Atom-level Adaptive Receptive Fields: A Pruning-Based Encoder for 2D Molecular Graphs (Student Abstract)

Yuhao Zhang[1], Ningkang Peng[1], Yafei Liu[1], Lin Li[3], Masaru Kitsuregawa[2], Yanhui Gu[1]*

[1] Nanjing Normal University, Nanjing, China
[2] The University of Tokyo, Tokyo, Japan
[3] Wuhan University of Technology, Wuhan, China
{yuhaozhang,nkpeng,yafeiliu}@nnu.edu.cn, cathylilin@whut.edu.cn, kitsure@tkl.iis.u-tokyo.ac.jp, gu@njnu.edu.cn

Abstract

The two-dimensional (2D) graph structure of a molecule encodes abundant latent property information. A well-designed molecular graph encoder can capture informative low-dimensional dense representations of molecules, which can subsequently be applied to a wide range of downstream tasks. To achieve fine-grained and discriminative molecular representations that capture localized structural information, we propose an novel atom-level adaptive receptive field encoder, enabling each atomic node in the molecular graph to dynamically adjust its receptive field size. **To the best of our knowledge, we are the first to introduce an effective rank-guided pruning strategy for 2D molecular graphs.**

Introduction

Raw 2D molecular graph data can be processed by deep learning models to generate continuous and dense embedding vectors. These embedding vectors can subsequently be applied to downstream tasks, including molecular property prediction for applications such as drug discovery, reaction optimization, and materials design.

However, existing methods face challenges in effectively enabling models to learn discriminative representations for structurally similar molecules, which in turn results in suboptimal performance in predicting properties that are highly sensitive to local structural variations. Meanwhile, unlike typical non-Euclidean data structures, such as social networks or knowledge graphs, 2D molecular graphs exhibit a strong coupling between local structural motifs and global functional properties. In graph neural networks, the inherent risk of oversmoothing can result in the loss of local substructural features of molecules.

In this work, we propose **A**tom-level adaptive **R**eceptive **F**ield **E**ncoder (**ARFE**) for 2D molecular graph representation learning. By incorporating a innovative effective-rank pruning strategy, ARFE is able to more effectively capture property-related information embedded in subtle local molecular structures, while simultaneously mitigating the adverse effects of oversmoothing on atom embeddings, thereby producing fine-grained and discriminative molecular representations.

*Corresponding author.

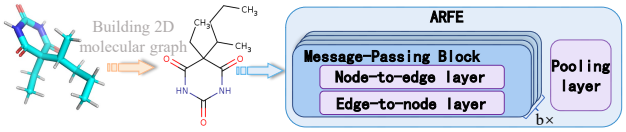

Figure 1: Overall Architecture of ARFE.

Proposed Method

To better capture the property-related information residing in the subtle local structures of molecules, while mitigating the potential impact of oversmoothing on atom embeddings, thereby obtaining fine-grained and discriminative molecular representations, we designed an atom-level adaptive receptive field encoder. Inspired by Diff-eRank (Wei et al. 2024), we propose a pruning-based two-stage message passing mechanism. First, the node-to-edge message passing mechanism is performed. At each iteration of the message passing process, the effective rank of the non-zero edge embeddings matrix $S_{\mathcal{N}} \in \mathbb{R}^{|\mathcal{E}|\times d}$ is defined as:

$$\mathrm{EdgeRank}(S_{\mathcal{N}}) = \exp\left(-\sum_{i=1}^{T}\frac{\sigma_i}{\sum_{i=1}^{T}\sigma_i}\log\frac{\sigma_i}{\sum_{i=1}^{T}\sigma_i}\right), \tag{1}$$

where $T = \min\{|\mathcal{E}|, d\}$, and $\sigma_1, \sigma_2, \ldots, \sigma_T$ are the singular values of matrix $S_{\mathcal{N}}$. We then conduct message passing over the edge embeddings as follows:

$$e^{\mathcal{N}+1}_{(u,v)} = \mathrm{MLP}\left(x^{\mathcal{N}}_u \parallel x^{\mathcal{N}}_v \parallel e^{0}_{(u,v)}\right), \tag{2}$$

where $x^{\mathcal{N}}_u$ and $x^{\mathcal{N}}_v$ denote the node embeddings of the nodes connected this edge, $e^{0}_{(u,v)}$ denotes the initial embeddings of this edge, $e^{\mathcal{N}+1}_{(u,v)}$ denotes the embeddings of the edge after the current round of message passing. The detection matrix $S_{\mathcal{N}+1}$ can then be constructed. Subsequently, we compute the ΔEdgeRank as follows:

$$\Delta\mathrm{EdgeRank} = \mathrm{EdgeRank}(S_{\mathcal{N}+1}) - \mathrm{EdgeRank}(S_{\mathcal{N}}). \tag{3}$$

