<u>Original Paper</u>

Drug-Drug Interaction Predictions via Knowledge Graph and Text Embedding: Instrument Validation Study

Meng Wang[1,2*], PhD; Haofen Wang[3*], PhD; Xing Liu[4*], PhD; Xinyu Ma[1*], MA; Beilun Wang[1*], PhD

[1]School of Computer Science and Engineering, Southeast University, Nanjing, China

[2]Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, China

[3]College of Design and Innovation, Tongji University, Shanghai, China

[4]Third Xiangya Hospital, Central South University, Changsha, China

[*]all authors contributed equally

Corresponding Author:

Haofen Wang, PhD
College of Design and Innovation
Tongji University
No. 281 Fuxin Road
Yangpu District
Shanghai, 200092
China
Phone: 86 139 1858 6855
Email: carter.whfcarter@gmail.com

Abstract

Background: Minimizing adverse reactions caused by drug-drug interactions (DDIs) has always been a prominent research topic in clinical pharmacology. Detecting all possible interactions through clinical studies before a drug is released to the market is a demanding task. The power of big data is opening up new approaches to discovering various DDIs. However, these data contain a huge amount of noise and provide knowledge bases that are far from being complete or used with reliability. Most existing studies focus on predicting binary DDIs between drug pairs and ignore other interactions.

Objective: Leveraging both drug knowledge graphs and biomedical text is a promising pathway for rich and comprehensive DDI prediction, but it is not without issues. Our proposed model seeks to address the following challenges: data noise and incompleteness, data sparsity, and computational complexity.

Methods: We propose a novel framework, Predicting Rich DDI, to predict DDIs. The framework uses graph embedding to overcome data incompleteness and sparsity issues to make multiple DDI label predictions. First, a large-scale drug knowledge graph is generated from different sources. The knowledge graph is then embedded with comprehensive biomedical text into a common low-dimensional space. Finally, the learned embeddings are used to efficiently compute rich DDI information through a link prediction process.

Results: To validate the effectiveness of the proposed framework, extensive experiments were conducted on real-world data sets. The results demonstrate that our model outperforms several state-of-the-art baseline methods in terms of capability and accuracy.

Conclusions: We propose a novel framework, Predicting Rich DDI, to predict DDIs. Using rich DDI information, it can competently predict multiple labels for a pair of drugs across numerous domains, ranging from pharmacological mechanisms to side effects. To the best of our knowledge, this framework is the first to provide a joint translation-based embedding model that learns DDIs by integrating drug knowledge graphs and biomedical text simultaneously in a common low-dimensional space. The model also predicts DDIs using multiple labels rather than single or binary labels. Extensive experiments were conducted on real-world data sets to demonstrate the effectiveness and efficiency of the model. The results show our proposed framework outperforms several state-of-the-art baselines.

XSL•FO
RenderX

Introduction

An increasing amount of research in clinical studies is focusing on drug-drug interactions (DDIs) because the majority of adverse drug reactions (ADRs) occur between pairs of drugs. ADRs may lead to patient morbidity and mortality, accounting for 3% to 5% of all in-hospital medication errors [1]. Furthermore, patients with 2 or more diseases (eg, older adult patients with chronic diseases) have a higher risk of an ADR if they take 5 or more different drugs simultaneously [2,3]. Detecting DDIs based on experimentation is a time-consuming and laborious process for clinicians. This signals the need for a more comprehensive and automated method of predicting unknown DDIs before a new drug can be released.

Traditional experimental approaches in vitro [4], in vivo [5], and in populo [6] focus on small sets of specific drug pairs and have laboratory limitations. Many machine learning approaches, such as similarity or feature-based approaches [7-9], have been proposed to predict DDIs. Recently, several graph neural networks and long short-term memory methods based on knowledge graphs (KGs), such as KG neural network [10] and KG-DDI [11], have significantly outperformed traditional shallow machine learning methods. The superior performance of these proposed methods can be attributed to their use of the prior knowledge and learning of higher-level representations for DDI detection. However, as these approaches only predict binary DDIs or those that have been predefined in structured databases, they may be hampered by robustness caused by data sparsity and vast computation requirements. Although several approaches [12-14] have used natural language processing techniques to extract DDIs from biomedical text, to the best of our knowledge, they have not employed drug KGs to improve performance.

With the increasing emergence of biomedical data, many world-leading biomedical researchers are now focusing on automatically populating and completing biomedical KGs using the huge volume of structured databases and text available to the public. HKG [15], Knowlife [16], and DrugBank [17] are just a few examples. Efforts such as Bio2RDF [18] and Linked Open Drug Data [19] have mapped similar entities in different KGs and built large heterogeneous graphs that contain an abundance of basic biomedical facts about drugs. SPARQL [20], a query language for KGs, supports the retrieval and manipulation of drug-related facts distributed over different KGs. Unfortunately, these biomedical KGs are affected by incomplete and inaccurate data that impede their application in the field of safe medicine development.

Existing KGs already include thousands of relation types, millions of entities, and billions of facts [19]. As noted, KG applications based on conventional graph-based algorithms are restricted by data sparsity and computational inefficiency. To address these problems, graph embedding techniques [9,21-26] based on representation learning for KGs have been proposed that embed both entities and relations into a continuous low-dimensional vector space. Among these methods,

translation-based models [9,22,24] are the most simple and effective. Currently, they represent the state-of-the-art in knowledge acquisition and inference and link prediction [9]. In light of these analogies, DDIs can be treated as a category of relations in a drug KG, and KG embedding techniques can be used to predict unknown DDIs. However, most translation-based methods only concentrate on predefined relations or unstructured text and fail to exploit the link between existing relations and rich unstructured text.

Leveraging both drug KGs and biomedical text is a promising pathway for rich and comprehensive DDI prediction, but it is not without issues. Our proposed model seeks to address the following challenges: data noise and incompleteness—real-world KGs are known to be inaccurate, incomplete, and unreliable for direct use; data sparsity—the potential DDI information in both KGs and biomedical text is sparse, and estimating the potential DDIs in such a long-tailed distribution is difficult; computational complexity—undoubtedly, this will be precluded from practice if graph-based algorithms are employed to process large-scale KGs or represent data objects with simple one-hot feature vectors.

Given these challenges, we propose a novel framework called Predicting Rich DDI (PRD). The framework is based on graph embedding techniques and treats specific DDI predictions as a linked prediction process. The proposed framework proceeds as follows: A large, high-quality drug KG is generated from distributed drug resources, which includes data on drug-target interactions, the impact of drugs on gene expression, the outcomes of drugs in clinical trials, and so on. A novel translation-based embedding model embeds the entities and relations in the drug KG into a low-dimensional space, and an autoencoder incorporates the descriptions of the DDIs from biomedical text as representations into the same semantic space. The decoder predicts the corresponding labels for potential DDIs based on the learned embeddings.

