literature	8	Sinusitis	CTD, PubChem
	4	Proteus Infections	CTD	9	Wound Infection	ClinicalTrials, PubChem
	5	Septicemia	CTD, DrugBank, ClinicalTrials	10	Otitis Media	ClinicalTrials, PubChem
ciprofloxacin	1	Streptococcal Infections	DrugBank, ClinicalTrials, PubChem	6	Bronchitis	CTD,DrugBank, ClinicalTrials
	2	Soft Tissue Infections	ClinicalTrials, PubChem	7	Gram-Negative Bacterial Infections	CTD,DrugBank, ClinicalTrials
	3	Pneumonia Bacterial	CTD, DrugBank, ClinicalTrials, PubChem	8	Serratia Infections	DrugBank
	4	Salmonella Infections	CTD, DrugBank, PubChem	9	Bone Diseases Infectious	CTD, ClinicalTrials
	5	Chlamydia Infections	CTD, DrugBank	10	Bacterial Infections	CTD, DrugBank, ClinicalTrials
doxorubicin	1	Neoplasms	DrugBank, ClinicalTrials, PubChem	6	Esophageal Neoplasms	CTD, DrugBank, ClinicalTrials
	2	Leukemia Myeloid Acute	DrugBank, ClinicalTrials	7	Carcinoma Pancreatic Ductal	CTD, DrugBank, ClinicalTrials, PubChem
	3	Head and Neck Neoplasms	CTD, DrugBank, ClinicalTrials	8	Leukemia Lymphoid	CTD, DrugBank, ClinicalTrials
	4	Soft Tissue Neoplasms	CTD, DrugBank, ClinicalTrials	9	Trophoblastic Neoplasms	DrugBank
	5	Carcinoma Islet Cell	CTD, DrugBank, ClinicalTrials, PubChem	10	Hodgkin Disease	CTD, DrugBank, ClinicalTrials
etoposide	1	Sarcoma	CTD, DrugBank, ClinicalTrials, PubChem	6	Breast Neoplasms	CTD, DrugBank, ClinicalTrials, PubChem
	2	Leukemia	CTD, DrugBank, ClinicalTrials, PubChem	7	Precursor Cell Lymphoblastic Leukemia-Lymphoma	CTD, DrugBank, ClinicalTrials, PubChem
	3	Neoplasms	DrugBank, ClinicalTrials, PubChem	8	Cystitis	CTD
	4	Lymphoma	CTD, DrugBank, ClinicalTrials, PubChem	9	Lymphoma Non-Hodgkin	CTD, DrugBank, ClinicalTrials, PubChem
	5	Leukemia Lymphoid	CTD, DrugBank, ClinicalTrials, PubChem	10	Hodgkin Disease	CTD, DrugBank, ClinicalTrials, PubChem
hydrocortisone	1	Dermatitis	CTD, DrugBank, ClinicalTrials, PubChem	6	Gastrointestinal Diseases	CTD, ClinicalTrials
	2	Skin Diseases	CTD, DrugBank, ClinicalTrials, PubChem	7	Photosensitivity Disorders	unconfirmed
	3	Pruritus	CTD, DrugBank, ClinicalTrials, PubChem	8	Collagen Diseases	CTD, DrugBank, ClinicalTrials
	4	Psoriasis	CTD, DrugBank, ClinicalTrials	9	Keratosis	CTD
	5	Rosacea	CTD	10	Scleroderma Systemic	ClinicalTrials

Key Points

- Three bi-layer heterogeneous networks are established, and each of them is composed of the drug and disease nodes, and their related similarity connections and association connections. Each bi-layer network has its specific topology information, which is benefit for its subsequent topology representation learning.
- Multiple kinds of drug similarities reflect topologies among the drug nodes from different perspectives. At the same time, each drug node (or disease node) has neighboring topologies with multiple scales. Therefore, we design the strategy based on random walks in the networks to construct various topology embeddings.
- A new framework based on Bi-LSTM and convolutional networks is constructed to learn the neighboring topology representation and attribute representation of a pair of drug and disease nodes. The topology representation reveals the multi-scale topology structures of multiple bi-layer networks, while the attribute representation encodes the similarity and association attributes of the node pair.
- Due to the various contributions of different features and different topological scales to drug-disease association prediction, we establish the feature-level and scale-level attention mechanisms to discriminate their contributions.

References

[1]Karl B. Ashburn, Ted T.and Thor. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8):673–683, Aug 2004.

