How to Better Introduce Geometric Information in Equivariant Message Passing?

Yusong Wang[*]
Xi'an Jiaotong University;
Mirosoft Research AI4Science
t-yusongwang@microsoft.com

Shaoning Li[*]
Microsoft Research AI4Science
v-shaoningli@microsoft.com

Tong Wang[†]
Microsoft Research AI4Science
watong@microsoft.com

Zun Wang
Microsoft Research AI4Science
zunwang@microsoft.com

Xinheng He [*]
Microsoft Research AI4Science
v-xinhenghe@microsoft.com

Bin Shao
Microsoft Research AI4Science
binshao@microsoft.com

Tie-Yan Liu
Microsoft Research AI4Science
tyliu@microsoft.com

Abstract

Equivariant Graph Neural Network (EGNN) is currently prevailing approach for modeling molecular geometric structures. Previous works have demonstrated the effectiveness of elaborate geometric information, i.e., angles and dihedrals, to enhance the modeling capability. However, few of them shows a complete and reasonable framework for how to extract and exploit such latent information. To address this issue, we propose **Global-ViSNet** (short for "vector-scalar interactive neural network"), an equivariant graph neural network to make full use of the geometric information in message passing module with powerful vector-scalar interactive operations for fully connected molecular graph. With sufficiently employing the geometric features contained in 3D structures, Global-ViSNet instructs the prediction of homo-lumo gap from topology graphs.

1 Introduction

OGB-LSC is a large-scale Machine Learning (ML) challenge to exploit the power of ML for graph data. In this competition, we aim at the PCQM4Mv2 dataset, which is a quantum chemistry dataset for predicting DFT-calculated HOMO-LUMO energy gap of molecules. Compared with PCQM4M, PCQM4Mv2 provides additional 3D conformers for training but without validation and test. As a result, a collaborate strategy is needed to jointly learn from both topology and spatial knowledge. Interestingly, such problem could be considered as a multi-view representation learning task, in which different views (2D & 3D conformers) share one encoding backbone and contribute to the final prediction together. Previous work Transformer-M [4] has verified its feasibility and gained

[*] Work done during an internship at Microsoft Research.

[†] Corresponding author.

Preprint. Under review.

incredible performance. Therefore, we adopt the similar strategy and propose a more powerful model for learning the geometric information.

2 Methodology

2.1 Runtime Geometry Calculation

We first demonstrate how we extract the geometric information in an elegant manner. Inspired by the message aggregation, a sufficient way to calculate the surrounding angles between the target node and its neighbors is to compute the summation of the angles as messages passed to itself. Such method enable to reduce the computational complexity from $\mathcal{O}(\mathcal{N}^2)$ to $\mathcal{O}(\mathcal{N})$. And it is proved to be effective by PaiNN [5] but solely considering the angles. However, GemNet [2] and SphereNet [3] has testified the great influence of dihedrals for molecular modeling, but they suffer from high computational overhead due to explicitly extract the dihedrals in quadruplet atoms, which reaches $\mathcal{O}(\mathcal{N}^3)$ complexity. To alleviate such intolerable computational overhead and maintain the utilization of dihedrals features, we propose a strategy to calculate the dihedrals in linear time complexity. Together with the angle calculation, we name it the **Runtime Geometry Calculation** (RGC). The details of RGC can be referred to the preprint version of ViSNet.

2.2 Vector-Scalar Interaction for Intersecting Space

Intuitively, the most common approach for combining scalar and vector features is to treat the scalar features as the scale for vectors. When extended to high dimensionality, we utilize the tensor product to conduct scalar and vector interaction:

$$\tilde{\mathbf{v}} = \mathbf{s} \bigotimes v \tag{1}$$

The bold symbols $\tilde{\mathbf{v}}$, $\mathbf{s}$ denote the features with high dimension, i.e., $\tilde{\mathbf{v}} \in \mathrm{R}^{V \times F}$, $\mathbf{s} \in \mathrm{R}^{F}$ and $v \in \mathrm{R}^{V}$. Here F represents the size of hidden channel and V represents the dimension of the space (e.g., 3 in Cartesian coordinate or $2l+1$ in spherical space). It is worthy noting that in this task we solely leverage the vectors in the Cartesian coordinate, but could be extended to high dimensional space by spherical harmonics [1].