To the best of our knowledge, our PRD approach is the first method that is able to predict comprehensive and specific DDIs based on large-scale drug KGs and comprehensive biomedical text on pharmacology and ADRs. Our method further includes a joint translation-based embedding model that encodes the KG and rich DDI information from biomedical text into a shared low-dimensional space. The DDI predictions are then translated into a linked prediction process from the learned embeddings. Extensive experiments on real-world data sets were conducted to evaluate the framework. The results show that the framework can be powerful in predicting rich DDIs and outperforms several state-of-the-art baselines in terms of both capability and accuracy.

Methods

Figure 1 shows the architecture of the proposed framework. It consists of 3 key phases: drug KG generation, joint embedding learning, and DDI relations prediction.

Figure 1. Overview of the framework. DDI: drug-drug interaction.
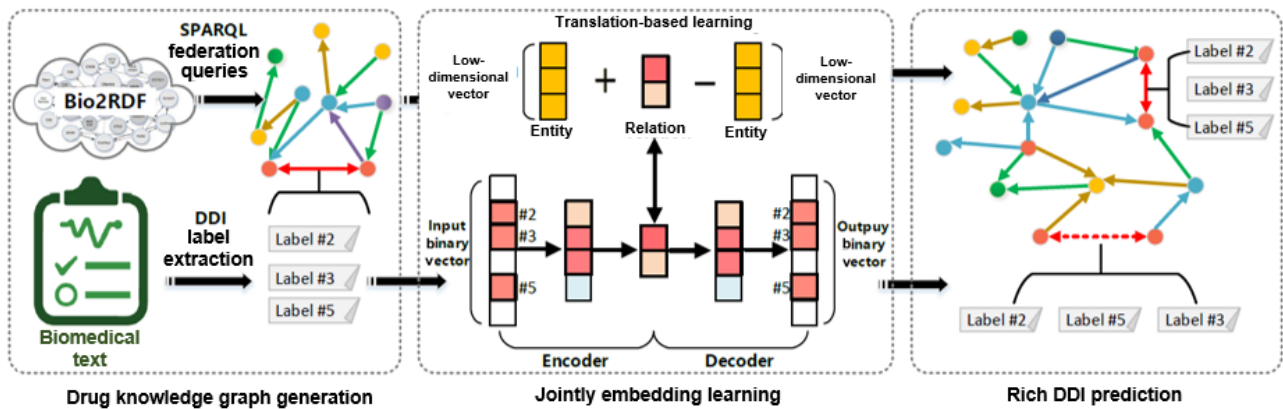
Drug KG Generation

A typical KG usually arranges knowledge as a triple set of facts that indicates the relation between 2 entities, and thus comprises a head entity, a relation, and a tail entity. These are denoted as (h, r, t).

First, a basic drug KG is constructed by collecting drug-related entities and relations among these entities. We follow the data model of drug-related extraction settings defined in the work of Kamdar and Musen [27], in which the types of entities or relations are summarized in the fashion depicted in Table 1. Specifically, we use SPARQL federation queries [20] to extract triples that contain 4 types of drug-related entities ($E_1 \sim E_4$) and

5 types of biological relations ($R_1 \sim R_4$) from a variety of biomedical sources (eg, Bio2RDF [18]). These extracted triples are defined as basic triples in our drug KG according to definition 1: (basic triple) $B = (E, R)$ is a set of basic triples in the form (h, r, t), where $E = E_1 \cup E_2 \ldots \cup E_4$ is a set of entities; and $R = R_1 \cup R_2 \ldots \cup R_5$ is a set of relations, $h, t \in E$, and $r \in R$.

For instance, we can extract "(etanercept, hasTarget, lymphotoxin-alpha)" as a basic triple in our drug KG, which indicates that there is a relationship "hasTarget" linking etanercept to lymphotoxin-alpha, meaning that lymphotoxin-alpha is one of the targets of etanercept.

Table 1. Entities and relations of basic triples in Kamdar and Musen [27].

Variable	Entity or relation interpretation
E	
E_1	Drugs
E_2	Drugs
E_3	Pathways
E_4	Phenotypes
R	
R_1	Drug, hastarget, protein
R_2	Drug, hasenzyme, protein
R_3	Drug, hastransporter, protein
R_4	Protein, ispresentin, pathway
R_5	Pathway, isimplicatedin, phenotype

A specific DDI between 2 drugs can be captured by multiple key phrases extracted from biomedical text, as shown in Figure 2. Hence, we collect biomedical DDI text documenting drug pairs (eg, DDI corpus [28], MEDLINE abstracts, and DrugBank DDI documents). We remove all stop words from raw text and use an entity linking method [29] to align the drug names in the

biomedical text with the KG. The top-n labels (n=5) are then selected from the biomedical text for each DDI based on the term frequency-inverse document frequency (TF-IDF) features (some other textual features can be used to rank the labels instead).

