Towards Efficient Fusion for Graph Neural Networks

Oleksii Tsepa
Department of Computer Science
University of Toronto
oleksii.tsepa@mail.utoronto.ca

Bohdan Naida
Department of Computer Science
University of Toronto
bohdan.naida@mail.utoronto.ca

Abstract

Humans experience the world around them from multiple sensory modalities. Since real-world data is quite often multimodal, understanding of different modalities is meaningful for ML models. Although both Multimodal deep learning and Graph Machine Learning have significantly developed in recent years, multimodal learning with graphs was not been fully explored. In this paper, we transfer the idea of fusion to graphs and create a universal framework for fusion among multiple entities. Since interaction between a few graphs commonly occurs in Drug Discovery, we evaluate our framework on drug pair scoring tasks (drug synergy prediction and drug-drug interaction). Our method reached state-of-the-art results on the O'Neil dataset on Bliss, Loewe, HSA, and ZIP synergy scores with much fewer parameters in comparison with the previous solutions.

1 Introduction

Multimodal sensations greatly influence human perception. A human uses vision, hearing, touch, taste, and smell to gather complete information about the surrounding world. Multimodal Deep Learning has made a lot of progress influenced by how humans experience the world around them. Even though Graph Machine Learning is one of the most innovative domains in DL which pushes forward a lot of scientific discoveries, including Drug Discovery, a lot of multimodal techniques have not been created or discovered in this domain yet. In this paper, we want to transfer the idea of fusion to graphs. In tasks that require a pair (or multiple) of graphs on input, it is highly important to effectively utilize the information about each graph separately and their interaction as well.

We have tested our approach on drug synergy and drug-drug interaction prediction tasks, which play a crucial role in drug discovery. Nowadays, drug combination therapy is widely applied all around the world due to its various advantages. In comparison to monotherapy, a combination treatment can lead to improved efficacy. Drug combinations may reduce the side effects, reduce toxicity, and overcome drug resistance. Multi-drug therapy has the potential to treat various complex diseases such as cancer or human immunodeficiency virus (HIV). Nevertheless, some drug combinations can have unfavourable or even harmful effects. Consequently, it is vital to accurately predict synergetic drug pairs and possible side effects between pairs of drugs.

2 Related works

Fusion stages. [1] introduced the idea of early, mid, and late fusion. Early fusion implies that information from different modalities is fused from the first encoder layer (all encoding layers are cross-modal). Late fusion - modalities are encoded separately, and no information is shared until the predictive head (all encoding layers are unimodal). While the mid-fusion allows extracting unimodal patterns first, and later layers encode cross-modal patterns. [1] showed that the mid-fusion improved accuracy metrics compared to early and late fusion.

Preprint. Under review.

Fusion on graphs. DDS [2], DeepDDS [3], and MR-GNN [4] encode graphs separately (late fusion), concatenate their representations and pass them to MLP. DSN-DDI [5], MHCADDI [6] performed both inter-graph and graph-graph message passing on the early-fusion stage. GMPNN-CS [7], DSN-DDI [5], SA-DDI [8] extract substructure information from different MPNN layers and measures their relevance with the co-attention mechanism.

3 Dataset

Our approach was evaluated on OncoPolyPharmacology [9] (also known as the O'Neil dataset). To follow the evaluation strategy of SDCNet [10], we used a subset of the dataset containing 31 cell lines and 38 drugs. The synergy score between pairs of drugs conditionally on a cell line was calculated using four different methods: Loewe additivity (Loewe) [11], bliss independence (Bliss) [12], zero interaction potency (ZIP) [13] and highest single agent (HSA) [14]. To stratify the synergy scores into synergistic (positive) and antagonistic (negative) samples, we used the preset thresholds as in [10]. Specifically, the thresholds are (0, 10) for Loewe, (-3.37, 3.68) for the Bliss score, (-3.02, 3.87) for the HSA score, and (-4.48, 2.64) for the ZIP score.

The dataset provided SMILES [15] representations of drugs - a form of a textual notation that specifies the molecular properties (atoms, chemical bonds, etc.). SMILES were derived from DrugBank [16], which we further converted to graphs. Nodes in graphs were atoms and edges - bonds. We used only atomic numbers as node features and bond types as edge features.

