Simple is the Best: Pooling Methods for Quantum Chemical Property Prediction

Mingwei Ge
Department of Electrical Engineering
Yale University
New Haven, CT 06511
mingwei.ge@yale.edu

Abstract

Pooling method in graph neural networks (GNN) is crucial for graph level tasks including quantum chemical property prediction. In this work, 4 pooling approaches ranging from global pooling to hierarchical pooling with learned or precomputed cluster matrix are evaluated on molecular orbital prediction tasks. RingPool, proposed in this work for the first time, showed a comparable performance to DiffPool. Surprisingly, the global sum pool has the best performance in light of the quantum chemical nature alignment in such tasks.

1 Introduction

Molecular property prediction, an extremely important task in drug discovery and materials design, is experiencing a huge improvement in light of the advancement of GNNs in recent years [1]. Among them, quantum mechanical property is one of the most pressing issue, since they contains rich information about the reactivity, optical and electrical property, but classically, precise calculations require highly time-consuming computations at density functional theory (DFT, $O(N^3)$) or post-Hatree-Fock ($O(N^{5-7})$) level [2].

Generally, pooling in GNN aggregates node embeddings to a graph-level latent vector, usually locates in the middle of message passing layers and the read-out MLP. Plenty of pooling methods are introduced and widely applied. They can be done either globally or hierarchically. The former aggregates all node at once by sum, mean or max operations, while later first pools it to a coarse-grained graph, do more message passing and globally pool it finally. It is also proved that the choice of pooling layer play a important role in GNN and it should be task-specific to reach the highest performance, since the inappropriate choice would aggravate over-smoothing and loss expressiveness [3, 4]. However, there are neither systematic evaluation on pooling for quantum chemical tasks, nor task-specific pooling for such challenging problem.

Hence in this work, I made the following contributions to such issue:

- Proposed and implemented a hierarchical pooling based on pre-computed clusters for molecules graph, RingPool, for the first time.
- Systematically evaluated four pooling methods on quantum chemical property prediction, ranging from global pooling (sum and mean) to hierarchical pooling (DiffPool and self-designed RingPool).
- In-deep analysis of pooling methods based on principle and result of ab-initio quantum chemical calculations.

RingPool, showed a comparable performance to DiffPool and global mean pooling on PM9 and OPV dataset. Surprisingly, global sum pool has the best performent because the LUMO/HOMO

Preprint. Under review.

wavefunctions consist of unconnected small local sections. Global sum pool GNN structure aligns with the process of ab-initio quantum chemical calculation for these values.

2 Related Works

Pooling of GNNs. A wide variety of pooling methods have been introduced in recent years. Simple Sum and mean pooling have been widely applied in or beyond molecular property prediction. Advanced approaches have also been proposed ranging from model-free methods that pre-compute the cluster to form coarse-gained graph, to model-based methods that perform pooling through a learnable function of the node features[5]. For example, some method pre-compute pooled graphs by spectral graph theory [6, 7]. Set2set approach employs a long short-term memory (LSTM) architecture designed for unordered and size-variant input sets [8]. DiffPool leverages another GNNs to compute the cluster assignment matrix that progressively coarsens the input graph in each layer until a single graph representation is obtained. Graph U-Nets employed gPool layer that adaptively selects some nodes to form a smaller graph based on their scalar projection values on a trainable projection vector[9]. SortPool arranges the learned node representation in a consistently ordered tensor which is then truncated or extended to a user-defined fixed size [10]. There are also some works refined such methods to make them more computational efficient [11].

Molecular property prediction benchmarks Molecular property prediction have became an popular benchmark for evaluation the performance of GNN on graph regression and classification tasks for a long time, thanks to the graph-structured nature of molecules. For example, QM7 and QM9 contains properties of small molecules calculated by quantum chemical method for graph regression. FreeSolv, ESOL and Lipophilicity are physical chemistry dataset for regression. HIV, PDBbind and Tox21 also involve the property in biophysics or physiology. There are also multi-dataset benchmarks that systematically evaluate GNNs on graph-level supervised learning [12]. Most of these datasets have been already built within major GNN packages including PyTorch Geometric and TorchDrug.