Figure 2. A drug knowledge graph is shown on the left with missing relations represented as dotted lines. There is usually no direct DDI relation between drugs. DDI descriptions from the biomedical text are shown on the right. The words in red represent concerns regarding DDI information in terms of both adverse DDIs and in-depth ways drugs can interact in pharmacology. DDI: drug-drug interaction.
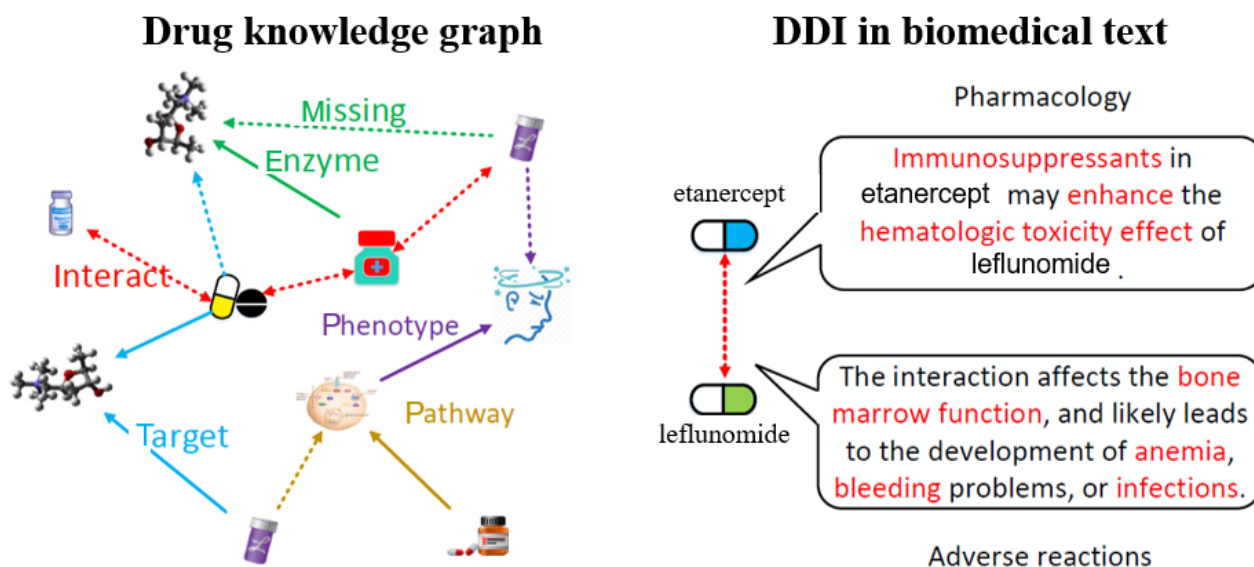
Based on this, the DDI relations between drug entities are defined as a set of labels rather than as a single label according to definition 2: (rich DDI triple) $T = (E_1, L)$ is a set of rich DDI triples in the form (u, l, v), where E_1 is a set of drug entities; L is a fixed label vocabulary from biomedical text; and $u, v \in E_1$ and $l = \{n_1, n_2, \ldots\} \subseteq L$ is the set of labels to describe the DDI information.

For instance, the following is an example of a rich DDI triple: (etanercept, {immunosuppressants, enhancetoxicity, anemia, infections}, leflunomide), where "enhancetoxicity" means etanercept can enhance the toxicity of leflunomide. Note that the DDI relations between 2 drugs are bidirectional; hence, our method replaces each rich DDI relation with 2 directed triples of opposing directions in the drug KG.

Formally, the generated drug KG is defined according to definition 3 (drug KG): the drug KG, G, is denoted as (E, B, T), where $E = E_1 \cup E_2 \ldots \cup E_4$ is a set of entities, B is a set of basic triples, and T is a set of rich DDI triples.

Joint Embedding Learning

KG embedding mainly consists of 3 steps: representation of entities and relations, definition of a scoring function, and encoding of the entity and relation into dense vectors. This section introduces the translation-based KG embedding model that learns representations from the drug KG, G = (E, B, T) and the optimization described in the following sections.

Basic Triple Encoder

For a set of basic triples, B, the method aims to encode entities and relations into a continuous vector space. This paper, without loss of generality, uses the bold letters $\mathbf{h}$, $\mathbf{r}$, $\mathbf{t}$ to denote the embedding vectors h, r, t. We adopt the translation-based mechanism $\mathbf{h}+\mathbf{r}\approx\mathbf{t}$ to capture the correlations between entities and relations. Translation in this context refers to a translation operation $\mathbf{r}$ between 2 entity vectors $\mathbf{h}$ and $\mathbf{t}$ in the low-dimensional space. We follow the TransR model in Lin et

al [22] to represent entities and relations in distinct vector spaces bridged by relation-specific matrices so as to learn more thorough graph representations. Specifically, for each triple, (h, r, t) $\in B$, h and t are embedded into h, t $\in R^k$, and r is embedded into $r \in R^d$. For each relation r, a projection matrix $M_r \in R^{(k \times d)}$ projects entities from the entity space to the relation space. The energy function $z_{bte}(\mathbf{h, r, t})$ is then defined as follows:

$$z_{bte}(\mathbf{h, r, t}) = b_1 - \| \mathbf{hM_r} + \mathbf{r} - \mathbf{tM_r} \|_{(L1/L2)} \quad (1)$$

where b_1 is a bias constant.

The conditional probability of a triple h, r, t is defined as follows:

$$P(h|r,t) = \frac{\exp\{z_{bte}(\mathbf{h,r,t})\}}{\sum_{\bar{h} \in E} \exp\{z_{bte}(\mathbf{h,r,t})\}}$$

(2)

$P(t|h, r)$, $P(r|h, t)$ can be defined in an analogous manner. The likelihood of observing a triple (h, r, t) is defined as follows:

$$L(h, r, t) = \log P(h \mid r, t) + \log P(t \mid h, r) + \log P(r \mid h, t) \quad (3)$$

By maximizing the conditional likelihoods of all existing triples in B, the objective function is defined as follows:

$$\mathcal{L}_{bte} = \sum_{(h,r,t) \in B} \mathcal{L}(h, r, t)$$

(4)

It is worth mentioning that other graph embedding models, such as HOLE [23], can also be easily adopted for basic triple encoding. In the interest of brevity, this paper only explores the effectiveness of TransR.

Rich DDI Triple Encoder

The interaction l between 2 drug entities, u and v, in rich DDI triples (u, l, v), $\in T$, can also be represented as translations in low-dimensional space. We set $\mathbf{u}, \mathbf{v} \in R^k$, $\mathbf{l} \in R^d$. The energy function $z_{dte}(u, l, v)$ is defined as follows:

$$z_{dte}(\mathbf{u}, \mathbf{l}, \mathbf{v}) = b_2 - \parallel \mathbf{u}\mathbf{M}_r + \mathbf{l} - \mathbf{v}\mathbf{M}_l \parallel_{(L1/L2)} \quad \textbf{(5)}$$

where b_2 is a bias constant and $\mathbf{M}_l = \mathrm{R}^{\times d}$ is the projection matrix. Following the analogous method in the basic triple encoder, the conditional likelihoods of all existing triples are maximized as follows:

$$\mathcal{L}_{dte} = \sum_{(u,l,v)\in T} \mathcal{L}(u, l, v)$$

(6)

Note, in equation 5, l is the relation representation obtained from $l = \{n_1, n_2,\dots\}$. This will be introduced in-depth next.

A deep autoencoder is employed to construct the relation representation $l \in R^d$ for a rich DDI triple $(u, l, v) \in$ T. Specifically, a DDI relation, l, is described by a set of labels $l = \{n_1, n_2,\dots\} \subseteq L$. The corresponding binary vector for l is initialized as $\mathbf{s} = \{s_i\}_{i=1}^{|L|}$, where $s_i = 1$ if $n_i \in l$, and $s_i = 0$ otherwise. The deep autoencoder then takes the binary vector $\mathbf{s}$ as input and uses the following nonlinear transformation layers to transform the label set into the low-dimensional space R^k:

$$h^{(1)} = f(\mathbf{W}^{(1)}\,\mathbf{s} + \mathbf{b}^{(1)})$$
$$h^{(i)} = f(\mathbf{W}^{(i)}h^{(i-1)} + b^{(i)}), \; I = 2, \dots, K \quad \textbf{(7)}$$

where f is the activation function and K is the number of layers. Here, $h^{(i)}$, $\mathbf{W}^{(i)}$, and $\mathbf{b}^{(i)}$ represent the hidden vector, transformation matrix, and the bias vector in the i-th layer, respectively.

