Methods xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning

Yang Zhang[a], Yang Qiu[a], Yuxin Cui[a], Shichao Liu[a,*], Wen Zhang[a,*]

[a] College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China

ARTICLE INFO

Keywords:
Drug-drug interactions
Network embedding
Deep auto-encoders
Missing link prediction
Positive-unlabeled learning

ABSTRACT

Drug-drug interactions (DDIs) are crucial for public health and patient safety, which has aroused widespread concern in academia and industry. The existing computational DDI prediction methods are mainly divided into four categories: literature extraction-based, similarity-based, matrix operations-based and network-based. A number of recent studies have revealed that integrating heterogeneous drug features is of significant importance for developing high-accuracy prediction models. Meanwhile, drugs that lack certain features could utilize other features to learn representations. However, it also brings some new challenges such as incomplete data, non-linear relations and heterogeneous properties. In this paper, we propose a multi-modal deep auto-encoders based drug representation learning method named DDI-MDAE, to predict DDIs from large-scale, noisy and sparse data. Our method aims to learn unified drug representations from multiple drug feature networks simultaneously using multi-modal deep auto-encoders. Then, we apply four operators on the learned drug embeddings to represent drug-drug pairs and adopt the random forest classifier to train models for predicting DDIs. The experimental results demonstrate the effectiveness of our proposed method for DDI prediction and significant improvement compared to other state-of-the-art benchmark methods. Moreover, we apply a specialized random forest classifier in the positive-unlabeled (PU) learning setting to enhance the prediction accuracy. Experimental results reveal that the model improved by PU learning outperforms the original method DDI-MDAE by 7.1% and 6.2% improvement in AUPR metric respectively on 3-fold cross-validation (3-CV) and 5-fold cross-validation (5-CV). And in F-measure metric, the improved model gains 10.4% and 8.4% improvement over DDI-MDAE respectively on 3-CV and 5-CV. The usefulness of DDI-MDAE is further demonstrated by case studies.

1. Introduction

With the number of drugs developed increasing rapidly over the past decade, combination drug therapies have a bright future to treat diseases. Simultaneous administration of two or more drugs would lead to unexpected influence on each other which is termed drug-drug interactions (DDIs). The reactions to influenced drugs could be divided into three categories: synergistic, antagonistic and no reaction [1]. The synergistic reaction is the best result for patients' treatment which means that the efficacy compared to being administered separately is greater. The type of no reaction can neither enhance nor weaken the efficacy of drugs. Antagonistic reaction not only results in reducing the efficacy, but also may lead to adverse effects which could even threaten patients' lives. Hence, detecting DDIs has aroused considerable attention in public health safety and drug safety surveillance.

Identifying DDIs through traditional clinical trials and vitro methods is considerably costly and time-consuming. However, the rapid development of technology is conducive to the collection of different types of drug data, which offers promising possibilities in the field of computational drug discovery and drug safety research. Hence, plenty of computational prediction methods have been proposed, which contributes to safer and more effective prescription to patients during drug development and helps to reduce the time and cost of clinical trials. Unstructured biomedical literature such as medical reports, documents, scientific journals and freely available drug-specific databases provide rich and fresh information about drugs. Hence, drug data are mainly from literature and databases.

Due to the successful applications of NLP related techniques, computational methods based on literature extract drug-drug DDIs through statistical and text mining techniques to make accurate predictions. FBK-irst team [2] proposed a method applying a hybrid kernel based RE classifier which achieves an F-score of 0.651 ranking top in DDIExtraction 2013 challenge. Kim et al. [3] developed a two-stage method gaining an F-score of 0.670 which is based on linear SVM utilizing rich

* Corresponding authors.
E-mail addresses: scliu@mail.hzau.edu.cn (S. Liu), zhangwen@mail.hzau.edu.cn (W. Zhang).

https://doi.org/10.1016/j.ymeth.2020.05.007
Received 9 February 2020; Received in revised form 6 May 2020; Accepted 13 May 2020

features on the same corpus. The application of the deep convolutional neural network is promising as well. Sun et al. [1] proposed DCNN which could generate suitable feature set automatically with a deep architecture with multiple layers of small convolutions. Zhang et al. [4] adopted a hierarchical RNNs model to concatenate the sentence sequence and the shortest dependency path (SDP). Jiang et al. [5] considered a skeleton-LSTM network to extract the intrinsic structure of DDI. Xu et al. [6] utilized user-generated content (UGC) resource to obtain fresh and rich information through a full-attention mechanism.

Similarity-based methods utilize data from databases, which are based on the assumption that similar drugs may have interactions with the same drug. Vilar et al. [7] proposed a simple and applicable matrix calculation method utilizing the molecular structural similarity to predict DDIs. Gottlieb et al. [8] introduced Inferring Drug Interactions (INDI) method, which considers seven drug-drug similarities allowing the inference of pharmacokinetic and pharmacodynamic DDIs. Ma et al. [9] presented multi-view graph auto-encoders (GAE) considering each type of feature as a view and employing attention mechanism on integrated drug similarities, then extended the model to semi-supervised and transductive settings.

Matrix operations-based methods are powerful to identify DDIs of the new drugs without any known interactions. Zhang et al. [10] proposed the sparse feature learning ensemble method with linear neighborhood regularization (SFLLN) combining four drug features and the known drug-drug interactions. Some methods were proposed to predict comprehensive DDIs further. For instance, Shi et al. [11] presented TMFUF using triple matrix factorization and a unified predicting model. Yu et al. [12] developed DDINMF method based on semi-nonnegative matrix factorization.

Network-based methods infer potential interactions through the network structure based on databases. Zhang et al. [13] developed a label propagation approach considering high-order similarity to make predictions. Park et al. [14] simulated signaling propagation on the protein–protein interaction network through random walk with restart algorithm. Sridhar et al. [15] proposed a probabilistic approach to infer missing interactions from a network of multiple drug-based similarities and known interactions jointly. Huang et al. [16] developed a metric termed “S-score” to measure the strength of network connections for predictions.

In addition, ensemble learning combines several machine learning models in a certain way, and generally could achieve better performances than separate models. Deepika et al. [17] adopted a semi-supervised learning framework with network representation learning and meta-learning from four drug datasets to predict DDIs. Zhang et al. [18] proposed a sparse feature learning ensemble method combing similarity-based methods and matrix perturbation method.

Although these computational methods have performed well, it has been revealed in recent studies that integrating heterogeneous drug features is of great significance for DDI prediction [19–21], which brings great challenges on learning from multiple biomedical data sources. First of all, statistical properties vary from drug datasets. Hence, shallow models may achieve poor performance on extracting high-level concepts from multiple data sources. Secondly, relations between biomedical events are usually nonlinear across all types of drug features. Third, incompleteness of some drug datasets (e.g. missing labels, noise data) may have adverse influence on prediction models. In addition, most existing methods regard known drug-pairs with interactions as positive instances while unknown drug pairs as negative pairs [21,22]. However, there may exist a considerable number of real positives unknown yet (false negatives). Hence, the performances of prediction models would be affected negatively if the method learning from positives and negatives is directly applied to positives and unlabeled data.

To address the aforementioned problems, we present a multi-modal deep auto-encoders based drug representation learning method for DDI prediction, named DDI-MDAE. In this paper, five drug data sources are regarded as respective forms of drug feature networks including drug-drug interaction network, drug–chemical substructure network, drug-target network, drug-enzyme network and drug-pathway network. Then, we learn unified representations of drugs from these five drug feature networks simultaneously by multi-modal deep auto-encoders. We apply several operators to the learned drug embeddings for representing drug-drug pairs, and then adopt a random forest classifier to train models to predict accurate drug-drug interactions. Furthermore, we improve our method by PU learning. To summarize, our investigation makes four primary contributions:

- We introduce DDI-MDAE, an efficient multi-modal deep auto-encoders based drug representation learning method, which provides a new insight for learning unified representations of drugs with multiple drug features simultaneously. Further DDI-MDAE could predict for drugs with incomplete features even faced with large-scale, noisy and sparse data.
- We adopt several operators on the learned drug embeddings to represent drug-drug pairs, and a random forest classifier is trained to predict potential drug-drug interactions.
- We extensively evaluate our models on several tasks, and experimental results show that DDI-MDAE has high efficiency and robustness, even trained with sparse drug features.
- We improve the proposed DDI-MDAE by applying a specialized random forest classifier in the positive-unlabeled learning setting to enhance predictive accuracy. And experimental results show that the improved method outperforms the original method DDI-MDAE in most cases.