$\mathrm{EdgeRank}(S_{\mathcal{N}+1})$ and $\mathrm{EdgeRank}(S_{\mathcal{N}})$ quantify the uncertainty in the edge embeddings matrix before and after message passing, respectively. ΔEdgeRank thus represents the

reduction in uncertainty within the edge embeddings matrix resulting from the message passing in that iteration. If $\Delta\text{EdgeRank}$ exceeds 0, pruning is considered necessary. Sequentially compute the contribution of each edge to the magnitude of change in the effective rank of the edge embedding matrix before and after message passing, denoted as ΔEdge, as follows:

$$\Delta\text{Edge} = \Delta\text{EdgeRank} - \left|\text{EdgeRank}(S^*_{\mathcal{N}+1}) - \text{EdgeRank}(S^*_{\mathcal{N}})\right| \tag{4}$$

$\text{EdgeRank}(S^*_{\mathcal{N}})$ and $\text{EdgeRank}(S^*_{\mathcal{N}+1})$ denote the EdgeRank values computed after removing a specific edge from the edge embeddings matrix before and after the current message passing round, respectively. Then, we apply a masking operation to the message updates of the top k edges, while performing residual updates on the other edges.

Building on this, we further implement the edge-to-node message passing mechanism. Similar to Equation (2), we can compute the effective rank of non-zero node embeddings matrix $X_{\mathcal{N}} \in \mathbb{R}^{|\mathcal{V}|\times d}$ as follows:

$$\text{NodeRank}(X_{\mathcal{N}}) = \exp\left(-\sum_{i=1}^{T}\frac{\sigma_i}{\sum_{i=1}^{T}\sigma_i}\log\frac{\sigma_i}{\sum_{i=1}^{T}\sigma_i}\right), \tag{5}$$

where $T = \min\{|\mathcal{V}|, d\}$, and $\sigma_1, \sigma_2, \ldots, \sigma_T$ are the singular values of matrix $X_{\mathcal{N}}$. We then conduct message passing over the node embeddings as:

$$x_u^{\mathcal{N}+1} = \text{MLP}\left(x_u^{\mathcal{N}} \parallel e_{(u,v)}^{\mathcal{N}}\right). \tag{6}$$

Similar to Equation (4), we compute the $\Delta\text{NodeRank}$ as:

$$\Delta\text{NodeRank} = \text{NodeRank}(X_{\mathcal{N}+1}) - \text{NodeRank}(X_{\mathcal{N}}). \tag{7}$$

Similarly, if $\Delta\text{NodeRank}$ is greater than zero, pruning is performed. ΔEdge can be derived following an approach similar to Equation (5):

$$\Delta\text{Node} = \Delta\text{NodeRank} - \left|\text{NodeRank}(X^*_{\mathcal{N}+1}) - \text{NodeRank}(X^*_{\mathcal{N}})\right| \tag{8}$$

where $\text{NodeRank}(X^*_{\mathcal{N}})$ and $\text{NodeRank}(X^*_{\mathcal{N}+1})$ represent the NodeRank values calculated before and after the current message passing round, respectively, following the removal of a specific node from the node embeddings matrix. Then, we mask the message updates for the top k nodes while applying residual updates to the remaining nodes. Subsequently, we obtain the complete molecular representation through a specific graph pooling method.

Experiment

Performance on Downstream Tasks

We selected seven molecular activity cliff prediction tasks from MoleculeACE (Van Tilborg, Alenicheva, and Grisoni 2022) and one molecular property prediction task (Wu et al. 2018) to evaluate ARFE, with the results presented in Figure 2a and 2b, respectively. For the molecular activity cliff prediction tasks, we adopt the models provided in MoleculeACE as our baseline models. For the molecular property prediction task, we select MolGT (Chen et al. 2025), MoleBLEND (Yu et al. 2023), GraphMVP (Liu et al. 2022), and PretrainGNN (Hu et al. 2020) as our baseline models. All experiments are conducted following the same data splitting protocol as used for the baseline models. Figure 2c shows the visualized predictions for randomly selected samples from the test set on the Blood-Brain Barrier Penetration (BBBP) task.