There are 2 parts to the autoencoder: an encoder and a decoder. The encoder employs the *tanh* activation function to obtain the DDI relation representation $\mathbf{l} = h^{(K/2)}$. The decoder deciphers the embedding vector of $\mathbf{l}$ to obtain a reconstructed vector $\tilde{\mathbf{s}}$. Intuitively, PRD should then minimize the distance $\|\mathbf{s} - \tilde{\mathbf{s}}\|_{L1/L2}$ because the reconstructed vector $\tilde{\mathbf{s}}$ should be similar to $\mathbf{s}$. However, the number of zero elements in $\mathbf{s}$ is usually much larger than that of nonzero elements due to data sparsity. This leads the decoder to tend to reconstruct zero elements rather than nonzero elements, which conflicts with our purpose. To overcome this obstacle, different weights are set for different elements, and the following objective function is maximized:

$$z_l(\mathbf{s}, \tilde{\mathbf{s}}) = b_3 - \|(\mathbf{s} - \tilde{\mathbf{s}}) \odot \mathbf{x}\|_{L1/L2}$$

(8)

where b_3 is a bias constant, $\mathbf{x}$ is a weight vector, and $\odot$ is denoted as the Hadamard product. For $\mathbf{x} = \{x_i\}_{i=1}^{|L|}$, $\mathbf{x}_i = 1$, if $s_i = 0$, and $\mathbf{x}_i = \beta > 1$ otherwise. According to equation 8, the probability of $P(\mathbf{s}|\tilde{\mathbf{s}})$ can be defined as follows:

$$P(\mathbf{s}|\tilde{\mathbf{s}}) = \frac{\exp\{z_l(\mathbf{s},\tilde{\mathbf{s}})\}}{\sum_{\hat{\mathbf{s}}\in S} \exp\{z_l(\mathbf{s},\tilde{\mathbf{s}})\}}$$

(9)

where S is the set of binary vectors of all DDI relations. The likelihood of reconstructing the binary vector s of a relation l can be defined as follows:

$$L(l) = log P(\mathbf{s}|\tilde{\mathbf{s}}) \quad \textbf{(10)}$$

By maximizing the likelihoods of the encoding and the decoding for all described relations l, the objective function can be defined as follows:

$$\mathcal{L}_{rcl} = \sum_{l\in T} \mathcal{L}(l)$$

(11)

Joint Learning and Optimization

The goal of the framework PRD is to not only represent the basic triples (drug KG *B*) but also the rich DDI triples (biomedical text *T*) in a unified joint embedding model. Considering the above 3 objective functions (equations 4, 6, and 11) together, the optimization function is defined as follows:

$$O(X) = \mathrm{L}_{bte} + \mathrm{L}_{dte} + \mathrm{L}_{rcl} + \gamma C(X) \quad \textbf{(12)}$$

where X represents the embeddings of entities and relations, and γ is a hyper-parameter that weights the regularization factor $C(X)$, which is defined as follows:

$$C(X) = \sum_{e\in E}[\|\mathbf{e}\| - 1]_+ + \sum_{r\in R}[\|\mathbf{r}\| - 1]_+ + \sum_{l\in L}[\|\mathbf{l}\| - 1]_+ \quad \textbf{(13)}$$

where $[x]_+ = max(0, x)$ denotes the positive part of x. The regularization factor will normalize the embeddings during learning. We adopted the approach by Srivastava et al [30] to prevent deep neural networks from overfitting and used the Adam algorithm [31] to maximize the objective function.

It is impractical to directly compute the normalizers in $P(h \mid r, t)$, $P(t \mid h, r)$, $P(r \mid h, t)$, and $P(\mathbf{s}|\tilde{\mathbf{s}})$, as calculating them would require summing the complete set of entities and relations. To address this problem, we use the negative sampling method from Mikolov et al [32] to transform the objective functions. Taking $P(h \mid r, t)$ as an example, the following objective function is maximized instead of using its original form:

$$log\sigma(z_{bte}(\mathbf{h}, \mathbf{r}, \mathbf{t}))$$
$$+ \sum_{m=1}^{c} E_{\tilde{h}_m \sim z_{neg}(\{(\tilde{h},r,t)\})}[-\sigma(z_{bte}(\tilde{\mathbf{h}}_m, \mathbf{r}, \mathbf{t}))]$$

(14)

where c is the number of negative examples, $\sigma(x) = 1/(1 + exp(-x))$ is the sigmoid function, $\{(\tilde{h}, r, t)\}$ is the invalid triple set, and z_{neg} is a function randomly sampling instances from $\{(\tilde{h}, r, t)\}$. When a positive triple $(h, r, t) \in B$ is selected to maximize equation 14, c-negative triples are constructed by sampling entities from a uniform distribution over E and replacing the head of (h, r, t). The objective functions of $P(r \mid h, t)$, $P(t \mid h, r)$, $log P(\mathbf{s}|\tilde{\mathbf{s}})$, and $L(u. l. v)$ are transformed and maximized in an equivalent manner. Finally, PRD iteratively selects random mini-batches from the training set to learn the embeddings efficiently until convergence.

DDI Relations Prediction

The DDI prediction task can be defined as a link prediction problem on KG; that is, with the learned deep autoencoder and the embedding vectors of all entities and relations, the

framework PRD can leverage the translation mechanism to predict the missing DDI relations between 2 drug entities. To be more specific, given 2 drug entities u, $v \in E_1$, the following equation predicts the potential relation embedding **l** between **u** and **v.**

$$\mathbf{l} = \mathbf{v}\mathbf{M}_l - \mathbf{u}\mathbf{M}_l \ (15)$$

Finally, with the decoder part of the learned deep autoencoder, PRD can obtain the label set l by decoding the embedding vector l.

Results

To examines the effectiveness of the DDI prediction framework PRD, we performed 2 types of experiments. First, we compared the performance of our model to several baseline methods on binary-type DDI predictions. We then investigated PRD's strengths in modeling rich DDI relations between drug entities. The results demonstrate that PRD significantly outperformed the baselines in terms of both accuracy and capability.