2. Background

In this section, we introduce the background of the proposed method, including network embedding, deep neural networks and positive-unlabeled learning.

2.1. Network embedding

Network embedding (a.k.a. graph embedding) learning that aims to automatically learn low-dimensional node representations, has aroused increasing attention in the biomedical network analysis in recent years [23].

Some earlier methods such as Linear Embedding [24], IsoMAP [25] construct the affinity graph based on the feature vectors and embed the affinity graph into a low-dimension space. Recently, DeepWalk [26] samples the graph structure into a stream of random walks, with which a skip-gram model is trained to predict the path of the random walk. LINE [27] defines a loss function to capture both 1-step and 2-step local structure information. Node2vec [28] presents a bias random walk to explore neighborhoods incorporating breadth-first search and depth-first search strategies to sample the network. There are several deep learning based methods applied to network embedding. SDNE [29] utilizes deep auto-encoders to preserve the first-order and second-order network proximities. GCN [30] designs the variant of convolutional neural networks to encode both local graph structure and features of nodes.

Additionally, there are several studies for learning representations in the scenario of heterogeneous network [31–33] and attributed network [34–36]. These models embed the complex networks with multiple modalities into a shared, low-dimensional space. Inspired by pioneer work, this work utilizes structural network embedding to learn the unified representations of drugs from multiple drug datasets.

2.2. Deep neural networks

Deep neural networks (DNN) have been applied to feature learning in many fields [37–39] successfully. Different from the previous work,

we adopt deep neural networks to manage network representation learning. DNN like auto-encoders are able to learn highly non-linear projects which map data from a high dimensional space into a lower dimensional space. Auto-encoder is a non-linear unsupervised neural network model that has become one of the commonly used building blocks [40]. An auto-encoder is composed of two parts including the encoder and decoder. In the encoder part, a deterministic mapping f_θ transforms an input vector x into hidden representation y. σ is a non-linear activation function (e.g. sigmoid function). The decoder maps the resulting hidden representation y to a reconstructed dimension $\hat{x}$. $\theta = (W, b)$ and $\hat{\theta} = (\hat{W}, \hat{b})$ are the parameter sets to be determined in the training process.

$$y = f_\theta(x) = \sigma(Wx + b)$$
$$\hat{x} = g_{\hat{\theta}}(y) = \sigma(\hat{W}y + \hat{b}) \quad (1)$$

The goal of the auto-encoder is to minimize the reconstruction error of the input and the output. The loss function of auto-encoder is given as below:

$$L\left(\hat{x}, x\right) = \sum_{i=1}^{n} \left|\left|\hat{x}_i - x_i\right|\right|_2^2 \quad (2)$$

where n denotes the dimension size of the input. Parameters θ and $\hat{\theta}$ can be optimized by gradient descent methods. Stacked auto-encoders contain multiple layers of such auto-encoders and learn suitable representations of each layer using the layer-wise training method [41].

2.3. Positive-unlabeled learning

Traditional binary classifiers in machine learning are learned from positive and negative data. However, negative samples are usually not available in real-world domains. Faced with the real-world problem: identifying positive data from a large amount of unlabeled data containing positive and negative samples, an advanced learning technique named positive-unlabeled learning (PU learning) was proposed to address it.

In PU learning task, there are two sets of samples: the positive set P and the unlabeled set U. Its ultimate goal is to identify appropriate positive data from U. For learning a classifier in PU learning settings, three kinds of methods are proposed basically: the direct method, the two-step method and bagging PU learning method.

As for the direct methods, Liu et al. [42] proposed a method based on a biased formulation of SVM. Elkan et al. [43] proposed LPU method to consider making learning from positive and unlabeled samples essentially equivalent to learning from positive and negative samples. And there are some other methods which aim to assign weights to unlabeled samples in certain rules and regard these unlabeled samples as weighted negative samples to train a classifier [44,45]. For instance, Lee et al. [44] proposed to learn with noise by regarding all unlabeled samples as negative and applied a linear function to learn from the noisy samples through weighted logistic regression algorithm.

Most methods adopt the two-step strategy. The first step is to extract a subset from the unlabeled samples as reliable negative samples. The second step is to build classifiers with an iterative classification algorithm. Different strategies could be adopted in the two steps. For the first step, 1-DNF [46], Spy technique [47], Rocchio [48], and Naïve Bayesian [42] are wildly applied. For the second step, the expectation–maximization and support vector machines are usually applied for iteratively learning.

Bagging PU learning methods are divided into two categories: inductive PU learning and transductive PU learning [49]. Inductive PU learning comprises aggregating classifiers trained to discriminate positive samples from a small random subsample of U and the predictions are averaged. Transductive PU learning, an extension of inductive PU learning, aggregates the predictions of the classifiers trained on subsamples excludes the current sample.

Most traditional methods mentioned above are wildly applied for text classification. Promisingly, PU learning technique has been applied in some bioinformatics research fields successfully such as disease gene identification [50–52], drug-target interaction prediction [53] and drug-drug interaction prediction [54,55]. For instance, Mordelet et al. [52] proposed a bagging method from positive and unlabeled samples named ProDiGe for disease gene prediction. Deepika et al. [17] adopted a PU learning-based classifier named the bagging SVM classifier to act as a meta-classifier in their work. Hameed et al. [55] considered clustering approach Self Organizing Map to predict drug-drug interactions. Zheng et al. [54] employed a two-step strategy using two techniques (i.e. one-class support vector machine and k-nearest neighbors) for drug-drug interactions prediction. Lan et al. [53] adopted three strategies (random walk with restarts, k-nearest neighbors and heat kernel diffusion) to classify unlabeled samples into two sets: reliable negative samples and likely negative samples. Then they employed different penalty factors to predict drug-target interactions.

In our paper, we collect our data from DrugBank [56] where labeled data can be accessed. However, the information provided by DrugBank is mostly about positive labels while the number of available negative labels is significantly small. Hence, positive-unlabeled learning could be used to address the problem. We adopt a typical direct method named LPU mentioned above. The proof which underlies this approach is that a classifier trained on positive and unlabeled examples predicts probabilities that differ by only a constant factor from the true conditional probabilities of being positive under the “selected completely at random” assumption [43]. We build a specialized random forest classifier in the PU learning setting to predict potential DDIs. More details will be illustrated in the following section.

3. Materials and methods

3.1. Datasets

DrugBank [56] is a comprehensive public database containing extensive drug information such as drug-drug interactions, drug substructures, drug targets, drug enzymes and drug transporters. In this paper, the application scenario that DDI-MDAE focuses on is to predict DDIs from large-scale, noisy and sparse data. Hence, we collect drugs that contain at least one feature or one interaction to develop our data set from DrugBank (version 5.1.0, released in April 2018). Drug targets are mapped to KEGG database [57] to obtain drug pathways. As shown in Table 1, our datasets have 2,367 drugs characterized by four different dimensions of features: 699 chemical substructures, 2,411 targets, 285 enzymes and 314 pathways.

Different drug features have different degrees of influence on DDI prediction. Hence, our paper integrates multiple drug features to learn drug representations and constructs an effective model to make DDI predictions. Each drug dataset is regarded as a view of drug feature network. And the known drug-drug interactions extracted from DrugBank are considered as a drug feature network as well. Consequently, we generate five drug feature networks. The details of these networks are shown in Table 2. For example, the drug-target network contains drug nodes and target nodes which are associated with the original datasets, and edges which denote associations between drugs and targets. Especially, We obtained 209,494 unique DDIs

Table 1
The descriptions about drug datasets.

