Multi-relational Contrastive Learning Graph Neural Network for Drug-drug Interaction Event Prediction

Zhankun Xiong[1*], Shichao Liu[1*], Feng Huang[1*], Ziyan Wang[1], Xuan Liu[1], Zhongfei Zhang[2], Wen Zhang[1†]

[1]College of Informatics, Huazhong Agricultural University
[2]Computer Science Department, Binghamton University
xiongzk@webmail.hzau.edu.cn, scliu@mail.hzau.edu.cn, {fhuang233, wangziyan, lx666}@webmail.hzau.edu.cn,
zzhang@binghamton.edu, zhangwen@mail.hzau.edu.cn

Abstract

Drug-drug interactions (DDIs) could lead to various unexpected adverse consequences, so-called DDI events. Predicting DDI events can reduce the potential risk of combinatorial therapy and improve the safety of medication use, and has attracted much attention in the deep learning community. Recently, graph neural network (GNN)-based models have aroused broad interest and achieved satisfactory results in the DDI event prediction. Most existing GNN-based models ignore either drug structural information or drug interactive information, but both aspects of information are important for DDI event prediction. Furthermore, accurately predicting rare DDI events is hindered by their inadequate labeled instances. In this paper, we propose a new method, **M**ulti-**R**elational **C**ontrastive learning **G**raph **N**eural **N**etwork, **MRCGNN** for brevity, to predict DDI events. Specifically, MRCGNN integrates the two aspects of information by deploying a GNN on the multi-relational DDI event graph attributed with the drug features extracted from drug molecular graphs. Moreover, we implement a multi-relational contrastive learning with a designed dual-view negative counterpart augmentation strategy, to capture implicit information about rare DDI events. Extensive experiments on two datasets show that MRCGNN outperforms the state-of-the-art methods. Besides, we observe that MRCGNN achieves satisfactory performance when predicting rare DDI events.

Introduction

The combinatorial therapy with a concurrent use of multiple drugs is a promising strategy to treat patients with complicated diseases (Bansal et al. 2014). However, drug-drug interactions (DDIs) could bring about unexpected adverse consequences such as the reduction in efficacy or the increased toxicity of the drugs, and thus incur injuries and huge medical costs (Vilar et al. 2014; Lyu et al. 2021). Identifying DDIs has been deeply concerned in public health security and medicine safety surveillance.

In the past decade, plenty of computational prediction methods have been widely developed for a binary prediction task that determines whether interactions exist in pairwise drugs or not (Kastrin, Ferk, and Leskošek 2018; Zhang et al. 2019). In reality, DDIs lead to different types of biological consequences, so-called DDI events. Compared to the binary DDI prediction, predicting the event types of DDIs (we called predicting DDI events in following text for simplicity) is more helpful for investigating the mechanism hidden behind the consequences of polypharmacy. However, the binary prediction methods can hardly be generalized to multi-type DDI event prediction without a loss of accuracy. Recently, researchers have paid more attention to DDI event prediction and developed a number of specialized deep learning-based prediction methods. For example, Ryu, Kim, and Lee (2018) and Deng et al. (2020) utilized one or multiple drug biochemical features to build DNN models; Jin et al. (2017) formulated the DDI event data as tensors, and built the tensor factorization-based model; Lin et al. (2021) integrated multi-source data and drug features to build the transformer-based model.

Recent years have witnessed the strong power of graph neural networks (GNNs) in graph representation learning. GNNs have also been utilized for DDI event prediction in two implementation ways. Some works applied GNNs to learn drug structural information from drug molecular graphs with atoms as nodes and bonds as edges (Nyamabo, Yu, and Shi 2021; Nyamabo et al. 2021), and others are devoted to learning drug interactive information from drug association graphs that represent relationships among biological entities including drugs (Ma et al. 2018; Zitnik, Agrawal, and Leskovec 2018; Yu et al. 2021; Lyu et al. 2021). Although these methods have achieved satisfactory results, most of them consider either structural information or interactive information. Moreover, there exist some DDI events with a extremely low occurrence, called rare events. Since rare events involve exceedingly inadequate labeled instances that fail to provide sufficient supervisory signals, it remains challenging for most models to predict rare DDI events.

To alleviate the above limitations, we propose a novel DDI event prediction method, named MRCGNN (**M**ulti-**R**elational **C**ontrastive learning **G**raph **N**eural **N**etwork). In order to integrate drug structural information and drug interactive information, we apply R-GCN (Schlichtkrull et al. 2018) on the multi-relational DDI event graph attributed

[*]These authors contributed equally.
[†]Corresponding authors.

with the drug features extracted from drug molecular graphs, which is inspired by (Wang et al. 2020, 2021). Moreover, we consider contrastive learning over the DDI event graph to handle the challenge of predicting rare DDI events. Despite many advances in graph contrastive learning, there are still few studies on deploying contrastive learning over multi-relational graphs like the DDI event graph. For this work, we enable the multi-relational contrastive learning with a dual-view negative counterpart augmentation strategy under the mutual information maximization scheme (Velickovic et al. 2019), in which node features and edge relations are randomly shuffled in the DDI event graph respectively to form two views of negative counterparts.

In summary, the main contributions of this paper are described as follows:

- We integrate drug structural information from the drug molecular graphs and drug interactive information from the DDI event graph hierarchically to improve the DDI event prediction.
- We enable the multi-relational contrastive learning on the DDI event graph with a dual-view negative counterpart augmentation strategy, to capture implicit information about rare events.
- Extensive experiments on two datasets show that the proposed MRCGNN outperforms state-of-the-art baselines, and more importantly enhances the performance in predicting rare DDI events.