Data Construction

Experiments in this paper were performed on 2 real drug-related data sets, Bio2RDF [18] and DDI Corpus [28].

Bio2RDF (version 4) is an open-source project that provides 11 billion triples from 35 biological and pharmacological KGs across a wide variety of drug-related entities, such as proteins, targets, and diseases. It is accessible online via the SPARQL endpoint.

DDI Corpus (2013 version) is a semantically annotated corpus of documents describing DDIs from the DrugBankdatabase and MEDLINE abstracts. It contains 233 MEDLINE abstracts and 784 DrugBank texts on the DDIs subjects. There are a total of 5021 annotated DDIs in 18,491 pharmacological sentences.

Following the federation queries in Kamdar and Musen [27], we extracted basic triples for our drug KG from 4 different KGs in Bio2RDF: (1) DrugBank [17] provides comprehensive data about drug, disease, and target information; (2) Kyoto Encyclopedia of Genes and Genomes [33] offers pathways, proteins, and drugs information; (3) PharmGKB [34] contains protein-drug-disease relations; (4) Comparative Toxicogenomics Database ([35] provides data about protein interactions and pathway-disease relations.

For the rich DDI triples, we collected 4694 DrugBank DDI sentences about 8197 drugs from the DDI corpus. The top 5 labels from each sentence were selected based on TF-IDF to construct rich DDI triples and build the DDI label vocabulary. To overcome the issue of inconsistent drug names between basic triples and rich DDI triples, we applied the entity linking method [29] to align the drug aliases.

The drug KG we constructed contains 71,460 basic triples, 4694 rich DDI triples, 8197 drug entities, 305,642 other entities, and 1053 distinct labels in the DDI vocabulary.

Baselines

For the baseline approaches, DDI prediction and state-of-the-art KG embedding methods were used. Three DDI methods were used:

1. Tiresias [8] is a large-scale similarity-based framework that predicts DDIs through link prediction. It takes various sources of drug-related data and knowledge as inputs and generates binary DDI predictions as outputs.
2. Syntax Convolutional Neural Network (SCNN) [36] represents a DDI extraction method based on a SCNN to extract 4 predefined DDI types (ADVICE, EFFECT, INT, and MECHANISM) from the biomedical literature.
3. Multitask dyadic DDI prediction (MDDP) [37] defines the DDI type prediction problem as a multitask dyadic regression problem. It can predict the specific DDI type between 2 drugs.

Two state-of-the-art KG embedding methods were used:

1. TransE [9] is the most representative translational distance model to embed components of a KG, including entities and relations, into continuous vector spaces. These embeddings can also be used for link prediction.
2. TransR [22] shares a similar approach with TransE, but represents entities and relations in distinct vector spaces bridged by relation-specific matrices.

Evaluation Method and Metrics

Given a drug KG with some DDI relations removed, rich DDI prediction aims to predict the occurrence of DDI relations among drug entities. DDI relations with a rate of 0.3 chosen randomly as the ground truths for the test set were removed, and the remaining KG was used as the training set. We also randomly sampled an equal number of drug pairs with no DDI relations to serve as the negative sample in the test set.

To make an unbiased comparison, we first treated DDI prediction as a binary classification task. Tiresias is already a binary classification model. For SCNN and MDDP, we defined the 2 DDI types as yes and no in the training model. For TransE, TransR, and our PRD method, we concatenated the representations of the entities of a candidate drug pair to form the feature vector and used logistic regression to train classifiers. We then treated multiple DDI type predictions as a multilabel classification task. For Tiresias, SCNN, and MDDP, we used their feature representation methods and adopted one-versus-rest logistic regression to train a multilabel classifier. For TransE and TransR, we separated each training triple (u, l, v) where $l = \{n_1, n_2,…\}$ into several triples (ie, $[u, n_i, v]$ for $n_i \in l$), which could be directly used to train the models.

We used 10-fold cross-validation on the training set to tune PRD's embedding model. We determined the optimal parameters using a grid search strategy. The search ranges for the various parameters were as follows: the learning rate λ for the Adam algorithm {0.1, 0.01, 0.001}; γ for the soft constraints {0.1, 0.01, 0.001}; the vector dimension k {20, 50, 80, 100}; and all bias constants b_1, b_2, b_3, c were 10 to 10. The training instances were conducted over 1000 iterations. The running time per iteration was 391 seconds. The best configurations for

the joint model were $\lambda=0.001$, $\gamma=0.01$, $k=100$, $b_1=5$, $b_2=5$, $b_3=1$, $c=10$, and $K=3$, with L_1 being used as a dissimilarity metric.

We used receiver operator characteristic curves and precision-recall curves to evaluate the proposed method on binary DDI-type predictions. For multiple DDI- type predictions, we followed the setting in TransE [9] and report the 2 measures as evaluation metrics: the average rank of all correct relations (MeanRank) and the proportion of correct relations ranked in top k (Hits@k). The above metrics may be biased for methods that rank other correct labels higher in the same label set. Hence, all other correct labels were filtered out before ranking. The filtered version is denoted as "Filter," and the unfiltered version is denoted as "Raw."

Experiment Results

As shown in Figure 3a and Figure 3b, the proposed framework PRD outperformed all baselines. In terms of the receiver operator characteristic curve, PRD outperformed Tiresias by 6.69%, TransR by 7.13%, and MDDP and TransE by 12%; meanwhile, SCNN had a relatively low predictive ability. According to the precision-recall curve, PRD learned 14.2% better than did Tiresias (which was at the top among the 3 DDI prediction baselines), 16.8% better than TransR, 21.57% better than MDDP, 25.33% better than TransE, and 37.89% better than SCNN.

Table 2 shows the evaluation results for rich DDI relation predictions according to the different evaluation metrics for both the raw and filter tests.

Figure 3. ROC and PR results of binary drug-drug interaction-type predictions. MDDP: multitask dyadic drug-drug interaction (DDI) prediction; ROC: receiver operator characteristic; PR: precision-recall.

