John (Ziyuan) Zhou

Week 2 Summary

UW ID: 2150115

In the article “NodeAug: Semi-Supervised Node Classification with Data Augmentation”, the author emphasizes the importance of the Data Augmentation (DA) for graph data as the representation of each node’s information. DA for different nodes influence each other and lead to undesired results. In addition, it is a challenge to ensure the labels for each node are not altered significantly while their features are augmented effectively. To conquer this challenge, the author proposes the NodeAug, which is a methodology consisting of a “parallel universe” scheme to conduct DA for each node that happens independently and in a parallel and consistent training scheme to minimize the classification divergence between the predictions of the original nodes and augmented nodes. To complement NodeAug, the author describes the three different DA strategies – replacing attributes, removing edges, and adding edges. After that, he proposes a subgraph mini-batch training method to tackle the resource problem for large graphs. For the experiment, the author successfully validated his work and results on the Cora, Citeseer, and Pubmed Dataset. The future works relate to better data augmentation with higher accuracy.