

Predicting Rich Drug-Drug Interactions via Biomedical Knowledge Graphs and Text Jointly Embedding*

Meng Wang[†]

Abstract

Minimizing adverse reactions caused by drug-drug interactions has always been a momentous 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 discover various drug-drug interactions. However, these discoveries contain a huge amount of noise and provide knowledge bases far from complete and trustworthy ones to be utilized. Most existing studies focus on predicting binary drug-drug interactions between drug pairs and ignore other interactions. In this paper, we propose a novel framework, called PRD, to predict drug-drug interactions. The framework uses the graph embedding that can overcome data incompleteness and sparsity issues to achieve multiple DDI label prediction. First, a large-scale drug knowledge graph is generated from different sources. Then, the knowledge graph is 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. To validate the effectiveness of the proposed framework, extensive experiments were conducted on real-world datasets. The results demonstrate that our model outperforms several state-of-the-art baseline methods in terms of capability and accuracy.

1 Introduction

An increasing amount of research in clinical studies focuses 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, and they account for 3-5% of all in-hospital medication errors [20]. To make matters worse, patients with two or more diseases, e.g., elderly patients with chronic diseases, have a higher risk of an ADR if they take five or more different drugs simultaneously [14, 34]. 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 drug is released.

Many existing methods, such as [31, 1], leverage the existing properties of the drugs and use similarity measurements to predict DDIs. Independent similarity scores for a candidate pair of drugs are computed and used as the input for a machine learning classifier (e.g., a logistic regression model), which outputs a binary prediction. However, DDI predictions should not be as narrow as a simple true or false question. The interactions between drugs in real-world data resources exhibit rich and variant meanings (e.g., increasing hematologic toxicity and high-risk infections). Capturing detailed DDI information may help caregivers make more accurate and safe medical prescriptions. Recently, Jin et al. [13] proposed a multitask learning framework to detect multiple types of DDIs among developed drugs. Unfortunately, this framework relies on adverse DDIs (e.g., heart rate increase) that have been pre-defined in a structured database and largely ignores pharmacological DDI information in unstructured biomedical text. As a result, this method is not able to generate in-depth insights into the underlying mechanisms of DDIs. More critically, this multitask learning framework suffers from sparsity issues and has limitations on large-scale data sets because it employs a one-hot feature vector and treats each drug as a unique symbol.

Motivation. Many biomedical knowledge bases are published as knowledge graphs (KGs), as illustrated in Figure 1. In KGs, nodes are entities, which might include drugs, diseases, protein targets, substructures, side effects, pathways, etc. The edges (or links) represent the various relations between the nodes, such as drug-target interactions. Data items in different knowledge bases can therefore be linked to form large heterogeneous graphs, such as the Bio2RDF [2] and the Linked Open Drug Data [27]. Recent works [15, 16] leverage these graphs and Semantic Web techniques to analysis drug reactions. However, these heterogeneous graphs face data incompleteness problem. Furthermore, as shown in Figure 1, DDIs are frequently described in unstructured text, such as pharmacology literature. Obviously, the increasing emergence of biomedical text offers an opportunity to uncover more than multiple adverse reactions, such as specific synergies or antagonistic effects in pharmacology.

*Supported by the National Key Research and Development Program of China, grant number 2016YFB1000903; National Science Foundation of China, grant Nos.61672420, 61672419, 61532015, 61532004; Innovative Research Group of the National Natural Science Foundation of China, No.61721002; Ministry of Education Innovation Research Team No.IRT17R86; and the Project of China Knowledge Centre for Engineering Science and Technology.

[†]MOEKLINNS Lab, Xi'an Jiaotong University, China

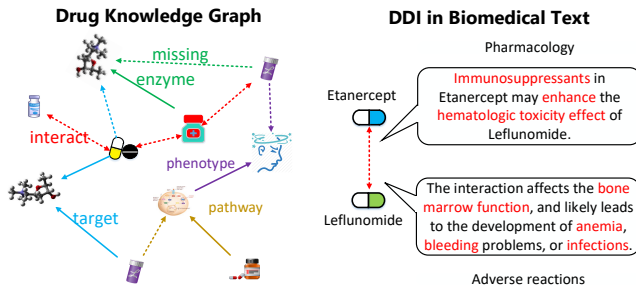


Figure 1: A drug knowledge graph is shown on the left with missing relations represented as dotted lines. There is usually no direct DDI relation between drugs. DDI descriptions from biomedical text are shown on the right. The words colored red represent concerns regarding DDI information in terms of both adverse DDIs and in-depth ways drugs can interact in pharmacology.

Challenges. Leveraging both drug KGs and biomedical text is a promising pathway for rich and comprehensive DDI prediction, but it is not without issues. Our model confronts the following challenges: 1. *Data Noise and Incompleteness*. Real-world KGs are known to be inaccurate, incomplete, and unreliable for direct use. 2. *Data Sparsity*. The potential DDI information in both KGs and biomedical text is sparse. It is a tough task to estimate potential DDIs in such long-tail distribution. 3. *Computational Efficiency*. Undoubtedly, computational complexity 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.

