DPDDI: a Deep Predictor for Drug-Drug Interactions

Yue-Hua Feng[1], Shao-Wu Zhang[1], Jian-Yu Shi[2]
1 Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China
2 School of Life Sciences, Northwestern Polytechnical University, Xi'an, China

Abstract

Background—The treatment of complex diseases taking multiple drugs becomes popular. However, drug-drug interactions (DDIs) may give rise to the risk of unanticipated adverse effects and even unknown toxicity. As DDI detection in the wet lab is expensive and time-consuming, computational DDIs prediction based on machine learning becomes a promising approach due to its low cost and fast running. Generally, most of the existing computational approaches construct drug features from diverse drug properties, which are costly obtained and not available in many cases.

Results—To address this issue, by organizing DDIs as a network, we propose a novel predicting approach, which can work without drug property. It consists of a feature extractor based on graph convolution network (GCN) as well as a predictor based on deep neural network (DNN). The former characterizes drugs in a graph embedding space, where each drug was represented as a low-dimensional latent feature vector capturing the topological relationship to its neighborhood drugs by GCN. The latter concatenates latent feature vectors of any two drugs as the feature vector of the corresponding drug pairs and trains a DNN to predict potential interactions. In the experiments, we first demonstrate that our DNN-based predictor greatly outperforms the inner product-based predictor in the original GCN, and our network-derived latent feature greatly outperforms other features derived from chemical, biological or anatomical properties of drugs. Then, we indicate the over-optimistic prediction caused by down-sampling unlabeled drug pairs and validate the robustness of our approach to different datasets w.r.t. drug number, DDI number, and network sparsity. Moreover, the comparison with four state-of-the-art approaches using drug properties demonstrates the significant superiority of our approach under 5-fold cross-validation. Finally, a

novel prediction validates its potentials in a real predicting scenario with finding 13 verified DDI out of the top 20 unlabeled candidates.

Conclusion—We propose a simple but robust method DPDDI to predicting novel DDIs, which can work without drug property. It can be expected that DPDDI can be helpful in other DDI-related scenarios, such as the detection of unexpected side effects, and the guidance of drug combination.

Keywords—Drug-drug interaction, DDI prediction, graph convolution network (GCN), feature extraction, deep neural network

Background

Due to synergistic effects caused by drug-drug interactions (DDIs), the combinational treatment of multiple drugs for complex diseases are popular nowadays [1]. However, unexpected DDI can also trigger side effects, adverse reactions, and even serious toxicity. They lead to patients in danger [2]. As the need of multi-drug treatments is increasing, the identification of DDIs is urgent. Nevertheless, it is expensive and time-consuming to detect DDIs among a large scale of drug pairs both in vitro and in vivo. To assist to screen DDIs, computational approaches have been developed to deduce candidate interactions.

Existing computational approaches can be roughly classified into two categories: text mining-based and machine learning-based. The text mining-based approaches discover and collect annotated DDIs from scientific literature, electronic medical records[3, 4], insurance claim databases, and the FDA Adverse Event Reporting System[5]. They are very useful in building DDI-related databases. Nevertheless, since those approaches cannot detect unannotated DDIs, they cannot give an alert to potential DDIs before a combinational treatment is made [2]. In contrast, machine learning-based approaches provide a promising way to identify unannotated potential interactions for downstream experimental validations.

Usually, a machine learning-based approach consists of a feature extractor and a

supervised predictor. The feature extractor represents drugs in a form of feature vector according to drug chemical (e.g. molecular structure[2, 6-14]), biological (e.g. targets[2, 8-11]), anatomical properties(e.g. ATC[8-10, 12]), phenotypic (e.g. side effects[8, 9, 11, 13, 14]) properties, medication and/or clinical observations[11], etc. Sometimes, drug feature vectors are transformed into one or more drug similarity matrices [2, 6, 7] [8] [9] [11, 13] [14-16] . However, drug properties can be not obtained in many cases.

The supervised predictor is usually implemented by classifiers (e.g. KNN[12], SVM[12], logistic regression[2, 8, 10], decision tree[10], naïve Bayes[10]), network propagation (e.g. reasoning over drug-drug network structure[6-8], label propagation[13], random walk[11, 16], probabilistic soft logic[9, 10]) or matrix factorization[14]. Usually, the predictor first trains a model with both feature vectors/similarity matrices and annotated DDI labels, then deduces potential DDIs with the well-trained model. Most approaches utilize a single predictor [2, 5-8, 13-16], while some of them integrate multiple predictors to obtain the final prediction [10, 12].

In general, the predicting performance of existing approaches heavily relies on the quality of handcrafted features, which are generated by their feature extractors when drug properties are available. Deep learning shows its power in many applications, such as speech recognition, computer vision, natural language processing and bioinformatics [17-19]. One of the advantages of deep learning is its ability to extract features automatically without handcraft[20]. Another is its good representation of non-linear structural relationships between samples (e.g. drugs) in non-Euclidean domains[21].

