[Arxiv 2020] Deep Graph Contrastive Representation Learning [paper][code]
Node/Graph Tasks: Node classification on citation networks

Training Type: Firstly, training the GNN encoder by contrastive learning under two graph views generated by feature and topology corruption. Secondly, applying the trained GNN encoder to extract node embeddings and feed them to a logistic regression classifier for downstreaming classification.

Pretext task data: graph topology, node features
The pretext task here is to optimize a two-layer GCN encoder or a three-layer GraphSAGE-GCN encoder so that the mutual information (similarity) between positive pairs is maximized and between the negative pairs is minimized. The positive pairs come from the same nodes in different graph views while the negative pairs come from either different nodes in the same graph view or different nodes in different graph views.

Brief Summary:
In view that existing graph representation models are established mostly in a supervised manner requring abundant annotated labels for training, and unsupervised models such as node2vec and DeepWalk are yet to be inductive and fail to general to new added nodes, Deep Graph InfoMax (DGI) is proposed with the objective to maximize the mutual information between node embeddings and the graph embedding by discriminating nodes in the original graph from nodes in a corrupted graph. However, the requirement that the readout function in DGI should be injective is too restrictive to fulfill. Besides, the mean-pooling readout function does not guarantee that the graph embedding can distill useful information from nodes. Moreover, the scheme to apply feature shuffling to generate corrupted views of graphs is insufficient to generate different neighborhoods for nodes in sparse feature matrix. To this end, this paper introduces a contrastive framework for unsupervised graph representation learning, GRACE, where two correlated graph views are generated by randomly performing corruption on attributes (masking node features) and topology (removing graph edges). Then the GNN-based encoders are trained using a contrastive loss to maximize the agreement between node embeddings in these two views.

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 GNN-based encoder, we optimize the similar objective to the InfoNCE objective [] at each iteration:

$$\mathcal{L} = \frac{1}{2N}\sum_{i=1}^{N}[l(\mathbf{u}_i, \mathbf{v}_i) + l(\mathbf{v}_i, \mathbf{u}_i)] \tag{12}$$

$$l(\mathbf{u}_i,\mathbf{v}_i) = \log \frac{e^{\theta(\mathbf{u}_i,\mathbf{v}_i)/\tau}}{\underbrace{e^{\theta(\mathbf{u}_i,\mathbf{v}_i)/\tau}}_{\text{positive pair}} + \underbrace{\sum_{k\neq i} e^{\theta(\mathbf{u}_i,\mathbf{v}_k)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{k\neq i} e^{\theta(\mathbf{u}_i,\mathbf{u}_k)/\tau}}_{\text{intra-view negative pairs}}} \tag{13}$$

where τ is a temperature parameter and $\theta(\mathbf{u},\mathbf{v})$ is the cosine 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.

Two graph corruptions are performed by randomly select subset of edges and randomly masking a fraction of dimensions with zeros in node features by Bernoulli experiment with probability as a hyperparameter.

Experiments are performed under two learning settings. In transductive learning, a two-layer GCN is employed as the encoder and the performance on Cora, Citeseer, Pubmed and DBLP is compared with the traditional methods: DeepWalk and node2vec, and with the deep learning methods: GAE, VGAE, SGC, GCN, and DGI. In inductive learning, a three-layer GraphSAGE-GCN with residual connections is employed as the encoder and the performance on Reddit and PPI is compared with the traditional methods: FastGCN and GaAN-mean. Results demonstrate the consisting outperformance of GRACE over all other baseline models. Also, the sensitivity analysis on critical hyperparameters is performed and the ablation studies are done to show the necessity of jointly considering corruption at both graph topology and node feature levels.

Bibtex:

@articlezhu2020deep, title=Deep Graph Contrastive Representation Learning, author=Zhu, Yanqiao and Xu, Yichen and Yu, Feng and Liu, Qiang and Wu, Shu and Wang, Liang, journal=arXiv preprint arXiv:2006.04131, year=2020