Related Work

Graph Neural Networks for DDI Event Prediction

In the DDI event prediction, GNNs are targeted at drug feature learning from two types of graph structure data: drug molecular graphs and drug association graphs.

A line of works apply GNNs to generate drug features by learning drug chemical structures from drug molecular graphs with atoms as nodes and bonds as edges. SSI-DDI (Nyamabo, Yu, and Shi 2021) employs the graph attention network (GAT) (Velickovic et al. 2018) on drug molecular graphs and integrated different combinations of multiple GAT layers from a pair of drugs by a co-attention mechanism to obtain prediction of the drug pair. Nyamabo et al. (2021) proposed a novel gated message passing neural network for drug feature extraction. MUFFIN (Chen et al. 2021) obtains two types of drug features from GNN on drug molecular graphs and KG embeddings, and develops a multi-scale fusion model to integrate the drug features for DDI event prediction. These models only take DDI events as supervised labels, but overlook rich interactive information in DDI event data.

The other line of works employ GNNs to extract drug features from drug association graphs, which represent the relationship among biological entities including drugs. The model in (Ma et al. 2018) learns multiple drug-drug similarity graphs through attentive multi-view graph auto-encoders with graph convolutional networks (Kipf and Welling 2017) as the backbones. Zitnik, Agrawal, and Leskovec (2018) developed a graph convolutional neural network for multi-relational link prediction in a multi-modal graph which contains interactions among drugs and target proteins. Feeney et al. (2021) considered an extension of the R-GCN on drug association graphs to model the importance of relation types for neighborhood sampling. Yu et al. (2021) and Lyu et al. (2021) built GNNs on drug association knowledge graphs (KGs) to extract knowledge-enriched drug features for predicting DDI events. However, in these drug association graphs, GNNs could not provide sufficiently high-quality representations for those drugs with relatively weak links to their neighbors, and thus it will influence the performances of prediction models on the rare events.

Different from above methods, we integrate both aspects of information from drug molecular graphs and the DDI event graph hierarchically to predict DDI events.

Contrastive Learning on Graphs

Contrastive learning is a kind of self-supervised learning paradigm that allows models to learn meaningful knowledge from pseudo labels generated from data themselves. Recent advances in graph contrastive learning, such as maximizing mutual information between local representations and global representation (Velickovic et al. 2019), have achieved the state-of-the-art results in unsupervised graph representation learning. Most graph contrastive learning methods are built on the underlying graphs, and few studies focus on multi-relational graphs. Generally, it is pivotal to construct positive and negative sample pairs in graph contrastive learning, which is commonly implemented by graph augmentation strategies, including but not limited to global-view augmentation by corrupting graph structure or shuffling initial node features (You et al. 2020) and local-view augmentation by subgraph sampling (Qiu et al. 2020; Hassani and Khasahmadi 2020). Here, we design a dual-view negative counterpart augmentation strategy with a modification on (Velickovic et al. 2019), in which two views of negative counterparts of DDI event graph are generated by shuffling edge relations and node features.

Method

In this section, we first formulate the DDI event prediction problem. After that, we elaborately enumerate all components of the proposed method MRCGNN that is shown in Figure 1. At last, we provide an exposition of model training.

Problem Formulation

DDI event data can be formulated as a multi-relational DDI event graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, where $\mathcal{V}$ is the set of nodes representing drugs, $\mathcal{R}$ is the set of relations representing event types, and $\mathcal{E}$ is the set of relational edges among drugs representing DDI events. Each node v in $\mathcal{V}$ can be viewed as a drug molecular graph, which is denoted as $\mathcal{M}_v$ with atoms as nodes and bonds as edges. Based on $\mathcal{G}$ and $\{\mathcal{M}_v\}_{v\in\mathcal{V}}$, our goal is to learn a model which predicts the specific DDI event of each drug pair.