In general, great advancement has been made in both pooling method and molecular property prediction tasks of GNN. However, in-deep task-specific study of pooling on quantum chemical property still remains rare. This is because most works either focus on generalization ability or include model with too much difference that hinders the comparison on their pooling. Besides, most GNNs for quantum chemistry are only evaluated on dataset (e.g. QM9) of molecules much smaller than the real-world ones with practical applications, which may hidden their serious artifacts on pooling.

3 Methods

3.1 Preliminaries

Task definition. I represent molecular graph G as (A, F) where $A \in \mathbb{R}^{n\times n}$ is the adjacency matrix, and $F \in \mathbb{R}^{n\times d}$ is the node feature matrix that each node has d features. Given a set of labeled graphs G, our goal is to learn a mapping $f : G \to y$ that maps graphs to the set of labels. In this work, I only focus on regression tasks where y are continuous values represent quantum chemical properties of molecules. The evaluation metrics is the mean absolute error (MAE) like other molecular property prediction benchmarks[1].

GNN architecture In this work, All GNNs begin with the general message-passing layer. Since we only interested in pooling part, I implemented message-passing by the classical Graph Convolutional Network (GCN) architecture with ReLU activation to simplify the problem:

$$H^{(k)} = M(A, H^{(k-1)}; W^{(k)}) = ReLU(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H^{(k-1)}W^{(k-1)})$$

where $H^{(k)} \in \mathbb{R}^{n\times d}$ are the node embeddings k steps of message-passing, $W^{(k)} \in \mathbb{R}^{d\times d}$ is a trainable weight matrix, $\widetilde{A} = A + I$ and $\widetilde{D}$ is the degree matrix based on $\widetilde{A}$. After the conducting message-passing on original molecular graph, different pooling approach transfer node embeddings to a finite dimensional vector in $\mathbb{R}^D$ (See details at Figure 1a and section 3.2). A MLP take the vector as input and output the label y.

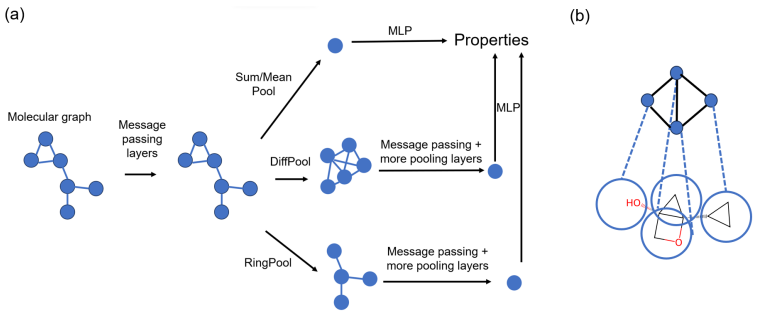

Figure 1: (a) Schematic diagram of the models implemented in this work. All 4 models begin with GCN layers, pooled by max, mean, RingPool and DiffPool respectively and readout the property by MLP. (b)Schematic diagram of the RingPool that implemented in this work for the first time.

3.2 Pooling layer choices

RingPool. As shown in Figure 1b, Because cycles or rings exist in most molecules and determined the conformation and various quantum chemical properties, the key idea of RingPool is to consider simple cycles and groups not on rings clusters (noted as $\mathcal{X}_|$) inspired by some molecular generation model like JT-VAE [13]. . Each row of $S^(l)$ corresponds to one of the n_l nodes (or clusters) at layer l, and each column of $S^(l)$ corresponds to one of the n_{l+1} clusters at the next layer $l+1$ The coarse-gained graph is precomputed before the training and constructed by adding edges between all intersecting clusters. The RingPool layer in GNN was implemented by the cluster matrix S inspired from DiffPool [3].

$$X^{(l+1)} = {S^{(l)}}^T Z^{(l)}, \; A^{(l+1)} = {S^{(l)}}^T A^{(l)} S^{(l)}$$

But Unlike in DiffPool, S is precomputed in RingPool by:

$$S_{ij} = \begin{cases} 1, & F_i \in \mathcal{X}_j \\ 0, & else \end{cases}$$

Although similar ideas are common in other works, RingPool in this work is the first implementation of it on pooling layers of GNN.

DiffPool. According to the original literature [3], cluster matrix S and is calculated by another GNN separate from the one for the node embeddings. The pooling is perform as previous.

$$S^{(l)} = Softmax(GNN_{pool}(A^{(l)}, X^{(l)}))$$

Max and mean pool. Max and mean pool simply apply max or mean over all node embeddings of the molecular graph as implemented in PyG.