4 Method

In this section, we introduce the general framework for graph fusion. First, we show the motivation of the research. Next, we list the layer update equations. Finally, we present its characteristics in terms of modularity.

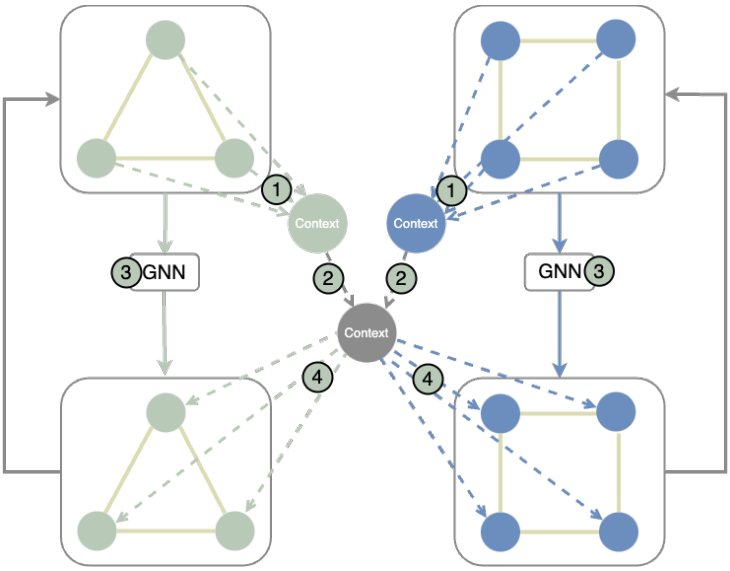

Figure 1: **Fusion part of the framework.** First, the message is propagated from graph nodes to the local contexts. Second, the local contexts are averaged to form the global context. Third, the MPNN encodes each graph separately. Fourth, the global context is propagated back to nodes.

4.1 Motivation

[1] introduced the idea of early, mid, and late fusion. Early fusion implies that information from different modalities is fused from the first encoder layer (all encoding layers are cross-modal). Late fusion - modalities are encoded separately, and no information is shared until the predictive head (all encoding layers are unimodal). While the mid-fusion allows extracting unimodal patterns first, and

later layers encode cross-modal patterns. [1] showed that the mid-fusion improved accuracy metrics compared to early and late fusion. We believe that late fusion does not fully capture cross-modal relations, while early fusion does not have enough capability to learn about each separate modality. In drug synergy predictions, drugs in combination have common effects and interact. Consequently, we believe that sharing information in the encoding part is beneficial. However, in the domain, most approaches like [2], [3], [4] are late-fusion-based, while only [5] uses early fusion. To the best of our knowledge, we are the first to implement middle fusion on graphs and investigate its effect on model performance.

4.2 Framework structure

In this paper, we consider the case of operating on pairs of graphs, but the framework can be easily extended to multiple graphs. Further, we will use the notation of context - an optional feature set associated with a pair of graphs used for the predictions and the context node - the node with context features. The first F_L layers of L-layer Message-Passing Neural Networks (MPNN) encode each graph separately (all MPNNs are applied with shared weights to both graphs). From F_L to the $L-1$ layer, fusion is performed between a pair of graphs (Figure 1). See Algorithm 1 below for the pseudo-code of our proposed framework.

Algorithm 1: Fusion algorithm for a pair of graphs

Input: Pair of graphs G_A, G_B with N_A, N_B nodes, node features X_A, X_B and E_A, E_B edges, respectively; Context node C with features X_C; Graph G_A, G_B to C edges E_{C2A}, E_{C2B}, and vice versa E_{A2C}, E_{B2C}; Message passing model instance $\texttt{MPNN}_G$ for node to node update in G_A, G_B; Message passing model instances $\texttt{MPNN}_{N2C}$, $\texttt{MPNN}_{C2N}$ for node to context and context to node updates; Layer $l \in [0; L-1]$; Layer L_F in which fusion is injected.