Solution. Given these challenges, we propose a novel framework called PRD. The framework is based on graph embedding techniques and treats specific DDI predictions as a linked prediction process. The procedure includes the following:

1. A large, high-quality drug KG is generated from distributed drug resources, including drug-target interactions, the impact of drugs on gene expression, the outcomes of drugs in clinical trials, and so on.
2. 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.
3. The decoder predicts the corresponding labels for potential DDIs based on the learned embeddings.

Contributions. In summary, our method PRD makes the following contributions.

1. To the best of our knowledge, this is the first method that is able to predict comprehensive and specific DDIs

based on a large-scale drug KGs and comprehensive biomedical text from pharmacology to ADRs.

2. The method includes a joint translation-based embedding model that encodes the KG and rich DDI information from biomedical text into a shared low-dimensional space. Then, the DDI predictions are translated into a linked prediction process from the learned embeddings.
3. Extensive experiments on real-world datasets 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.

The remainder of the paper is organized as follows. Related work is discussed in Section 2. Section 3 presents the details of the proposed framework. The evaluation of the framework is reported in Section 4. Finally, our conclusions and future work are presented in Section 5.

2 Related Work

This section discusses existing related research, including DDI detection, biomedical KGs, and graph embedding.

DDI detection is a major issue in pharmacological research. Traditional experimental approaches *in vitro* [12], *in vivo* [25], and *in populo* [28] focus on small sets of specific drug pairs and have laboratory limitations. As mentioned in Section 1, many computational approaches have been proposed to predict DDIs in recent years. However, similarity or feature-based approaches [31, 1, 13] only predict binary DDIs or those that have been pre-defined in structured databases and suffer from robustness caused by data sparsity and vast computation requirements. Although several approaches [11, 4, 22] have used natural language processing techniques to extract DDIs from biomedical text, to the best of our knowledge, they have not employed drug knowledge graphs 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 [26], Knowlife [7] and Drug-Bank [19] are just a few examples. Efforts such as Bio2RDF [2] and Linked Open Drug Data [27] have mapped similar entities in different KGs and built large heterogeneous graphs that contain an abundance of basic biomedical facts about drugs. SPARQL [8], 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 [27]. As noted, KG applications based on conventional graph-based algorithms are compromised by data sparsity and computational inefficiency. To address these problems, graph embedding techniques [3, 21, 33, 24, 36, 30, 5] based on representation learning for KGs have been proposed to embed both entities and relations into a continuous low-dimensional vector space. Among these methods, translation-based models [3, 21, 30] are the most simple and effective. Currently, they represent state-of-the-art performance for knowledge acquisition and inference and link prediction [3]. Inspired by these analogies, DDI can be treated as a category of relations in a drug KG, and KG embedding techniques could be used to predict unknown DDIs. However, most translation-based methods only concentrate on pre-defined relations or unstructured text, ignoring to bridge the existing relations and rich unstructured text together.

3 The Proposed Framework

Figure 2 shows the architecture of the proposed framework. It consists of three key phases: drug KG generation (described in 3.1), jointly embedding learning (described in 3.2) and DDI prediction (described in 3.3).

3.1 Drug KG Generation A typical KG usually arranges knowledge as a triple set of facts that indicate the relation between two entities, i.e., 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 [15], in which the types of entities/relations are summarized as shown in Table 1. Specifically, we use SPARQL federation queries [8] to extract triples that contain four types of drug-related entities ($E_1 \sim E_4$) and five types of biological relations ($R_1 \sim R_5$) from a variety of biomedical sources (Bio2RDF [2]). These extracted triples are defined as basic triples in our drug KG:

Definition 1. (BASIC TRIPLE) $B = (E, R)$ is a set of basic triples in the form (h, r, t) , where $E = E_1 \cup E_2 \cdots \cup E_4$ is a set of entities, $R = R_1 \cup R_2 \cdots \cup R_5$ 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*.

A specific DDI between two drugs can be captured by multiple key phrases extracted from biomedical text, as shown in Figure 1. Hence, we collect biomedical DDI text documenting drug pairs (e.g., DDI corpus [9], MedLine

Table 1: Entities and relations of basic triples in drug KG.

Variable	Entity/Relation Interpretation
$E_1 \sim E_4$	E_1 : drugs, E_2 : proteins, E_3 : pathways, E_4 : phenotypes.
$R_1 \sim R_5$	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).

abstracts¹, and DrugBank DDI documents). We remove all stop words from raw text and use an entity linking method [32] to align the drug names in the biomedical text with the KG. Then, the top- n labels² are selected from the biomedical text for each DDI based on *TF-IDF* features (some other textual features can be used to rank the labels instead). Based on this, the DDI relations between drug entities are defined as a set of labels, rather than as a single label.

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 $l = \{n_1, n_2, \dots\} \subseteq L$ is the set of labels to describe the DDI information.

