Paper Review Summary

Yingheng Wang
Feb/03/2022

Pre-training GNNs:

(1) GROVER [1]: This paper proposed a molecule representation pre-training model with a Graph-Transformer architecture via two self-supervised learning tasks: (i) contextual properties prediction and (ii) graph-level motif prediction. The message-passing paradigm used in this paper is as general as MPNNs [2], which have the equivalent expressive power as the 1-WL test and preserves the invariance property. This paper reached good performance after fine-tuning on several different small-size molecular property prediction datasets. I believe the contribution of the achieved good performance of GROVER can be attributed to three parts: *(i) powerful Transformer architecture, (ii) large-scale pre-training datasets, and (iii) easy tasks and small sizes of downstream datasets*. For (i), the powerful expressivity of Transformers on pre-training molecule representations has also been proven by our method MolCloze [3]. The multi-layer multi-head self-attention blocks with a massive number of parameters guarantee a good fitting ability of the Transformer-based models. For (ii), large-scale pre-training datasets ensure good generalizability of the model on downstream datasets. However, a severe problem remains that *the pre-training dataset may have covered those molecules in downstream datasets*. Besides, our methods, MolCloze [3] and MolCLE [4], show that our models can achieve better performance with a smaller pre-training dataset, proving that *such a large-scale pre-training is unnecessary*. For (iii), since the fine-tuning datasets in this paper are too small and too simple to evaluate the effectiveness of the pre-training model, I believe the results somewhat lack credibility. The model should be *evaluated on more complicated and more extensive fine-tuning datasets* for effectiveness verification. Moreover, several critical spatial features can be added to the pre-training model to improve its expressive power on molecule data, like distance encodings [5], conformer features [6], and spatial encodings [7]. Graphormer [7] applied three encodings into their similar Graph-Transformer architecture and achieved SOTA performance on OGB-LSC [8] dataset, which is more complicated due to its quantum chemistry prediction task.

Equivariant GNNs:

(2) E(3) Equivariant GNNs [9]: This paper proposed an equivariant GNNs for learning interatomic potentials (IP) from reference first principle *ab-initio* calculations, achieving E(3) symmetries. The results outperformed other most recent baselines on many IP prediction tasks with higher efficiency and better accuracy. The advantages of this model mainly lie in two parts: *(i) the usage of higher-order*

geometric tensors and (ii) implicitly satisfying symmetry constraints (i.e., rotational equivariance and permutational invariance). The first point guarantees *fast convergence and consistent performance improvement via higher tensor ranks*, highlighting a crucial role of the equivariance. The same idea is also introduced by UNiTE [10], which presents an N-body equivariant GNNs architecture that can process 2-body and high-order tensor representations of molecules. The second point leads to *better data efficiency*, which means that the model requires much fewer data points to achieve a given accuracy. A similar idea is also introduced by PAINN [11], which is one of many possible ways to incorporate more physical information into GNNs architecture. There are also other approaches that can incorporate mathematical structures and the physical boundaries relevant to electronic structure methods to enhance data efficiency and model generalizability like reproducing kernels optimized for long-range intermolecular forces [12]. However, despite the good performance, this method also includes several disadvantages: *(i) high computational complexity from the introduced Clebsh-Gordan coefficients and spherical harmonics, and (ii) lack of evaluation on large-scale datasets like QM9 [13] and OGB-LSC*.

(3) **E(n) Equivariant GNNs [14]:** This paper proposed a new GNNs architecture to learn the equivariance to rotations, translations, reflections, and permutations, achieving a so-called E(n)-Equivariance. This paper mainly contributed to *(i) extending existing methods from 3d space to higher-dimensional space, (ii) avoiding expensive computations like spherical harmonics, and (iii) good performance on large-scale datasets like QM9*. EGNN can be served as a combination of E(n) equivariant update [15] and SchNet [16], which benefits from both the high bias of E(n) methods and the flexibility of GNNs. Compared to (2), this method is *more computationally efficient (through more straightforward standard basis rather than harmonic basis functions) and powerful (through the extension to higher-order dimensional space)* when keeping equivariant to Euclidean symmetries, realizing continuous symmetry transformations. Thus, it can also be used to model the first-order derivative of the continuous-time flow, which can help to generate E(n) equivariant data like molecules in 3d [17]. This work can still be extended and further explored, e.g., utilizing EGNNs for the non-flow molecule or material generation and chirality-equivariant generative model.

References

[1] Rong Y, Bian Y, Xu T, et al. Self-supervised graph transformer on large-scale molecular data[J]. Advances in Neural Information Processing Systems, 2020, 33: 12559-12571.
[2] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry[C]//International conference on machine learning. PMLR, 2017: 1263-1272.
[3] Wang Y, Chen X, Min Y, et al. MolCloze: A Unified Cloze-style Self-supervised Molecular Structure Learning Model for Chemical Property Prediction[C]//2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2021: 2896-2903.
[4] Wang Y, Min Y, Shao E, et al. Molecular Graph Contrastive Learning with Parameterized Explainable Augmentations[C]//2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2021: 1558-1563.
[5] Li P, Wang Y, Wang H, et al. Distance encoding: Design provably more powerful neural networks for graph representation learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 4465-4478.
[6] Addanki R, Battaglia P W, Budden D, et al. Large-scale graph representation learning with very deep GNNs and self-supervision[J]. arXiv preprint arXiv:2107.09422, 2021.
[7] Ying C, Cai T, Luo S, et al. Do Transformers Really Perform Badly for Graph Representation?[J]. Advances in Neural Information Processing Systems, 2021, 34.
[8] Hu W, Fey M, Ren H, et al. Ogb-lsc: A large-scale challenge for machine learning on graphs[J]. arXiv preprint arXiv:2103.09430, 2021.
[9] Batzner S, Musaelian A, Sun L, et al. Se (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials[J]. arXiv preprint arXiv:2101.03164, 2021.
[10] Qiao Z, Christensen A S, Welborn M, et al. Unite: Unitary n-body tensor equivariant network with applications to quantum chemistry[J]. arXiv preprint arXiv:2105.14655, 2021.
[11] Schütt K, Unke O, Gastegger M. Equivariant message passing for the prediction of tensorial properties and molecular spectra[C]//International Conference on Machine Learning. PMLR, 2021: 9377-9388.
[12] Cvitaš M T, Soldán P, Hutson J M. Long range intermolecular forces in triatomic systems: connecting the atom–diatom and atom–atom–atom representations[J]. Molecular Physics, 2006, 104(1): 23-31.
[13] Ruddigkeit L, Van Deursen R, Blum L C, et al. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17[J]. Journal of chemical information and modeling, 2012, 52(11): 2864-2875.
[14] Satorras V G, Hoogeboom E, Welling M. E (n) equivariant graph neural networks[C]//International Conference on Machine Learning. PMLR, 2021: 9323-9332.
[15] Köhler J, Klein L, Noé F. Equivariant flows: sampling configurations for multi-body systems with symmetric energies[J]. arXiv preprint arXiv:1910.00753, 2019.
[16] Schütt K, Kindermans P J, Sauceda Felix H E, et al. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions[J]. Advances in neural information processing systems, 2017, 30.
[17] Garcia Satorras V, Hoogeboom E, Fuchs F, et al. E (n) Equivariant Normalizing Flows[J]. Advances in Neural Information Processing Systems, 2021, 34.