Data type	Database	Description
Drug	DrugBank [56]	2,367 drug types
Target	DrugBank [56]	2,411 target types
Enzyme	DrugBank [56]	285 enzyme types
Pathway	KEGG [57]	314 pathway types
Substructure	PubChem [58]	699 substructure types

Table 2
The properties of the five feature networks.

Network Type	Nodes	Edges	Average Degree	Density
DDI network	2,367	209,494	219.251	0.075
Drug-target network	4,347	10,885	5.008	0.001
Drug-enzyme network	1,501	4,199	5.595	0.004
Drug-pathway network	1,892	20,633	21.811	0.012
Drug-structure network	2,806	240,695	171.557	0.061

(positive instances) for the 2,367 drugs. There are 2,590,667 possible drug pairs (unlabeled instances) for these collected drugs. The positive DDIs to unlabeled DDIs ratio is approximately 1 to 12 (1:12).

3.2. Overview of method

In this section, we introduce our deep learning architecture for DDI predictions from five drug feature networks. Fig. 1 illustrates the overall architecture of our proposed model. With the purpose of capturing the multiple non-linear network structures, we present a structural network embedding model for learning drug representations using multi-modal deep auto-encoders, named DDI-MDAE. In our model, each drug feature network transformed as input, is trained by a deep auto-encoder channel (marked in different colors respectively in Fig. 1). We use different colors and shapes to represent the nodes of different feature networks. The representations of drugs in the hidden layer are shared to learn concurrently, which may contribute to capturing the correlations among the different networks. Then, we adopt four operators on the learned drug representations and then feed them into a random forest classifier to train models. Furthermore, we develop DDI-MDAE (PU learning) through adopting a random forest classifier in the positive-unlabeled learning setting to improve our original method for DDI prediction. Therefore, it's crucial to learn drug representations in our deep architecture. We will introduce the DDI-MDAE method for further details in the following section.

3.3. Multi-modal deep auto-encoders

As shown in Fig. 1, DDI-MDAE consists of several auto-encoder channels and a shared hidden layer, which aims at learning unified representations from five drug feature networks. It is expected to obtain the low-dimensional space in the shared hidden layer after the training process. The objective function of DDI-MDAE method is constrained by three parts:

- Reconstruction error of multiple auto-encoders

Owing to the sparsity of drug feature networks, we employ more penalty to reconstruction error of the non-zero elements than that of zero elements recommended in [29]. Therefore, we describe a revised loss function of reconstruction error as below:

$$H_1 = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \left\| \left(\hat{x}_i^t - x_i^t \right) \odot b_i \right\|_2^2 \tag{3}$$

where T denotes the number of input feature networks. Here, $T = 5$. We take the adjacency matrix $A = \{A^{(1)}, A^{(2)}, ..., A^{(T)}\}$ as the input of our model. n_t is the dimension size of $A^{(t)}$. x_i^t is the input of the auto-encoder. $\hat{x}_i^t$ is the reconstructed data by the auto-encoder. $\odot$ is the Hadamard product, and $b_i = \{b_{i,j}\}_{j=1}^{n}$. If node v_i and v_j are connected (i.e. $a_{i,j} = 1$), $b_{i,j} = \beta > 1$, else $b_{i,j} = 1$. If these nodes have similar neighborhood structure, nodes should have similar representations in the hidden layer to minimize the H_1 loss.

- Constrain of known drug-drug interactions

Our model is designed to learn drug representations from five drug feature networks containing a small amount of known drug-drug interactions. Inspired by [59,60], we assume that drugs with interactions are expected to be mapped near in the hidden embedding space. The loss function can be described as follows:

$$H_2 = \sum_{(i,j) \in A^{(ddi)}} a_{i,j} \left\| y_i^{(K)} - y_j^{(K)} \right\|_2^2 \tag{4}$$

where $y_i^{(K)}$ is the K-th layer hidden representation of drug v_i. K is the length of the projection layers. $A^{(ddi)}$ denotes the adjacency matrix of drug-drug interaction network. $a_{i,j} = 1$ if drug v_i and v_j are connected, otherwise $a_{i,j} = 0$.

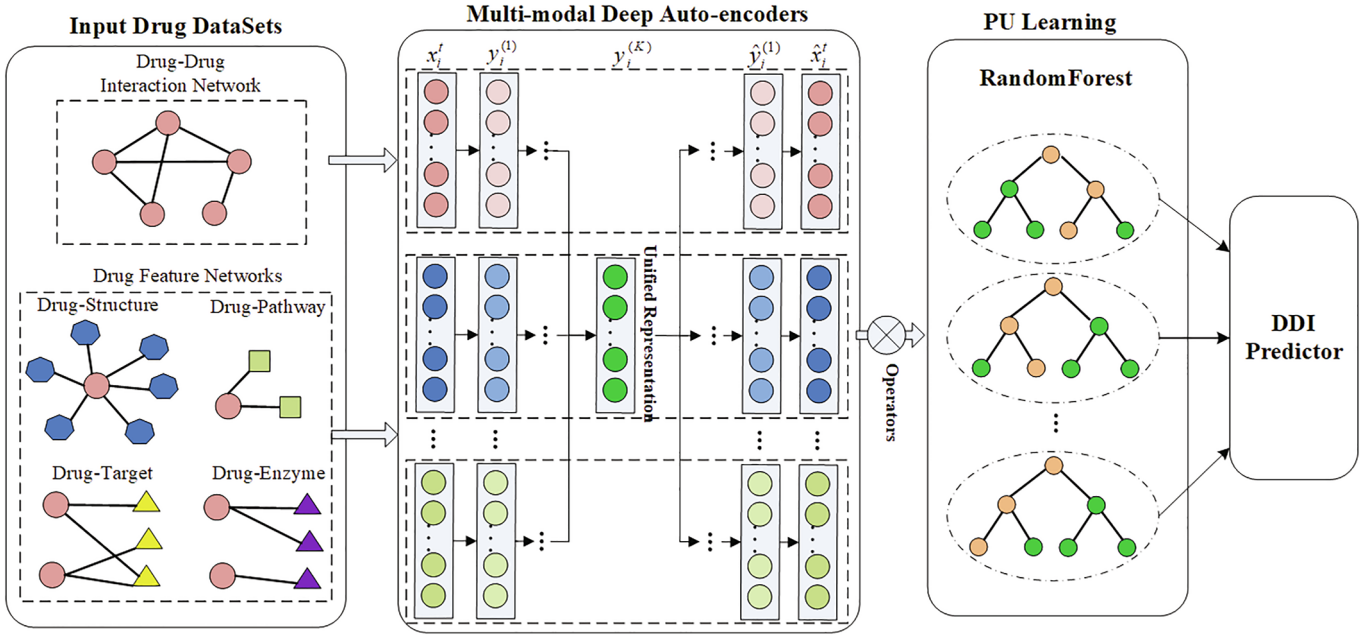


Fig. 1. Overview of MDAE with PU learning.

- Similarities constrain of drug feature networks

It is noteworthy that there may exist potential interactions between the drugs without observed links as they have some similar drug features. Different from the loss H_2 which pays attention to the constrain of known drug-drug interactions, we focus on the drug similarities constrained on drug pairs. We first construct five binary vectors for each drug v_i according to the five networks as d_i^{dd}, d_i^{dt}, d_i^{de}, d_i^{dp} and d_i^{ds} respectively. Then we concatenate the five vectors and obtain a comprehensive vector $D_i = [d_i^{dd}, d_i^{dt}, d_i^{de}, d_i^{dp}, d_i^{ds}]$ to represent drug v_i containing five drug feature networks. The dimension of drug feature vectors equals to the sum of all node types (i.e. drugs, targets, enzymes, pathways and substructures). We calculate the drug similarities based on the feature vectors using the Jaccard formula:

$$\mathscr{J}\left(D_i, D_j\right) = \frac{S_{11}}{S_{01} + S_{10} + S_{11}} \tag{5}$$

where S_{11} is the number of dimensions where D_i and D_j both have a value of 1; S_{01} is the number of dimensions where D_i has a value of 0 and D_j has a value of 1; S_{10} is the number of dimensions where D_i has a value of 1 and D_j has a value of 0. The loss function of this goal can be defined as follows:

$$H_3 = \sum_{i,j=1}^{n_d} \mathscr{J}\left(D_i, D_j\right) \left|\left| y_i^{(K)} - y_j^{(K)} \right|\right|_2^2 \tag{6}$$

where n_d is the number of drugs, $\mathscr{J}(D_i, D_j)$ is the Jaccard similarities of drug v_i and v_j.