For instance, (*Etanercept*, {immunosuppressants, enhancetoxicity, anemia, infections}, *Leflunomide*) is a rich DDI triple. Note that the DDI relations between two drugs are bidirectional; hence, our method replaces each rich DDI relation with two directed triples of opposite directions in the drug KG.

Formally, the generated drug knowledge graph is defined as follows.

Definition 3. (DRUG KNOWLEDGE GRAPH) The drug knowledge graph G is denoted as (E, B, T) , where $E = E_1 \cup E_2 \cdots \cup E_4$ is a set of entities, B is a set of basic triples and T is a set of rich DDI triples.

3.2 Model Description This section introduces the translation-based embedding model that learns representations from the drug knowledge graph $G = (E, B, T)$ and the optimization method as follows.

3.2.1 Basic Triple Encoder (BTE) 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 ***h***, ***r***, ***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

¹<https://www.medline.com/>

²We set $n=5$ in this paper.

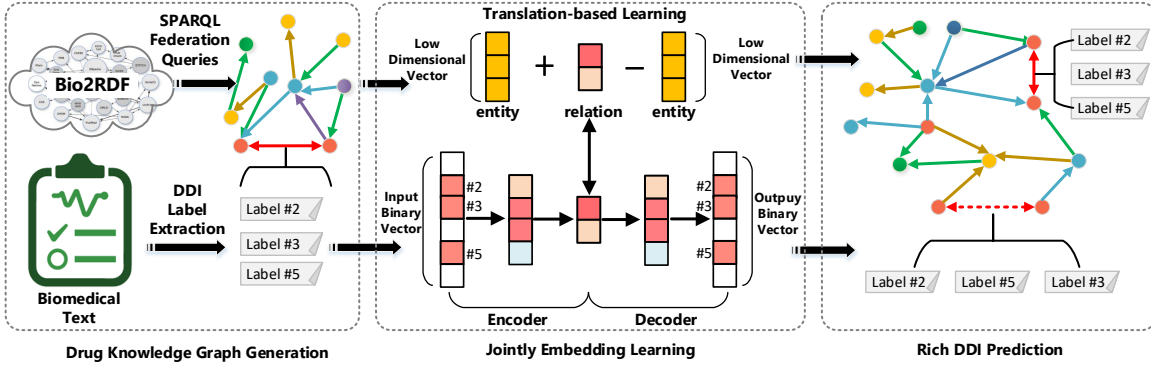


Figure 2: Overview of the framework.

translation operation \mathbf{r} between two entity vectors \mathbf{h} and \mathbf{t} in the low-dimensional space. We follow the TransR model in [21] 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 $\mathbf{h}, \mathbf{t} \in \mathbb{R}^k$, and r is embedded into $\mathbf{r} \in \mathbb{R}^d$. For each relation r , a projection matrix $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ projects entities from the entity space to the relation space. Then, the energy function $z_{bte}(h, r, t)$ is defined as:

$$(3.1) \quad z_{bte}(\mathbf{h}, \mathbf{r}, \mathbf{t}) = b_1 - \|\mathbf{h}\mathbf{M}_r + \mathbf{r} - \mathbf{t}\mathbf{M}_r\|_{L1/L2}$$

where b_1 is a bias constant.

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

$$(3.2) \quad P(h|r, t) = \frac{\exp\{z_{bte}(\mathbf{h}, \mathbf{r}, \mathbf{t})\}}{\sum_{\hat{h} \in E} \exp\{z_{bte}(\hat{\mathbf{h}}, \mathbf{r}, \mathbf{t})\}}$$

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

$$(3.3) \quad \mathcal{L}(h, r, t) = \log P(h|r, t) + \log P(t|h, r) + \log P(r|h, t)$$

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

$$(3.4) \quad \mathcal{L}_{bte} = \sum_{(h, r, t) \in B} \mathcal{L}(h, r, t)$$

It is worth mentioning that other graph embedding models, such as HOLE [24], can also be easily adopted for basic triple encoding. Due to the space limit, in this paper we only explore the effectiveness of TransR.

3.2.2 Rich DDI Triple Encoder (RDTE) The interaction l between two drug entities u, v in rich DDI triples $(u, l, v) \in T$ can also be represented as translations in low-dimensional

space. We set $\mathbf{u}, \mathbf{v} \in \mathbb{R}^k$, $\mathbf{l} \in \mathbb{R}^d$. The energy function $z_{dte}(u, l, v)$ is defined as:

$$(3.5) \quad z_{dte}(\mathbf{u}, \mathbf{l}, \mathbf{v}) = b_2 - \|\mathbf{u}\mathbf{M}_l + \mathbf{l} - \mathbf{v}\mathbf{M}_l\|_{L1/L2}$$

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

$$(3.6) \quad \mathcal{L}_{dte} = \sum_{(u, l, v) \in T} \mathcal{L}(u, l, v)$$

Note, in Eq.(3.5), \mathbf{l} is the relation representation obtained from $l = \{n_1, n_2, \dots\}$. This is demonstrated in-depth next.