Output: node representations $X_A^L \in R^{N_A \times D}$ and $X_B^L \in R^{N_B \times D}$ of G_A and G_B, respectively.

for $l = 0, 1, ..., L_F - 1$ **do**
$X_A^{l+1} \leftarrow \texttt{MPNN}_G(X_A^l, E_A)$
$X_B^{l+1} \leftarrow \texttt{MPNN}_G(X_B^l, E_B)$

if $!\exists X_C$ **then**
$X_C \leftarrow (\texttt{Pool}(X_A) + \texttt{Pool}(X_B))/2$

for $l = L_F, L_F + 1..., L - 1$ **do**
$X_{C_A}^{l+1} \leftarrow \texttt{MPNN}_{N2C}((X_C^l, X_A^l), E_{C2A})$
$X_{C_B}^{l+1} \leftarrow \texttt{MPNN}_{N2C}((X_C^l, X_B^l), E_{C2B})$
$X_C^{l+1} \leftarrow (X_{C_A}^{l+1} + X_{C_B}^{l+1})/2$
$X_A^{l+1} \leftarrow \texttt{MPNN}_G(X_A^l, E_A)$
$X_B^{l+1} \leftarrow \texttt{MPNN}_G(X_B^l, E_B)$
$\hat{X}_A^{l+1} \leftarrow \texttt{MPNN}_{C2N}((X_A^{l+1}, X_C^{l+1}), E_{A2C})$
$\hat{X}_B^{l+1} \leftarrow \texttt{MPNN}_{C2N}((X_B^{l+1}, X_C^{l+1}), E_{B2C})$
$X_A^{l+1} \leftarrow \hat{X}_A^{l+1} + X_A^{l+1}$
$X_B^{l+1} \leftarrow \hat{X}_B^{l+1} + X_B^{l+1}$

return $X_A^L \in R^{N_A \times D}$, $X_B^L \in R^{N_B \times D}$

4.3 Customizability

Customizability is achieved by the ability to modify the modules, including node-to-node-MPNN, Graph Pooling, context-to-node-MPNN, and node-to-context-MPNN. The framework provides the

following hyperparameters: the number of total node-to-node-MPNN layers and the injection layer, starting from which the fusion is injected.

5 Results

We compare our method versus foundational dense neural network models for Drug Synergy prediction (Deep Synergy [17], Matchmaker [18]), and recent GNN-based models with SOTA performance (DeepDDS [3], DTF [19], Jiang's method [20], and SDCNet [10]). We used 10-fold Stratified Cross Validation with a training/validation ratio of 90/10. Our method outperformed other approaches on the O'Neil dataset on Loewe, Bliss, and ZIP synergy scores, while on HSA, it performs as well as the previous SOTA. We trained our model with GINEConv [21] as node-to-node MPNN with 7 layers and fusion after the 5^{th} layers, SAGEConv [22] as context-to-node MPNN, and GATConv [23] as a node-to-context MPNN.

Method	Loewe		Bliss		ZIP		HSA	
	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC
Deep Synergy	0.80 ±0.01	0.78 ±0.01	0.87 ±0.01	0.91 ±0.01	0.88 ±0.01	0.94 ±0.01	0.81 ±0.02	0.95 ±0.01
MatchMaker	0.81 ±0.01	0.79 ±0.01	0.88 ±0.01	0.91 ±0.01	0.89 ±0.01	0.95 ±0.01	0.81 ±0.02	0.95 ±0.01
Jiang's method	0.86 ±0.01	0.86 ±0.01	0.92 ±0.01	0.94 ±0.01	0.91 ±0.01	0.96 ±0.01	0.90 ±0.01	0.96 ±0.01
DeepDDS	0.89 ±0.01	0.86 ±0.01	0.92 ±0.01	0.93 ±0.01	0.92 ±0.01	0.95 ±0.01	0.89 ±0.01	0.96 ±0.01
DTF	0.91 ±0.01	0.90 ±0.01	0.94 ±0.01	0.96 ±0.01	0.94 ±0.01	0.97 ±0.01	0.92 ±0.01	0.97 ±0.01
SDCNet	0.93 ±0.01	0.92 ±0.01	0.96 ±0.01	0.97 ±0.01	0.95 ±0.01	0.98 ±0.01	**0.94 ±0.01**	0.98 ±0.01
Ours	**0.94 ±0.01**	**0.93 ±0.01**	**0.97 ±0.00**	**0.98 ±0.00**	**0.96 ±0.03**	**0.99 ±0.00**	**0.94 ±0.01**	**0.99 ±0.00**

Table 1: Model evaluation scores on the O'Neil dataset

References

[1] Arsha Nagrani, Shan Yang, Anurag Arnab, Aren Jansen, Cordelia Schmid, and Chen Sun. Attention bottlenecks for multimodal fusion. *Advances in Neural Information Processing Systems*, 34:14200–14213, 2021.