To preserve the three objective functions above simultaneously, we define a unified loss function for DDI-MDAE as following:

$$H = H_1 + \alpha H_2 + \gamma H_3 + \lambda L_r \tag{7}$$

where $L_r$is a regularization term to address the over-fitting problem, defined as follows:

$$L_r = \sum_{k=1}^{K} \left(\left|\left| W^{(k)} \right|\right|_2^2 + \left|\left| \hat{W}^{(k)} \right|\right|_2^2 \right) \tag{8}$$

where K is the length of projection layers. $W^{(k)}$ denotes the weight matrix of the layer k in the encoder part of the auto-encoder while $\hat{W}^{(k)}$ denotes the weight matrix of the layer k in the decoder part.

To minimize the total loss H with parameter sets $\theta = (W, b)$ and $\hat{\theta} = (\hat{W}, \hat{b})$, the objective function (7) is optimized using Adam optimization method by Tensorflow [45]. We will describe the hyper parameters for our method in the experiment setting section.

3.4. Predicting drug-drug interactions

The DDI prediction task can be considered as a binary classifier problem, which regards drug-drug pairs as positive instances or negative instances. We obtain drug representations from the shared hidden layer of DDI-MDAE directly. Consequently, we design four operators on the drug representations to represent drug-drug pairs. Given two drugs v_i, v_j and their representations $\Phi(v_i)$, $\Phi(v_j)$, We can obtain different representations of drug pair respectively by four operators demonstrated in Table 3, where d is the dimension of drug representations. We intend to represent each drug pair even though there is no edge between them because our test set contains positive and negative edges in the built classifier.

Table 3
The descriptions of four operators to represent drug pairs.

Operator	Dimension	Description
Average	d	$\Phi\left(v_i, v_j\right) = \frac{1}{2}\left(\Phi(v_i) + \Phi(v_j)\right)$
Hadamard	d	$\Phi(v_i, v_j) = \Phi(v_i) \odot \Phi(v_j)$
L1-norm	d	$\Phi(v_i, v_j) = \lvert\Phi(v_i) - \Phi(v_j)\rvert$
Concatenate	$2d$	$\Phi(v_i, v_j) = \Phi(v_i) \oplus \Phi(v_j)$

After calculating the representations of drug pairs, we train a random forest classifier [61] to predict potential drug-drug interactions. A random forest is a meta estimator that fits many decision tree classifiers on various sub-samples of the datasets and uses averaging to improve the accuracy of prediction and control over-fitting problem.

To explain our approach in more detail, we describe DDI-MDAE method in Algorithm 1.

Algorithm 1 The algorithm of DDI-MDAE

Input: Drug feature networks with adjacency matrix $A = \{A^{(1)}, A^{(2)}, ..., A^{(T)}\}$ the parameters dimension d, α, β, γ, λ, learning rate
Output: newly predicted DDIs
1: Initialize parameter sets $\theta = (W, b)$, $\hat{\theta} = (\hat{W}, \hat{b})$.
2: Learn drug representations Φ:
3: **repeat**
4: Compute the loss function H based on Eq. 7.
5: Calculate $\partial H/\partial\theta$, $\partial H/\partial\hat{\theta}$ and use Adam optimizer to update θ, $\hat{\theta}$.
6: **until** convergence
7: Obtain the representations of drugs according to Eq. 1.
8: **for each** drug pairs (v_i, v_j) **do**
9: Representations $\Phi(v_i, v_j) = \Phi(v_i)$ operators $\Phi(v_j)$.
10: **end for**
11: Build the random forest classifier based on the representations of drug pairs and then predict unobserved DDIs.

3.5. Enhancing the DDI prediction using PU learning

For the training data that consists of positive instances and numerous unlabeled instances, most previous work and the proposed original method regard all unlabeled instances as negative ones, which would lead to adverse influence on the performance of prediction models. In our dataset, the positive DDIs to unlabeled DDIs ratio is approximately 1 to 12 (1:12). Hence, we improve DDI-MDAE by applying a specialized random forest classifier in the positive-unlabeled learning setting to enhance predictive performance. For trying to make learning from positive and unlabeled instances essentially equivalent to learning from positive and negative instances in a certain way, we adopt LPU method whose core concept is that a classifier trained on positive and unlabeled examples predicts probabilities that differ by only a constant factor from the true conditional probabilities of being positive [43]. The function of learning from positive and unlabeled instances are defined as below:

$$g(x) = p(\hat{y} = 1|x) \tag{9}$$

where x denotes the instance. $\hat{y} = 1$ if the instance x is labeled, else $\hat{y} = 0$.

Our goal is to learn a function $f(x) = p(y = 1|x)$. The core concept could be stated as:

$$\begin{aligned} p(\hat{y} = 1|x) &= p(y = 1 \cap \hat{y} = 1|x) \\ &= p(y = 1|x)p(\hat{y} = 1|y = 1, x) \\ &= p(y = 1|x)p(\hat{y} = 1|y = 1) \\ &= f(x)p(\hat{y} = 1|y = 1) \end{aligned} \tag{10}$$

The result follows by dividing each side by $p(\hat{y} = 1|y = 1)$. Hence, the constant probability $c = p(\hat{y} = 1|y = 1)$.

We feed the representations and label matrix of drug pairs (positive and unlabeled instances) into a random forest classifier [61] with 0.1 ratios of positive instances removed randomly to train the model. Let P be the selected positive instances set. Then instances in P are fed into the trained random forest classifier to obtain $g(x)$ for each instance. The estimator of c is calculated as below:

$$c = \frac{1}{n} \sum_{x \in P} g(x) \tag{11}$$

where n is the cardinality of P. Finally we apply the trained classifier into test set and then generate $f(x)$ according to Eq. 11.

4. Experiment

In this section, we will illustrate our experiment settings in details and present the performances of our proposed model compared to other benchmark methods.

4.1. Experiment setting

In this paper, we evaluate the performances of DDI prediction model using k-fold cross-validation (i.e. 3-CV, 5-CV). We split known DDIs into k subsets with equal size at random. Then We choose one subset as the testing set and the rest as the training set. In order to implement the DDI prediction, the existing links in the testing set are hidden and the left network is used to train DDI-MDAE model. Then, we utilize the obtained drug representations to predict the unobserved links. We apply both 3-fold cross-validation and 5-fold cross-validation to all methods respectively to evaluate the performances of our deep architecture compared to other models.

Several widely-used evaluation metrics are adopted to measure performances of prediction models including area under ROC curve (AUC), precision and recall, accuracy (ACC), F-measure (F1), and the area under the precision-recall curve (AUPR).

4.2. Parameter discussion

We adjust the parameters of benchmark methods in order to achieve optimal performances. Here, we consider different combinations of several key parameters. DDI-MDAE model contains 3 hidden layers. $d \in \{32, 64, 128, 192, 256\}$, is the embedding size of the shared layer, $\alpha \in \{0.01, 0.1, 0.2, 0.3, 0.4\}$, $\beta \in \{2, 5, 10, 15, 20\}$, $\gamma \in \{0, 0.001, 0.002, 0.005, 0.01\}$, regularization term $\lambda \in \{1e-6, 1e-5, 1e-4, 1e-3, 1e-2\}$ and learning rate $\in \{0.005, 0.01, 0.02, 0.03, 0.04\}$. As for the settings of the random forest classifier, we choose to keep the default settings of sklearn.ensemble.RandomForestClassifier [62] to make DDI predictions. We evaluate the four operators on the representations of drug pairs in the experiments.