A deep autoencoder is employed to construct the relation representation $\mathbf{l} \in \mathbb{R}^d$ 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 $\mathbf{s} = \{\mathbf{s}_i\}_{i=1}^{|L|}$, where $\mathbf{s}_i = 1$ if $n_i \in l$, and $\mathbf{s}_i = 0$ otherwise. The deep autoencoder then takes the binary vector \mathbf{s} as input and uses the following non-linear transformation layers to transform the label set into the low-dimensional space \mathbb{R}^k :

$$(3.7) \quad \begin{aligned} \mathbf{h}^{(1)} &= f(\mathbf{W}^{(1)}\mathbf{s} + \mathbf{b}^{(1)}), \\ \mathbf{h}^{(i)} &= f(\mathbf{W}^{(i)}\mathbf{h}^{(i-1)} + \mathbf{b}^{(i)}), i = 2, \dots, K. \end{aligned}$$

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

There are two parts to the autoencoder: an encoder and a decoder. The encoder employs the \tanh activation function to obtain the DDI relation representation $\mathbf{l} = \mathbf{h}^{(K/2)}$. 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 non-zero elements due to data sparsity. This leads the decoder tends to reconstruct zero elements rather than non-zero elements, which conflicts with our purpose. To overcome this obstacle, different weights are set for different elements, and the following objective function is maximized:

$$(3.8) \quad z_l(\mathbf{s}, \tilde{\mathbf{s}}) = b_3 - \|(\mathbf{s} - \tilde{\mathbf{s}}) \odot \mathbf{x}\|_{L1/L2}$$

where b_3 is a bias constant, \mathbf{x} is a weight vector and \odot is denoted as the Hadamard product. For $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^{|L|}$, $\mathbf{x}_i = 1$ if $\mathbf{s}_i = 0$, and $\mathbf{x}_i = \beta > 1$ otherwise. According to Eq.(3.8), the probability of $P(\mathbf{s}|\tilde{\mathbf{s}})$ can be defined as follows:

$$(3.9) \quad P(\mathbf{s}|\tilde{\mathbf{s}}) = \frac{\exp\{z_l(\mathbf{s}, \tilde{\mathbf{s}})\}}{\sum_{\tilde{\mathbf{s}} \in S} \exp\{z_l(\tilde{\mathbf{s}}, \tilde{\mathbf{s}})\}}$$

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

$$(3.10) \quad \mathcal{L}(l) = \log P(\mathbf{s}|\tilde{\mathbf{s}})$$

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

$$(3.11) \quad \mathcal{L}_{rcl} = \sum_{l \in T} \mathcal{L}(l)$$

3.2.3 Joint Learning and Optimization The goal of the framework PRD is to not only represent the basic triples (drug KG B) but also rich DDI triples (biomedical text T) in a unified joint embedding model. Considering the above three objective functions Eq.(3.4) (3.6) (3.11) together, the optimization function is defined as:

$$(3.12) \quad O(X) = \mathcal{L}_{bte} + \mathcal{L}_{dte} + \mathcal{L}_{rcl} + \gamma C(X)$$

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:

$$(3.13) \quad C(X) = \sum_{e \in E} [\|\mathbf{e}\| - 1]_+ + \sum_{r \in R} [\|\mathbf{r}\| - 1]_+ + \sum_{l \in L} [\|\mathbf{l}\| - 1]_+$$

where $[x]_+ = \max(0, x)$ denotes the positive part of x . The regularization factor will normalize the embeddings during learning. We adopt the approach in [29] to prevent deep neural networks from overfitting, and use the Adam algorithm [18] to maximize the objective function.

It is impractical to directly compute the normalizers in $P(h|r, t)$, $P(t|h, r)$, $P(r|h, t)$ 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 [23] to transform the objective functions. Taking $P(h|r, t)$ as an example, the following objective function is maximized instead of using its original form:

$$(3.14) \quad \log \sigma(z_{bte}(\mathbf{h}, \mathbf{r}, \mathbf{t})) + \sum_{m=1}^c E_{\tilde{h}_m \sim z_{neg}(\{\tilde{h}, r, t\})} [\sigma(z_{bte}(\tilde{\mathbf{h}}_m, \mathbf{r}, \mathbf{t}))]$$

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 Eq.(3.14), 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 $P(r|h, t)$, $P(t|h, r)$, $\log P(\mathbf{s}|\tilde{\mathbf{s}})$ and $\mathcal{L}(u, l, v)$ 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.

3.3 Predicting DDI Relations With the learned deep autoencoder and the embedding vectors of all entities and relations, the framework PRD can now leverage the translation mechanism to predict the DDI relations between two drug entities. To be more specific, given two drugs entities $u, v \in E_1$, the following equation predicts the potential relation embedding \mathbf{l} between \mathbf{u} and \mathbf{v} .

$$(3.15) \quad \mathbf{l} = \mathbf{v}\mathbf{M}_l - \mathbf{u}\mathbf{M}_l$$

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

4 Experiments and Evaluation

To scrutinize the effectiveness of the DDI prediction framework PRD, we performed two types of experiments. First, we compared the performance of our model to several baseline methods on binary-type DDI predictions. Then we investigated PRDs 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.