[2]Michael Dickson and Jean Paul Gagnon. Key factors in the rising cost of new drug discovery and development. *Nature reviews Drug discovery*, 3(5):417–429, 2004.

[3]Nihad AM Tamimi and Peter Ellis. Drug development: from concept to marketing! *Nephron Clinical Practice*, 113(3):c125–c131, 2009.

[4]Maryam Lotfi Shahreza, Nasser Ghadiri, Sayed Rasoul Mousavi, Jaleh Varshosaz, and James R Green. A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics*, 19(5):878–892, 02 2017.

[5]BM Padhy, YK Gupta, et al. Drug repositioning: re-investigating existing drugs for new therapeutic indications. *Journal of Postgraduate Medicine*, 57(2):153, 2011.

[6]Hanqing Xue, Jie Li, Haozhe Xie, and Yadong Wang. Review of drug repositioning approaches and resources. *International journal of biological sciences*, 14(10):1232–1244, Jul 2018.

[7]Sudeep Pushpakom, Francesco Iorio, Patrick A. Eyers, K. Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, Alan Norris, Philippe Sanseau, David Cavalla, and Munir Pirmohamed. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1):41–58, Jan 2019.

[8]Nicola Nosengo. Can you teach old drugs new tricks? *Nature News*, 534(7607):314, 2016.

[9]Jayne-Louise E Pritchard, Tracy A O'Mara, and Dylan M Glubb. Enhancing the promise of drug repositioning through genetics. *Frontiers in pharmacology*, 8:896, 2017.

[10]MR Hurle, L Yang, Q Xie, DK Rajpal, P Sanseau, and P Agarwal. Computational drug repositioning: from data to therapeutics. *Clinical Pharmacology & Therapeutics*, 93(4):335–341, 2013.

[11]Jiao Li, Si Zheng, Bin Chen, Atul J Butte, S Joshua Swamidass, and Zhiyong Lu. A survey of current trends in computational drug repositioning. *Briefings in bioinformatics*, 17(1):2–12, 2016.

[12]Xing Chen, Biao Ren, Ming Chen, Quanxin Wang, Lixin Zhang, and Guiying Yan. Nllss: predicting synergistic drug combinations based on semi-supervised learning. *PLoS computational biology*, 12(7):e1004975, 2016.

[13]G. Ceddia, P. Pinoli, S. Ceri, and M. Masseroli. Matrix factorization-based technique for drug repurposing predictions. *IEEE Journal of Biomedical and Health Informatics*, 24(11):3162–3172, Nov 2020.

[14]Philippe Sanseau, Pankaj Agarwal, Michael R. Barnes, Tomi Pastinen, J. Brent Richards, Lon R. Cardon, and Vincent Mooser. Use of genome-wide association studies for drug repositioning. *Nature Biotechnology*, 30(4):317–320, Apr 2012.

[15]Kevin Bleakley and Yoshihiro Yamanishi. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics*, 25(18):2397–2403, 07 2009.

[16]Feixiong Cheng, Chuang Liu, Jing Jiang, Weiqiang Lu, Weihua Li, Guixia Liu, Weixing Zhou, Jin Huang, and Yun Tang. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol*, 8(5):e1002503, 2012.

[17]Shobeir Fakhraei, Bert Huang, Louiqa Raschid, and Lise Getoor. Network-based drug-target interaction prediction with probabilistic soft logic. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(5):775–787, 2014.

[18]Hee Sook Lee, Taejeong Bae, Ji-Hyun Lee, Dae Gyu Kim, Young Sun Oh, Yeongjun Jang, Ji-Tea Kim, Jong-Jun Lee, Alessio Innocenti, Claudiu T Supuran, et al. Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug. *BMC systems biology*, 6(1):1–10, 2012.

[19]Liang Yu, Jianbin Huang, Zhixin Ma, Jing Zhang, Yapeng Zou, and Lin Gao. Inferring drug-disease associations based on known protein complexes. *BMC medical genomics*, 8(2):1–13, 2015.

[20]Victor Martinez, Carmen Navarro, Carlos Cano, Waldo Fajardo, and Armando Blanco. Drugnet: network-based drug–disease prioritization by integrating heterogeneous data. *Artificial intelligence in medicine*, 63(1):41–49, 2015.

[21]Yongcui Wang, Shilong Chen, Naiyang Deng, and Yong Wang. Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data. *PloS one*, 8(11):e78518, 2013.