[2] Kyriakos Schwarz, Alicia Pliego-Mendieta, Lara Planas-Paz, Chantal Pauli, Ahmed Allam, and Michael Krauthammer. Ddos: A graph neural network based drug synergy prediction algorithm. *arXiv preprint arXiv:2210.00802*, 2022.

[3] Jinxian Wang, Xuejun Liu, Siyuan Shen, Lei Deng, and Hui Liu. Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics*, 23(1):bbab390, 2022.

[4] Nuo Xu, Pinghui Wang, Long Chen, Jing Tao, and Junzhou Zhao. Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions. *arXiv preprint arXiv:1905.09558*, 2019.

[5] 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), 2023.

[6] Yunsheng Bai, Ken Gu, Yizhou Sun, and Wei Wang. Bi-level graph neural networks for drug-drug interaction prediction. *arXiv preprint arXiv:2006.14002*, 2020.

[7] 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):bbab441, 2022.

[8] Shenggeng Lin, Yanjing Wang, Lingfeng Zhang, Yanyi Chu, Yatong Liu, Yitian Fang, Mingming Jiang, Qiankun Wang, Bowen Zhao, Yi Xiong, et al. 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*, 23(1):bbab421, 2022.

[9] Jennifer O'Neil, Yair Benita, Igor Feldman, Melissa Chenard, Brian Roberts, Yaping Liu, Jing Li, Astrid Kral, Serguei Lejnine, Andrey Loboda, et al. An unbiased oncology compound screen to identify novel combination strategies. *Molecular cancer therapeutics*, 15(6):1155–1162, 2016.

[10] Peng Zhang, Shikui Tu, Wen Zhang, and Lei Xu. Predicting cell line-specific synergistic drug combinations through a relational graph convolutional network with attention mechanism. *Briefings in Bioinformatics*, 23(6), 2022.

[11] S Loewe. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung*, 3:285–290, 1953.

[12] Chester I Bliss. The toxicity of poisons applied jointly 1. *Annals of applied biology*, 26(3): 585–615, 1939.

[13] Bhagwan Yadav, Krister Wennerberg, Tero Aittokallio, and Jing Tang. Searching for drug synergy in complex dose–response landscapes using an interaction potency model. *Computational and structural biotechnology journal*, 13:504–513, 2015.

[14] Xu Tan, Long Hu, Lovelace J Luquette III, Geng Gao, Yifang Liu, Hongjing Qu, Ruibin Xi, Zhi John Lu, Peter J Park, and Stephen J Elledge. Systematic identification of synergistic drug pairs targeting hiv. *Nature biotechnology*, 30(11):1125–1130, 2012.

[15] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28 (1):31–36, 1988.

[16] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research*, 46(D1):D1074–D1082, 2018.

[17] Kristina Preuer, Richard PI Lewis, Sepp Hochreiter, Andreas Bender, Krishna C Bulusu, and Günter Klambauer. Deepsynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics*, 34(9):1538–1546, 2018.

[18] Halil Ibrahim Kuru, Oznur Tastan, and A Ercument Cicek. Matchmaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(4):2334–2344, 2021.

[19] Zexuan Sun, Shujun Huang, Peiran Jiang, and Pingzhao Hu. Dtf: deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics*, 36(16):4483–4489, 2020.

[20] Peiran Jiang, Shujun Huang, Zhenyuan Fu, Zexuan Sun, Ted M Lakowski, and Pingzhao Hu. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Computational and structural biotechnology journal*, 18:427–438, 2020.

[21] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265*, 2019.

[22] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[23] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.