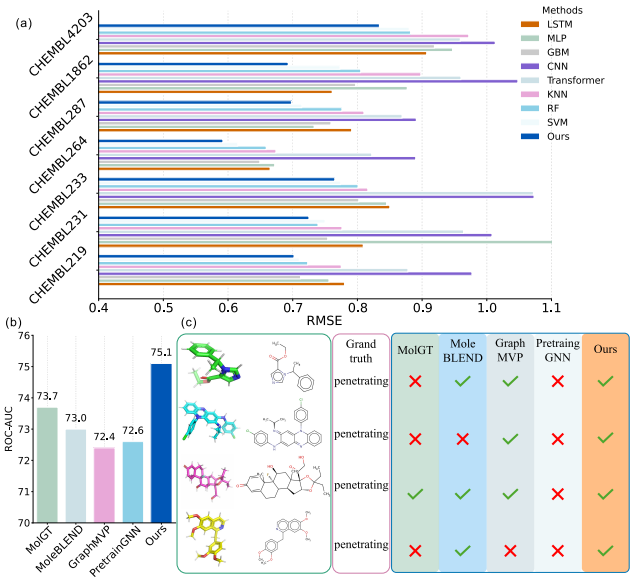

Figure 2: Results on the downstream tasks.

ARFE's strong performance on both molecular activity cliff prediction and the BBBP task demonstrates its ability to generate fine-grained molecular representations and its robustness for molecular property prediction in drug design and optimization.

Results When Used as a Plug-and-Play

Leveraging the lightweight advantage of ARFE, it can be integrated as a plug-and-play module into existing molecular representation learning models. For the BBBP task, integrating ARFE yields a performance improvement of 17.2% and 26.3% over the original MGSSL (Zhang et al. 2021) and Uni-Mol (Zhou et al. 2023), respectively.

Conclusion

Obtaining fine-grained and discriminative molecular embeddings from 2D molecular graph data is crucial for a variety of downstream prediction tasks. In this work, we propose ARFE, which introduces a pruning strategy based on the effective rank of atom feature and bond feature matrices, enabling the encoder to generate fine-grained representations of 2D molecular graphs. Meanwhile, we are committed to further improving ARFE to facilitate protein-ligand docking and other real-world applications.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant No.22033002 and Grant No.92370127).

References

Chen, R.; Li, C.; Wang, L.; Liu, M.; Chen, S.; Yang, J.; and Zeng, X. 2025. Pretraining graph transformer for molecular representation with fusion of multimodal information. *Inf. Fusion*, 115: 102784.

Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V. S.; and Leskovec, J. 2020. Strategies for Pre-training Graph Neural Networks. In *8th International Conference on Learning Representations, ICLR 2020,*.

Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; and Tang, J. 2022. Pre-training Molecular Graph Representation with 3D Geometry. In *The Tenth International Conference on Learning Representations, ICLR*.

Van Tilborg, D.; Alenicheva, A.; and Grisoni, F. 2022. Exposing the limitations of molecular machine learning with activity cliffs. *Journal of chemical information and modeling*, 62(23): 5938–5951.

Wei, L.; Tan, Z.; Li, C.; Wang, J.; and Huang, W. 2024. DiffeRank: A Novel Rank-Based Metric for Evaluating Large Language Models. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024*.

Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; and Pande, V. 2018. MoleculeNet: a benchmark for molecular machine learning. *Chemical science*, 9(2): 513–530.

Yu, Q.; Zhang, Y.; Ni, Y.; Feng, S.; Lan, Y.; Zhou, H.; and Liu, J. 2023. Unified Molecular Modeling via Modality Blending. *CoRR*, abs/2307.06235.

Zhang, Z.; Liu, Q.; Wang, H.; Lu, C.; and Lee, C.-K. 2021. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34: 15870–15882.

Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; and Ke, G. 2023. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. In *International Conference on Learning Representations*.