(a) ROC curve

(b) PR curve
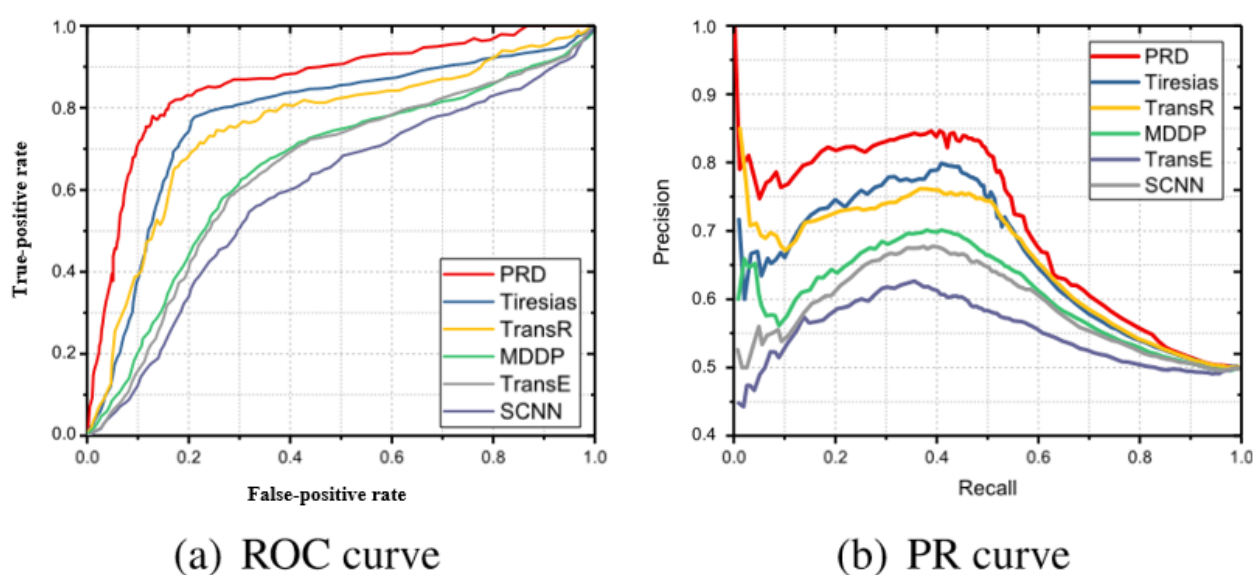
Table 2. Evaluation results for multiple drug-drug interaction relation predictions (×100 for Hits@k).

Framework	Raw				Filter			
	Hits@1[a]	Hits@5	Hits@10	MeanRank[b]	Hits@1	Hits@5	Hits@10	MeanRank
Tiresias	14.23	33.18	50.61	21.89	19.21	45.29	52.94	17.93
SCNN[c]	12.19	26.31	39.02	37.91	16.82	27.03	40.78	37.06
MDDP[d]	20.95	58.66	79.48	13.53	43.19	68.57	84.12	7.85
TransE	26.61	70.23	83.97	8.01	57.88	79.99	87.27	7.02
TransR	31.33	75.80	87.63	6.89	69.58	84.01	89.01	6.25
PRD[e]	45.11	85.57	91.01	6.11	75.11	88.60	92.85	5.45

[a]Hits@x: accuracy of real values contained in the top x rank.

[b]MeanRank: the average rank of all correct relations.

[c]SCNN: Syntax Convolutional Neural Network.

[d]MDDP: multitask dyadic drug-drug interaction prediction.

[e]PRD: Predicting Rich Drug-Drug Interaction.

Case Study

To further demonstrate PRD's ability for rich DDI predictions, we selected the drug acetylsalicylic acid (aspirin) as a test case. The DDI predictions and rich labels relations are shown in Table 3. According to the usefulness and diversity of the predicted labels, a professional pharmacist evaluated and annotated the practical useful predictions (labels in italics in Table 3). Observe that both TransR and PRD were able to recommend reasonable DDI labels for the drug interactions, representative of detailed DDI information. However, TransR sometimes recommended similar labels for a specific drug because it is based on a similarity method. Conversely, PRD was able to recommend discriminative labels because it uses a decoder.

We also present a case study to visualize the effectiveness of binary DDI types of prediction on a DDI network sample. We constructed drug-drug networks to indicate whether any 2 drugs would result in a binary DDI. A node in the network denotes a drug. An edge between 2 nodes denotes the existence of a DDI. Intuitively, the more drugs interact, the more risk there is. In the network, the size of the node specifies the degree of risk of a drug. We classified the degree of risk into various levels using different colors (ie, high risk is shown in dark green, and low risk is shown in light green). The red nodes denote forecasting errors of DDI drugs. As shown in Figure 4a to Figure 4f, PRD predicts DDIs mostly accurately. The ID of the drug with a high risk is shown on the node.

Table 3. Rich drug-drug interaction predictions for acetylsalicylic acid.

Interacted drug	TransR[a]	PRD[b]
Ibritumomab	*enhance*[c] *adverse*, *toxic*, risk, bleeding	*enhance*, *toxic*, *bleeding*, *platelet*, *antiplatelet*
Alteplase	*enhance*, increase, *adverse*, toxic, effect	*enhance*, *toxic*, *bleeding*, *thrombolytic*, *adverse*
Anistreplase	*enhance*, effect, *thrombolytic*, *agents*, *anticoagulant*	*enhance*, *anticoagulant*, *antiplatelet*, *thrombolytic*, *agents*
Ramipril	*diminish*, *antihypertensive*, effect, treatment, affect	*diminish*, *antihypertensive*, *inhibitor*, *doses*, affect

[a]TransR: a knowledge graph embedding model, which performs translation in the corresponding relation space.

[b]PRD: Predicting Rich Drug-Drug Interaction.

[c]Labels in italics indicate those annotated by a professional pharmacist.

Figure 4. Case visualization of the binary drug-drug interaction-type prediction on a drug-drug interaction network sample. MDDP: multitask dyadic drug-drug interaction prediction; PRD: Predicting Rich Drug-Drug Interaction; SCNN: Syntax Convolutional Neural Network.