4.1 Dataset Construction Experiments in this paper were performed on two real drug-related datasets, Bio2RDF [2] and DDI Corpus [9].

- Bio2RDF (version 4) is an open-source project that provides 11 billion triples from 35 biological and pharma-

Table 2: Statistics of the generated drug KG

#Basic triples	71,460	#Drug entities	8,197
#Rich DDI triples	4,694	#Other entities	305,642
#Distinct labels in the DDI vocabulary	1,053		

cological KGs across a wide variety of drug-related entities, such as proteins, targets, and diseases. It is accessible on the Web via the SPARQL endpoint³.

- DDI Corpus (2013 version) is a semantically annotated corpus of documents describing DDIs from the DrugBank database and MedLine abstracts. It contains 233 MedLine abstracts and 784 DrugBank text on the DDIs subjects. There are a total of 5,021 annotated DDIs in 18,491 pharmacological sentences.

We used SPARQL federation queries to extract basic triples for our drug KG from four different KGs in Bio2RDF: 1) DrugBank [19] provides comprehensive data about drug, disease and target information. 2) Kyoto Encyclopedia of Genes and Genomes (KEGG) [17] offers biological data of pathways, proteins, and drugs. 3) PharmGKB [10] renders us protein-drug-disease-relations. 4) Comparative Toxicogenomics Database (CTD) [6] furnishes data about protein interactions and pathway-disease relations.

For the rich DDI triples, we collected the 4,694 DrugBank DDI sentences about 8,197 drugs from DDI corpus. Top-5 labels from each sentence were selected based on *TF-IDF* to construct rich DDI triples and build the DDI label vocabulary. Since we undergo the issue on inconsistent drug names between basic triples and rich DDI triples, we apply the entity linking method [32] to align the drug aliases.

The details of the drug KG we constructed are shown in Table 2.

4.2 Baselines The following three DDI prediction approaches and two state-of-the-art KG embedding methods were employed as baselines:

- Tiresias [1] 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.
- SCNN [35] represents a DDI extraction method based on a syntax convolutional neural network to extract four pre-defined DDI types (*ADVICE*, *EFFECT*, *INT* and *MECHANISM*) from the biomedical literature.

- Multitask dyadic DDI prediction (MDDP) [13] defines the DDI type prediction problem as a multitask dyadic regression problem. It can predict the specific DDI type between two drugs.
- TransE [3] is the most representative translational distance model to embed components of a KG, including entities and relations, into continuous vector spaces. Those embeddings can also be used for link prediction.
- TransR [21] shares a similar approach with TransE, but represents entities and relations in distinct vector spaces bridged by relation-specific matrices.

4.3 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. Thirty percent of DDI relations, chosen randomly as the ground truths for the test set, were removed and the remaining KG was used 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 two DDI types as *yes* and *no* in the training model. For TransE, TransR and our method PRD, 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 multi-label classification task. For Tiresias, SCNN, and MDDP, we used their feature representation methods and adopted one-vs-rest logistic regression to train a multi-label classifier. For TransE and TransR, we separated each training triple (u, l, v) , where $l = \{n_1, n_2, \dots\}$, into several triples, i.e., (u, n_i, v) for $n_i \in l$, which can be directly used to train the models.

We used 10-fold cross-validation on the training set to tune PRDs embedding model. We determined the optimal parameters using a grid search strategy. The search ranges for the various parameters follow: 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 constant 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 391s. The best configurations for the joint model were: $\lambda = 0.001$, $\gamma = 0.01$, $k = 100$, $b_1 = 5$, $b_2 = 5$, $b_3 = 1$, $c = 10$, $K = 3$ and taking L_1 as dissimilarity metric.

We used receiver operator characteristic (ROC) curves and precision-recall (PR) curves to evaluate the proposed method on binary DDI-type predictions. For multiple DDI-type predictions, we followed the setting in TransE [3] and report the two measures as evaluation metrics: the average

³<http://bio2rdf.org/>

Table 3: Evaluation results for multiple DDI relation predictions.($\times 100$ for Hits@ k)

Metric	Raw				Filter			
	Hits@1	Hits@5	Hits@10	MeanRank	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	12.19	26.31	39.02	37.91	16.82	27.03	40.78	37.06
MDDP	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	45.11	85.57	91.01	6.11	75.11	88.60	92.85	5.45

Table 4: Rich DDI predictions for *Acetylsalicylic acid*. (underlined labels are annotated by a professional pharmacist)

