CALMOL: DISENTANGLED CAUSAL GRAPH LLM FOR MOLECULAR RELATIONAL LEARNING

Anonymous authors
Paper under double-blind review

0.1 EXPERIMENT SETTINGS

In this section, we introduce our experimental setups in detail with descriptions of used datasets along with baseline models to benchmark the performance of our proposed method.

0.1.1 DATASETS

In our experiment, 7 diverse datasets are employed, encompassing both drug-drug interaction and solute-solvent interaction tasks. Here we provide a brief overview of the original datasets. Detailed statistical information on the datasets used in this study can be found in Table **??**.

ZhangDDI (Zhang et al. (2017)). It consists of 548 drugs and 48,548 drug-drug interaction pairs in total, along with multiple types of similarity information between drug pairs.

ChChMiner (Marinka Zitnik & Leskovec (2018)). This dataset contains 1,322 drugs and their labeled DDIs, all of which have been extracted from official drug labels and validated through scientific research.

DeepDDI (Ryu et al. (2018)). It collects 1704 various drugs with their labeled DDIs. The collection is gathered from DrugBank which features detailed DDI data alongside associated side-effect annotations.

FreeSolv (Mobley & Guthrie (2014)). The dataset includes 643 hydration free energy measurements for small molecules in water, both experimental and calculated. For our study, we focus on 560 experimental values, consistent with previous work.

CompSol (Moine et al. (2017)). This dataset aims to demonstrate the influence of hydrogen-bonding interactions on solvation energies. It includes a total of 3,548 combinations involving 442 unique solutes and 259 solvents, as referenced in earlier studies.

Abraham (Grubbs et al. (2010)). It compiles information published by the Abraham research group at University College London. It includes 6,091 combinations of 1,038 unique solutes and 122 solvents, in accordance with prior studies.

CombiSolv (Vermeire & Green (2021)). It integrates data from the MNSol, FreeSolv, CompSol, and Abraham datasets, resulting in 8780 unique pairings between 1,415 solutes and 309 solvents

0.1.2 BASELINES

In this section, we provide introduction of the baseline models utilized in our experiment. Both traditional deep learning based methods and the recent biochemical LLMs are employed. For qualitative tasks, we use the following baselines:

CIGIN (Pathak et al. (2020)). This model uses a three-phase framework—message passing, interaction, and prediction—to achieve high accuracy in solvation free energy predictions and provides chemically interpretable insights into electronic and steric factors governing solubility.

SSI-DDI (Nyamabo et al. (2021)). This method applies a 4-layer GAT model to uncover substructures across different layers, while the co-attention mechanism handles the final prediction.

DSN-DDI (Li et al. (2023)). It persents a dual-view drug representation learning network that integrates local and global drug substructure information from both individual drugs ('intra-view') and drug pairs ('inter-view').

CMRL (Lee et al. (2023b)). The approach reveals the main substructure driving chemical reactions through a conditional intervention model that adapts its intervention based on the paired molecule.

CGIB (Lee et al. (2023a)). It adapts the detected substructure depending on the paired molecule to mimick real chemical reactions, based on the conditional graph information bottleneck theory.

DeepDDI (Ryu et al. (2018)). In this method, the structural similarity profile of the two drugs is first evaluated against other drugs, after which a deep neural network is used to complete the prediction.

MHCADDI (Deac et al. (2019)). The model utilizes a gated information transfer neural network to manage substructure extraction, and interactions are guided by an attention mechanism.

MolTC (Fang et al. (2024)). It introduces a novel multi-modal framework that integrates molecular graph structures and LLMs using Chain-of-Thought (CoT) theory.

As for quantitative tasks, the following baselines are employed besides CIGIN, CMRL, CGIB and MolTC which are mentioned above:

D-MPNN (Vermeire & Green (2021)). This technique combines the fundamentals of quantum calculations with the experimental precision of solvation free energy measurements, using a transfer learning approach with the CombiSolv-QM and CombiSolv-Exp databases.