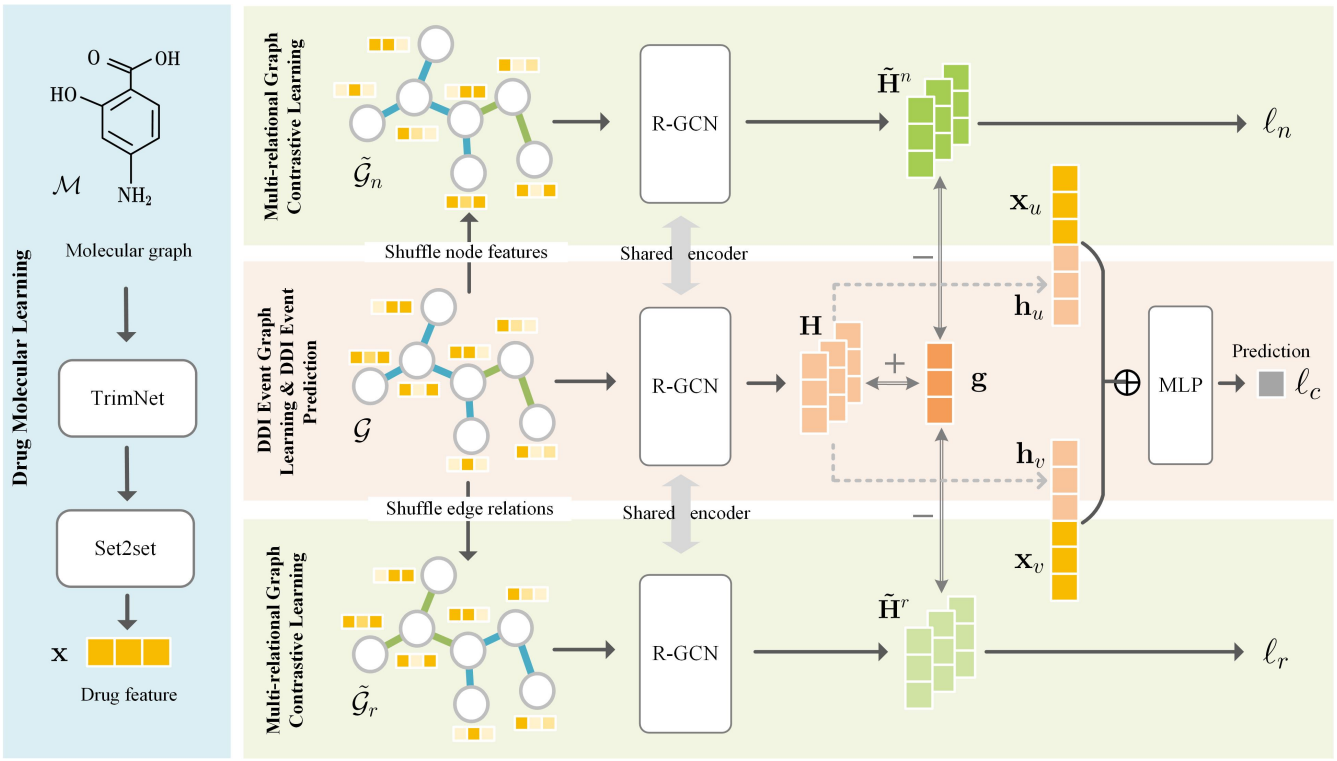

Figure 1: Overview of the proposed MRCGNN.

Drug Molecular Learning

For each drug molecular graph $\mathcal{M}_v$, we utilize TrimNet (Li et al. 2020), an advanced variant of the generic message passing neural network (Gilmer et al. 2017), to extract drug feature. TrimNet operates in two phases: a message passing phase and a readout phase.

More specifically, the message passing phase includes T iterations. In each iteration t, hidden state $\mathbf{s}_i^{(t)}$ associated with each atom i of $\mathcal{M}_v$ is updated by:

$$\mathbf{s}_i^{(t+1)} = \mathrm{U}(\mathbf{s}_i^{(t)}, \sum\nolimits_{j \in \mathcal{N}_i} \mathbf{m}_{ij}^{(t+1)})$$
$$\mathbf{m}_{ij}^{(t+1)} = \mathrm{M}(\mathbf{s}_i^{(t)}, \mathbf{s}_j^{(t)}, \mathbf{e}_{ij}^{(t)}). \quad (1)$$

Note that $\mathcal{N}_i$ represents the neighbors of atom i, $\mathbf{m}_{ij}^{(t+1)}$ is the message from atom j to i, $\mathbf{e}_{ij}^{(t)}$ denotes the hidden state of the edge between i and j, and the update function U, a gated recurrent unit (Chung et al. 2014) followed by a layer normalization, aggregates neighborhood messages. The message function M is defined by a multi-head attention mechanism:

$$\mathrm{M}(\mathbf{s}_i, \mathbf{s}_j, \mathbf{e}_{ij}) = ||_{k=1}^{K} \alpha_{ij}^k \odot \mathbf{W}_s^k \mathbf{s}_j \odot \mathbf{W}_e^k \mathbf{e}_{ij}$$
$$\alpha_{ij} = \mathrm{Softmax}(\sigma(\mathbf{u}^{\mathrm{T}}[\mathbf{W}_s \mathbf{s}_i || \mathbf{W}_e \mathbf{e}_{ij} || \mathbf{W}_s \mathbf{s}_j])), \quad (2)$$

where $\|$ represents vector concatenation operation, $\odot$ is the element-wise product, σ is the LeakyReLU nonlinear function, α_{ij} is a scalar measuring attention, and the vector $\mathbf{u}$ as well as all feature transformation matrices $\mathbf{W}_s$ and $\mathbf{W}_e$ are learnable parameters.

In the readout phase, TrimNet summarizes all atom embeddings after the former phase into a drug feature, denoted as $\mathbf{x}_v$, by using a readout function named Set2Set (Vinyals, Bengio, and Kudlur 2016). Thereafter, we obtain drug features $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times F}$.

Drug-drug Interaction Event Graph Learning

Due to the simplicity yet considerable effectiveness of the relational graph convolutional network (R-GCN) (Schlichtkrull et al. 2018) in modeling multi-relational graphs, we employ an R-GCN encoder to learn representations of drugs from the DDI event graph $\mathcal{G}$ with the drug features $\mathbf{X}$ as node attributes. Concretely, at the l-th layer of R-GCN, the intermediate embedding $\mathbf{h}_v^{(l)}$ of each drug v is updated by:

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{N}_v^r} \frac{1}{c_{vr}} \mathbf{W}_r^{(l)} \mathbf{h}_u^{(l)} + \mathbf{W}_o^{(l)} \mathbf{h}_v^{(l)}\right), \quad (3)$$

where $\sigma(\cdot)$ denotes an activation function, such as ReLU, $\mathcal{N}_v^r$ denotes the set of neighbor nodes of v under relation $r \in \mathcal{R}$, c_{vr} is a problem-specific normalization constant that can either be learned or chosen in advance (such as $c_{vr} = |\mathcal{N}_v^r|$), $\mathbf{h}_v^{(0)} = \mathbf{x}_v$ and $\mathbf{W}_r^{(l)}$ and $\mathbf{W}_o^{(l)}$ denote trainable weight matrices. Considering that the embeddings at different layers of R-GCN contain interactive information of different levels from the DDI event graph, we utilize a layer-

attention mechanism to combine these embeddings and obtain the final representation of each drug v:

$$\mathbf{h}_v = \sum_{l=1}^{L} \alpha_l \mathbf{h}_v^{(l)}, \tag{4}$$

where α_l is a trainable parameter as the adaptive contribution of the l-th layer embedding to the final drug representation $\mathbf{h}_v \in \mathbb{R}^Q$. We denote the representations for all drugs as $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}| \times Q}$.