[22]Huimin Luo, Jianxin Wang, Min Li, Junwei Luo, Xiaoqing Peng, Fang-Xiang Wu, and Yi Pan. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics*, 32(17):2664–2671, 2016.

[23]Hailin Chen and Zuping Zhang. Prediction of Drug-Disease Associations for Drug Repositioning Through Drug-miRNA-Disease Heterogeneous Network. *IEEE Access*, 6:45281–45287, 2018. Conference Name: IEEE Access.

[24]Huimin Luo, Jianxin Wang, Min Li, Junwei Luo, Peng Ni, Kaijie Zhao, Fang-Xiang Wu, and Yi Pan. Computational Drug Repositioning with Random Walk on a Heterogeneous Network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(6):1890–1900, November 2019. Conference Name: IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[25]Xujun Liang, Pengfei Zhang, Lu Yan, Ying Fu, Fang Peng, Lingzhi Qu, Meiying Shao, Yongheng Chen, and Zhuchu Chen. Lrssl: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics*, 33(8):1187–1196, 2017.

[26]Wen Zhang, Xiang Yue, Weiran Lin, Wenjian Wu, Ruoqi Liu, Feng Huang, and Feng Liu. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics*, 19(1):1–12, 2018.

[27]Ming Zhang, Gerold Schmitt-Ulms, Christine Sato, Zhengrui Xi, Yalun Zhang, Ye Zhou, Peter St George-Hyslop, and Ekaterina Rogaeva. Drug Repositioning for Alzheimer's Disease Based on Systematic 'omics' Data Mining. *PloS One*, 11(12):e0168812, 2016.

[28]Han-Jing Jiang, Yu-An Huang, and Zhu-Hong You. Predicting drug-disease associations via using gaussian interaction profile and kernel-based autoencoder. *BioMed research international*, 2019, 2019.

[29]Ping Xuan, Yilin Ye, Tiangang Zhang, Lianfeng Zhao, and Chang Sun. Convolutional neural network and bidirectional long short-term memory-based method for predicting drug–disease associations. *Cells*, 8(7):705, 2019.

[30]Ping Xuan, Ling Gao, Nan Sheng, Tiangang Zhang, and Toshiya Nakaguchi. Graph Convolutional Autoencoder and Fully-Connected Autoencoder with Attention Mechanism Based Method for Predicting Drug-Disease Associations. *IEEE journal of biomedical and health informatics*, 25(5):1793–1804, May 2021.

[31]David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, and Michael Wilson. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1):D1074–D1082, January 2018.

[32]Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. Pubchem substance and compound databases. *Nucleic acids research*, 44(D1):D1202–D1213, 2016.

[33]Alex Mitchell, Hsin-Yu Chang, Louise Daugherty, Matthew Fraser, Sarah Hunter, Rodrigo Lopez, Craig McAnulla, Conor McMenamin, Gift Nuka, Sebastien Pesseat, et al. The interpro protein families database: the classification resource after 15 years. *Nucleic acids research*, 43(D1):D213–D221, 2015.

[34]UniProt Consortium et al. Uniprot: the universal protein knowledgebase. *Nucleic acids research*, 46(5):2699, 2018.

[35]Dong Wang, Juan Wang, Ming Lu, Fei Song, and Qinghua Cui. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, 26(13):1644–1650, 05 2010.

[36]Charles X. Ling, Jin Huang, and Harry Zhang. AUC: A Better Measure than Accuracy in Comparing Learning Algorithms. In Yang Xiang and Brahim Chaib-draa, editors, *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, pages 329–341, Berlin, Heidelberg, 2003. Springer.

[37]Takaya Saito and Marc Rehmsmeier. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3):e0118432, March 2015. Publisher: Public Library of Science.

[38]Ping Xuan, Yilin Ye, Tiangang Zhang, Lianfeng Zhao, and Chang Sun. Convolutional neural network and bidirectional long short-term memory-based method for predicting drug–disease associations. *Cells*, 8(7), 2019.

[39]Wenhui Wang, Sen Yang, Xiang Zhang, and Jing Li. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics*, 30(20):2923–2930, 2014.

[40]Allan Peter Davis, Cynthia J. Grondin, Robin J. Johnson, Daniela Sciaky, Jolene Wiegers, Thomas C. Wiegers, and Carolyn J. Mattingly. Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Research*, 49(D1):D1138–D1143, January 2021. Publisher: Oxford Academic.