This work proposes a deep learning-inspired approach for predicting DDIs, named Deep Predictor for Drug-Drug Interaction (DPDDI), which consists of a feature extractor based on graph convolution network (GCN) [22, 23] as well as a predictor based on deep neural network (DNN)[19]. The former characterizes drugs in a graph embedding space, where each drug was represented as a low-dimensional latent feature vector capturing the topological relationship to its neighborhood drugs by GCN. The latter concatenates latent feature vectors of any two drugs as the feature vector of the corresponding drug pairs and trains a DNN to predict potential interactions. In the

experiments, we first validate the superiority of both the GCN-derived feature and the DNN-based predictor. Then, we investigate the performance of DPDDI under different datasets and sampling strategies. Last, we demonstrate the effectiveness of DPDDI by both the comparison with four state-of-the-art approaches and the novel prediction in a real predicting scenario.

Results

Parameter tuning

With the training data, we performed a grid search of different parameters by seeking for both the minimum value of the loss function and the best accuracy. Both the GCN-based feature extractor and the DNN-based predictor need to tune the values of Learning rate, Epochs, Batch size, Dropout rate, as well as neuro numbers (dimensions) in hidden layers. Specifically, with the full batch size, the GCN-based feature extractor tuned the learning rate from the list of {0.1, 0.01, 0.001, 0.005, 0.0001}, the epoch from {200, 500, 800, 1000, 1200, 1400, 1600}, the dropout rate from {0.01, 0.001, 0.0001}, and the dimensions from {[800,512], [800,256], [800,128], [512,256], [512,128], [512,64], [256,64], [128,32]}. The DNN-based predictor tuned the learning rate from {0.1, 0.05, 0.01, 0.005}, the epoch from {20, 40, 60, 80,100,140,160}, the batch size from {10, 20, 40, 50, 60, 80}, the dropout rate from {0.01, 0.001, 0.0001} and the dimensions from {[128, 32], [128, 64], [64, 32], [128,64,32], [128, 32, 16], [64, 32, 16], [128, 64, 32, 16], [64, 32, 16, 4]}. The optimal values of these parameters are shown in Table 1.

Table 1. The optimal values of parameters in DPDDI

Parameters	L-rate	Epochs	Dropout	Batch-size	Input-dim	Hidden layers	Output-dim
Feature extractor	0.001	1400	0.0001	Full-batch	1562	[512,128]	128
Predictor	0.01	140	0.001	50	256	[128,64,32]	2

Comparison of feature aggregations

Table 2 shows predicting result of three feature aggregations. Taking the inner product as the input of the predictor, the first aggregation makes the discrimination by a DNN but not the inner product itself. The second aggregation sums up two drug latent vectors into the feature vector of the drug pair, which is taken as the input of DNN. The third aggregation takes the concatenation of two drug latent vectors as the input of DNN.

Table 2. The comparisons of feature aggregation strategies in the DNN-based predictor

Feature Aggregation	AUROC	AUPR	Recall	Precision	ACC	F_1
Inner product	0.938	0.810	0.709	0.761	0.927	0.734
Summation	0.970	0.898	0.774	0.854	0.949	0.812
Concatenation	0.983	0.925	0.844	0.836	0.955	0.840

Under the same concatenation approach, other traditional features derived from three heterogeneous sources are also considered, including chemical structure, drug-binding proteins (DBP), and Anatomical Therapeutic Chemical Classification labels (ATC). Chemical structures of the drugs are characterized by 881 PubChem fingerprints, their DBPs are represented by 1121-dimensional binary vectors, of which each bit indicates the binding occurrence of a specific DBP across the drugs. Their 118-dimensional ATC features were converted from the 7-bit ATC code by one-hot code. See also Datasets Section for the details. Taking DNN as the predictor, the performance of different features is showed in Table 3.

Table 3. Performance of different features with DNN in DDI prediction

Feature	AUROC	AUPR	Recall	Precision	ACC	F_1
Sub-structure	0.904	0.635	0.668	0.554	0.876	0.605
DBP	0.874	0.616	0.602	0.584	0.882	0.593
ATC	0.901	0.656	0.659	0.576	0.882	0.615
GCN	0.983	0.925	0.844	0.836	0.955	0.840

Influence of sample numbers on prediction

So far, all the interactions are considered as positive samples and all the unlabeled drug pairs as negative samples. The ratio of the number of positive samples to that of negative samples is imbalanced (about 1:6). The training of DDI prediction under different ratio schemes of sample numbers including 1:1, 1:3, and 1:6 is tested. In the first ratio scheme, we randomly selected the same number of negative samples as that of positive samples to train our DNN. In the second scheme, the number of selected negative samples is three times to that of positive samples. In the last one, all the negative samples were chosen. From the Fig. 1, the balanced sample scheme achieves the best and the imbalanced sample scheme according actual ratio of the number of positive samples to that of negative samples gets the worst.

