

Appendix

A Details about Experimental Setup

Our model is implemented with PyTorch 1.6.0(Paszke et al. 2019) and pytorch-geometric 1.7.0. Table 1 demonstrates all the hyper-parameters of MRCGNN. We develop all codes on the machine with Intel(R)Core(TM)i9-7900X CPU @ 3.30GHz and 2 GPUs(NVIDIA GeForce 1080Ti).

B Details and results of visualization analysis

Here, we use t-SNE (Laurens and Hinton 2008) to visualize drug pair representations. Specifically, we first extracted the drug representations learned by MRCGNN and baselines, and obtained the representations of drug pairs by concatenating the drug representations according to the instances in the dataset. Since the length of representations learned by different methods is inconsistent, and some of them are very long (for example, MUFFIN), we used Principal Components Analysis (PCA) (Hotelling 1933) method to reduce the dimension of drug pair representations as (Laurens and Hinton 2008) did. After that, we used t-SNE method to project the representations into the two-dimensional space, and finally drew the visualization for representations of drug pairs. Since there are dozens of DDI event types (65 in Deng’s dataset and 86 in Ryu’s dataset), we choose 20 events with the lowest frequency and 5 events with the highest frequency for visualization. It is worth noting that 5 events with the highest frequency have a large number of instances, so we selected a maximum of 500 for each type of event for visualization.

Figure 1 shows the visualization on Ryu’s dataset. Compared with the visualization analysis experiment on Deng’s dataset, the visualization results of each method is better on Ryu’s dataset than that on Deng dataset, which may be due to the fact that in 20 events with the lowest frequency, the proportion of rare DDI events on Deng’s dataset is higher than that on Ryu’s dataset (in Deng dataset, rare events account for 20%, And the proportion in Ryu dataset is 5.8%). In addition, we can still clearly observe that drug pairs are more tightly clustered in MRCGNN compared with baselines, 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.

C Case study

In this section, we conduct a case study to validate the usefulness of MRCDDI in practice. We use all DDIs and their events in Ryu’s dataset which were originally from DrugBank to train the prediction model, and then make predictions for the other drug–drug pairs. We pay attention to two events with high frequency in the dataset, and check up on the top 10 predictions related to these events. We use the Interactions Checker tool provided by drugs.com or searched on Drugbank to validate these predictions. The case study results are shown in Table 2, where #1 denotes the DDI event

“DRUG1 may increase the central nervous system depressant (CNS depressant) activities of DRUG2”. #2 denotes the DDI event “The risk or severity of adverse effects can be increased when DRUG1 is combined with DRUG2”. There are 90% of the predictions related to #1 can be confirmed. For example, the central nervous system and/or respiratory-depressant effects may be additively or synergistically increased in patients taking both **trimethobenzamide** and **paraldehyde**, especially in elderly or debilitated patients. And there are 50% of the predictions related to #2 can be confirmed. For example, according to drugs.com, both **dextroamphetamine** and **mephentermine** can increase blood pressure and heart rate, and combining them may enhance these effects.

References

- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6): 417.
- Laurens, V. D. M.; and Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(2605): 2579–2605.
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Table 1: The hyper-parameters of MRCGNN.

Hyper-parameter	Description	Value
T	the number of iterations in the message passing phase of TrimNet	3
K	the number of heads of multi-head attention	4
Trimnet_node_hidden	the node hidden size for Trimnet	64
F	the feature size of drug features learned by Trimnet	128
dropout_Trimnet	dropout rate of dropout layer in Trimnet	0.1
R-GCN_node_hidden1	the node hidden size for the first layer of R-GCN	64
R-GCN_node_hidden2	the node hidden size for the second layer of R-GCN	32
Q	the representations of drugs	96
l	the number of layers of R-GCN	2
dropout_R-GCN	dropout rate of dropout layer in R-GCN	0.5
dropout_MLP	dropout rate of dropout layer in MLP	0.1
weight_decay	weight_decay for Adam optimizer	0.0005
epoch	the number of training epochs	100
batch_size	the input batch size	256

Table 2: Top 10 drug pairs with interaction type #1 and top 10 drug pairs with interaction type #2

Event	Drug names	Drug names	Evidence
#1	Delorazepam	Paraldehyde	confirmed
	Doxylamine	Delorazepam	confirmed
	Doxylamine	Pirenzepine	confirmed
	Doxylamine	Trospium	confirmed
	Trimethobenzamide	Paraldehyde	confirmed
	Delorazepam	Zolpidem	confirmed
	Nabilone	Delorazepam	confirmed
	Delorazepam	Hydrocodone	confirmed
	Magnesium salicylate	Paraldehyde	N.A
	Perampanel	Trospium	confirmed
#2	Nicergoline	Alitretinoin	N.A
	Dextroamphetamine	Mephentermine	confirmed
	Nicergoline	Triprolidine	confirmed
	Remoxipride	Camazepam	N.A
	Nicergoline	Chloral hydrate	confirmed
	Nicergoline	Desloratadine	confirmed
	Nicergoline	Belinostat	N.A
	Fluphenazine	Camazepam	N.A
	Hydroxyamphetamine	Oladaterol	confirmed
	Fluphenazine	Delorazepam	N.A

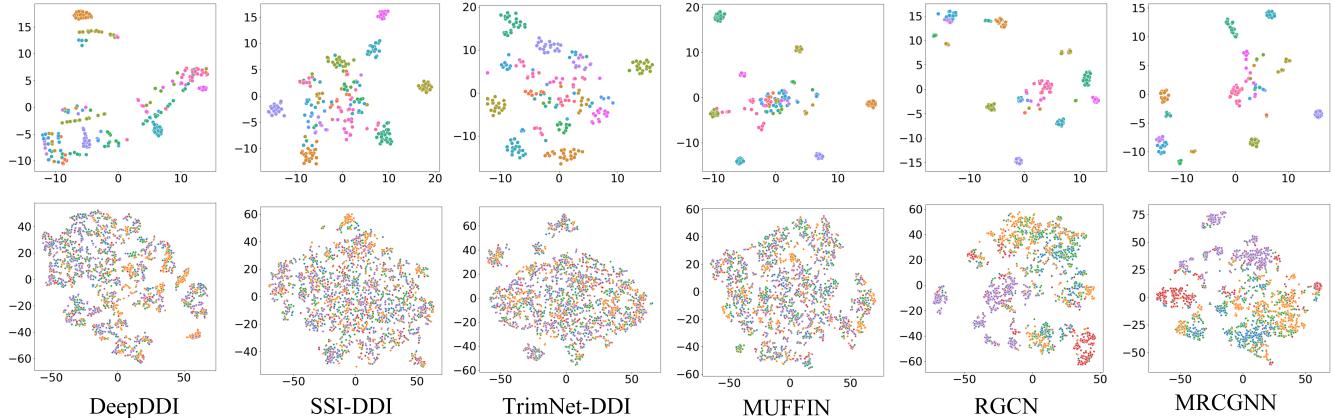


Figure 1: Visualization on Ryu's dataset using the t-SNE. Each point represents a drug pair, and the color represents the DDI event. Upper: 20 events with the lowest frequency. Lower: 5 events with the highest frequency.