Multi-Relational Contrastive Learning

Constructing positive and negative sample pairs by data augmentation is commonly crucial for contrastive learning. DGI (Velickovic et al. 2019) treats the 'fake' graphs generated by corrupting the original graph as negative counterparts, and provides several optional corruption operations such as shuffling initial node features and dropping some edges. Different from DGI that focuses on underlying graphs, for the multi-relational DDI event graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, we design a dual-view negative counterpart augmentation strategy that creates two views of corrupted graphs respectively by shuffling nodes features and shuffling edge relations. Then, we implement the multi-relational contrastive learning on $\mathcal{G}$ under the local-global mutual information maximization scheme (Velickovic et al. 2019).

In detail, we corrupt $\mathcal{G}$ to obtain two views of corrupted graphs $\tilde{\mathcal{G}}_n$ and $\tilde{\mathcal{G}}_r$ by: $\mathrm{C}_n : \mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R}) \rightarrow \tilde{\mathcal{G}}_n = (\tilde{\mathcal{V}}, \mathcal{E}, \mathcal{R})$ that shuffles the drug features $\mathbf{X}$, and $\mathrm{C}_r : \mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R}) \rightarrow \tilde{\mathcal{G}}_r = (\mathcal{V}, \mathcal{E}, \tilde{\mathcal{R}})$ that shuffles the edge relations. We employ the shared R-GCN encoder on $\mathcal{G}_n$ and $\mathcal{G}_r$, and obtain the corresponding 'fake' drug representations $\tilde{\mathbf{H}}^n$ and $\tilde{\mathbf{H}}^r$. Given the original drug representations $\mathbf{H}$, we use a readout function Γ to get global representation $\mathbf{g} \in \mathbb{R}^Q$ formulated by $\mathbf{g} = \Gamma(\mathbf{H})$. Then, the training objective of contrastive learning is to maximize the consistency between $\mathbf{H}$ and $\mathbf{g}$, and the difference between $\tilde{\mathbf{H}}^n$/$\tilde{\mathbf{H}}^r$ and $\mathbf{g}$, which can be formulated as the following loss functions:

$$\begin{aligned}
\ell_n = -\frac{1}{|\mathcal{V}| + |\tilde{\mathcal{V}}|} \Bigg(& \sum_{v \in \mathcal{V}} \mathbb{E}_{(\mathcal{V}, \mathcal{E}, \mathcal{R})} [\log \mathrm{D}(\mathbf{h}_v, \mathbf{g})] \\
& + \sum_{u \in \tilde{\mathcal{V}}} \mathbb{E}_{(\tilde{\mathcal{V}}, \mathcal{E}, \mathcal{R})} [\log(1 - \mathrm{D}(\tilde{\mathbf{h}}_u^n, \mathbf{g}))] \Bigg) \\
\ell_r = -\frac{1}{|\mathcal{V}| + |\mathcal{V}|} \Bigg(& \sum_{v \in \mathcal{V}} \mathbb{E}_{(\mathcal{V}, \mathcal{E}, \mathcal{R})} [\log \mathrm{D}(\mathbf{h}_v, \mathbf{g})] \\
& + \sum_{u \in \mathcal{V}} \mathbb{E}_{(\mathcal{V}, \mathcal{E}, \tilde{\mathcal{R}})} [\log(1 - \mathrm{D}(\tilde{\mathbf{h}}_u^r, \mathbf{g}))] \Bigg),
\end{aligned} \tag{5}$$

where $D(\mathbf{h}_v, \mathbf{g}) = \sigma(\mathbf{h}_v^{\mathrm{T}} \mathbf{W} \mathbf{g})$ and $\mathbf{W}$ is a trainable parameter matrix.

DDI Event Prediction

For each drug pair (u, v), we now have their features $\mathbf{x}_u$ and $\mathbf{x}_v$, and their final representations $\mathbf{h}_u$ and $\mathbf{h}_v$. Then, we concatenate them to form the drug pair representation $\mathbf{h}_{(u,v)} = \mathbf{h}_u || \mathbf{x}_u || \mathbf{h}_v || \mathbf{x}_v$. After that, $\mathbf{h}_{(u,v)}$ is fed into a Multi-Layer Perceptron (MLP) followed by a Softmax function to get the multi-class prediction of the drug pair:

$$\hat{\mathbf{y}}_{(u,v)} = \mathrm{Softmax}(\mathrm{MLP}(\mathbf{h}_{(u,v)})), \tag{6}$$

where $\hat{\mathbf{y}}_{(u,v)} \in \mathbb{R}^{|\mathcal{R}|}$. The training objective of DDI event prediction is to minimize the loss function:

$$\ell_c = -\sum_{(u,v) \in \Omega} \sum_{r \in \mathcal{R}} y_{(u,v)}^r \log \hat{y}_{(u,v)}^r, \tag{7}$$

where Ω is the training set, $\hat{y}_{(u,v)}^r$ indicates the predicted probability that the drug pair (u, v) belongs to relation type r and $y_{(u,v)}^r$ is the corresponding true label.