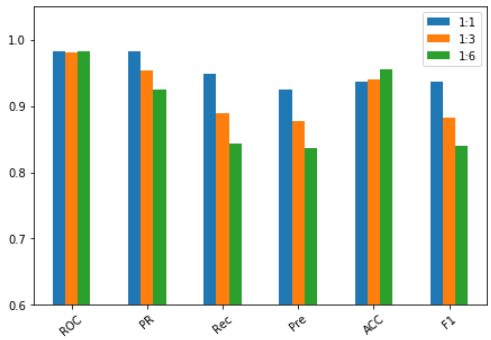

Fig. 1. Performance of DPDDI on unbalance negative samples

Moreover, the generalization of DPDDI is tested by investigating its performance on two other datasets containing different drugs and DDIs. The result is showed in Table 4. The first additional dataset (DB2) collected by [11] from TWOSIDES [24] contains 548 drugs and 48,548 DDIs while the second additional dataset (DB3) is the updated version of our dataset based on DrugBank[25] and contains 1934 drugs and 230,887 DDIs. They have different sparsity in terms of the network. More details can be found in Datasets Section.

Table 4. Performance of DPDDI in different datasets

Dataset	Type	Sparsity	AUROC	AUPR	Recall	Precision	ACC	F_1
DB2	Small	32.4%	0.956	0.907	0.810	0.754	0.940	0.840
DB1	Medium	14.8%	0.983	0.925	0.844	0.836	0.955	0.840
DB3	Large	12.4%	0.981	0.932	0.835	0.876	0.960	0.855

Comparison with the state-of-the-art approaches

In order to validate the effectiveness, we compared our DPDDI under 5-CV with four state-of-the-art approaches, including two Vilar's models[6, 7] (denoted Vilar 1 and Vilar 2 respectively), label propagation-based approach[13] (denoted LP) and Zhang et al's approach [11] (denoted Zhang). Vilar 1 integrates a Tanimoto similarity matrix of molecular structures with known DDI matrix by a linear matrix transformation to identify new DDIs [6]. Vilar 2 uses drug interaction profile fingerprints (IPFs) to measure similarity and predict new DDIs [7]. LP applies Label Propagation to assign labels from known DDIs to previously unlabeled nodes by computing drug similarity-derived weights of edges on the DDI network. Zhang is an integrated approach [11]. After collecting multiple types of drug features, including drug substructures, proteins, pathways, indications, and side effects, it integrates 29 sub-classifier results to obtain the final prediction.

To make a fair comparison, the dataset in [11] is adopted to run the prediction. The DPDDI achieves the best and outperforms other state-of-the-art approaches significantly across all the metrics (Table 5). Specifically, DPDDI achieves the improvements of 0.2~24.9%, 6.6~64.5%, 2.2~31.5%, 2.5~50.1%, 0.6~22.1%, 8.9~50.6% against all the approaches in terms of AUROC, AUPR, Recall, Precision, Accuracy, and F1-score respectively.

Table 5. Performance of comparing state-of-art methods

Model	AUROC	AUPR	Recall	Precision	ACC	F_1
Vilar 1[6]	0.707	0.262	0.495	0.253	0.719	0.334
Vilar 2[7]	0.826	0.533	0.569	0.515	0.862	0.540

LP[13]	0.851	0.799	0.685	0.729	0.809	0.706
Zhang[11]	0.954	0.841	0.788	0.717	0.934	0.751
DPDDI	**0.956**	**0.907**	**0.810**	**0.754**	**0.940**	**0.840**

Novel prediction

In this section, we investigate the generalized performance of our model in a real scenario by performing a novel prediction. In such a scenario, we aim to deduce potential DDIs among unlabeled drug pairs in hand. In other words, we infer possible interactions among drugs, which are not connected to each other in the DDI network. To solve the task, we first mapped all the drugs in the network into the latent space by GCN and reconstructed the network by DDN by the procedure in Predictor Section.

In the reconstruction network, each edge has a predicting score assigned by the predictor. The greater the score, the higher the probability of being an interaction. Then, we ranked the previously unlabeled edges according to predicting scores and focused on the top-20 drug pairs in the ranking list (Table 6). Last, we validated them by the current version of DrugBank and the website of Drug Interaction Checker (Drugs.com). The validation reveals that 13 out of the top 20 drug pairs have been found in DrugBank and Drugs.Com.