Interacted Drug	TransR	PRD
Ibuprofen	<u>enhance</u> , <u>adverse</u> , <u>toxic</u> , risk, <u>bleeding</u>	<u>enhance</u> , <u>toxic</u> , <u>bleeding</u> , platelet, antiplatelet
Alteplase	<u>enhance</u> , increase, <u>adverse</u> , toxic, effect	<u>enhance</u> , <u>toxic</u> , <u>bleeding</u> , <u>thrombolytic</u> , <u>adverse</u>
Anistreplase	<u>enhance</u> , effect, <u>thrombolytic</u> , agents, <u>anticoagulant</u>	<u>enhance</u> , <u>anticoagulant</u> , <u>antiplatelet</u> , <u>thrombolytic</u> , agents
Ramipril	<u>diminish</u> , <u>antihypertensive</u> , effect, treatment, affect	<u>diminish</u> , <u>antihypertensive</u> , <u>inhibitor</u> , <u>doses</u> , affect

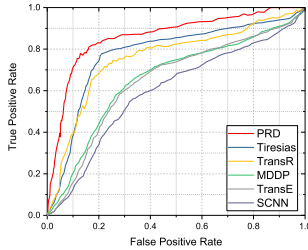


Figure 3: ROC curve

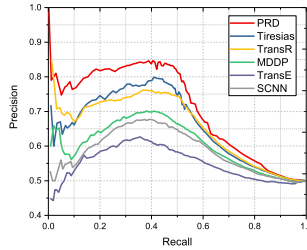


Figure 4: PR curve

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.”

4.4 Experiment Results As shown in Figures 3 and 4, the proposed framework PRD outperformed all baselines. In terms of the ROC curve, PRD outperformed Tiresias by 6.69%, 7.13% than TransR, around 12% than MDDP and TransE, while SCNN had the relatively low predictive ability. According to the PR curve, PRD learned 14.2% better than Tiresias, which was at the top among three DDI prediction baselines, 16.8% than TransR, 21.57% than MDDP, 25.33% than TransE, and 37.89% than SCNN.

Table 3 shows the evaluation results for rich DDI relation predictions according to the different evaluation metrics

for both the Raw and Filter tests. From these tables, we make the following observations:

- 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 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 representations learning and also models multiple DDI labels of a DDI relation simultaneously. This accounts for its promising results in rich DDIs prediction.

4.5 Case Study To further demonstrate PRDs ability on rich DDI predictions, we select the drug *Acetylsalicylic acid* (*Aspirin*) as a test case. The DDI predictions and rich labels relations are shown in Table 4. According to the usefulness and diversity of the predicted labels, a professional pharmacist have evaluated and annotated the practical useful predictions (underlined labels in Table 4). 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.

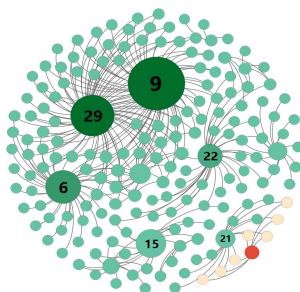


Figure 5: PRD

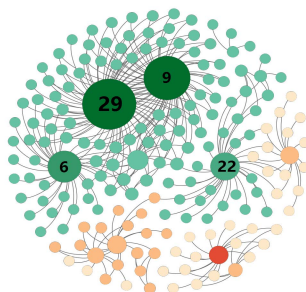


Figure 6: Tiresias

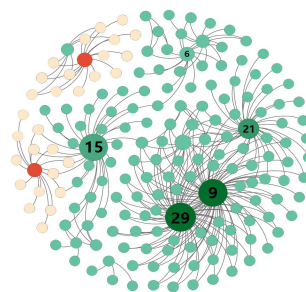


Figure 7: TransR

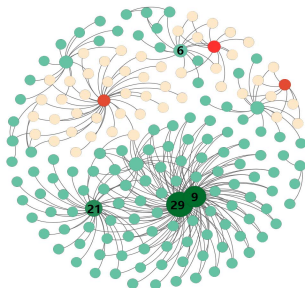


Figure 8: MDDP

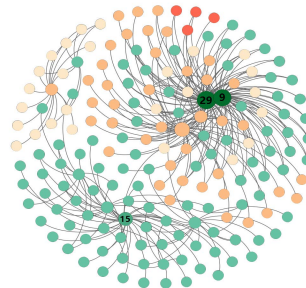


Figure 9: TransE

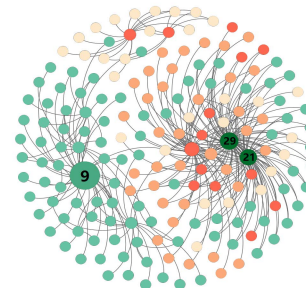


Figure 10: SCNN

Here we also present case study to visualize the effectiveness of binary DDI types prediction on a DDI network sample. We construct drug-drug networks to indicate whether any two drugs would result in a binary DDI. A node in the network denotes a drug. An edge between two nodes denotes the existence of a DDI. Intuitively, the more drugs interact, the more risk they will burn. 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, i.e., high-risk is shown as dark green and low-risk is shown as light green. The red nodes denote forecasting errors of DDI drugs. As shown in Figure 5 to Figure 10, PRD predicts DDIs mostly accurate. The ID of the drug with high risk is shown on the node.