Model Training

For training our model MRCGNN, we optimize the total loss that combines Eq.(5) and Eq.(7):

$$\ell = \ell_c + \alpha \ell_r + \beta \ell_n, \tag{8}$$

where α and β are hyper-parameters which balance the contributions of different tasks.

Experiments

In this section, we first introduce the experimental settings, and then compare our model MRCGNN with baselines on DDI event prediction, prediction for rare DDI events and representation visualization. After that, we investigate the effectiveness of each component and conduct the hyper-parameter sensitivity analysis in our model. Besides, we also conduct a case study which can be found in the Appendix.C.

Experimental Settings

Datasets. We evaluate our method MRCGNN on two datasets: (1) *Deng's dataset* (Deng et al. 2020) contains a total of 37,264 DDIs between 570 drugs with 65 types of DDI events. (2) *Ryu's dataset* (Ryu, Kim, and Lee 2018) contains a total of 191,570 DDIs between 1,700 drugs with 86 types of DDI events. Specifically, we count the number of DDI instances involving each DDI event, called event frequency, and then we classify DDI events into five groups based on their frequencies. Table 1 lists the proportion of events in each group to all events. It is worth mentioning that there are a number of events with extremely limited labeled instances, as indicated by the group [1,10]. In our subsequent experiments, we take the events in the group [1,10] as rare events, and we evaluate the performances of prediction models on all the events and the rare events.

Table 1: Proportions of events in five groups to all events.

Datasets	Five groups				
	[1,10]	(10,50]	(50,100]	(100,300]	(300,+∞)
Deng's	20.0%	21.5%	24.6%	15.4%	18.5%
Ryu's	5.8%	21.0%	11.6%	14.0%	47.6%

Table 2: Results of MRCGNN and baselines for DDI event prediction on two datasets.

Methods	Deng's dataset				Ryu's dataset			
	Acc.	Macro-F1	Macro-Rec.	Macro-Prec.	Acc.	Macro-F1	Macro-Rec.	Macro-Prec.
DeepDDI	0.7807	0.6055	0.5839	0.6611	0.9323	0.8643	0.8512	0.8928
SSI-DDI	0.7866	0.4216	0.3896	0.5139	0.9008	0.6663	0.6287	0.7507
TrimNet-DDI	0.8570	0.6548	0.6363	0.7046	0.9353	0.8288	0.8128	0.8627
MUFFIN	0.8269	0.5245	0.4844	0.6204	0.9510	0.8566	0.8339	0.8980
R-GCN	0.8695	0.7026	0.6878	0.7500	0.9284	0.8487	0.8291	0.8881
GoGNN	0.8766	0.6938	0.6841	0.7316	0.9424	0.8589	0.8451	0.8949
MRCGNN	**0.8979**	**0.7791**	**0.7688**	**0.8101**	**0.9566**	**0.8894**	**0.8727**	**0.9221**

Baselines. We compare our model with several baselines, which can be categorized as follows.

- *DeepDDI* (Ryu, Kim, and Lee 2018) is the first event prediction method, which uses drug structural similarity as input and builds a deep neural network to predict the interaction types of drug pairs.
- *SSI-DDI* (Nyamabo, Yu, and Shi 2021) employs a GAT on drug molecular graphs, and combines embeddings from multiple GAT layers with a co-attention mechanism to obtain the predictions of drug pairs .
- *TrimNet-DDI* learns drug features from drug molecular graphs through a TrimNet (Li et al. 2020), and feeds the concatenated features of two drugs into a MLP to yield the prediction of the drug pair.
- *MUFFIN* (Chen et al. 2021) is a multi-scale feature fusion deep learning model for DDI event prediction, which fuses the drug features learned from molecular graphs and the pre-trained KG embeddings.
- *R-GCN* (Schlichtkrull et al. 2018) is the classic GNN designed for multi-relational graphs. We directly employ R-GCN on the DDI event graph to learn drug representations, and then feed the concatenated representations of drug pairs into a MLP to predict DDI events. '
- *GoGNN* (Wang et al. 2020) models molecular interaction network as an interaction graph of molecular graphs. It adopts GNNs to learn graph-level representations which are then updated by another GNN deployed on the interaction graph for molecular interaction prediction.

Implementation details. For model evaluation, we split DDIs in each dataset into training, validation and test sets with a 7:1:2 ratio, and ensure that training/validation/test sets contain DDIs from all types. Since predicting DDI events is a multi-class classification task, we adopt several evaluation metrics, including Accuracy, Macro-F1, Macro-Recall and Macro-Precision under a multi-class setting. For training our model, we set learning rate $lr = 0.001$ and the coefficients $\alpha = 0.1$ and $\beta = 0.05$, and the best performing models based on the Accuracy and Macro-F1 on the validation set with a maximum epoch of 100 are selected for model testing on test set. For each model, we conduct five independent runs with different data splits, and the average metrics of these five independent runs are adopted as the final results. All experiments are conducted on the same machine with Intel(R)Core(TM)i9-7900X CPU @ 3.30GHz and 2 GPUs(NVIDIA GeForce 1080Ti). More details of our model configuration can be found in the Appendix.A. Our code, data and appendix are publicly available[1].

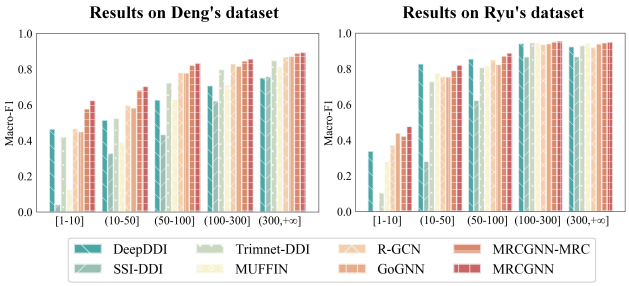

Figure 2: Results of MRCGNN and baselines on five groups of events.