Table 6. Top 20 novel interactions predicted by our approach

Number	Drug 1	Drug 2	Validation source	Description
1	Doxycycline	Bleomycin	DrugBank	Doxycycline may decrease the excretion rate of Bleomycin which could result in a higher serum level.
2	Doxycycline	Rifapentine	N/A	N/A
3	Doxycycline	Fusidic acid	N/A	N/A
4	Pramipexole	Paroxetine	DrugBank	Paroxetine may increase the sedative activities of Pramipexole.
5	Luliconazole	Doxycycline	N/A	N/A
6	Netupitant	Doxycycline	DrugBank	The metabolism of Netupitant can be decreased when combined with Doxycycline.
7	Tenoxicam	Minocycline	N/A	N/A
8	Etoperidone	Tenoxicam	DrugBank	Tenoxicam may decrease the excretion rate of Etoperidone which could result in a higher serum level.
9	Pramipexole	Minocycline	N/A	N/A
10	Ropinirole	Pramipexole	DrugBank	Ropinirole may increase the sedative activities of Pramipexole.
11	Minocycline	Ropinirole	DrugBank	Minocycline may increase the central nervous system depressant (CNS depressant) activities of Ropinirole.
12	Bleomycin	Doxycycline	DrugBank	Doxycycline may decrease the excretion rate of Bleomycin which could result in a higher serum level.
13	Pramipexole	Metyrosine	drugs.com	Using metyroSINE together with pramipexole may increase side effects such as dizziness, drowsiness, confusion, and difficulty concentrating.
14	Osimertinib	Doxycycline	DrugBank	The metabolism of Osimertinib can be decreased when combined with Doxycycline.
15	Dronabinol	Pramipexole	DrugBank	Dronabinol may increase the sedative activities of Pramipexole.
16	Rufinamide	Tenoxicam	N/A	N/A
17	Phenobarbital	Pramipexole	drugs.com	Using PHENobarbital together with pramipexole may increase side effects such as dizziness, drowsiness, confusion, and difficulty concentrating.
18	Bleomycin	Mitotane	N/A	N/A
19	Fosaprepitant	Doxycycline	DrugBank	The metabolism of Fosaprepitant can be decreased when combined with Doxycycline.
20	Duloxetine	Rufinamide	drugs.com	Using DULoxetine together with rufinamide may increase side effects such as dizziness, drowsiness, confusion, and difficulty concentrating.

Discussion

One of the key factors in DDI prediction is the predictor. The original GCN maps the nodes in the DDI network into an embedding space to obtain drug latent feature vectors and discriminates whether there is a DDI between two drugs by directly calculating the inner product of their latent vectors. However, since such a linear discrimination cannot reflect the non-linear relationship between drugs, we consider non-linear discrimination based on deep neural network (DNN) and provide three ways to aggregate the feature vectors of two individual drugs into the feature vector of their drug pair. See also Section of Feature aggregation for drug pairs. Through comparing three feature aggregations (Table 2), the result shows that the concatenation is the best whereas the inner product is the worst. Therefore, we adopt the concatenation as the feature aggregation in sequential experiments.

Another key factor in DDI prediction is the feature. we compared the GCN-derived structure feature with other three traditional features. The result (Table 3) show the great superiority of our GCN-derived feature across all the performance metrics. Especially, our GCN-derived feature achieves >20% improvement in terms of AUPR, Precision, and F-score. In summary, DNN is a better predictor for capturing the non-linear relationship between drug features while GCN generates a better feature for capturing the structural relationship between drugs.

In addition, we paid attention to how to apply negative samples in the training phase. Although many former works in other similar areas (e.g. [26-28]) adopted the same number of negative samples as that of positive samples to avoid the computational challenge caused by the sample imbalance. The result (Fig. 1) shows that the balance sample scheme achieves the best and the unbalance sample scheme (sample ratio 1:6) achieves the worst. Obviously, the training with a smaller number of negative samples causes the over-optimistic prediction. We suggest that the training should adopt all the negative samples although the down-sampling method can generate a good-looking prediction.

In addition, we consider the generalization of our approach by investigating its performance on two other datasets containing different drugs and DDIs. From the result showed in Table 4, the approximated predicting performances on these datasets show that our approach is robust to different datasets with respect to drug numbers and DDI numbers.

Finally, we validated the performance of our DPDDI. The comparison with the state-of-the-art approaches shows that DPDDI achieves the best and outperforms other state-of-the-art approaches significantly across all the metrics (Table 5). Moreover, the potentials of DPDDI in practice is demonstrated by the novel prediction with the finding that 13 DDI out of top 20 unlabeled candidates are verified (Table 6).