4 Experiments

I evaluated 4 models with different pooling approaches on QM9 and OPV dataset on LUMO/HOMO energy prediction task, aiming at following questions:

- Performance difference of such pooling on these 2 datasets.
- How to interpret the result from quantum mechanical nature.

4.1 Datasets Statistics

Models evaluated on QM9 and OPV dataset on LUMO/HOMO energy prediction task, using the version from TorchDrug package. QM9 also contain coordinates of 3D conformations but I didn't make use of it to make it align with OPV. As shown in Figure 2a, QM9 contains 133,885 small molecules with maximum of 9 atoms while the OPV have 94,576 larger molecules with atom ranging from 10 to 122. This enables the comparison of the model on different level of graph complexity.

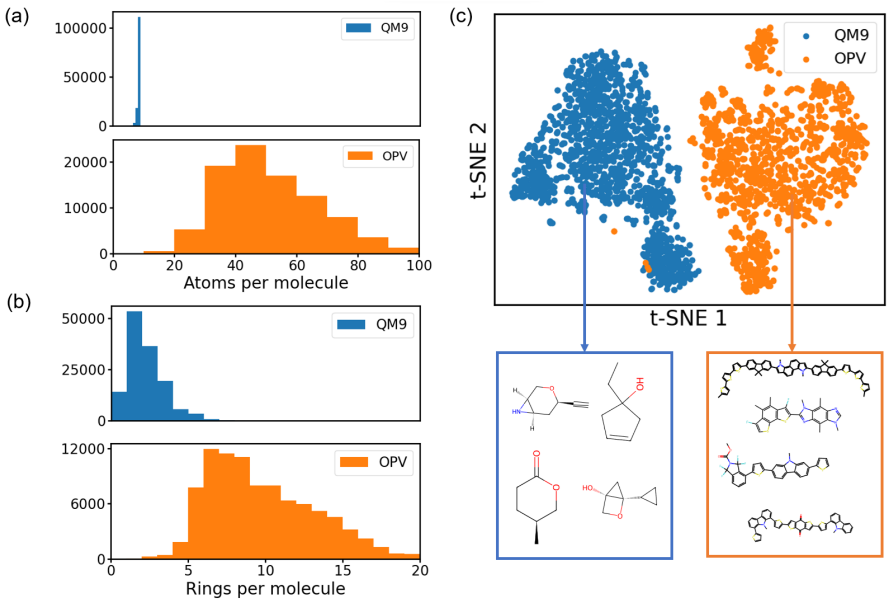

Figure 2: Datasets Statistics. (a) Number of atoms per molecules. (b) Number of simple rings per molecules. (c) t-SNE visualization based on Morgan fingerprint in chemical space.

Table 1: Test MAE on LUMO/HOMO energy. Lumo and HOMO MAE are averaged at equal weight. All Values are present in form of the mean of 3 experiments ± standard deviation.

Method	**Dataset**	
	QM9	OPV
Mean Pool	0.0078 ± 0.0003	0.124 ± 0.008
Sum Pool	**0.0076 ± 0.0002**	**0.093 ± 0.007**
RingPool	0.0079 ± 0.0008	0.137 ± 0.004
DiffPool	0.0080 ± 0.0001	0.130 ± 0.007

The t-SNE visualization based on Morgan fingerprints (Figure 2c) confirms that molecules from this 2 datasets are at different location in the chemical space. As shown in Figure 2b, molecules in both datasets have considerable amount of rings, which may suitable for RingPool.

4.2 Training Details

Mean and sum pool have a 4-layer GCN and a 2-layer MLP. RingPool and DiffPool start with a 2-layer GCN , followed by another 2-layer GCN, a global sum pool, and the 2-layer MLP. The hidden size of all model are set to 32 to make the total size of weight comparable to each other. The output size is 2 correspond to the LUMO and HOMO energy respectively. Since LUMO and HOMO energy are equally important continuous values at same magnitude, I used the L2 loss with equal weight on them. Models were optimized using Adam optimizer with learning rate of 0.01 and batch size of 128 for the first 50 epochs, and reduced to 0.005 and 32 for rest epochs. The datasets were randomly separated into train, validation and test set with ratio of 0.8:0.1:0.1. All experiments are conducted 3 times with different random seeds (1222,324 and 10086). For DiffPool, the column number of S is setted as 0.25 times the max nodes (same as the original literature).