Comparison with Baselines

Table 2 shows the performances of MRCGNN and the baselines on the two datasets, where the best results are bold. According to the results shown in Table 2, MRCGNN acquires the best performance on both datasets. We also have the following observations: (1) Compared with SSI-DDI, TrimNet-DDI and MUFFIN that only consider drug structural information, MRCGNN makes improvements of 14.15%, 4.77% and 8.59% on Deng's datasets, as well as 6.19%, 2.28% and 0.59% on Ryu's dataset in terms of Accuracy, which indicates that drug interactive information benefits DDI event prediction. (2) Compared with DeepDDI and R-GCN that only considered drug interactive information, MRCGNN surpasses them by 15.01% and 3.27% on Deng's datasets, and 2.61% and 3.04% on Ryu's dataset in terms of Accuracy, which implies that considering information from drug molecular graph is also beneficial to DDI event prediction. (3) Among all baselines, GoGNN exhibits better performance than most others, implying the advantage of integrating both aspects of information from drug molecular graphs and the DDI event graph. Our model MRCGNN still makes improvements of 2.4% on Deng's datasets and 1.5% on Ryu's dataset in terms of Accuracy over GoGNN. It may be attributed to the reason that MRCGNN can not only effectively integrate drug structural information and drug interac-

[1]https://github.com/Zhankun-Xiong/MRCGNN

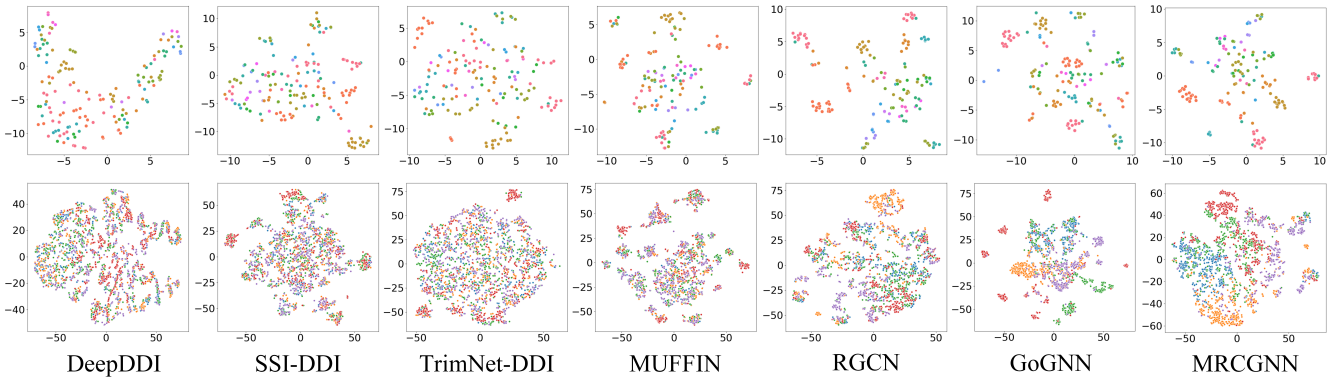

Figure 3: Visualization on Deng's dataset using the t-SNE. Each point represents a drug pair, and the color represents the DDI event. Upper: 20 events with the lowest frequency. Lower: 5 events with the highest frequency.

tive information as GoGNN does, but also capture implicit information behind the multi-relational DDI event graph by the multi-relational contrastive learning.

To further compare MRCGNN with baselines, we investigate their Macro-F1 scores on the five DDI event groups listed in Table 1, and especially focus on their performances on rare events. To verify the effectiveness of multi-relational contrastive learning in predicting rare DDI events, we extra include a variant of MRCGNN here, named MRCGNN-MRC, which removes the multi-relational contrastive learning from MRCGNN. As shown in Figure 2, the performances of all models prominently decline with decreasing event frequencies, and MRCGNN outperforms all baselines on each group of DDI events, especially on rare events with a significant improvement, which demonstrates that MRCGNN has considerable advantages in predicting rare DDI events. In addition, we also have the following observations: (1) SSI-DDI, TrimNet-DDI and MUFFIN that only use drug structural information achieve relatively unsatisfactory performances on rare event group, likely due to the ignorance of drug interactive information. (2) Compared to DeepDDI and R-GCN that only take in drug interactive information, MRCGNN-MRC has better performance, which suggests that integrating drug structural information and drug interactive information can effectively enhance the prediction for rare DDI events. (3) The superiority of MRCGNN over GoGNN and MRCGNN-MRC illustrates the multi-relational contrastive learning can further help the prediction for rare DDI events. To sum up, integrating two aspects of information and considering contrastive learning both contribute to the best performance of our model MRCGNN.

To better understand the superiority of MRCGNN over baselines, we use t-SNE (Laurens and Hinton 2008) to visualize drug pair representations. Since there are dozens of DDI event types, we choose 20 events with the lowest frequency and 5 events with the highest frequency for visualization. Figure 3 shows the visualization on Deng's dataset (the illustration on Ryu's dataset can be found in the Appendix.B). We clearly observe that drug pairs are more tightly clustered in MRCGNN compared with baselines (the silhouette coefficients of MRCGNN and baselines can be found in our Appendix.B), which implies that MRCGNN can learn more high-quality representations for drug pairs by effectively integrating drug structural information from drug molecular graphs and drug interactive information from the DDI event graph. It is worth noting that more compact clusters in MRCGNN on 20 events with the lowest frequency (see the upper part in Figure 3) can well illustrate the excellent ability of MRCGNN on rare DDI event prediction.