Conclusions

Aiming at the preliminary screening of DDIs, this work presents a novel predicting approach DPDDI by organizing a large set of DDIs as a network. It consists of a feature extractor based on graph convolution network (GCN) as well as a predictor based on deep neural network (DNN). The former characterizes drugs in a graph embedding space, where each drug was represented as a low-dimensional latent feature vector capturing the topological relationship to its neighborhood drugs by GCN. The latter concatenates latent feature vectors of any two drugs as the feature vector of the corresponding drug pairs and trains a DNN to predict potential interactions. Designated experiments for DPDDI bring several observations: (1) the concatenation of two drug feature vector into the feature of the corresponding drug pair is better than other feature aggregations, including the inner product and the summation; (2) the GCN-derived latent feature greatly outperforms other features derived from chemical, biological or anatomical properties of drugs; (3) the down-sampling of unlabeled drug pairs generates over-optimistic predictions; (4) DPDDI is robust to different datasets with regard to drug number, DDI number, and network sparsity; (5) the superiority of DPDDI is significantly superior to four state-of-the-art approaches using drug properties; (6) the finding 13 verified DDI out of top 20 unlabeled candidates in a novel

prediction reveals DPDDI's potentials in a real predicting scenario. To summarize, the proposed DPDDI is an effective approach for predicting DDIs. It can be expected that DPDDI can be helpful in other DDI-related scenarios, such as the detection of unexpected side effects, and the guidance of drug combination.

Methods

Datasets

We extracted approved small molecular drugs and their DDIs from DrugBank 4.0 [29] and build a dataset (denoted as DB1) containing 1562 drugs and 180,576 annotated interactions between them. It shall be used to validate the performance of our GCN-based feature extractor and DNN-based predictor. Also, in order to compare with other state-of-the-art approaches, we adopted a smaller dataset collected by Zhang et al. [11], which is denoted as DB2. Moreover, we collected a new and larger dataset from DrugBank 5.0[25], denoted as DB3.

From the point of view of the network, these three datasets are treated as DDI networks, and their properties (e.g. the number of drugs, the number of interactions, and degree) are summarized in Table 7. We shall investigate the robustness of our approach to different datasets w.r.t. drug number, DDI number, and network sparsity.

Table 7. Summary of drug-drug interaction networks

Dataset	#Drug	#Interaction	No-link number	Sparsity	Max degree	Min degree
DB1	1562	180,576	1,038,565	14.8%	903	1
DB2	548	48,584	101,294	32.4%	512	1
DB3	1934	230,887	1,637,357	12.4%	1049	1

Moreover, in order to compare our network-based feature with other drug features derived from diverse drug properties, we also download chemical structures, Anatomical Therapeutic Chemical classification (ATC) codes and drug-binding proteins (DBPs) from DrugBank.

The chemical structure-based feature represents each drug by an 881-dimensional binary vector. Each bit in the vector indicates the occurrence or nonoccurrence of a specific substructure according to Pubchem fingerprints.

ATC codes are released by the World Health Organization [30], and they categorize drug substances at different levels according to organs they affect, application area, therapeutic properties, chemical and pharmacological properties. It is generally accepted that compounds with similar physicochemical properties exhibit similar biological activity. As 138 of 1,562 drugs in DB1 have no ATC code, we adopted their predicted codes by SPACE [31], which deduce ATC codes from chemical structures. To feed the 7-bit ATC code into DNN, we convert them into a one-hot code with 118 bits.

We also used drug-binding proteins (DBP) collected by [15], including 899 drug targets and 222 non-target proteins, which play important roles in DDI prediction. Similarly, each drug is represented as a binary DBP-based feature vector, of which each bit indicates whether the drug binds to a specific protein.

Problem formulation

Our task is to deduce DDI candidates among those unannotated drug pairs base on annotated DDIs in the form of a network. Technically, let $G(D, E)$ be a DDI network, where $D = \{d_1, d_2, \ldots, d_m\}$ is the set of m approved drugs and E denotes the interactions between them. This network can be usually represented by an $m \times m$ symmetric binary adjacent matrix $A_{m \times m} = \{a_{ij}\}$, where $a_{ij} = 1$ indicates an annotated interaction between drug d_i and drug d_j, and otherwise no annotated interaction between drug d_i and drug d_j.

DDI prediction can be solved by a three-step approach. First, function $f_1(A)$ to obtain the latent feature vector Z_i of each drug in A, where $Z_i \in R^{1 \times k} (k \ll m)$. Next, the latent vectors (Z_i and Z_j) of two drugs are aggregated to the feature vector of a

drug pair. Last, function $f_2(Z_i, Z_j)$ ($Z_i, Z_j \in Z$) reconstructs the network $\breve{A}$. Function f_1 is called the feature extractor, while function f_2 is named as the predictor in our model. In this work, by implementing the solution based on deep learning, we provide a Deep Predictor for Drug-Drug Interactions (DPDDI). Its overall framework is illustrated in Fig.2.

The loss of DPDDI contains two parts as follows:

$$Loss = L_f(p, q) + L_p(p, q) \quad (1)$$

where L_f is the loss of its feature extractor and L_p is the loss of its predictor. The first part adopts a binary weighted-cross-entropy as follows:

$$L_f(p, q) = -\sum_{i,j} p(a_{ij}) \log(q(a_{ij})) * W_{pos} + (1 - p(a_{ij}))(1 - \log(q(a_{ij})) \quad (2)$$

where $p(a_{ij})$ is the true label of the training interaction a_{ij}, $q(a_{ij}) = \sigma(z_i \cdot z_j^T)$ is the predicting probability computed by the inner product of latent vectors of two nodes generated by the GCN, and W_{pos} is the weight equal to the number of negative samples over the number of positive samples. The second part is defined by a binary cross-entropy as follows:

$$L_p(p, q) = -\sum_{i,j} p(a_{ij}) \log(s(a_{ij})), \quad (3)$$

where $s(a_{ij})$ is the predicting probability generated by the DNN.

