[Arxiv 2020] Graph Contrastive Learning with Adaptive Augmentation [paper]
Node/Graph Tasks: Node Classification

Training Type: pre-training in an unsupervised manner and the resulting embeddings are used to train and test a logistic regression classifier

Pretext task data: graph topology, node features
The pretext task here is to optimize a two-layer GCN encoder so that the agreement between extracted node embeddings under two correlated graph views is maximized. Different graph views are generated by performing stochastic corruption on the graph topology and the node features.

Brief Summary:
In view that GNN models are mostly performed in a supervised manner which require abundant label information for training, Deep Graph InfoMax(DGI) is proposed, which first augments the original graph and then maximize the mutual information between node embeddings and the global graph embedding. However, existing augmentation schemes fail to generate diverse neighborhoods and yet to consider the impact of different nodes and edges when performing data augmentation. To this end, this paper proposes a novel contrastive framework for unsupervised graph representation learning with adaptive augmentation.

The goal is to maximize the agreement within representations between different graph views generated by performing stochastic corruption on the graph topology and the node features. Assuming that at each iteration, two graph views $\widetilde{G}_1 = v_1(\widetilde{G}), \widetilde{G}_2 = v_2(\widetilde{G})$ are generated where two augmentation function v_1, v_2 are randomly sampled from the set of all possible augmentation functions $\mathcal{T}$. Denoting node embeddings in the two views as $U = f(\widetilde{X}_1, \widetilde{A}_1), V = f(\widetilde{X}_2, \widetilde{A}_2)$ where $\widetilde{X}_*, \widetilde{A}_*$ are the feature and adjacency matrices of the views and f is a two-layer GCN encoder, we optimize the similar objective to the InfoNCE objective [] at each iteration:

$$\mathcal{L} = \frac{1}{2N} \sum_{i=1}^{N} [l(\mathbf{u}_i, \mathbf{v}_i) + l(\mathbf{v}_i, \mathbf{u}_i)] \tag{14}$$

$$l(\mathbf{u}_i, \mathbf{v}_i) = \log \frac{e^{\theta(\mathbf{u}_i, \mathbf{v}_i)/\tau}}{\underbrace{e^{\theta(\mathbf{u}_i, \mathbf{v}_i)/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k \neq i} e^{\theta(\mathbf{u}_i, \mathbf{v}_k)/\tau}}_{\text{inter}-\text{view negative pairs}} + \underbrace{\sum_{k \neq i} e^{\theta(\mathbf{u}_i, \mathbf{u}_k)/\tau}}_{\text{intra}-\text{view negative pairs}}} \tag{15}$$

where τ is a temperature parameter and $\theta(\mathbf{u}, \mathbf{v})$ is the osine similarity between the transformed node features through a two-layer perceptron. Given a node $\mathbf{u}_i$, its only positive pair is $\mathbf{v}_i$ and all other nodes are its negative samples.

The topology-level augmentation is to randomly select subset of edges by performing Bernoulli experiment on every edge with removing probability proportional to its corresponding edge centrality. Three node centrality measures, degree centrality, eigenvector centrality and the PageRank centrality are used and edge centrality is calculated based on the centrality of its two connected nodes. Similarly, the node-attribute-level augmentation is to randomly masking a fraction of dimensions with zeros in node features by performing Bernoulli experiment with probability computed based on the occurence of corresponding dimenstion in each node and node centrality.

By performing node classification on 5 datasets, the proposed model GCA consistently outperforms existing state of the art methods and even surpasses some supervised counterparts. Furthermore, ablation studies and the sensitivity analysis are performed to learn the impact of each component and the critical hyperparameters in GCA.

Bibtex:

@articlezhu2020graph, title=Graph Contrastive Learning with Adaptive Augmentation, author=Zhu, Yanqiao and Xu, Yichen and Yu, Feng and Liu, Qiang and Wu, Shu and Wang, Liang, journal=arXiv preprint arXiv:2010.14945, year=2020