A BRIEF TECHNICAL REPORT OF HFAGNN

A PREPRINT

Can Xu
Research Center for Graph Computing
Zhejiang Lab
Hangzhou, Zhejiang Province
leoxc1571@163.com

Hu Chen
Research Center for Graph Computing
Zhejiang Lab
Hangzhou, Zhejiang Province
huchen@mail.sdu.edu.cn

Ruifeng Li
Research Center for Graph Computing
Zhejiang Lab
Hangzhou, Zhejiang Province
lirf@zju.edu.cn

Jinsong Lv
Research Center for Graph Computing
Zhejiang Lab
Hangzhou, Zhejiang Province
jingsonglv@zhejianglab.com

October 31, 2022

ABSTRACT

To predict molecule properties, we proposed *Hybrid Feature-Aware Graph Neural Network* (HFAGNN), a model that combine 2D topology structure and 3D conformation information into the message passing scheme in graph neural networks. In this technical report, model structure and training details of HFAGNN on PCQM4Mv2 dataset are reported.

Keywords Molecular Graph · Graph Neural Networks · HOMO-LUMO Gap

1 Data preprocessing

In this challenge, the purpose is to predict the homo-lomo gap for each 2D molecule. The PCQM4M-v2 dataset consists of over three million molecules. Additional 3D structure are provided for molecules in the training set.

In the proposed model, we make use of the 3D information provided in the training set. In terms of those molecules in valid and test set without 3D conformation, we adopt rdkit.Chem.AllChem.MMFFOptimizeMolecule() to generate 3D conformations. However, this method does not guarantee 100% success rate when it comes to complex molecules. In this scenario, we generate the 2D position alternatively. In a molecule, we are able to extract 9 types of features for each atom and 3 for each bond.

2 Model structure

In the proposed model, we leverage both spatial information and topology structure of a given molecule. For spatial information extraction, we inherit the design of Wang et al. [2022] for 2-hop and 3-hop information extraction. Besides, we implement GIN-Virtual Gilmer et al. [2017] for topology structure learning.

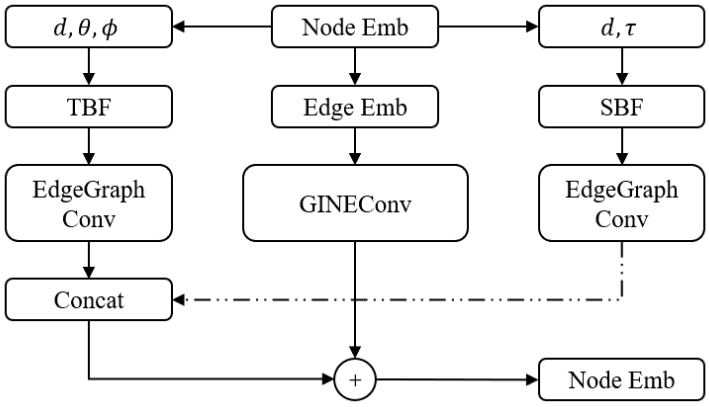

Figure 1: Structure of Hybrid Block

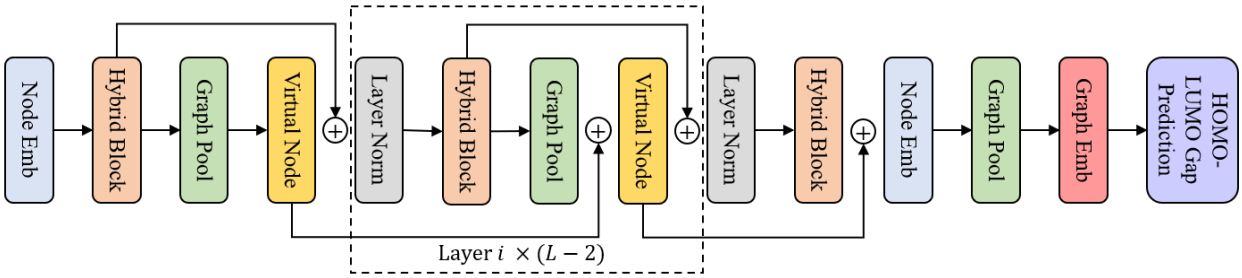

Figure 2: Structure of HFAGNN

3 Result

We train our models with one NVIDIA Titan V GPU (12GB memory) and Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz (256GB memory). The model is trained with the given train subset and validated with the given valid subset. Hyperparameter settings are shown in Table 1. Training one model with one seed on a single GPU takes 25 hours. Inference with one model and one seed on the testdev takes 1 hour. Experiment results are shown in Table 2.

Table 1: Hyperparameter Settings

Learning Rate	Number of layers	Hidden Size	Dropout Ratio
0.0005	6	256	0.1

Table 2: Experiment Result

Model	#Param	Valid MAE
HFAGNN	3949959	0.1005

References

Limei Wang, Yi Liu, Yuchao Lin, Haoran Liu, and Shuiwang Ji. ComENet: Towards Complete and Efficient Message Passing for 3D Molecular Graphs, June 2022. URL `http://arxiv.org/abs/2206.08515`. arXiv:2206.08515 [cs].

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 06–11 Aug 2017. URL `https://proceedings.mlr.press/v70/gilmer17a.html`.