We research how the different combinations of parameters affect the performance of our proposed method. We search for the suitable parameters with the initial values $d = 128$, $\alpha = 0.01$, $\beta = 10$, $\gamma = 0.01$, $\lambda = 1e-5$ and learning rate $= 0.01$. When analyzing the influence of a certain tested parameter, the remaining parameters are fixed with default values. We report the results of AUPR metric on 3-CV with different combinations of hyper parameters. As Fig. 2 shows, DDI-MDAE achieves the best AUPR score when $d = 128$ or 256. Parameter α is designed to measure the influence of known DDIs on DDI-MDAE. Our method is only constrained on H_1 and H_3 when $\alpha = 0$. It can be observed that the performance of the model using $\alpha = 0.01$ is better than that of $\alpha = 0$ and the AUPR score decreases when α increases, which demonstrates that α is essential for our method and should be carefully determined. As for parameter β, it is used to control the reconstruction of zero elements. The AUPR score increases as β increases from 2 to 10 and reaches a peak when $\beta = 10$ and drops after that. Therefore, we choose $\beta = 10$ as our experiment evaluation. Parameter γ aims at measuring the effect of drug similarities of multiple drug feature networks (i.e. H_3 defined in Section 3.3). The AUPR score of DDI-MDAE gains improvement as $\gamma = 0.001$ than that of $\gamma = 0$. As for parameter λ and learning rate have a slight effect on our method. It could be inferred that our model is robust to γ, λ and learning rate within the range of parameters we choose.

After analysis above, we adopt $d = 128$, $\alpha = 0.01$, $\beta = 10$, $\gamma = 0.001$, $\lambda = 1e-5$ and learning rate $= 0.01$ in our following evaluation experiments.

4.3. Benchmark methods

In this section, we introduce several state-of-the-art DDI prediction methods for comparison.

- **Collaborative Filtering** [7] is an intuitive method based on the assumption that similar drugs may have interactions with the same drug. We employ collaborative filtering on each drug feature network and obtain five collaborative filtering models, i.e. structure-based CF, target-based CF, enzyme-based CF, pathway-based CF and DDI-based CF.
- **Random Walk** [14] adopts a random walk with restart strategy to simulate signaling propagation in drug feature networks in order to capture the possibility of their distant interference. Same as collaborative filtering methods, we build five random walk models, i.e. structure-based RW, target-based RW, enzyme-based RW, pathway-based RW and DDI-based RW.
- **MP** [18] presents a matrix perturbation method based on the assumption that the network structure will not be changed by the random removal of a small portion of links. MP method will be used to compare with our proposed model in the sparsity experiment, which will be described in the following section.
- **LPMS** [13] proposes an integrative label propagation framework for DDI prediction with three similarities from drug chemical substructure, label side effects and off label side effects.
- **SFLLN** [10] is a sparse feature learning ensemble method with linear neighborhood regularization, which combines multiple drug features including chemical substructures, targets, enzymes, pathways and known DDIs for DDI prediction.
- **Node2vec** [28] is a network embedding method by training a series of tunable walks. We adopt this method for drug representations learning and use a random forest classifier to make predictions.

4.4. Results

In this section, we report the empirical results of our method compared to the benchmark methods on 3-fold and 5-fold cross-validation first. We adopt four operators (i.e. Average, Hadamard, $L1$-norm and Concatenate) and evaluate their effects on the performances of our model. As is shown in Table 4, the method with the concatenate operator on drug representation is highly stable and achieves the best result compared to other three operators on both 3-fold and 5-fold cross-validation. Table 5 reveals that our method (with concatenate operator) produces better results than other benchmark methods in most cases where we gain 16.6% improvement over SFLLN method in F-measure metric and 12.4% improvement over DDI-based CF method in AUPR metric on 3-fold cross validation. As Table 6 illustrates, for 5-fold cross-validation, we have 19.5% improvement over node2vec method in F-measure metric and 11.8% improvement over node2vec method in AUPR metric. It can be observed from the results that drug feature (exclude DDI feature) based collaborative filtering (CF) methods, the LPMS method as well as random walk methods perform poorly. That is because the drug feature networks presented in Table 2 are very noisy and sparse.

We conduct experiments to analyze the influence of different thresholds on performances of DDI-MDAE (PU Learning). As shown in Table 7, we find DDI-MDAE (PU learning) could achieve the best predictive performances on most evaluation metrics when threshold equals to 0.5. We also compare DDI-MDAE (PU learning) with the original method DDI-MADE. As Table 8 illustrates, DDI-MDAE (PU learning) achieves better performances than DDI-MDAE. DDI-MDAE (PU

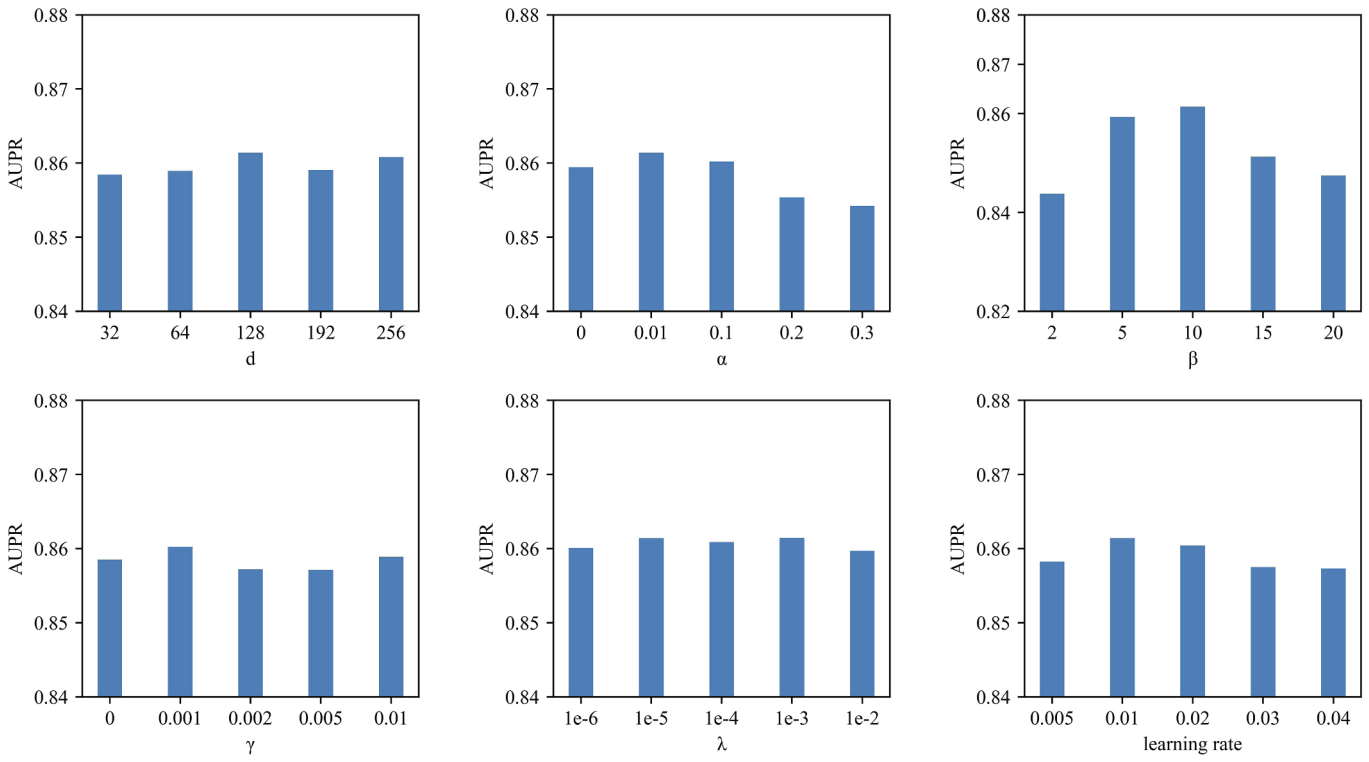


Fig. 2. Parameter sensitivity.

learning) gains 7.1% improvement in AUPR metric and 10.4% improvement in F-measure metric over DDI-MDAE for 3-fold cross-validation. For 5-fold cross-validation, DDI-MDAE (PU learning) has 6.2% improvment in AUPR metric and 8.4% improvement in F-measure metric over DDI-MDAE. Therefore, PU learning strategy could enhance the predictive performances of our model.

