On Self-Distilling Graph Neural Network

Zepeng Zhang

This paper [1] proposed the first teacher-free knowledge distillation method for GNNs, termed GNN Self-Distillation (GNN-SD), that serves as a drop-in replacement of the standard training process. The self-distillation perform knowledge extraction and transfer between layers of a single network without the assistance from auxiliary models. The method is built upon the neighborhood discrepancy rate (NDR), which quantifies the non-smoothness of the embedded graph in an efficient way. The NDR of layer l is defined as

$$\mathbf{s}^{(l)} = \left(s_1^{(l)}, \ldots, s_N^{(l)}\right),$$

where

$$s_v^{(l)} = 1 - \frac{\mathbf{X}_v^{(l)} \left(\mathbf{A}\mathbf{X}^{(l)}\right)_v^T}{\left\|\mathbf{X}_v^{(l)}\right\|_2 \left\|\left(\mathbf{A}\mathbf{X}^{(l)}\right)_v\right\|_2}, \quad v = 1, \ldots, N.$$

Based on this metric, the adaptive discrepancy retaining (ADR) regularizer is proposed to empower the transferability of knowledge that maintains high neighborhood discrepancy across GNN layers. The ADR loss is computed by matching the NDR of deep layer to the target one of its previous layer. Besides, the ADR regularizes the GNN only when the magnitude of NDR of the target layer is larger then the online layer.

[3] also tries to control the distance between $\mathbf{X}$ and $\mathbf{AX}$. However, they actually want them to be close and claims it acts as an infinite-depth GCN. Besides, [2] considers preserving the node similarity in GNNs as well. These two papers' claims seem contradicting with the one in [1]. The relationship between these approaches and the effect of node feature distances are worth exploring.

References

[1] Yuzhao Chen, Yatao Bian, Xi Xiao, Yu Rong, Tingyang Xu, and Junzhou Huang. On self-distilling graph neural network. *IJCAI*, 2021. (document)

[2] Wei Jin, Tyler Derr, Yiqi Wang, Yao Ma, Zitao Liu, and Jiliang Tang. Node similarity preserving graph convolutional networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 148–156, 2021. (document)

[3] Han Yang, Kaili Ma, and James Cheng. Rethinking graph regularization for graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4573–4581, 2021. (document)

Preprint. Under review.