1 Ablation Study

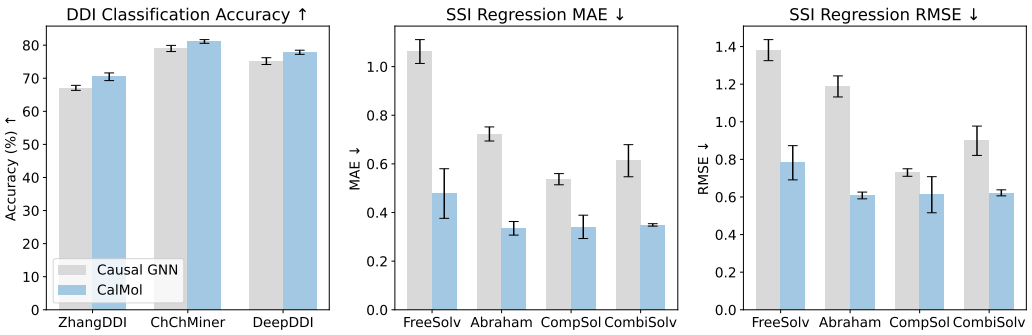

Figure 1: Ablation study.

The ablation studies we conducted are shown in Figure 1. A detailed evaluation was performed to compare our approach with the standalone use of Causal Molecule Motif-interaction Disentangling module proposed in Section **??** across diverse task settings and datasets. It is evident that full **CalMol** excels in the context of both DDI classification and SSI regression tasks on a wide range of datasets. The most striking difference can be seen in quantitative SSI tasks, where its superiority are most pronounced.

REFERENCES

Andreea Deac, Yu-Hsiang Huang, Petar Veličković, Pietro Liò, and Jian Tang. Drug-drug adverse effect prediction with graph co-attention. *arXiv preprint arXiv:1905.00534*, 2019.

Junfeng Fang, Shuai Zhang, Chang Wu, Zhengyi Yang, Zhiyuan Liu, Sihang Li, Kun Wang, Wenjie Du, and Xiang Wang. Moltc: Towards molecular relational modeling in language models. *arXiv preprint arXiv:2402.03781*, 2024.

Laura M Grubbs, Mariam Saifullah, E Nohelli, Shulin Ye, Sai S Achi, William E Acree Jr, and Michael H Abraham. Mathematical correlations for describing solute transfer into functionalized alkane solvents containing hydroxyl, ether, ester or ketone solvents. *Fluid phase equilibria*, 298 (1):48–53, 2010.

Namkyeong Lee, Dongmin Hyun, Gyoung S Na, Sungwon Kim, Junseok Lee, and Chanyoung Park. Conditional graph information bottleneck for molecular relational learning. In *International Conference on Machine Learning*, pp. 18852–18871. PMLR, 2023a.

Namkyeong Lee, Kanghoon Yoon, Gyoung S Na, Sein Kim, and Chanyoung Park. Shift-robust molecular relational learning with causal substructure. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1200–1212, 2023b.

Zimeng Li, Shichao Zhu, Bin Shao, Xiangxiang Zeng, Tong Wang, and Tie-Yan Liu. Dsn-ddi: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Briefings in Bioinformatics*, 24(1):bbac597, 2023.

Sagar Maheshwari Marinka Zitnik, Rok Sosič and Jure Leskovec. BioSNAP Datasets: Stanford biomedical network dataset collection. `http://snap.stanford.edu/biodata`, August 2018.

David L Mobley and J Peter Guthrie. Freesolv: a database of experimental and calculated hydration free energies, with input files. *Journal of computer-aided molecular design*, 28:711–720, 2014.

Edouard Moine, Romain Privat, Baptiste Sirjean, and Jean-Noël Jaubert. Estimation of solvation quantities from experimental thermodynamic data: Development of the comprehensive compsol databank for pure and mixed solutes. *Journal of Physical and Chemical Reference Data*, 46(3), 2017.

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

Yashaswi Pathak, Siddhartha Laghuvarapu, Sarvesh Mehta, and U Deva Priyakumar. Chemically interpretable graph interaction network for prediction of pharmacokinetic properties of drug-like molecules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 873–880, 2020.

Jae Yong Ryu, Hyun Uk Kim, and Sang Yup Lee. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national academy of sciences*, 115(18):E4304–E4311, 2018.

Florence H Vermeire and William H Green. Transfer learning for solvation free energies: From quantum chemistry to experiments. *Chemical Engineering Journal*, 418:129307, 2021.

Wen Zhang, Yanlin Chen, Feng Liu, Fei Luo, Gang Tian, and Xiaohong Li. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC bioinformatics*, 18:1–12, 2017.