To evaluate the performances of our method with a deep learning architecture on the above sparse and noisy drug feature networks, we design an experiment on different sparsity of the network by removal of a certain proportion of links to evaluate its performance. Fig. 3 reports the results of different prediction methods on 3-fold and 5-fold cross-validation. As we can see, DDI-MDAE significantly improves the performance than other benchmark methods as the ratios of removed links ranging from 10% to 50%. Even when half of its links removed, our method still achieves an AUPR score higher than 0.5. In this task, we find that increasing the threshold value of MDAE with PU learning (to about 0.7 or 0.8) to some extent, could enhance the performance. It can be seen that our method improved by PU learning outperforms the original method DDI-MDAE for 3-fold and 5-fold cross-validation. Both our method and Node2vec are network embedding based approaches, it could be observed that network embedding based methods can achieve better performance when the network is sparser, which could be emerged as an effective strategy to make DDI predictions in a sparse network. As for our previous work SFLLN [10], the experiment result shows that the AUPR scores of SFLLN are not satisfactory. It needs to note that the application scenario of SFLLN is not exactly the same as

Table 5
Experiment results of different models on 3-fold cross-validation.

Method	AUPR	AUC	F1	Accuracy	Precision	Recall
Structure-based CF	0.3274	0.8512	0.3662	0.9416	0.3327	0.4077
Target-based CF	0.4310	0.8200	0.4386	0.9513	0.4193	0.4597
Enzyme-based CF	0.3157	0.7485	0.3526	0.9574	0.4760	0.2804
Pathway-based CF	0.3895	0.8359	0.3958	0.9553	0.4491	0.3538
DDI-based CF	0.7667	**0.9758**	0.7000	0.9773	0.7741	0.6390
Structure-based RW	0.3250	0.8501	0.3642	0.9407	0.3282	0.4100
Target-based RW	0.4274	0.8022	0.4378	0.9516	0.4214	0.4555
Enzyme-based RW	0.3139	0.7446	0.3524	0.9573	0.4741	0.2805
Pathway-based RW	0.3846	0.8320	0.3926	0.9546	0.4407	0.3540
DDI-based RW	0.7517	0.9736	0.6880	0.9764	0.7586	0.6296
LPMS	0.3113	0.8440	0.3533	0.9391	0.3160	0.4015
SFLLN	0.7574	0.9502	0.7149	0.9535	0.7070	0.7237
Node2vec	0.7460	0.7394	0.6465	0.9863	**0.9995**	0.4789
DDI-MDAE	**0.8614**	0.8582	**0.8339**	**0.9925**	0.9989	**0.7165**

Table 4
Experiment results of different operators on drug representations on 3-fold and 5-fold cross-validation.

	Method	AUPR	AUC	F1	Accuracy	Precision	Recall
3-CV	DDI-MDAE (Average)	0.7921	0.7875	0.7271	0.9888	0.9980	0.5750
	DDI-MDAE (Hadamard)	0.7950	0.7903	0.7316	0.9890	0.9983	0.5807
	DDI-MDAE (Weighted-L1)	0.6466	0.6382	0.4287	0.9810	0.9978	0.2764
	DDI-MDAE (Concatenate)	**0.8614**	**0.8582**	**0.8339**	**0.9925**	**0.9989**	**0.7165**
5-CV	DDI-MDAE (Average)	0.8149	0.8138	0.7683	0.9940	0.9963	0.6275
	DDI-MDAE (Hadamard)	0.8196	0.8185	0.7752	0.9942	0.9964	0.6371
	DDI-MDAE (Weighted-L1)	0.6685	0.6652	0.4917	0.9893	0.9959	0.3304
	DDI-MDAE (Concatenate)	**0.8714**	**0.8703**	**0.8498**	**0.9958**	**0.9980**	**0.7407**

Table 6
Experiment results of different models on 5-fold cross-validation.

Method	AUPR	AUC	F1	Accuracy	Precision	Recall
Structure-based CF	0.2478	0.8514	0.3006	0.9601	0.2699	0.3393
Target-based CF	0.3665	0.8232	0.3807	0.9695	0.3908	0.3713
Enzyme-based CF	0.2587	0.7483	0.3186	0.9724	0.4230	0.2556
Pathway-based CF	0.3187	0.8363	0.3522	0.9711	0.4063	0.3111
DDI-based CF	0.7045	**0.9760**	0.6643	0.9846	0.7416	0.6016
Structure-based RW	0.2456	0.8503	0.2987	0.9602	0.2693	0.3356
Target-based RW	0.3642	0.8046	0.3794	0.9689	0.3826	0.3764
Enzyme-based RW	0.2572	0.7446	0.3178	0.9724	0.4229	0.2546
Pathway-based RW	0.3139	0.8324	0.3487	0.9711	0.4060	0.3060
DDI-based RW	0.6870	0.9738	0.6506	0.9840	0.7276	0.5884
LPMS	0.2337	0.8442	0.2877	0.9595	0.2591	0.3236
SFLLN	0.6286	0.8792	0.5834	0.9056	0.5982	0.5693
Node2vec	0.7793	0.7763	0.7109	0.9929	**0.9988**	0.5527
DDI-MDAE	**0.8714**	0.8703	**0.8498**	**0.9958**	0.9980	**0.7407**

Table 7
Experiment results of DDI-MDAE (PU learning) with different thresholds on 5-fold cross-validation.

Threshold	AUPR	AUC	F1	Accuracy	Precision	Recall
0.1	0.6197	**0.9689**	0.4047	0.9539	0.2547	**0.9844**
0.2	0.8401	0.8376	0.8060	0.9948	**0.9998**	0.6751
0.3	0.9084	0.9590	0.9076	0.9970	0.8958	0.9197
0.4	0.9104	0.9372	0.9096	0.9096	0.8971	0.9225
0.5	**0.9252**	0.9604	**0.9215**	**0.9976**	0.9736	0.8748
0.6	0.9055	0.9070	0.8950	0.9970	0.9940	0.8140
0.7	0.8775	0.8761	0.8581	0.9960	0.9989	0.7521
0.8	0.8744	0.8729	0.8540	0.9959	0.9990	0.7459
0.9	0.8344	0.8318	0.7977	0.9946	**0.9998**	0.6636

Table 8
Experiment results of improved model and original model on 3-fold and 5-fold cross-validation.

	Method	AUPR	AUC	F1	Accuracy	Precision	Recall
3-CV	DDI-MDAE	0.8614	0.8582	0.8339	0.9925	0.9989	0.7165
	DDI-MDAE (PU learning)	**0.9229**	**0.9486**	**0.9210**	**0.9960**	**0.9446**	**0.8985**
5-CV	DDI-MDAE	0.8714	0.8703	0.8498	0.9958	0.9980	0.7407
	DDI-MDAE (PU learning)	**0.9252**	**0.9604**	**0.9215**	**0.9976**	**0.9736**	**0.8748**

that of DDI-MDAE. SFLLN integrates multiple drug features and drugs are supposed to have all features. As mentioned before, our dataset in this paper is large-scale and sparse.

