Simple Orthogonal Graph Representation Learning (Student Abstract)

Taoyong Cui [1], **Yuhan Dong** [1]

[1]Shenzhen International Graduate School, Tsinghua University,Shenzhen, China
cty21@mails.tsinghua.edu.cn, dongyuhan@sz.tsinghua.edu.cn

Abstract

Graph neural networks (GNNs) have attracted significant interests recently since they can effectively process and analyze graph-structured data commonly found in real-world applications. However, the predicament that GNNs become difficult to train with as the number of layers increase. The essence of this problem is that stacking layers will reduce the stability of forward propagation and gradient back-propagation. As the size of models increase (measured by the number of parameters), how to efficiently and effectively adapt it to particular downstream tasks becomes an intriguing research issue. In this work, we propose a simple orthogonal training framework to impose the orthogonality constraints on GNNs, which can help models find a solution vector in a specific low dimensional subspace and stabilize the signaling processes at both the forward and backward directions. Specifically, we propose a novel polar decomposition-based orthogonal initialization (PDOI-R) algorithm to identify the low intrinsic dimension within the Stiefel manifold and stabilize the training process. Extensive experiments demonstrate the effectiveness of the proposed method in multiple downstream tasks, showcasing its generality. This simple method can help existing state-of-the-art models achieve better performance.

Introduction

In the past decades, graph neural networks (GNNs) have become increasingly popular in artificial intelligence research due to the prevalence of graph-structured data in various domains such as e-commerce, traffic, chemistry, and the knowledge base field. While these models can automatically learn hidden deep features from graphs and are highly effective, several problems still exist. A key limitation of GNNs is that the learned node representations are highly indistinguishable with layer stacking. Previous researchers attribute this problem to the over-smoothing issue (Chen et al. 2020a). Because of the recursive neighbor aggregation on the graph structure, the node representation becomes indistinguishable. To alleviate the over-smoothing issue, increasing efforts have been made in succession, such as skip connection (Chen et al. 2020a) and graph augmentation (Rong et al. 2019). However, it is hard to explain why shallow GNNs still have the same issue with deep GNNs. The authors of (Guo et al. 2022) conducted experiments on the feature transformation module and concluded that the vanilla feature transformation damages the steady model signaling at both the forward and backward passes, which in turn degrades the performance of GNNs. Another thing worth noting is that with the continuous development of the large model, how to efficiently transfer the information learned to downstream tasks has become a very important research direction, especially in pre-training tasks.

To efficiently adapt GNNs to downstream tasks, we impose approximately orthogonality constraints on the GNNs to find a low intrinsic dimension on the Stiefel manifold, which can help learn the essence features of data representation. From the view of norm-preserving, we find that keeping the orthonormality within a layer weight during the learning process is a sufficient and necessary condition to ensure the stability of forward activation norm and backward errors. And we map the weight matrices from high-dimensional space to low-dimensional orthogonal space (the Stiefel manifold) without losing information to ensure diversity and essentiality of learned features (Cui et al. 2022). In summary, three contributions can be summarized as follows: **(1)** We propose a simple training method that aims to enhance the performance of downstream tasks by reducing parameter redundancy and improving feature expressiveness, especially in pre-training tasks. **(2)** We introduce a novel polar decomposition-based orthogonal initialization (PDOI-R) algorithm to identify the low intrinsic dimension within the Stiefel manifold, which not only facilitates the imposition of orthogonal constraints but also stabilizes the training of GNNs. **(3)** Extensive experiments have been conducted on various GNN tasks to evaluate the effectiveness of PDOI-R. Numerical results demonstrate that PDOI-R can significantly enhance the performance of existing models and ensure their stability.