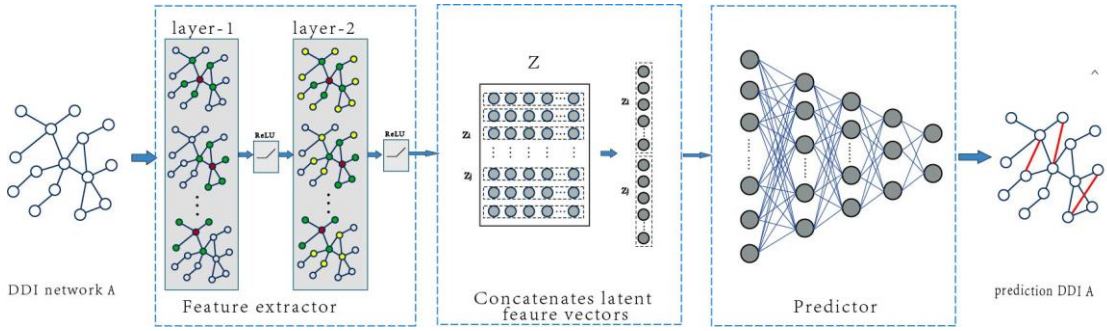

Fig. 2. The overall framework of DPDDI.

Feature extractor

We employ a two-layer auto-encoder of graph convolutional network(GCN)[22,

23] to obtain embedding representations of drug nodes. Each drug is represented as a latent feature vector, which contains the high-dimensional information about its neighborhood in the DDI network without manual feature engineering. Such a node embedding provides a promising way of representing the relationship between nodes in a complex network.

Technically, the GCN takes the adjacent matrix A as the input and outputs embedding vectors $\{\mathrm{Z_i} \in \mathrm{R}^{1\times \mathrm{H_p}}, i = 1,2,\dots,m\}$ for every drug in the DDI network, where $\mathrm{H_p}$ is the dimension of the last hidden layer. Like [32] recommended, our GCN adopts two layers as well. Suppose that $\mathrm{H}^{(0)}$ is the feature matrix, of which each row denotes the input feature vector of each node in the network. In the case of no input feature, $\mathrm{H}^{(0)}$ is just an identity matrix. Then, the output $\mathrm{H}^{(1)}$ of the first hidden layer is defined as:

$$\mathrm{H}^{(1)} = f\left(H^{(0)}, A\right) = \mathrm{ReLU}\left(\widehat{\mathrm{A}}\mathrm{H}^{(0)}W^{(0)}\right), \tag{4}$$

where $\widehat{\mathrm{A}} = \widetilde{\mathrm{D}}^{-\frac{1}{2}}\widetilde{\mathrm{A}}\widetilde{\mathrm{D}}^{-\frac{1}{2}}$ is the symmetrically normalized adjacent matrix, $\widetilde{\mathrm{D}}_{ii} = \sum_j \widetilde{\mathrm{A}}_{ij}$ and $\widetilde{\mathrm{A}} = \mathrm{A} + I_N$, $\mathrm{W}^{(0)} \in \mathrm{R}^{\mathrm{m}\times \mathrm{H_1}}$ is the weight matrix to be learned, and ReLU is the activation function. Similarly, the output $\mathrm{H}^{(2)}$ of the second hidden layer is recursively defined as:

$$\mathrm{H}^{(2)} = \mathrm{f}\left(\mathrm{H}^{(1)}, \mathrm{A}\right) = \mathrm{ReLU}(\widehat{\mathrm{A}}\mathrm{H}^{(1)}\mathrm{W}^{(1)}), \tag{5}$$

where $\mathrm{W}^{(1)} \in \mathrm{R}^{\mathrm{H_1}\times \mathrm{H_2}}$. Because our GCN contains only two layers, $\mathrm{H}^{(2)}$ is just the final embedding matrix $\mathrm{Z} \in \mathrm{R}^{\mathrm{m}\times \mathrm{H_2}}$.

Feature aggregation for drug pairs

So far, latent feature vectors of single drugs in the embedding space are obtained. The next task is to obtain feature vectors of drug pairs. Given two drugs d_i and d_j, and their latent vectors $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$ obtained by GCN. We consider three ways to aggregate the latent feature vectors of two drugs into a feature vector of the corresponding drug pair.