Ablation Study

To investigate the importance of various components to our model, we consider the following variants of MRCGNN:

- *MRCGNN without the DDI event graph learning* (w/o DEG) removes multi-relational DDI event graph learning and directly uses drug features learned from drug molecular for prediction.
- *MRCGNN without drug molecular learning* (w/o DM) removes drug molecular learning and replaces drug features with randomly generated drug features.
- *MRCGNN without multi-relational contrastive learning* (w/o MGC) removes the multi-relational contrastive learning from MRCGNN.
- *MRCGNN without shuffling node features* (w/o SNF) removes the view of the corrupted graph generated by shuffling node features in DDI event graph.
- *MRCGNN without shuffling edge relations* (w/o SER) removes the view of the corrupted graph generated by shuffling edge relations in DDI event graph.

As shown in Figure 4, all variants of MRCGNN produce the decreased performances, verifying that all components can contribute to DDI event prediction. Besides, we have the following observations: (1) MRCGNN(w/o MRC) outperforms the MRCGNN(w/o DEG) and MRCGNN(w/o DM), which demonstrates that integrating drug structural information and drug interactive information is helpful for DDI event prediction. Note that MRCGNN(w/o DEG) has more decreased performance than MRCGNN(w/o DM), which indicates drug interactive information is more pivotal for

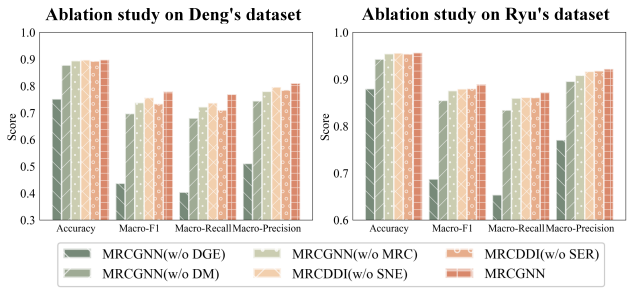

Figure 4: Results of MRCGNN and its variants in ablation study.

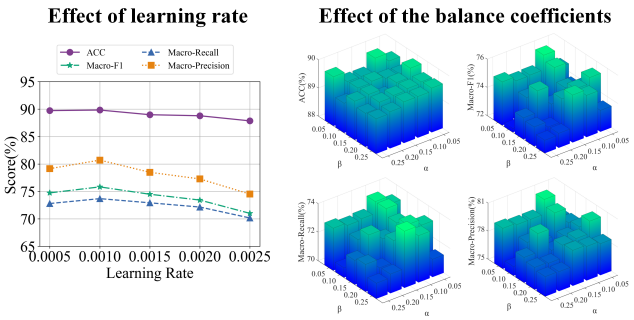

Figure 5: Hyper-parameter Sensitivity Analysis

DDI event prediction than drug structural information. (2) The comparison of MRCGNN and MRCGNN(w/o MRC) reveals the capability of the multi-relational contrastive learning mining implicit information from DDI event data. (3) MRCGNN outperforms MRCGNN(w/o SNF) and MRCGNN(w/o SER), which demonstrates the effectiveness of our designed dual-view negative counterpart augmentation strategy on the DDI event graph. (4) MRCGNN(w/o SNF) performs better than MRCGNN(w/o SER) on Deng's dataset, while the results are reversed on Ryu's dataset. Note that the proportion of rare DDI events on Deng's dataset is higher than that on Ryu's dataset. The reason of this observation may be that the corrupted graph generated by shuffling edge relations enhances multi-relational constrastive learning in mining implicit information of rare events.

Hyper-parameter Sensitivity Analysis

In this section, we conduct hyper-parameter sensitivity analysis on Deng's dataset to study the influence of several hyper-parameters on the performance of MRCGNN. We choose the learning rate lr and the coefficients α and β in Eq.(8) for hyper-parameter sensitivity analysis here.

Effect of learning rate lr. We conduct experiments by varying the learning rate lr to be 0.0005, 0.001, 0.0015, 0.002 and 0.0025. Results in Figure 5 show that MRCGNN achieves the best results when lr=0.001 and suffers from degrading performance after that. As a result, we set 0.001 as default learning rate.

Effect of the coefficients α and β. We vary the coefficients α and β in Eq.(8), to investigate the contribution of different contrastive learning tasks. In particular, the coefficients α and β are searched in the range of $\{0.05, 0.10, 0.15, 0.20, 0.25\}$, and MRCGNN achieves the best performance when $\alpha = 0.10$ and $\beta = 0.05$. We summarize the results in Figure 5, and have the following observations: (1) Larger values of α and β will lead to degradation of model performance, because the high weight of the contrast learning task will make the model pay too much attention to the contrastive learning task during model training, and thus make the model poorly fit the DDI event prediction task. (2) α is greater than β when MRCGNN achieves the best performance, which demonstrates that the corrupted graph generated by shuffling edge relations may be more important for the multi-relational contrastive learning on Deng's dataset.

Conclusion

In this paper, we propose a multi-relational contrastive learning graph neural network (MRCGNN) to predict DDI events. MRCGNN hierarchically integrates the drug structural information from drug molecular graphs and the drug interactive information from the DDI event graph. To enable the multi-relational contrastive learning on the DDI event graph, we design a dual-view negative counterpart augmentation strategy, to capture implicit information about rare events and improve the prediction for rare events. The experimental results on the two benchmark datasets show the superior performances of MRCGNN over baselines, and MRCGNN can achieves satisfactory performance when predicting rare DDI events.

