ADAPTIVE DISTANCE MESSAGE PASSING FROM THE MULTI-RELATIONAL EDGE VIEW

Anonymous authors
Paper under double-blind review

ABSTRACT

Message-passing based graph neural network is a leading prominent paradigm for performing deep learning tasks on graphs. Despite its success in various tasks, it is limited by passing information to the fixed length k distance neighbouring nodes, where k is the number of layers. In reality, different types of edges (alternatively relations) may influence nodes at different lengths and should not be treated universally. This paper addresses the problem by considering the relations that edges represent accordingly and uncovering the roles of the types of edges in producing better results from adaptive distance message passing. Experiments on real-world datasets validate the effectiveness of our method.

1 INTRODUCTION

Message-passing based graph neural networks (MP-GNNs) have proven to be effective for graph representation learning (Sanchez-Lengeling et al., 2021), which aggregate the information from the neighbouring nodes iteratively and update the representation of nodes based on passed information and their previous states (Gilmer et al., 2017).

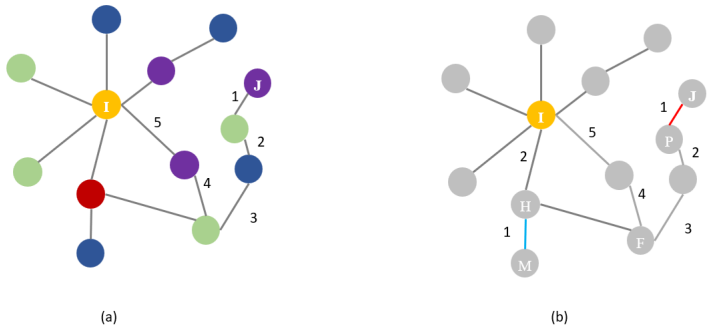

Figure 1: Local aggregation

However, MP-GNNs are limited by local message passing (He et al., 2022). Specifically, if information from $k-hop$ neighbouring nodes needs to be aggregated, the neural network must stack k layers, which we named it as *uniform length message passing*. As illustrated in Figure 1 (a), we need five layers to pass information from node J to node I, so that all nodes in the graph can capture information from $5-hop$ neighbouring nodes. However, MP-GNN loses the generalisation ability after stacking more than four layers, and overly compressed information leads to over-squashing (Alon & Yahav, 2020). Although advanced methods such as non-local update mechanisms(Balcilar et al., 2021) or transformers are recently applied with MP-GNNs to capture long-range dependencies in the graph(Min et al., 2022), they always focus more on the influence of nodes instead of multiple types of edges over nodes, which is not optimal for tasks like molecule learning, where different types of relations exist. Other studies (Vashishth et al., 2019; Zheng et al., 2022) investigated message passing on multi-relational graphs but still were limited to *uniform length message passing*.

Addressing the above problems, we propose a novel method, an adaptive distance message passing from edge view (ADMPE), where edges with significant impacts can be passed to distant nodes rather than fixed $k - hop$ neighbouring nodes in *uniform length message passing*. As shown in Figure 1 (b), the red colour edge between source node J and target node P can be passed to five hop node I while the blue colour edge connected with source node M and target node H will be passed to two hop nodes I and F.

2 METHODOLOGY

We investigate the edge attributes and define how far the information of edges can be passed to nodes. For instance, there are four types of edges (0,1,2,3) between atoms (nodes) in the ZINC dataset representing *no bond*, *single*, *double* and *aromatic bonds*, respectively. Double bonds are more reactive than single bonds in chemical reactions, and aromatic bonds are more stable than single or double bonds, leading to increased electron density. Our mechanism allows the 1 type of edges to be passed to $k - hop$ where k is the number of layers, and 3 type of edges can be passed to $k + n - hop$ of nodes. n is a learnable parameter. We follow the standard MP-GNN steps (Gilmer et al., 2017): initialisation, aggregation, update and readout, but change the aggregation to include features of different types of edges to nodes at different hops as follows:

$$\mathbf{h}_i' = f_u(\mathbf{h}_i, \sum_{j \in N(i)} \mathbf{W}_1 \mathbf{x}_j \sum_{i \in M(r)} \mathbf{W}_2 \mathbf{e}_{jp}^r) \tag{1}$$

$\mathbf{h}_i'$ and $\mathbf{h}_i$ represent the updated representation and previous state of node i, respectively. $\mathbf{x}_j$ denotes the initial features for neighbour node j and $\mathbf{e}_r$ represents edge feature of relation r. f_u is the update function and $\mathbf{W}_1$, $\mathbf{W}_2$ are learnable matrices. $N(i)$ denotes the neighbours of node i and $M(r)$ is a hop count set indicating that node i is within the maximum hop length $k + n - hop$ of a type of r edge which connected with node j and p, as shown in Figure 1 (b).