Methods

In this work, we propose a novel algorithm PDOI-R in 1, to find a local optimum on the Stiefel manifold as the starting point. We use the Randomized Singular Value Decomposition (RSVD) to compute the polar decompositions. RSVD (Erichson et al. 2016) is a computationally efficient algorithm for computing the truncated SVD of large data matrices. To improve the quality of RSVD, the power iteration

Model	BBBP ↑	Tox21 ↑	ToxCast ↑	SIDER ↑	ClinTox ↑	MUV ↑	HIV ↑	BACE ↑
GraphMAE (Hou et al. 2022)	72.0±0.6	75.5±0.6	64.1±0.3	60.3±1.1	82.3±1.2	76.3±2.4	77.2±1.0	83.1±0.9
PDOIR-GraphMAE	**73.0±0.5**	**77.25±0.4**	**65.3±0.4**	**64.7±1.0**	**86.1±1.2**	**79.2±1.8**	**79.9±0.7**	**85.0±0.7**

Table 1: Comparison with SOTA on transfer learning on molecular property prediction benchmarks. We report ROC-AUC scores (%). The models are first pre-trained on ZINC15 (Sterling and Irwin 2015) and then finetuned on the above datasets

Method	Cora ↑	Cite. ↑	Pumb. ↑	Corn. ↑	Texa. ↑	Wisc. ↑
GCNII (Chen et al. 2020b)	88.01	77.13	90.30	76.49	77.84	81.57
Ortho-GCNII (Guo et al. 2022)	88.69	77.21	90.42	77.84	80.54	83.53
PDOIR-GCNII	**90.95**	**79.58**	**90.59**	**78.38**	**81.28**	**84.27**

Table 2: Accuracy (%) comparisons with SOTA (GCNII (Chen et al. 2020b)) on full-supervised node classification tasks. The PDOIR-GCNII achieves the best performance and is highlighted in the table.

Model	BBBP ↑	Tox21 ↑	ToxCast ↑	SIDER ↑	ClinTox ↑	MUV ↑	HIV ↑	BACE ↑
GraphCL (You et al. 2021)	69.7±0.7	73.9±0.7	62.4±0.6	60.5±0.9	76.0±2.7	69.8±2.7	78.5±1.2	75.4±1.4
GraphLoG (Hafidi et al. 2022)	72.5±0.8	75.7±0.5	63.5±0.7	61.2±1.1	76.7±3.3	76.0±1.1	77.8±0.8	83.5±1.2
GraphMAE (Hou et al. 2022)	72.0±0.6	75.5±0.6	64.1±0.3	60.3±1.1	82.3±1.2	76.3±2.4	77.2±1.0	83.1±0.9
PDOIR-GraphMAE	**73.0±0.5**	**77.25±0.4**	**65.3±0.4**	**64.7±1.0**	**86.1±1.2**	**79.2±1.8**	**79.9±0.7**	**85.0±0.7**

Table 3: Comparison with SOTA on transfer learning on molecular property prediction benchmarks. We report ROC-AUC scores (%). The models are first pre-trained on ZINC15 and then finetuned on the above datasets

Algorithm 1: PDOI-R algorithm

Input: a starting point θ_0.
Output: $\theta_k, k \geq 1$.

1: **for** k=1,2,..., until a stopping criterion is satisfied **do**
2: Compute Euclidean gradient $\nabla g\left(\theta_{k-1}\right)$.
3: Compute the RSVD decomposition ($U\Sigma V^T$) of

$$\nabla g\left(\theta_{k-1}\right). \tag{1}$$

4: Update θ_k to be the product of two orthogonal matrices UV^T.
5: **end for**

method can be used directly.

The PDOI-R exploits a low-dimensional manifold that can approximately represent the whole model parameters (Ding et al. 2023), and the optimization trajectory follows this manifold. The algorithm is an easier optimization method due to the non-reliance on the tangent space and can provide a better convergence. The parameter optimization of the pre-training model follows a low-dimensional manifold, which can be embedded in a low-dimensional representation of the solution vector. If this representation accurately captures the low-dimensional space, then fine-tuning the full parameters on the original model is equivalent to refining the model within this lower-dimensional parameter space. In this way, the fitting or fine-tuning process for large models can be accomplished with minimal computational costs, requiring only a small amount of data and parameter tuning.

The PDOI-R is a general module capable of augmenting the existing GNNs. Without loss of generality, we adopt the simple method by the polar decomposition iteration on the Stiefel manifold to achieve orthogonal initialization.

