[Arxiv 2020] Distance-wise Graph Contrastive Learning [paper] [code]

Node/Graph Tasks:
Node tasks - Node classification

Training Type:
Contrastive learning with joint training

Pretext task data: hybrid, structure - graph topology, feature - node features, labels - node types
The pretext tasks include perturbation of graph topology and sampling in contrastive learning based on **graph topology, node features and node labels**. In perturbation of graph topology, nodes with lower TIG values are augmented with a higher probability. In contrastive learning, positive and negative pairs are selected based on local/global topology distance and initial embedding distance. The TIG values and the global topology distance are calculated based on Group PageRank, which contains information of graph topology and node labels.

Initial short summary here
The observation that nodes with further topological distance to the labeled nodes are more likely to be misclassified indicates the uneven distribution of the ability of GCN to embed node features in the whole graph. However, existing GCL methods ignore this uneven distribution, which motivates the authors to propose the Distance-wise Graph Contrastive Learning (DwGCL) method that can adaptively augment the graph topology, sampling the positive and negative pairs, and optimizing the similarity.

The Group PageRank matrix $\mathbf{Z} \in \mathbb{R}^{n*k}$ is created by Eq. (6) where $\mathbf{Z}_{ij}$ measures the supervision influence of category j on node i. Further the topology information gain (TIG) is calculated based on Group PageRank and node features to describe the task information effectiveness that the node obtains from information source along the graph topology. By rank the performance of GCN encoder on nodes according to their TIG values with/without CL, the authors find that CL mainly improves the performance on nodes that are topologically far away from the labeled nodes.

$$\mathbf{Z} = \alpha(\mathbf{E} - (1-\alpha)\mathbf{A}')^{-1}\mathbf{I}^* \tag{6}$$

$$\mathbf{I_i^j} = \begin{cases} 1/\,|\mathbf{L}_c|\,, \text{if the node } j \text{ is a labeled node of class } c \\ 0, \text{otherwise} \end{cases} \tag{7}$$

where $\mathbf{E}$ is the unit matrix; $\mathbf{I}^* \in \mathbb{R}^{n*k}$ is the concatenation of k teleport vectors $\mathbf{I}_i \in \mathbb{R}^n, i \in \{1,2,...,k\}$; $\mathbf{L}_c$ denotes the set of labeled nodes with class c.

Based on the above finding, the authors 1) perturb the graph topology by augmenting nodes according to their TIG value, 2) sampling the positive and negative pairs considering local/global topology distance and node embedding distance, and 3) assigning different weights to the self-supervised loss for different nodes based on their TIG ranks. Results demonstrate that CL modules improve the model performance by leveraging the information from unsupervised(unlabeled) nodes by comparing the perturbed representations. The DwGCL increase the model performance by paying more attention to the nodes that re topologically far away from the labeled nodes in graph-based SSL.

Bibtex:
@miscchen2020distancewise, title=Distance-wise Graph Contrastive Learning, author=Deli Chen and Yanyai Lin and Lei Li and Xuancheng Ren. Peng Li and Jie Zhou and Xu Sun, year=2020, eprint=2012.07437, archivePrefix=arXiv, primaryClass=cs.LG