4.3 Results

The average MAE of the 4 models on these 2 dataset are shown in Table 1. Surprisingly, The simple sum pool reached the lowest MEA at both QM9 and OPV dataset. Learning curve of such experiments are shown in Figure 3. For QM9 dataset with small molecular graphs, 4 models have comparable validation MAE trend and sum pool exhibit a slightly better performance. But for OPV dataset with larger molecular graphs, the sum pool became obviously outstanding while the rest 3 models have

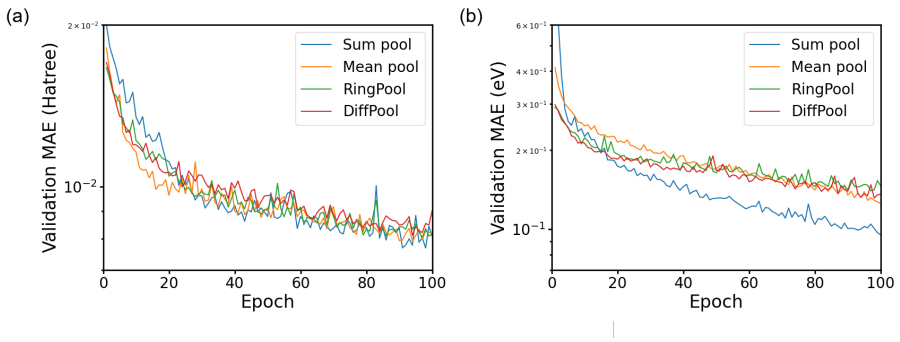

Figure 3: The MAE (averaged over 3 experiments) on the validation set of QM9 (a) and OPV (b).

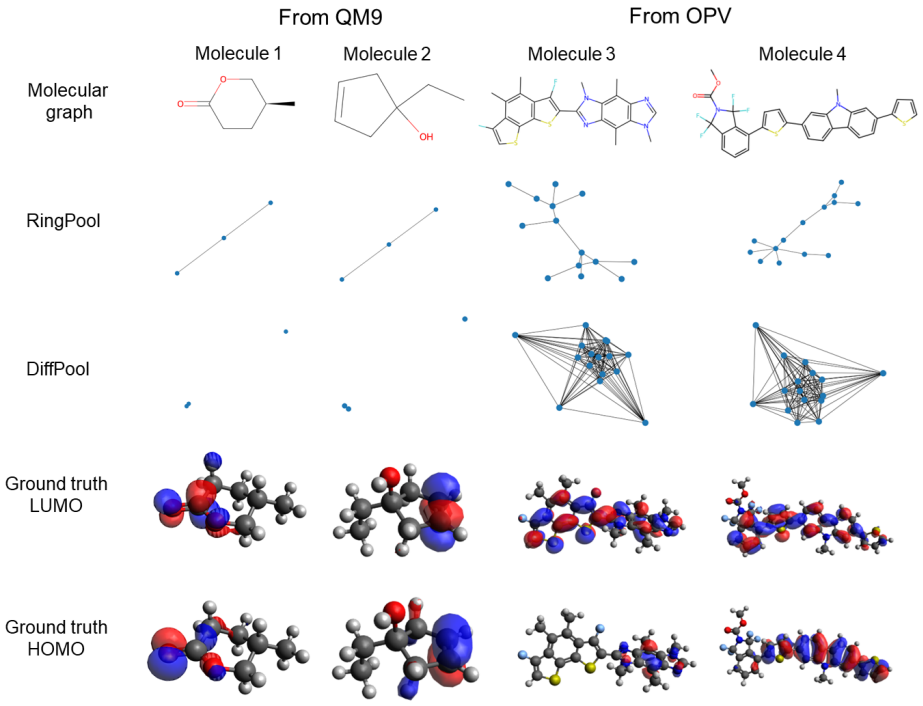

Figure 4: The visualization of 4 molecules on their molecular graphs, coarse-gained graph in DiffPool and RingPool, along with their ground truth LUMO/HOMO wavefunction plot calculated by DFT at B3LYP/6-31g(d) level.

similar performance. Hierarchically poolings with no matter precomputed (DiffPool) or learned clusters (RingPool) did not show their strength in such quantum chemistry graph regression task. RingPool also exhibits a comparable result to DiffPool, which indicate that it is possible to apply precomputed cluster in hierarchically poolings to reduce the complexity of the model without obvious performance reduction.