5 Conclusions and Future Work

This paper presents a framework for DDI prediction that is unlike existing models. It can competently predict multi-labels for a pair of drugs from rich DDIs information in many ways, 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 multi-labels, rather than single or binary labels. Extensive experiments were conducted on real 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 per-

formance of embedding model. Another direction for our research is to have the embedding model consider subgraph features composed in the generated drug knowledge graph during learning. This may make it possible to predict DDIs among three or more drugs.

References

- [1] I. ABDELAZIZ, A. FOKOUE, O. HASSANZADEH, P. ZHANG, AND M. SADOOGHI, *Large-scale structural and textual similarity-based mining of knowledge graph to predict drug-drug interactions*, Web Semantics: Science, Services and Agents on the World Wide Web, (2017).
- [2] F. BELLEAU, M.-A. NOLIN, N. TOURIGNY, P. RIGAUT, AND J. MORISSETTE, *Bio2rdf: towards a mashup to build bioinformatics knowledge systems*, Journal of biomedical informatics, 41 (2008), pp. 706–716.
- [3] A. BORDES, N. USUNIER, A. GARCIA-DURAN, J. WESTON, AND O. YAKHNENKO, *Translating embeddings for modeling multi-relational data*, in NIPS, 2013, pp. 2787–2795.
- [4] Q.-C. BUI, P. M. SLOOT, E. M. VAN MULLIGEN, AND J. A. KORS, *A novel feature-based approach to extract drug-drug interactions from biomedical text*, Bioinformatics, 30 (2014), pp. 3365–3371.
- [5] X. CHANG AND Y. YANG, *Semisupervised feature analysis by mining correlations among multiple tasks*, IEEE Transactions on Neural Networks and Learning Systems, 28 (2017), pp. 2294–2305.
- [6] A. P. DAVIS, C. G. MURPHY, R. JOHNSON, J. M. LAY, K. LENNON-HOPKINS, C. SARACENI-RICHARDS, D. SCI- AKY, B. L. KING, M. C. ROSENSTEIN, T. C. WIEGERS, ET AL., *The comparative toxicogenomics database: update 2013*, Nucleic acids research, 41 (2012), pp. D1104–D1114.

