DIRICHLET ENERGY CONSTRAINED LEARNING FOR DEEP GRAPH NEURAL NETWORKS

A NOTE

Zepeng Zhang

July 10, 2022

This paper [1] proposed a Dirichlet energy constrained principle to show the importance of regularizing the Dirichlet energy at each layer within reasonable lower and upper limits. Given a node embedding matrix $X^{(k)}$, the Dirichlet energy is defined as follows:

$$E\left(X^{(k)}\right)=\operatorname{tr}\left(X^{(k)\top}\tilde{L}X^{(k)}\right)=\frac{1}{2}\sum a_{ij}\left\|\frac{x_i^{(k)}}{\sqrt{1+d_i}}-\frac{x_j^{(k)}}{\sqrt{1+d_j}}\right\|_2^2.$$

Lemma 1. *The Dirichlet energy at the k-th layer is bounded as follows:*

$$0\leq E\left(X^{(k)}\right)\leq s_{\max}^{(k)}E\left(X^{(k-1)}\right),$$

where $s_{\max}^{(k)}$ is the squares of the maximum singular values of $W^{(k)}$. Note that to simplify the derivation process, the non-linear activations are neglected.

The Dirichlet energy constrained learning defines the lower and upper limits at layer k as

$$c_{\min}E\left(X^{(k-1)}\right)\leq E\left(X^{(k)}\right)\leq c_{\max}E\left(X^{(0)}\right)$$

with $c_{\min}\in(0,1)$ and $c_{\max}\in(0,1]$. Instead of directly constrain the Dirichlet energy in GNN training, an efficient model EGNN is proposed to satisfy the constrained learning from three perspectives: weight controlling, residual connection and activation function.

To satisfy the upper limits of Dirichlet energy, weight $W^{(1)}$ is initialized as $\sqrt{c_{\max}}I$ and the other weight matrices are initialized as I. To make sure the constraint is satisfied during model training, two regularization terms are added to the loss, that is, $\left\|W^{(1)}-\sqrt{c_{\max}}I\right\|_F+\sum_{k=2}^K\left\|W^{(k)}-I\right\|_F$. To satisfy the lower limits of Dirichlet energy, residual graph convolution is used:

$$X^{(k)}=\sigma\left(\left[\left(1-c_{\min}\right)\tilde{A}X^{(k-1)}+\alpha X^{(k-1)}+\beta X^{(0)}\right]W^{(k)}\right),$$

where $\alpha+\beta=c_{\min}$. However, the use of ReLU may violate the lower limit as it decreases the Dirichlet energy. Therefore, a shifted ReLU is applied:

$$\sigma\left(X^{(k)}\right)=\max\left(b,X^{(k)}\right),$$

where b is a trainable shift.

References

[1] Kaixiong Zhou, Xiao Huang, Daochen Zha, Rui Chen, Li Li, Soo-Hyun Choi, and Xia Hu. Dirichlet energy constrained learning for deep graph neural networks. *Advances in Neural Information Processing Systems*, 34: 21834–21846, 2021.