CS6208 Paper Review

Neo Yuan Rong Dexter
A0214975W

1 Review

Semi-Supervised Classification with GCN (ICLR17): In this paper, the authors [1] present a simple yet effective and scalable approach to learning on graph-structured data, based on convolution. The authors suggest that their method is an approximation of spectral convolution, and their layer-wise propagation can be expressed with the following:

$$H^{(l+1)} = \sigma\left(\hat{D}^{-0.5}\hat{A}\hat{D}^{-0.5}H^{l}W^{l}\right) \tag{1}$$

where $\hat{A}$ is the adjacency matrix with self-connections plus the identity matrix and $\hat{D}$ is the normalization of the adjacency matrix and W is a set of learnable weights. The authors show that their method works well in terms of computation speed and accuracy against SOTA methods. The authors benchmarked their method's accuracies on Citation networks (Citeseer, Cora, Pubmed), Knowledge graph NELL and random graphs for wall-clock time.

Their experiment results show that, their method GCN outperforms SOTA methods, with the Renormalization trick in Equation 1 the most effective variant propagation model. In the wall-clock experiments on random graphs, the authors show that their method scales linearly with the number of edges on a graph and that CPU or mini-batching can be used for large graphs that do not fit into the memory requirement.

A limitation to this work, is that there is an implicit assumption that nodes of the same neighbourhood are equally important, which may not be the case for some datasets. This issue is addressed in the following paper review below.

Graph Attention Networks (ICLR18): Authors of GAT [2] propose masked self-attention layers, where the idea is widely popular in the NLP literature. The attention mechanism is well-known to be useful in handling sequential tasks such as machine translation. The proposed method can be summarized as the derivation of the attention weights α:

$$e_{ij} = a\left(Wh_i, Wh_j\right) \tag{2}$$

where h is the set of input node features, a are the self-attention weights and e is the self-attention output. After applying masking, softmax and non-linearities such as LeakyReLU, the final set of attention weights α is applied to the input features. The authors argue that GAT is different from GCN because it learns different importances between self-connections vs edges of the same neighbourhood. This difference benefits the model's performance and is shown when evaluated on Cora, Citeseer and Pubmed datasets, when GAT improves the baseline performance of GCN by about 2-3% points.

2 Implementation: GCNConv + GAT

We implement GCNConv [1] and compared our implementation against Pytorch's implementation of GCNConv. We further show that the added attention layer from GAT [2] is able to improve GCN. We evaluate our methods using the dataset MNIST Superpixels [3]. Our findings are shown in Figure 1 and are aligned with the experiment results published in both review papers

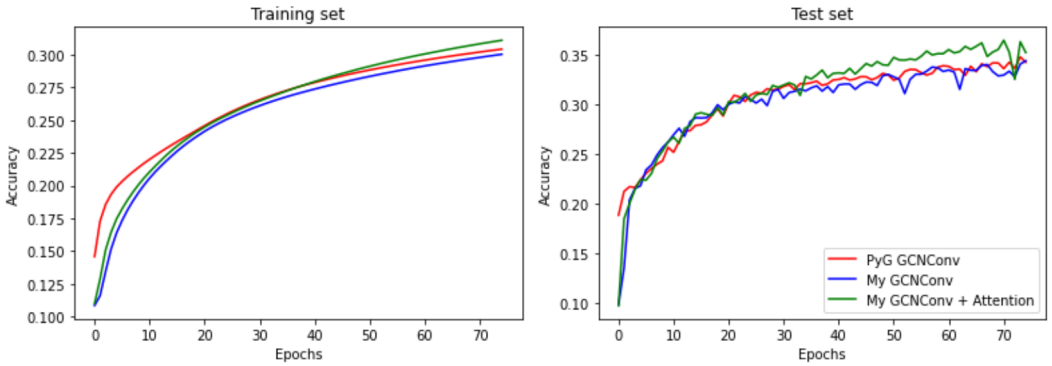

Figure 1: The standard learning curve of PyG's GCNConv (Red), our implementation matches closely (Blue). With the addition of an attention layer slightly better accuracy is achieved (Green).

References

[1] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[2] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[3] Federico Monti, D. Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5425–5434, 2016.