(a) PRD (b) Tiresias (c) TransR

(d) MDDP (e) TransE (f) SCNN
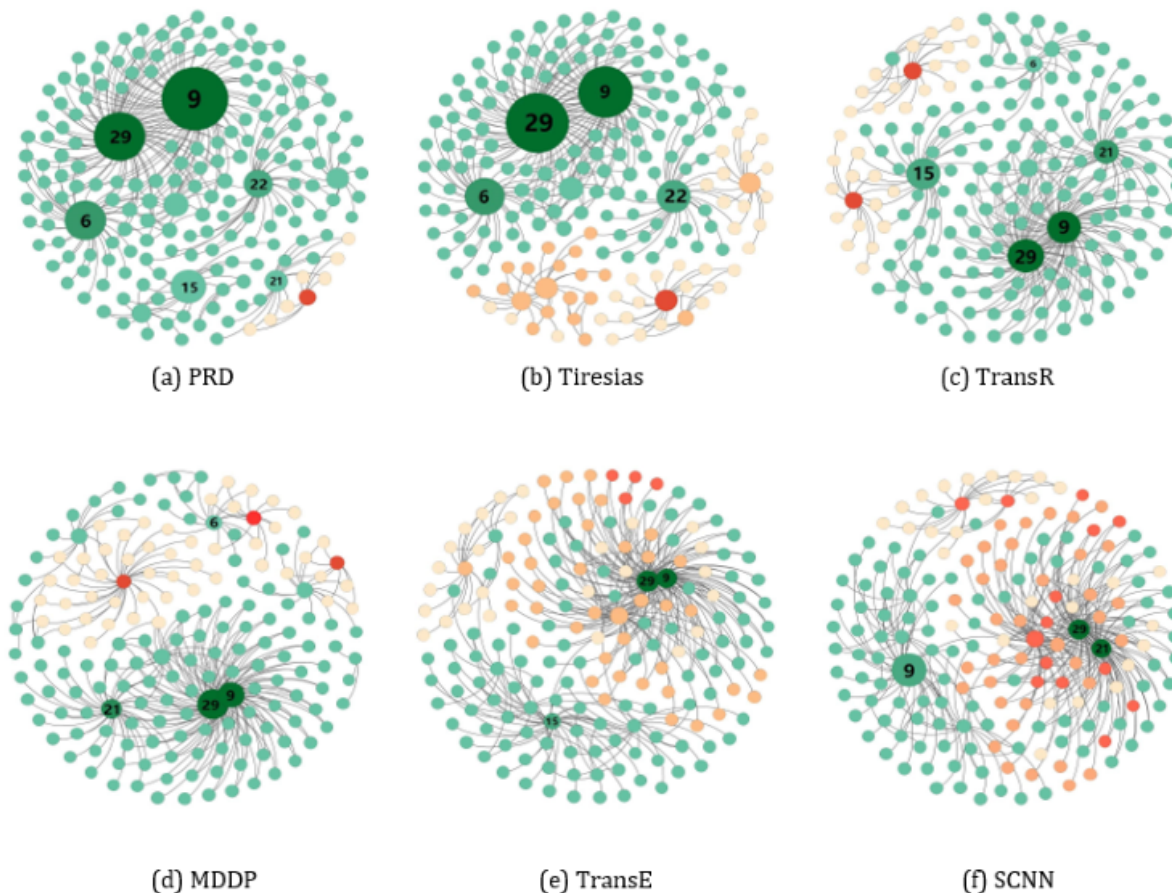
Discussion

Principal Findings

PRD achieved a significant improvement over all baselines. Specifically, PRD outperformed MDDP by around 10%. MDDP is currently considered to be the best DDI prediction baseline for multiple DDI type predictions. Tiresias and SCNN performed poorly because they neglect various types of semantic information concerning DDIs. These results demonstrate the effectiveness of PRD to predict rich DDI relations among drug entities.

Compared to TransR and TransE, PRD also performed better, as it incorporates binary DDI types into the relation representation learning and also models multiple DDI labels of a DDI relation simultaneously. This accounts for its promising results in rich DDI prediction.

Conclusions

PRD is unlike other existing models. Using rich DDI information, it can competently predict multiple labels for a pair of drugs across numerous domains, ranging from pharmacological mechanisms to side effects. To the best of our knowledge, this framework is the first to provide a joint translation-based embedding model that learns DDIs by integrating drug KGs and biomedical text simultaneously in a common low-dimensional space. The model also predicts DDIs using multilabels, rather than single or binary labels. Extensive experiments were conducted on real-world data sets to demonstrate the effectiveness and efficiency of the model. The results show PRD outperforms several state-of-the-art baselines. In future work, we intend to incorporate a convolutional neural network to encode the rich DDI text to improve the performance of the embedding model. Another direction for our research is to have the embedding model consider subgraph features composed in the generated drug KG during learning. This may make it possible to predict DDIs among 3 or more drugs.

Conflicts of Interest

None declared.

References

1. Leape LL. Systems analysis of adverse drug events. JAMA 1995 Jul 05;274(1):35. [doi: 10.1001/jama.1995.03530010049034]
2. Juurlink DN, Mamdani M, Kopp A, Laupacis A, Redelmeier DA. Drug-drug interactions among elderly patients hospitalized for drug toxicity. JAMA 2003 Apr 02;289(13):1652-1658. [doi: 10.1001/jama.289.13.1652] [Medline: 12672733]
3. Wang S, Li X, Chang* X, Yao L, Sheng QZ, Long G. Learning multiple diagnosis codes for ICU patients with local disease correlation mining. ACM Trans. Knowl. Discov. Data 2017 Apr 14;11(3):1-21. [doi: 10.1145/3003729]
4. Huang S, Strong JM, Zhang L, Reynolds KS, Nallani S, Temple R, et al. New era in drug interaction evaluation: US Food and Drug Administration update on CYP enzymes, transporters, and the guidance process. J Clin Pharmacol 2008 Jun;48(6):662-670. [doi: 10.1177/0091270007312153] [Medline: 18378963]
5. Quinney S, Zhang X, Lucksiri A, Gorski JC, Li L, Hall SD. Physiologically based pharmacokinetic model of mechanism-based inhibition of CYP3A by clarithromycin. Drug Metab Dispos 2010 Feb;38(2):241-248 [FREE Full text] [doi: 10.1124/dmd.109.028746] [Medline: 19884323]
6. Schelleman H, Bilker W, Brensinger C, Han X, Kimmel S, Hennessy S. Warfarin with fluoroquinolones, sulfonamides, or azole antifungals: interactions and the risk of hospitalization for gastrointestinal bleeding. Clin Pharmacol Ther 2008 Nov;84(5):581-588 [FREE Full text] [doi: 10.1038/clpt.2008.150] [Medline: 18685566]
7. Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripcsak G, Friedman C, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions. Nat Protoc 2014 Sep;9(9):2147-2163 [FREE Full text] [doi: 10.1038/nprot.2014.151] [Medline: 25122524]
8. Abdelaziz I, Fokoue A, Hassanzadeh O, Zhang P, Sadoghi M. Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions. Journal of Web Semantics 2017 May;44:104-117. [doi: 10.1016/j.websem.2017.06.002]
9. Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. 2013 Presented at: The 26th Annual Conference on Neural Information Processing Systems; 2013 Dec 3-10; Lake Tahoe, USA.
10. Lin X, Quan Z, Wang ZJ, Ma T, Zeng X. KGNN: Knowledge graph neural network for drug-drug interaction prediction. 2021 Presented at: The 29th International Joint Conference on Artificial Intelligence; 2012 Jan 7-15; Yokohama, Japan. [doi: 10.24963/ijcai.2020/380]
11. Karim MR, Cochez M, Jares JB, Uddin M, Beyan O, Decker S. Drug-Drug Interaction Prediction Based on Knowledge Graph Embeddings and Convolutional-LSTM Network. 2019 Presented at: The 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; 2019 Sept 7-10; Niagara Falls, USA.