What's more, after extracting the geometric information, the next confronted problem is how to effectively utilize these features. Previous works seldom consider the message transferring paths during message aggregation and simply deliver all the computed messages to the target nodes. To this end, we propose a Vector-Scalar interaction strategy for Intersecting Space, in terms of **ViS-IS** for short. The angular information is derived from the **intersecting node** i (labeled in red), which describes the spatial structure of the neighborhood $\mathcal{N}(i)$, i.e., the target node i and its 1-hop neighbors. From this view, we could treat the extracted angle features as necessary messages for target nodes. Meanwhile, the dihedral information is derived from the **intersecting edge** r_{ij}, which describes the relative positions between neighborhood $\mathcal{N}(i)$ and neighborhood $\mathcal{N}(j)$. Similarly, the edge feature of ij should not only contain the distance but be complemented with extracted dihedral features. Therefore, the process of ViS-IS can be summarized as:

$$\begin{aligned} h_i &\leftarrow \phi(\langle \tilde{\mathbf{v}}_i, \tilde{\mathbf{v}}_i \rangle) \\ f_{ij} &\leftarrow \phi(\langle \mathrm{Rej}_{\vec{r}_{ij}}(\tilde{\mathbf{v}}_i), \mathrm{Rej}_{\vec{r}_{ji}}(\tilde{\mathbf{v}}_j) \rangle) \end{aligned} \tag{2}$$

where h_i denotes the node features, f_{ij} denotes the edge features and ϕ denotes the non-linear update function. The rejection of vectors with high dimension can be treated as rejecting vectors from each dimension by $\vec{r}_{ij}$.

2.3 Model Architecture

The model architecture is based on Transformer-M, which integrates two modality (2D & 3D structures) in one same framework. Also, Transformer-M converts the molecules to fully connected graph and add multiple attention bias to encode the graph structure. However, it only involves the atomic distance and neglects the directional information to preserve equivariance. To this end, we can improve the geometry encoding process with the more powerful RGC and ViS-IS strategy.

We first modify the input features:

$$X^{(0)} = X + \Psi^{\text{2D}} + \Psi^{\text{3D RGC Angle}} \tag{3}$$

where X denotes the node with features, i.e., the embedding of atomic number. And Ψ^{2D} denotes the original extracted features from SMILES, $\Psi^{\text{3D Distance}}$ denotes the sum of Euclidean distance in Transformer-M. The initial vector features are calculated through tensor product:

$$\vec{V} = X \bigotimes \vec{R} \tag{4}$$

where $\vec{R}$ represents the matrix form of $\vec{r}_{ij}$ since the graph is fully connected. Then we can further compute the additional 3D angle features using RGC:

$$\Psi^{\text{3D RGC Angle}} = \langle W_{as}\vec{V}, W_{at}\vec{V} \rangle \tag{5}$$

with W_{as}, W_{at} as linear augmentation without bias to ensure equivariance. We follow the traditional Transformer Layer as Transformer-M, and revise the attention bias as:

$$A(X) = \text{softmax}(\frac{XW_Q(XW_K)^\intercal}{\sqrt{d}} + \Psi^{\text{2D}} + \Psi^{\text{3D Distance}} + \Psi^{\text{3D RGC Dihedral}}) \tag{6}$$

The revised 3D attention bias can be computed as:

$$\Psi^{\text{3D RGC Dihedral}} = \langle W_{ds}\text{Rej}_{\vec{R}}(\vec{V}), W_{dt}\text{Rej}_{-\vec{R}}(\vec{V}) \rangle \tag{7}$$

with W_{ds}, W_{dt} as linear augmentation without bias. In this manner, the messages from vector-scalar interactions are transferred following ViS-IS in Section 2.2. It extends the original 3D features with more geometric information and maintain the linear complexity.

3 Experiments

We conduct the experiments on 4 NVIDIA A100 GPUs with 256 batch-size on one single GPU. Since in this task, we treat the molecule graph to be fully connected, we name our model **Global-ViSNet** for discrimination. The layer of Transformer block is set to 18 and embedding dimension is set to 768.

Table 1: Experimental Results

Model	# of params	Valid MAE (single)	Inference time on valid (s)
Global-ViSNet	78,450,692	0.0784	37.2

We also report the results on QM9 dataset from our preprint manuscript ViSNet. QM9 consists of 12 kinds of quantum chemical properties of 133,385 small organic molecules with up to 9 heavy atoms, and the task is similar to PCQM4Mv2.

4 Conclusion

In this work, we propose Global-ViSNet, an equivariant graph neural network to better introduce the geometric information in message passing with powerful vector-scalar interaction for fully connected molecule graph. Global-ViSNet can instruct the prediction of quantum properties from topology graph by effectively employing geometric features.

References

[1] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):1–11, 2022.

[2] J. Klicpera, F. Becker, and S. Günnemann. Gemnet: Universal directional graph neural networks for molecules. In *Advances in Neural Information Processing Systems*, 2021.

[3] Y. Liu, L. Wang, M. Liu, X. Zhang, B. Oztekin, and S. Ji. Spherical message passing for 3d graph networks. *arXiv preprint arXiv:2102.05013*, 2021.

[4] S. Luo, T. Chen, Y. Xu, S. Zheng, T.-Y. Liu, L. Wang, and D. He. One transformer can understand both 2d & 3d molecular data. *arXiv preprint arXiv:2210.01765*, 2022.

[5] K. Schütt, O. Unke, and M. Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.