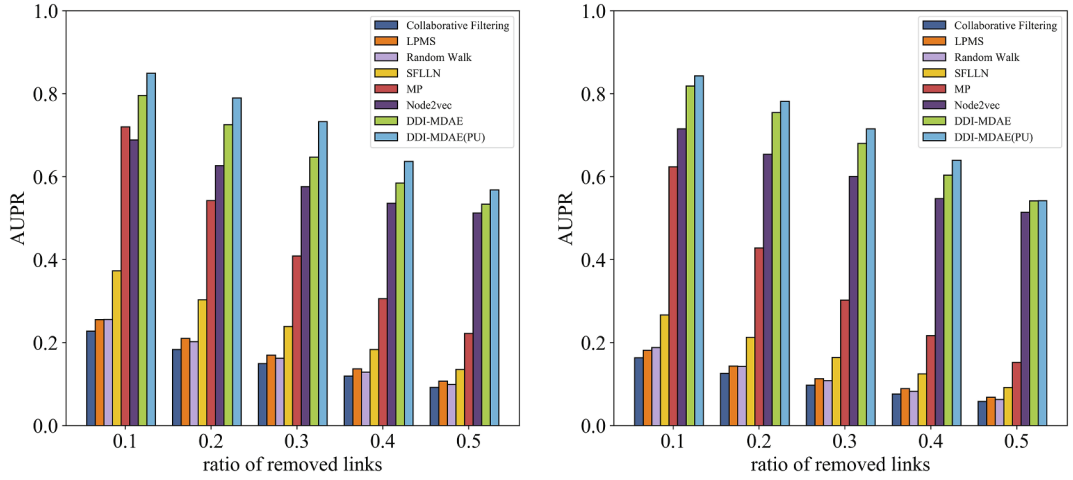
4.5. Case studies

In the experiments, our datasets have 2,367 drugs and 2,800,161 pairwise drug-drug interactions. In detail, there are 209,494 known DDIs between drugs, and 2,590,667 unlabeled drug pairs which may contain unobserved DDIs. We apply DDI-MDAE to predict unobserved DDIs by training with all known interactions, and higher scores of unobserved drug pairs indicate higher probabilities of having interactions. Table 9 shows the top 20 predicted drug-drug interactions by our method, which are not available in our datasets. We look up the evidence of newly predicted DDIs, and find that a significant fraction of DDIs (16 out of 20) is confirmed in the latest online version 5.1.4 of DrugBank database. For instance, the description of the interaction between drug "Clozapine" (DB00363) and drug "Lansoprazole" (DB00448) is "The metabolism of Clozapine can be decreased when combined with Lansoprazole.". The case studies demonstrate the usefulness of our method to detect the potential drug-drug interactions.

5. Conclusion

In this paper, we survey how to predict accurate drug-drug interactions using multiple drug datasets. We propose a deep architecture named DDI-MDAE to capture the multiple non-linear network structures. DDI-MDAE aims to learn unified drug representations using multi-modal deep auto-encoders, where each drug feature network is trained by a deep auto-encoder channel and the drug representations in the hidden layer are shared to learn simultaneously. Moreover, we apply four operators to the learned drug embeddings for representing drug-drug pairs. We adopt a random forest classifier to make DDI predictions. Evaluated on the benchmark dataset, our method enhances significantly in the performance compared to other existing DDI prediction models. Especially, our model is also suitable for very sparse networks. Furthermore, we improve the proposed DDI-MDAE method by applying a specialized random forest classifier in the PU learning setting to increase predictive accuracy.

In this paper, our model learns drug representations only from the structural topologies of drug feature networks. In our future work, we consider combing the network topologies and semantic information for drug representations, which may improve the predictive performance.



(a) Performance on 3-fold cross-validation. (b) Performance on 5-fold cross-validation.

Fig. 3. Performances on networks of different sparsity.

Table 9
Top 20 DDI predictions by our method.

Rank	Drug 1	Drug 2	Description
1	Clozapine	Lansoprazole	The metabolism of Clozapine can be decreased when combined with Lansoprazole.
2	Methylergometrine	Citalopram	N.A.
3	Spirapril	Sufentanil	Sufentanil may decrease the antihypertensive activities of Spirapril.
4	Dapsone	Haloperidol	The serum concentration of Haloperidol can be increased when it is combined with Dapsone.
5	Carbamazepine	Manidipine	The metabolism of Carbamazepine can be decreased when combined with Manidipine.
6	Acetylsalicylic acid	Clozapine	Acetylsalicylic acid may decrease the excretion rate of Clozapine which could result in a higher serum level.
7	Saquinavir	Tapentadol	The metabolism of Tapentadol can be decreased when combined with Saquinavir
8	Moclobemide	Pazopanib	The metabolism of Pazopanib can be decreased when combined with Moclobemide
9	Fluvoxamine	Rifapentine	The metabolism of Fluvoxamine can be increased when combined with Rifapentine.
10	Citalopram	Zafirlukast	The serum concentration of Citalopram can be increased when it is combined with Zafirlukast.
11	Celiprolol	Cilnidipine	The risk or severity of bradycardia can be increased when Cilnidipine is combined with Celiprolol.
12	Indacaterol	Sertraline	The serum concentration of Indacaterol can be increased when it is combined with Sertraline.
13	Nicergoline	Pantoprazole	N.A.
14	Atomoxetine	Primidone	The risk or severity of adverse effects can be increased when Atomoxetine is combined with Primidone.
15	Osimertinib	Pitavastatin	The serum concentration of Osimertinib can be increased when it is combined with Pitavastatin.
16	Cinnarizine	Saquinavir	The risk or severity of QTc prolongation can be increased when Cinnarizine is combined with Saquinavir.
17	Deferasirox	Valproic Acid	The metabolism of Valproic acid can be decreased when combined with Deferasirox.
18	Lapatinib	Ticagrelor	The metabolism of Ticagrelor can be decreased when combined with Lapatinib.
19	Ledipasvir	Telithromycin	N.A.
20	Bimatoprost	Estrone	N.A.

N.A.: evidence not available.

CRediT authorship contribution statement

Yang Zhang: Conceptualization, Methodology, Software. **Yang Qiu:** Data curation, Writing - original draft. **Yuxin Cui:** Software, Validation. **Shichao Liu:** Conceptualization, Methodology, Supervision, Project administration. **Wen Zhang:** Writing - review & editing.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61772381, 61572368), National Key Research and Development Program (2018YFC0407904), Huazhong Agricultural University Scientific & Technological Self-innovation Foundation, the Fundamental Research Funds for the Central Universities (2042017kf0219, 2042018kf0249, 2662019QD011). The funders have no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.ymeth.2020.05.007.