First, we adopt the inner product of two drug latent vectors Z_i and Z_j as the feature of the drug pair (d_i, d_j). Formally, $\boldsymbol{F}(d_i, d_j) = \boldsymbol{Z}_i \boldsymbol{Z}_j^{\boldsymbol{T}}$, which is the original way in GCN. Secondly, we adopt the summation of these drug latent vectors as the drug pair feature. That is $\boldsymbol{F}(d_i, d_j) = \boldsymbol{Z}_i + \boldsymbol{Z}_j$. Thirdly, we adopt their concatenation $\boldsymbol{F}(d_i, d_j) = [\boldsymbol{Z}_i, \boldsymbol{Z}_j]$.

Predictor

Given the feature vectors of drug pairs, we construct a deep neural network (DNN) as the predictor in DPDDI because of its proven excellent performance in classification. The predictor transforms DDI prediction as a binary classification, which is implemented by a five-layer DNN. The numbers of neurons in the layers in the DNN are 256, 128, 64, 32 and 2 respectively. See also Section 3.1 Parameter tuning. ReLU is adopted as the activation function in the first four layers while SoftMax is the activation function in the last layer, which outputs how likely drug pairs are potential DDIs.

Assessment

For all the edges in the DDI network, we randomly select 75% DDIs as the training set, 5% as the validation set, and the remaining 20% as the testing set. Both the training set and the validation set are used to train a model with a good learning performance while the testing set is used to measure the generalization performance of the trained model. The procedure is repeated 10 times under different random seeds. The average performance of 10 repetitions is reported.

The performance of DDI prediction is measured by multiple popular metrics, including the Area Under the Receiver Operating Characteristic Curve (AUROC), (the Area Under the Precision-Recall curve, AUPR), recall, precision, accuracy (ACC), and F1-score. Their definitions are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

$$\text{TPR} = \frac{TP}{TP + FN} \tag{10}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{11}$$

where TP, FP, TN and FN refer to the number of true positive samples, false positive samples, true negative samples and false negative samples, respectively; TPR denotes true positive rate that also known as sensitivity and Recall, which measures the proportion of actual positive samples that are correctly identified; FPR is false positive rate. Because the sum total of $(TP + FP + TN + FN)$ equals the number of all samples, Accuracy denotes the proportion of all samples predicted correctly, and Precision refers to the proportion of positive samples predicted correctly. F_1 score is the harmonic mean of precision and recall.

Availability of data and materials

All the code and data are openly available at the website of https://github.com/fenshiwuhen/DPDDI.

Abbreviations

DDIs: drug-drug interactions

GCN: graph convolution network

DNN: deep neural network

ATC: Anatomical Therapeutic Chemical classification

DBP: drug-binding protein

AUROC: Area Under the Receiver Operating Characteristic Curve

AUPR: Area Under the Precision-Recall curve

ACC: accuracy

References

1. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MCJNB: **Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions**. *Nat Biotechnol* 2017, **35**:463-474.
2. Takeda T, Hao M, Cheng T, Bryant SH, Wang YJJoC: **Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge**. *J Cheminform* 2017, **9**(16).
3. Pathak J, Kiefer RC, Chute CG: **Using linked data for mining drug-drug interactions in electronic health records**. *Stud Health Technol Inform* 2013, **192**:682-686.
4. Duke JD, Xu H, Zhiping W, Abhinita S, D. KS, Xiaochun L, D. HS, Yan J, Thomas CJ, Biology OMJJPC: **Literature Based Drug Interaction Prediction with Clinical Assessment Using Electronic Medical Records: Novel Myopathy Associated Drug Interactions**. *PloS cOMPUT biiol* 2012, **8**(8):e1002614.
5. Vilar S, Friedman C, Hripcsak GJBiB: **Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media**. *Brief Bioinform* 2018, **19**(5):863-877.
6. Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman CJJotAMIAJ: **Drug-drug interaction through molecular structure similarity analysis**. *Journal of the American Meidical Informatics Association* 2012, **19**(6):1066-1074.
7. Vilar S, Uriarte E, Santana L, Tatonetti NP, Friedman CJPo: **Detection of drug-drug interactions by modeling interaction profile fingerprints**. *PLOS ONE* 2013, **8**(3):e58321.
8. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan RJMSB: **INDI: a computational framework for inferring drug interactions and their associated recommendations**. *Mol Syst Biol* 2012, **8**(1).
9. Sridhar D, Fakhraei S, Getoor LJB: **A Probabilistic Approach for Collective Similarity-based Drug-Drug Interaction Prediction**. *Bioinformatics* 2016, **32**(20):3175-3182.
10. Cheng F, Zhao ZJJotAMIAJ: **Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties**. *Journal of the American Meidical Informatics Association* 2014, **21**(e2):e278-e286.
11. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li XJBB: **Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data**. *BMC bioinformatics* 2017, **18**(1):18.
12. Andrej K, Polonca F, Brane Le, One YJ-MJP: **Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning**. *J Cheminform*, **13**(5):e0196865-.
13. Zhang P, Wang F, Hu J, Sorrentino R: **Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects**. *Scientific Reports* 2015, **5**(1):12339.
14. Yu H, Mao K, Shi J, Huang H, Chen Z, Dong K, Yiu SJBSB: **Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization**. *BMC systerms biology* 2018, **12**(1):14.