XSL•FO

RenderX

12. Huang D, Jiang Z, Zou L, Li L. Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. Information Sciences 2017 Nov;415-416:100-109. [doi: 10.1016/j.ins.2017.06.021]

13. Bui Q, Sloot PMA, van Mulligen EM, Kors JA. A novel feature-based approach to extract drug-drug interactions from biomedical text. Bioinformatics 2014 Dec 01;30(23):3365-3371. [doi: 10.1093/bioinformatics/btu557] [Medline: 25143286]

14. Liu S, Tang B, Chen Q, Wang X. Drug-drug interaction extraction via convolutional neural networks. Comput Math Methods Med 2016;2016:6918381-6918388 [FREE Full text] [doi: 10.1155/2016/6918381] [Medline: 26941831]

15. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. Sci Rep 2017 Jul 20;7(1):5994 [FREE Full text] [doi: 10.1038/s41598-017-05778-z] [Medline: 28729710]

16. Ernst P, Siu A, Weikum G. KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences. BMC Bioinformatics 2015 May 14;16:157 [FREE Full text] [doi: 10.1186/s12859-015-0549-5] [Medline: 25971816]

17. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014 Jan;42(Database issue):D1091-D1097 [FREE Full text] [doi: 10.1093/nar/gkt1068] [Medline: 24203711]

18. Belleau F, Nolin M, Tourigny N, Rigault P, Morissette J. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. J Biomed Inform 2008 Oct;41(5):706-716 [FREE Full text] [doi: 10.1016/j.jbi.2008.03.004] [Medline: 18472304]

19. Samwald M, Jentzsch A, Bouton C, Kallesøe CS, Willighagen E, Hajagos J, et al. Linked open drug data for pharmaceutical research and development. J Cheminform 2011 May 16;3(1):19 [FREE Full text] [doi: 10.1186/1758-2946-3-19] [Medline: 21575203]

20. Harris S, Seaborne A. SPARQL 1.1 Query Language. W3C Working Draft. 2013 Mar 21. URL: https://www.w3.org/TR/2009/WD-sparql11-query-20091022/ [accessed 2021-05-01]

21. Chang X, Yang Y. Semisupervised feature analysis by mining correlations among multiple tasks. IEEE Trans. Neural Netw. Learning Syst 2017 Oct;28(10):2294-2305. [doi: 10.1109/tnnls.2016.2582746]

22. Lin Y, Jones P, Samatova NF. Learning entity and relation embeddings for knowledge graph completion. 2017 Presented at: ACM on Conference on Information and Knowledge Management; 2017 Nov 6-10; Singapore. [doi: 10.1145/3132847.3133095]

23. Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. 2016 Presented at: Proceedings of the 30th AAAI Conference on Artificial Intelligence; 2016 Feb 12-17; Phoenix, USA.

24. Tu C, Zhang Z, Liu Z, Sun M. TransNet: translation-based network representation learning for social relation extraction. 2017 Presented at: International Joint Conference on Artificial Intelligence; 2017 August 19-25; Melbourne, Australia.

25. Wang S, Chang X, Li X, Long G, Yao L, Sheng QZ. Diagnosis code assignment using sparsity-based disease correlation embedding. IEEE Trans. Knowl. Data Eng 2016 Dec 1;28(12):3191-3202. [doi: 10.1109/tkde.2016.2605687]

26. Zhu X, Li X, Zhang S, Ju C, Wu X. Robust joint graph sparse coding for unsupervised spectral feature selection. IEEE Trans. Neural Netw. Learning Syst 2017 Jun;28(6):1263-1275. [doi: 10.1109/tnnls.2016.2521602]

27. Kamdar M, Musen MA. PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data. 2017 Presented at: The 26th International Conference on World Wide Web; 2017 Apr 3-7; Perth, Australia. [doi: 10.1145/3038912.3052692]

28. Herrero-Zazo M, Segura-Bedmar I, Martínez P, Declerck T. The DDI corpus: an annotated corpus with pharmacological substances and drug-drug interactions. J Biomed Inform 2013 Oct;46(5):914-920 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.011] [Medline: 23906817]

29. Wang M, Zhang J, Liu J, Hu W, Wang S, Li X, et al. PDD graph: bridging electronic medical records and biomedical knowledge graphs via entity linking. 2017 Presented at: ISWC 2017: 16th International Semantic Web Conference; 2017 Oct 21–25; Vienna, Austria.

30. Srivastava N, Hinton J, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning 2014;15(1):1929-1958.

31. Kingma D, Ba JL. Adam: a method for stochastic optimization. 2015 Presented at: The 3rd International Conference for Learning Representation; 2015 May-79; San Diego, USA.

32. Mikolov T, Sutskever I, Chen K, Corrado K, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: The 26th Annual Conference on Neural Information Processing Systems; 2013 Dec 5-10; Lake Tahoe, USA. [doi: 10.4324/9780203776506-14]

33. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000 Jan 01;28(1):27-30 [FREE Full text] [doi: 10.1093/nar/28.1.27] [Medline: 10592173]

34. Hewett M, Oliver DE, Rubin DL, Easton KL, Stuart JM, Altman RB, et al. PharmGKB: the Pharmacogenetics Knowledge Base. Nucleic Acids Res 2002 Jan 01;30(1):163-165 [FREE Full text] [doi: 10.1093/nar/30.1.163] [Medline: 11752281]

35. Davis A, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, et al. The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res 2013 Jan;41(Database issue):D1104-D1114 [FREE Full text] [doi: 10.1093/nar/gks994] [Medline: 23093600]

36.   Zhao Z, Yang Z, Luo L, Lin H, Wang J. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. Bioinformatics 2016 Nov 15;32(22):3444-3453 [FREE Full text] [doi: 10.1093/bioinformatics/btw486] [Medline: 27466626]
37.   Jin B, Yang H, Xiao C, Zhang P, Wei X, Wang F. Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. 2017 Presented at: The 31st AAAI Conference on Artificial Intelligence; 2017 Feb 4-9; San Francisco, USA.

Abbreviations

ADR: adverse drug reaction
DDI: drug-drug interaction
KG: knowledge graph
MDDP: multitask dyadic drug-drug interaction prediction
PRD: Predicting Rich Drug-drug Interaction
SCNN: Syntax Convolutional Neural Network
TF-IDF: term frequency-inverse document frequency