References

[1] X. Sun, L. Ma, X. Du, J. Feng, and K. Dong, "Deep convolution neural networks for drug-drug interaction extraction," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018, pp. 1662–1668.
[2] M.F.M. Chowdhury, A. Lavelli, Fbk-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information, Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval, 2013, pp. 351–355.
[3] S. Kim, H. Liu, L. Yeganova, W.J. Wilbur, Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach, J. Biomed. Inform. 55 (2015) 23–30.
[4] Y. Zhang, W. Zheng, H. Lin, J. Wang, Z. Yang, M. Dumontier, Drug–drug interaction extraction via hierarchical rnns on sequence and shortest dependency paths, Bioinformatics 34 (5) (2017) 828–835.
[5] Z. Jiang, L. Gu, Q. Jiang, "Drug drug interaction extraction from literature using a skeleton long short term memory neural network," in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2017, pp. 552–555.
[6] B. Xu, X. Shi, Z. Zha, W. Zheng, H. Lin, Z. Yang, J. Wang, F. Xia, "Full-attention based drug drug interaction extraction exploiting user-generated content," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018, pp. 560–565.
[7] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, C. Friedman, Drug–drug interaction through molecular structure similarity analysis, J. Am. Med. Inform. Assoc. 19 (6) (2012) 1066–1074.
[8] A. Gottlieb, G.Y. Stein, Y. Oron, E. Ruppin, R. Sharan, "Indi: a computational framework for inferring drug interactions and their associated recommendations," Mol. Syst. Biol., vol. 8, no. 1, 2012.
[9] T. Ma, C. Xiao, J. Zhou, and F. Wang, "Drug similarity integration through attentive multi-view graph auto-encoders," arXiv preprint arXiv:1804.10850, 2018.
[10] W. Zhang, K. Jing, F. Huang, Y. Chen, B. Li, J. Li, J. Gong, Sflln: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions, Inf. Sci. 497 (2019) 189–201.
[11] J.-Y. Shi, H. Huang, J.-X. Li, P. Lei, Y.-N. Zhang, K. Dong, S.-M. Yiu, Tmfuf: a triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs, BMC Bioinform. 19 (14) (2018) 411.
[12] H. Yu, K.-T. Mao, J.-Y. Shi, H. Huang, Z. Chen, K. Dong, S.-M. Yiu, Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization, BMC Syst. Biol. 12 (1) (2018) 14.
[13] P. Zhang, F. Wang, J. Hu, R. Sorrentino, Label propagation prediction of drug-drug interactions based on clinical side effects, Sci. Rep. 5 (2015) 12339.
[14] K. Park, D. Kim, S. Ha, D. Lee, Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks, PloS one 10 (10) (2015) e0140816.
[15] D. Sridhar, S. Fakhraei, L. Getoor, A probabilistic approach for collective similarity-based drug–drug interaction prediction, Bioinformatics 32 (20) (2016) 3175–3182.
[16] J. Huang, C. Niu, C.D. Green, L. Yang, H. Mei, J.-D.J. Han, Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network, PLoS Comput. Biol. 9 (3) (2013) e1002998.
[17] S. Deepika, T. Geetha, A meta-learning framework using representation learning to predict drug-drug interaction, J. Biomed. Inform. 84 (2018) 136–147.
[18] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, X. Li, Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data, BMC Bioinform. 18 (1) (2017) 18.
[19] T. Takeda, M. Hao, T. Cheng, S.H. Bryant, Y. Wang, Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge, J. Cheminform. 9 (1) (2017) 16.
[20] P. Zhang, F. Wang, J. Hu, "Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity, " in AMIA Annual Symposium Proceedings, vol. 2014. American Medical Informatics Association, 2014, p. 1258.
[21] F. Cheng, Z. Zhao, Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties, J. Am. Med. Inform. Assoc. 21 (e2) (2014) e278–e286.
[22] A. Cami, S. Manzi, A. Arnold, B.Y. Reis, Pharmacointeraction network models predict unknown drug-drug interactions, PloS one 8 (4) (2013) e61468.
[23] X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S.M. Lin, W. Zhang, P. Zhang, H. Sun, Graph embedding on biomedical networks: methods, applications and evaluations, Bioinformatics 36 (4) (2020) 1241–1251.
[24] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[25] J.B. Tenenbaum, V. De Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[26] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 701–710.
[27] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, "Line: Large-scale information network embedding, Proceedings of the 24th International Conference on World

Wide Web. International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.
[28] A. Grover, J. Leskovec, "node2vec: Scalable feature learning for networks, Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2016, pp. 855–864.
[29] D. Wang, P. Cui, W. Zhu, "Structural deep network embedding, Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2016, pp. 1225–1234.
[30] T.N. Kipf, M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.
[31] S. Chang, W. Han, J. Tang, G.-J. Qi, C.C. Aggarwal, T.S. Huang, "Heterogeneous network embedding via deep architectures," in, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 119–128.
[32] Y. Dong, N.V. Chawla, A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks, Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, 2017, pp. 135–144.
[33] Y. Ma, Z. Ren, Z. Jiang, J. Tang, D. Yin, "Multi-dimensional network embedding with hierarchical structure, Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, ACM, 2018, pp. 387–395.
[34] C. Yang, Z. Liu, D. Zhao, M. Sun, E.Y. Chang, "Network representation learning with rich text information, Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires Argentina, 2015, pp. 2111–2117.
[35] L. Liao, X. He, H. Zhang, T.-S. Chua, Attributed social network embedding, IEEE Trans. Knowl. Data Eng. 30 (12) (2018) 2257–2270.
[36] S. Liu, S. Zhai, L. Zhu, F. Zhu, Z.M. Zhang, W. Zhang, "Efficient network representations learning: An edge-centric perspective," in International Conference on Knowledge Science, Engineering and Management. Springer, 2019, pp. 373–388.
[37] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.
[38] G.E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, IEEE Trans. Audio, Speech, Language Process. 20 (1) (2011) 30–42.
[39] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury et al., "Deep neural networks for acoustic modeling in speech recognition," IEEE Signal processing magazine, vol. 29, 2012.
[40] G.E. Hinton, R.S. Zemel, Autoencoders, minimum description length and helmholtz free energy, Adv. Neural Inform. Process. Syst. (1994) 3–10.
[41] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (Dec) (2010) 3371–3408.
[42] B. Liu, Y. Dai, X. Li, W.S. Lee, P.S. Yu, "Building text classifiers using positive and unlabeled examples," in Third IEEE International Conference on Data Mining. IEEE, 2003, pp. 179–186.
[43] C. Elkan, K. Noto, "Learning classifiers from only positive and unlabeled data, Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2008, pp. 213–220.
[44] W.S. Lee, B. Liu, "Learning with positive and unlabeled examples using weighted logistic regression," in ICML, vol. 3, 2003, pp. 448–455.
[45] Z. Liu, W. Shi, D. Li, Q. Qin, "Partially supervised classification: based on weighted unlabeled samples support vector machine, Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications, IGI Global, 2008, pp. 1216–1230.
[46] H. Yu, J. Han, K.-C. Chang, Pebl: Web page classification without negative examples, IEEE Trans. Knowl. Data Eng. 16 (1) (2004) 70–81.
[47] B. Liu, W.S. Lee, P.S. Yu, X. Li, "Partially supervised classification of text documents," in ICML, vol. 2. Citeseer, 2002, pp. 387–394.
[48] X. Li, B. Liu, Learning to classify texts using positive and unlabeled data, IJCAI 3 (2003) (2003) 587–592.
[49] F. Mordelet, J.-P. Vert, A bagging svm to learn from positive and unlabeled examples, Pattern Recogn. Lett. 37 (2014) 201–209.
[50] P. Yang, X. Li, H.-N. Chua, C.-K. Kwoh, S.-K. Ng, Ensemble positive unlabeled learning for disease gene identification, PloS one 9 (5) (2014) pp.
[51] P. Yang, X.-L. Li, J.-P. Mei, C.-K. Kwoh, S.-K. Ng, Positive-unlabeled learning for disease gene identification, Bioinformatics 28 (20) (2012) 2640–2647.
[52] F. Mordelet, J.-P. Vert, Prodige: Prioritization of disease genes with multitask machine learning from positive and unlabeled examples, BMC Bioinform. 12 (1) (2011) 389.
[53] W. Lan, J. Wang, M. Li, J. Liu, Y. Li, F.-X. Wu, Y. Pan, Predicting drug–target interaction using positive-unlabeled learning, Neurocomputing 206 (2016) 50–57.
[54] Y. Zheng, H. Peng, X. Zhang, Z. Zhao, X. Gao, J. Li, Ddi-pulearn: a positive-unlabeled learning method for large-scale prediction of drug-drug interactions, BMC Bioinform. 20 (19) (2019) 1–12.
[55] P.N. Hameed, K. Verspoor, S. Kusljic, S. Halgamuge, Positive-unlabeled learning for inferring drug interactions based on heterogeneous attributes, BMC Bioinform. 18 (1) (2017) 1–15.
[56] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, "Drugbank: a comprehensive resource for in silico drug discovery and exploration," Nucleic acids research, vol. 34, no. suppl_1, pp. D668–D672, 2006.
[57] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, "Kegg for representation and analysis of molecular networks involving diseases and drugs," Nucleic acids research, vol. 38, no. suppl_1, pp. D355–D360, 2009.
[58] Y. Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, S.H. Bryant, "Pubchem: a public information system for analyzing bioactivities of small molecules," Nucleic acids research, vol. 37, no. suppl_2, pp. W623–W633, 2009.
[59] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[60] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks," in, Proceedings of the fourth ACM conference on Recommender systems, ACM, 2010, pp. 135–142.
[61] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, R news 2 (3) (2002) 18–22.
[62] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, no. Oct, pp. 2825–2830, 2011.