15. Shi J-Y, Mao K-T, Yu H, Yiu S-MJJoC: **Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization**. *Journal of cheminformatics* 2019, **11**(1).
16. Park K, Kim D, Ha S, Lee DJPO: **Predicting Pharmacodynamic Drug-Drug Interactions through Signaling Propagation Interference on Protein-Protein Interaction Networks**. *PLOS ONE* 2015, **10**(10):e0140816.
17. Miotto R, Wang F, Wang S, Jiang X, Dudley JT: **Deep learning for healthcare: review, opportunities and challenges**. *Briefings in Bioinformatics* 2017, **19**(6):1236-1246.
18. Angermueller C, Pärnamaa T, Parts L, Stegle OJMsb: **Deep learning for computational biology**. *Molecular systems biology* 2016, **12**(7).
19. Ryu JY, Kim HU, Lee SYJPotNAoS: **Deep learning improves prediction of drug–drug and drug–food interactions**. *Proc Natl Acad Sci* 2018, **115**(18):E4304-E4311.
20. Zhou J, Cui G, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M: **Graph Neural Networks: A Review of Methods and Applications**. In: *arXiv e-prints arXiv:161107308.* 2018.
21. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS: **A Comprehensive Survey on Graph Neural Networks**. In: *arXiv e-prints.* 2019.
22. Kipf TN, Welling M: **Variational Graph Auto-Encoders**. In: *arXiv e-prints* 2016.
23. Kipf TN, Welling MJapa: **Semi-supervised classification with graph convolutional networks**. 2016.
24. Tatonetti NP, Ye PP, Daneshjou R, Altman RBJSTM: **Data-Driven Prediction of Drug Effects and Interactions**. *Science translational medicine* 2012, **4**(125):125ra131-125ra131.
25. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z *et al*: **DrugBank 5.0: a major update to the DrugBank database for 2018**. *Nucleic Acids Research* 2017, **46**(D1):D1074-D1082.
26. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J: **Distributed representations of words and phrases and their compositionality**. In: *Advances in neural information processing systems: 2013*. 3111-3119.
27. Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G: **Complex Embeddings for Simple Link Prediction**. 2017.
28. Zitnik M, Agrawal M, Leskovec JJB: **Modeling polypharmacy side effects with graph convolutional networks**. *Bioinformatics* 2018, **34**(13):i457-i466.
29. Vivian L, Craig K, Yannick D, Tim J, Guo AC, Liu Y, Adam M, David A, Michael W, Vanessa NJNAR: **DrugBank 4.0: shedding new light on drug metabolism**. *Nucleic acids research* 2013, **42**(D1):D1091-D1097.
30. Skrbo A, Begović B, Skrbo S: **Classification of drugs using the ATC system (Anatomic, Therapeutic, Chemical Classification) and the latest changes**. *Medicinski arhiv* 2004, **58**(1 Suppl 2):138-141.
31. Liu Z, Guo F, Gu J, Wang Y, Li Y, Wang D, Lu L, Li D, He F: **Similarity-based prediction for Anatomical Therapeutic Chemical classification of drugs by integrating multiple data sources**. *Bioinformatics* 2015, **31**(11):1788-1795.
32. Defferrard M, Bresson X, Vandergheynst P: **Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering**. *Advances in neural information processing systems*:3844-3852.

Declarations

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None of the authors has any competing interests.

Funding

This work has been supported by the National Natural Science Foundation of China (No. 61872297 and No. 61873202).

Affiliations

Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China

Yue-Hua Feng, Shao-Wu Zhang

School of Life Sciences, Northwestern Polytechnical University, Xi'an, China

Jian-Yu Shi

Authors' contributions

YHF collected the dataset, performed the experiments and drafted the manuscript. JYS analyzed the result. Both JYS and SWZ modified manuscript and they are the corresponding authors. All authors read and approved the final manuscript.

Correspondence

Shao-Wu Zhang, zhangsw@nwpu.edu.cn ; Jian-Yu Shi, jianyushi@nwpu.edu.cn;

Acknowledgements

We acknowledge anonymous reviewers for the valuable comments on the original manuscript.

Figure Legends

Fig. 1. The overall framework of DPDDI.

The main steps are as follows. First, the feature extractor of DPDDI constructs a two-layer graph convolutional network (GCN[22, 23]) to obtain drug latent features, which capture the complex relations between the drug nodes in the DDI network. Then, each pair of drugs is represented as a feature vector by concatenating the corresponding latent features of the drugs. Last, the predictor of DPDDI trains a regular deep neural network to deduce potential DDIs.

Fig. 2. Performance of DPDDI on unbalance negative samples.