3 EXPERIMENT

Datasets We consider two benchmark datasets ZINC (12K) (Dwivedi et al., 2020)and MolHIV (Hu et al., 2020) for molecule learning tasks, details shown in Table 2 in Appendix.

Baselines and Results. We compare two methods, GPS(Rampášek et al., 2022), state of the art on Zinc (12K) dataset and Graph MLP-Mixer(He et al., 2022) state of the art for MolHIV, with our method ADMPE. We keep the same implementations as the baselines and integrate them with the updated function in equation 1. Table 1 summarises the experimental results and proves the effectiveness of our method.

Table 1: Comparison of our model to baseline methods

Methods	ZINC	MolHIV
Metric	MAE	ROCAUC
GPS (Rampášek et al., 2022)	0.070	0.781
Graph MLP-Mixer	0.075	0.807
GPS with ADMPE	**0.069**	0.785
Graph MLP-Mixer with ADMPE	0.073	**0.809**

4 CONCLUSION

This paper introduces a novel adaptive distance message-passing method from the point of the multi-relational edge. Specifically, to address the limitations of *uniform length message passing*, ADMPE proposes a relation-aware rule to pass edges information to nodes at different distances. ADMPE can be easily integrated with existing work and enhance the representation learning ability.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. *arXiv preprint arXiv:2006.05205*, 2020.

Muhammet Balcilar, Pierre Héroux, Benoit Gauzere, Pascal Vasseur, Sébastien Adam, and Paul Honeine. Breaking the limits of message passing graph neural networks. In *International Conference on Machine Learning*, pp. 599–608. PMLR, 2021.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *arXiv preprint arXiv:2003.00982*, 2020.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017.

Xiaoxin He, Bryan Hooi, Thomas Laurent, Adam Perold, Yann LeCun, and Xavier Bresson. A generalization of vit/mlp-mixer to graphs. *arXiv preprint arXiv:2212.13350*, 2022.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.

Erxue Min, Runfa Chen, Yatao Bian, Tingyang Xu, Kangfei Zhao, Wenbing Huang, Peilin Zhao, Junzhou Huang, Sophia Ananiadou, and Yu Rong. Transformer for graphs: An overview from architecture perspective. *arXiv preprint arXiv:2202.08455*, 2022.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *arXiv preprint arXiv:2205.12454*, 2022.

Benjamin Sanchez-Lengeling, Emily Reif, Adam Pearce, and Alexander B Wiltschko. A gentle introduction to graph neural networks. *Distill*, 6(9):e33, 2021.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. Composition-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1911.03082*, 2019.

Xin Zheng, Miao Zhang, Chunyang Chen, Chaojie Li, Chuan Zhou, and Shirui Pan. Multi-relational graph neural architecture search with fine-grained message passing. In *2022 IEEE International Conference on Data Mining (ICDM)*, pp. 783–792. IEEE, 2022.

A APPENDIX

A.1 DATASETS

ZINC (12K) dataset consists of 28 types of nodes (atom types) and 4 types of edges (but type 0 represents *no bond* between atoms, so only 3 types of actual edges in the molecule graph). For Molhiv, edges represent chemical bonds, and nodes refer to atoms. The default setting of Molhiv in the Graph MLP-Mixer model is no types of nodes and edges. However, when loading the Molhiv dataset, the edge attribute still has four types of values (0,1,2,3), and we assign the same adaptive distance passing rule as for the ZINC dataset in the experiments. Then we apply the node and edge linear embeddings (hidden dimension is 128) to the raw input, same as (He et al., 2022).

Table 2: Dataset summary

Datasets	Graphs	Nodes	Avg Nodes	Avg Edges	Node Types	Edge Types	Class
ZINC	12000	9-37	23.2	24.9	28	4	Regression (1)
MolHIV	41127	2-222	25.5	54.9	None	None	2

A.2 EXPERIMENTAL SETTINGS.

Our model is implemented with Pytorch 1.12.1 and CUDA 11.3. We keep the same hyper-parameters setting as baselines Graph MLP-Mixer (He et al., 2022) and GPS Rampášek et al. (2022). The whole model is implemented on PyTorch Geometric.