In the future work, we have several directions to improve DDI event prediction, such as incorporating more relevant entities and relations into DDI event graphs, providing interpretable DDI event prediction models and modeling 3D structural information of drugs with geometrically equivariant graph neural networks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No.62072206, 62102158, 61772381), the Fundamental Research Funds for the Central Universities (Grant No.2662022JC004, 2662021JC008). Huazhong Agricultural University Scientific & Technological Self-innovation Foundation. The funders have no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

References

Bansal, M.; Yang, J.; Karan, C.; Menden, M. P.; Costello, J. C.; Tang, H.; Xiao, G.; Li, Y.; Allen, J.; Zhong, R.; et al. 2014. A community computational challenge to predict the activity of pairs of compounds. *Nature biotechnology*, 32(12): 1213–1222.

Chen, Y.; Ma, T.; Yang, X.; Wang, J.; Song, B.; and Zeng, X. 2021. MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics*, 37(17): 2651–2658.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv:1412.3555.

Deng, Y.; Xu, X.; Qiu, Y.; Xia, J.; Zhang, W.; and Liu, S. 2020. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*, 36(15): 4316–4322.

Feeney, A.; Gupta, R.; Thost, V.; Angell, R.; Chandu, G.; Adhikari, Y.; and Ma, T. 2021. Relation Matters in Sampling: A Scalable Multi-Relational Graph Neural Network for Drug-Drug Interaction Prediction. arXiv:2105.13975.

Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural Message Passing for Quantum Chemistry. In Precup, D.; and Teh, Y. W., eds., *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, 1263–1272. PMLR.

Hassani, K.; and Khasahmadi, A. H. 2020. Contrastive Multi-View Representation Learning on Graphs. In III, H. D.; and Singh, A., eds., *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 4116–4126. PMLR.

Jin, B.; Yang, H.; Xiao, C.; Zhang, P.; Wei, X.; and Wang, F. 2017. Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Kastrin, A.; Ferk, P.; and Leskošek, B. 2018. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PloS one*, 13(5): e0196865.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Laurens, V. D. M.; and Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(2605): 2579–2605.

Li, P.; Li, Y.; Hsieh, C.-Y.; Zhang, S.; Liu, X.; Liu, H.; Song, S.; and Yao, X. 2020. TrimNet: learning molecular representation from triplet messages for biomedicine. *Briefings in Bioinformatics*, 22(4). Bbaa266.

Lin, S.; Wang, Y.; Zhang, L.; Chu, Y.; Liu, Y.; Fang, Y.; Jiang, M.; Wang, Q.; Zhao, B.; Xiong, Y.; et al. 2021. MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics*.

Lyu, T.; Gao, J.; Tian, L.; Li, Z.; Zhang, P.; and Zhang, J. 2021. MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events. In Zhou, Z.-H., ed., *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 3536–3542. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Ma, T.; Xiao, C.; Zhou, J.; and Wang, F. 2018. Drug Similarity Integration Through Attentive Multi-view Graph Auto-Encoders. In Lang, J., ed., *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, 3477–3483. ijcai.org.

Nyamabo, A. K.; Yu, H.; Liu, Z.; and Shi, J.-Y. 2021. Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*.

Nyamabo, A. K.; Yu, H.; and Shi, J.-Y. 2021. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*.

Qiu, J.; Chen, Q.; Dong, Y.; Zhang, J.; Yang, H.; Ding, M.; Wang, K.; and Tang, J. 2020. GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '20, 1150–1160. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379984.

Ryu, J. Y.; Kim, H. U.; and Lee, S. Y. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the National Academy of Sciences*, 115(18): E4304–E4311.

Schlichtkrull, M. S.; Kipf, T. N.; Bloem, P.; van den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling Relational Data with Graph Convolutional Networks. In Gangemi, A.; Navigli, R.; Vidal, M.; Hitzler, P.; Troncy, R.; Hollink, L.; Tordai, A.; and Alam, M., eds., *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, 593–607. Springer.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Vilar, S.; Uriarte, E.; Santana, L.; Lorberbaum, T.; Hripcsak, G.; Friedman, C.; and Tatonetti, N. P. 2014. Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols*, 9(9): 2147–2163.

Vinyals, O.; Bengio, S.; and Kudlur, M. 2016. Order Matters: Sequence to sequence for sets. In Bengio, Y.; and LeCun, Y., eds., *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Wang, H.; Lian, D.; Zhang, Y.; Qin, L.; and Lin, X. 2020. GoGNN: Graph of Graphs Neural Network for Predicting Structured Entity Interactions. In Bessiere, C., ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, 1317–1323. ijcai.org.

Wang, Y.; Min, Y.; Chen, X.; and Wu, J. 2021. Multi-view Graph Contrastive Representation Learning for Drug-Drug

Interaction Prediction. In Leskovec, J.; Grobelnik, M.; Najork, M.; Tang, J.; and Zia, L., eds., *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, 2921–2933. ACM / IW3C2.

You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph Contrastive Learning with Augmentations. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 5812–5823. Curran Associates, Inc.

Yu, Y.; Huang, K.; Zhang, C.; Glass, L. M.; Sun, J.; and Xiao, C. 2021. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*, 37(18): 2988–2995.

Zhang, W.; Jing, K.; Huang, F.; Chen, Y.; Li, B.; Li, J.; and Gong, J. 2019. SFLLN: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Information Sciences*, 497: 189–201.

Zitnik, M.; Agrawal, M.; and Leskovec, J. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13): i457–i466.