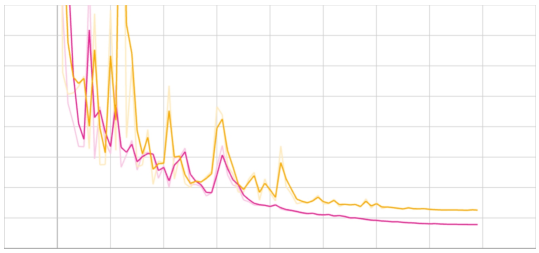

Figure 1: Verification set loss variation curve during the training process. The pink line represents the EGNN (Satorras, Hoogeboom, and Welling 2021)after PDOI-R initialization, while the yellow line represents the EGNN after Xavier initialization. The horizontal axis represents the training epochs, and the vertical axis represents the Mean absolute errors on ISO17 ($C_7O_2H_{10}$ isomers (Schütt et al. 2017)) in kcal/mol.

Conclusion

In this study, we propose a simple method to enhance the performance of GNNs in downstream tasks. Before the GNNs starts training, we utilize the PDOI-R algorithm for orthogonal initialization. The algorithm can realize the orthogonal transformation and converge to a locally optimal point on the Stiefel manifold. We evaluate the performance of our proposed method across various downstream tasks of GNNs. All experimental settings remain consistent with the original paper literature, ensuring a fair basis for comparison. The experimental results demonstrate that PDOI-R can effectively improve the model performance in numerous downstream tasks of GNNs.

Acknowledgements

This research is supported in part by Science, Technology and Innovation Commission of Shenzhen Municipality under grant KCXFZ20211020163813019 and WDZC20200818121348001.

References

Chen, D.; Lin, Y.; Li, W.; Li, P.; Zhou, J.; and Sun, X. 2020a. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3438–3445.

Chen, M.; Wei, Z.; Huang, Z.; Ding, B.; and Li, Y. 2020b. Simple and deep graph convolutional networks. In *Proceedings of the International Conference on Machine Learning*, 1725–1735. PMLR.

Cui, T.; Li, J.; Dong, Y.; and Liu, L. 2022. TAOTF: A Two-stage Approximately Orthogonal Training Framework in Deep Neural Networks. *arXiv preprint arXiv:2211.13902*.

Ding, N.; Qin, Y.; Yang, G.; Wei, F.; Yang, Z.; Su, Y.; Hu, S.; Chen, Y.; Chan, C.-M.; Chen, W.; et al. 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 1–16.

Erichson, N. B.; Voronin, S.; Brunton, S. L.; and Kutz, J. N. 2016. Randomized matrix decompositions using R. *arXiv preprint arXiv:1608.02148*.

Guo, K.; Zhou, K.; Hu, X.; Li, Y.; Chang, Y.; and Wang, X. 2022. Orthogonal graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 3996–4004.

Hafidi, H.; Ghogho, M.; Ciblat, P.; and Swami, A. 2022. Negative sampling strategies for contrastive self-supervised learning of graph representations. *Signal Processing*, 190: 108310.

Hou, Z.; Liu, X.; Dong, Y.; Wang, C.; Tang, J.; et al. 2022. GraphMAE: Self-Supervised Masked Graph Autoencoders. *arXiv preprint arXiv:2205.10803*.

Rong, Y.; Huang, W.; Xu, T.; and Huang, J. 2019. Dropedge: Towards deep graph convolutional networks on node classification. *arXiv preprint arXiv:1907.10903*.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E (n) equivariant graph neural networks. In *Proceedings of the International conference on machine learning*, 9323–9332. PMLR.

Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; and Müller, K.-R. 2017. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Proceedings of the Advances in neural information processing systems*, 30.

Sterling, T.; and Irwin, J. J. 2015. ZINC 15–ligand discovery for everyone. *Journal of chemical information and modeling*, 55(11): 2324–2337.

You, Y.; Chen, T.; Shen, Y.; and Wang, Z. 2021. Graph contrastive learning automated. In *Proceedings of the International Conference on Machine Learning*, 12121–12132. PMLR.