- [7] P. ERNST, A. SIU, AND G. WEIKUM, *Knowlife: a versatile approach for constructing a large knowledge graph for biomedical sciences*, BMC bioinformatics, 16 (2015), p. 157.
- [8] S. HARRIS, A. SEABORNE, AND E. PRUDHOMMEAUX, *Sparql 1.1 query language*, W3C recommendation, 21 (2013).
- [9] M. HERRERO-ZAZO, I. SEGURA-BEDMAR, P. MARTÍNEZ, AND T. DECLERCK, *The ddi corpus: An annotated corpus with pharmacological substances and drug–drug interactions*, Journal of biomedical informatics, 46 (2013), pp. 914–920.
- [10] M. HEWETT, D. E. OLIVER, D. L. RUBIN, K. L. EASTON, J. M. STUART, R. B. ALTMAN, AND T. E. KLEIN, *Pharmgkb: the pharmacogenetics knowledge base*, Nucleic acids research, 30 (2002), pp. 163–165.
- [11] D. HUANG, Z. JIANG, L. ZOU, AND L. LI, *Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks*, Information sciences, 415 (2017), pp. 100–109.
- [12] S.-M. HUANG, J. M. STRONG, L. ZHANG, K. S. REYNOLDS, ET AL., *New era in drug interaction evaluation: Us food and drug administration update on cyp enzymes, transporters, and the guidance process*, The Journal of clinical pharmacology, 48 (2008), pp. 662–670.
- [13] B. JIN, H. YANG, C. XIAO, P. ZHANG, X. WEI, AND F. WANG, *Multitask dyadic prediction and its application in prediction of adverse drug–drug interaction.*, in AAAI, 2017, pp. 1367–1373.
- [14] D. N. JUURLINK, M. MAMDANI, A. KOPP, A. LAUPACIS, AND D. A. REDELMEIER, *Drug–drug interactions among elderly patients hospitalized for drug toxicity*, Jama, 289 (2003), pp. 1652–1658.
- [15] M. R. KAMDAR AND M. A. MUSEN, *Phlegra: Graph analytics in pharmacology over the web of life sciences linked open data*, in WWW, ACM, 2017, pp. 321–329.
- [16] M. R. KAMDAR, M. TECH, AND M. A. MUSEN, *Mechanism-based pharmacovigilance over the life sciences linked open data cloud*, in AMIA, AMIA, 2017.
- [17] M. KANEHISA AND S. GOTO, *Kegg: kyoto encyclopedia of genes and genomes*, Nucleic acids research, 28 (2000), pp. 27–30.
- [18] D. KINGMA AND J. BA, *Adam: A method for stochastic optimization*, in ICLR, 2015, p. arXiv:1412.6980.
- [19] V. LAW, C. KNOX, Y. DJOUMBOU, T. JEWISON, A. C. GUO, Y. LIU, A. MACIEJEWSKI, D. ARNDT, M. WILSON, V. NEVEU, ET AL., *Drugbank 4.0: shedding new light on drug metabolism*, Nucleic acids research, 42 (2014), pp. D1091–D1097.
- [20] L. L. LEAPE, D. W. BATES, D. J. CULLEN, J. COOPER, H. J. DEMONACO, T. GALLIVAN, R. HALLISEY, J. IVES, N. LAIRD, G. LAFFEL, ET AL., *Systems analysis of adverse drug events*, Jama, 274 (1995), pp. 35–43.
- [21] Y. LIN, Z. LIU, M. SUN, Y. LIU, AND X. ZHU, *Learning entity and relation embeddings for knowledge graph completion.*, in AAAI, 2015, pp. 2181–2187.
- [22] S. LIU, B. TANG, Q. CHEN, AND X. WANG, *Drug–drug interaction extraction via convolutional neural networks*, Computational and mathematical methods in medicine, 2016 (2016).
- [23] T. MIKOLOV, I. SUTSKEVER, K. CHEN, G. S. CORRADO, AND J. DEAN, *Distributed representations of words and phrases and their compositionality*, in NIPS, 2013, pp. 3111–3119.
- [24] M. NICKEL, L. ROSASCO, T. A. POGGIO, ET AL., *Holographic embeddings of knowledge graphs.*, in AAAI, 2016, pp. 1955–1961.
- [25] S. K. QUINNEY, X. ZHANG, A. LUCKSIRI, J. C. GORSKI, L. LI, AND S. D. HALL, *Physiologically-based pharmacokinetic model of mechanism-based inhibition of cyp3a by clarithromycin*, Drug metabolism and disposition, (2009), pp. dmd–109.
- [26] M. ROTMENSCH, Y. HALPERN, A. TLIMAT, S. HORNG, AND D. SONTAG, *Learning a health knowledge graph from electronic medical records*, Scientific reports, 7 (2017).
- [27] M. SAMWALD, A. JENTZSCH, C. BOUTON, C. S. KALLESØE, E. WILLIGHAGEN, J. HAJAGOS, M. S. MARSHALL, E. PRUD’HOMMEAUX, O. HASSANZADEH, E. PICHLER, ET AL., *Linked open drug data for pharmaceutical research and development*, Journal of cheminformatics, 3 (2011), p. 19.
- [28] H. SCHELLEMAN, W. B. BILKER, C. M. BRENSINGER, X. HAN, S. E. KIMMEL, AND S. HENNESSY, *Warfarin with fluoroquinolones, sulfonamides, or azole antifungals: interactions and the risk of hospitalization for gastrointestinal bleeding*, Clinical pharmacology & therapeutics, 84 (2008), pp. 581–588.
- [29] N. SRIVASTAVA, G. E. HINTON, A. KRIZHEVSKY, I. SUTSKEVER, AND R. SALAKHUTDINOV, *Dropout: a simple way to prevent neural networks from overfitting.*, Journal of machine learning research, 15 (2014), pp. 1929–1958.
- [30] C. TU, Z. ZHANG, Z. LIU, AND M. SUN, *Transnet: translation-based network representation learning for social relation extraction*, in IJCAI, 2017, p. 28642870.
- [31] S. VILAR, E. URIARTE, L. SANTANA, T. LORBERBAUM, G. HRIPCSAK, C. FRIEDMAN, AND N. P. TATONETTI, *Similarity-based modeling in large-scale prediction of drug–drug interactions*, Nature protocols, 9 (2014), pp. 2147–2163.
- [32] M. WANG, J. ZHANG, J. LIU, W. HU, S. WANG, X. LI, AND W. LIU, *Pdd graph: Bridging electronic medical records and biomedical knowledge graphs via entity linking*, in ISWC, Springer, 2017, pp. 219–227.
- [33] S. WANG, X. CHANG, X. LI, G. LONG, L. YAO, AND Q. Z. SHENG, *Diagnosis code assignment using sparsity-based disease correlation embedding*, IEEE Transactions on Knowledge and Data Engineering, 28 (2016), pp. 3191–3202.
- [34] S. WANG, X. LI, L. YAO, Q. Z. SHENG, G. LONG, ET AL., *Learning multiple diagnosis codes for icu patients with local disease correlation mining*, ACM Transactions on Knowledge Discovery from Data, 11 (2017), pp. 31:1–31:21.
- [35] Z. ZHAO, Z. YANG, L. LUO, H. LIN, AND J. WANG, *Drug drug interaction extraction from biomedical literature using syntax convolutional neural network*, Bioinformatics, 32 (2016), pp. 3444–3453.
- [36] X. ZHU, X. LI, S. ZHANG, C. JU, AND X. WU, *Robust joint graph sparse coding for unsupervised spectral feature selection*, IEEE transactions on neural networks and learning systems, 28 (2017), pp. 1263–1275.