Graph Attention Networks Paper Review

Runze Cai (A0212755H), GitHub (https://github.com/RenzoTsai/GAT-Reproduce)

1 Introduction

Previous spectral graph ConvNets relied on Laplacian eigenvectors, making it difficult to apply models trained on a specific graph structure to other graphs with different structures. Spatial convolution mechanisms such as MoNet offered more flexible learning, allowing each node to learn from its close neighbors. However, there is still room for improvement in how groups of spatially close neighbors operate together.

The work, i.e., Graph Attention Networks (GAT), proposes a self-attention mechanism that better learns the weights of each node's neighbor and generates node representations for graphs. GAT can be applied to many applications, such as node classification as presented in this paper.

2 Method

GAT leverages a self-attention mechanism that enables each node to learn from the most relevant neighbors with learned weights, making it scalable to handle large graph sizes and neighbor connections. It also uses multi-heads to represent different types of node relationships, improving model complexity and efficiency through parallel computation. By making the weights for neighbor aggregation learnable and using deep learning to learn features, GAT has better mathematical soundness for learning graph representations.

3 Evaluation & Results

The GAT model was evaluated on four classification benchmarks, including Cora, Citeseer, and Pubmed for transductive learning, and the PPI dataset for inductive learning. Results showed that GAT outperformed all other methods, including MoNet on the Cora and Citeseer datasets and achieved comparable accuracy to GCN-64 on the Pubmed dataset. For inductive training, GAT improved by 20.5% compared to the SOTA GraphSAGE.

4 Conclusion & Improvement

The GAT model's attention mechanism enables it to operate on groups of neighbors in graphs using spatial learning, making it applicable to flexible graph structures and outperforming other techniques on node classification benchmarks.

While GAT provides SOTA performance on the four benchmarks at the moment of publishing, there is still room for improvement. For instance, larger datasets may require deeper or wider GAT architectures with more layers. In these cases, incorporating residual connections between layers may improve performance and efficiency. Even though the authors employed residual connections across the intermediate attentional layer, how to balance the relationship between the increased layer's number and the hidden layer's dimension and multi-head numbers is worth to further exploration. Moreover, GAT focuses on rather local sampling, but recent research has shown promising results using transformer-based architectures for GNN, which could be a future direction for improving GAT. Another approach to address this issue could be considering global representation using master node [1]. In addition, as the author also mentioned the possibility of applying edge features in future work, incorporating edge features with a self-attention mechanism would be helpful.

References

[1] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, *et al.*, "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.