Supplementary Materials in the inductive setting

Xinyu Zhu

November 2022

1 Experiments on SSI-DDI dataset

The dataset and splits used in the original work of SSI-DDI [1] are a little bit different from those used in our work. The dataset used in SSI-DDI contains 1,704 drugs rather than 1,706 as in our work. Our work adopts the dataset in GMPNN [2]. Also, SSI-DDI samples negative drug pairs at the beginning of each epoch, while we use the same negative drug pairs for all epochs. To provide an additional evaluation of our model in the harder inductive setting, we run MSAN-GCN on the dataset and splits from the original work of SSI-DDI, also following the same negative sampling strategy in SSI-DDI. The results are shown in the table below:

Table 1: Performance (mean±std) in the inductive setting using the dataset and splits from the original work of SSI-DDI

	$\mathcal{D}_{S_1}$			$\mathcal{D}_{S_2}$		
	ACC	**AUROC**	**AUPRC**	**ACC**	**AUROC**	**AUPRC**
DeepDDI	47.96±2.66	88.39±2.19	48.27±2.64	68.69±1.76	95.75±1.12	68.87±1.75
MR-GNN	49.93±2.32	91.70±0.90	50.23±2.30	72.38±0.73	97.52±0.31	72.54±0.73
MHCADDI	66.53±1.05	72.13±1.24	72.48±1.12	72.34±1.32	79.22±1.31	78.27±1.30
SSI–DDI	65.02±0.74	72.38±1.01	72.83±1.35	73.35±0.70	80.92±1.62	80.80±2.25
MSAN-GCN-NN-NS (Ours)	64.93±1.55	72.88±1.09	73.02±1.37	75.42±1.03	83.64±1.29	83.77±1.52
MSAN-GCN-NN-S (Ours)	68.97±1.07	77.09±0.83	77.88±0.31	77.84±0.44	86.37±0.60	86.95±0.53
MSAN-GCN-NS (Ours)	62.13±0.90	69.38±1.91	70.38±1.41	71.49±0.09	80.56±1.88	81.19±2.02
MSAN-GCN-S (Ours)	66.08±0.50	74.80±0.36	75.15±0.59	75.83±0.38	85.12±0.68	85.62±0.67

where the suffix "NN" means to use the nearest neighbor augmentation mentioned in the subsection 3.1 in our paper. "S" or "NS" means whether we have a strict range of negative sampling or not. A strict range of negative sampling means we can only replace old drugs with old drugs, and new drugs with new drugs.

From the table, we can conclude that the nearest neighbor augmentation and the range of negative sampling have a significant impact on the performance of our model. The results of the four baseline methods are collected from the original paper of SSI-DDI. I notice that some statistics of certain baseline methods are extremely high which are not in accordance with other statistics of the same

method. When those outliers are neglected, our model (with the nearest neighbor augmentation or strict range of negative sampling) generally outperforms all the baseline methods in most of the metrics.

The nearest neighbor augmentation strategy is now commonplace in many areas of machine learning. Many models interpolate the testing prediction with the predictions from the testing instance's k-nearest neighbors in the training set [3, 4, 5].

References

[1] Arnold K Nyamabo, Hui Yu, and Jian-Yu Shi. Ssi-ddi: substructure–substructure interactions for drug–drug interaction prediction. *Briefings in Bioinformatics*, 22(6), 5 2021. bbab133.

[2] Arnold K Nyamabo, Hui Yu, Zun Liu, and Jian-Yu Shi. Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, 23(1), 1 2022. bbab441.

[3] Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. Generalization through memorization: Nearest neighbor language models. In *International Conference on Learning Representations*, 2020.

[4] Xiang Chen, Lei Li, Ningyu Zhang, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. Relation extraction as open-book examination: Retrieval-enhanced prompt tuning. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 2443–2448, New York, NY, USA, 2022. Association for Computing Machinery.

[5] Xi'ao Su, Ran Wang, and Xinyu Dai. Contrastive learning-enhanced nearest neighbor mechanism for multi-label text classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 672–679, Dublin, Ireland, may 2022. Association for Computational Linguistics.