4.4 Interpretations

4 molecules from QM9 and OPV are chosen for analysis in order to get the answer why max pool works best and hierarchically poolings, generally considered more promising in preventing oversmoothing and bottle-neck effect fail in this task.

Failure of hierarchical pooling. As shown in Figure 4, the coarse-gained graphs in RingPool shows a line or tree-like morphology, since it aggregate all the rings. DiffPool, however, exhibits a highly dense graph in light of the learned S. In quantum chemistry, LUMO and HOMO energy are determined by their wavefunctions Ψ :$E_{HOMO} =< \Psi^*_{HOMO}|\hat{H}|\Psi_{HOMO} >$, $E_{HOMO} =< \Psi^*_{LUMO}|\hat{H}|\Psi_{LUMO} >$ where $\hat{H}$ is the Hamiltonian. Psi_{HOMO} and Psi_{LUMO} of 4 chosen molecules are visualized at the bottom of Figure 4. LUMO and HOMO are divided into a set of separate sections. Each section contains 1-3 atoms which are neighbors to each other, and there is nearly no connection among sections. RingPool aggregate rings but sections may locate between rings and one ring may contain multiple unconnected sections. DiffPool create dense coarse-gained graph. Redundant edge between non-interacting sections treat this problem over-globally and overwhelm the crucial local sections.

Advantages of global sum pooling. Since the physical nature that LUMO and HOMO consist of unconnected small local sections, few steps of simple one-hop message-passing over the original graph is enough to capture the contribution of local section to the target properties which were aggregated by global sum pooling to get the overall values. Besides, under the Hatree-Fock approximation, LUMO and HOMO wavefunctions can be represented by linear combinations of atomic orbitals: $\Psi = \sum_{i=1}^{n} c_i \Phi_i$ where i represent the node index in the molecular graph. Ab-initio quantum chemical calculation follow 3 steps to get the energy: perform self-consistent field (SCF) iterations to get c_i, sum over all $c_i\Phi_i$, and do integral with $\hat{H}$. Such process align with the global sum pooling GNN: do message passing to get single node contribution, sum over all nodes, and readout result via MLP. Global sum pooling enables the physical nature alignment, which makes its better performance compared to mean pool, DiffPool and RingPool.

5 Conclusion

Global sum pool has the expressive power on molecular orbital energy in light of the quantum chemical nature alignment. Task-specific pooling approach that is required for best expressiveness. Hierarchical pooling for quantum chemical task requires further development of approaches other than directly learned S or simply clustering over rings.

6 Reproducibility

All codes of the models, experiments, plotting, input files of the ab-initio calculations, and their dependencies are available at https://github.com/michaelge233/MoleculePooling

References

[1] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

[2] Peikun Zheng, Roman Zubatyuk, Wei Wu, Olexandr Isayev, and Pavlo O Dral. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nature communications*, 12(1):7022, 2021.

[3] Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. *Advances in neural information processing systems*, 31, 2018.

[4] Artur M Schweidtmann, Jan G Rittig, Jana M Weber, Martin Grohe, Manuel Dahmen, Kai Leonhard, and Alexander Mitsos. Physical pooling functions in graph neural networks for molecular property prediction. *Computers & Chemical Engineering*, 172:108202, 2023.

[5] Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. Spectral clustering with graph neural networks for graph pooling. In *International conference on machine learning*, pages 874–883. PMLR, 2020.

[6] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[7] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

[8] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391*, 2015.

[9] Hongyang Gao and Shuiwang Ji. Graph u-nets. In *international conference on machine learning*, pages 2083–2092. PMLR, 2019.

[10] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An end-to-end deep learning architecture for graph classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[11] Cătălina Cangea, Petar Veličković, Nikola Jovanović, Thomas Kipf, and Pietro Liò. Towards sparse hierarchical graph classifiers. *arXiv preprint arXiv:1811.01287*, 2018.

[12] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

[13] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pages 2323–2332. PMLR, 2018.