[NeurIPS 2020] Graph Contrastive Learning with Augmentations [paper] [code]

Node/Graph Tasks: Node classification(mentioned in the Appendix G, although the Appendix G does not appear in the arXiv version of the paper) and graph classification

Training Type: self-supervised pretraining by completing different tasks and then supervised fine-tuning (or supervised SVM)

Pretext task data: graph topology, node features The pretext task here is to train a GNN-based encoder to maximize the mutual information between positive pairs and minimize the mutual information between negative pairs. The positive pairs refer to the same graph generated by different augmentations while the negative pairs refer to the different graphs generated by different augmentations. The augmentations include node dropping (randomly selecting nodes to drop), edge perturbation (randomly adding or removing edges), attribute masking (randomly masking node attributes) and subgraph extraction (randomly sampling subgraph using random walk)

Initial short summary here To tackle the challenge of less annotated labels required in the traditional supervised learning GNN methods, this paper develops contrastive learning with augmentations for GNN pretraining to address the challenge of data heterogeneity in graphs, among which four graph augmentations are proposed and their priors imposed on the graph data and their performance improvement are systematically analyzed and demonstrated on four datasets including two social networks and two biological networks. Furthermore, a novel graph contrastive learning framework (GraphCL) for GNN pretraining is raised and its performance is demonstrated to be comparable to the state-of-the-art performance in the settings of semi-supervised learning, unsupervised representation learning and transfer learning. Besides, the GraphCL boosts the robustness of GNN-based learning models against adversarial attacks.

Following the traditional procedures of feature aggregation and transformation schemes in GNNs, the propagation of the kth layer is represented as:

$$\mathbf{a}_n^k = \text{AGGREGATION}^k(\{\mathbf{h}_{n'}^{k-1} : n' \in \mathcal{N}(n)\}), \tag{28}$$

$$\mathbf{h}_n^k = \text{COMBINE}^k(\mathbf{h}_n^{k-1}, \mathbf{a}_n^k), \tag{29}$$

$$f(\mathcal{G}) = \text{READOUT}(\{\mathbf{h}_n^k : v_n \in \mathcal{V}, k \in K\}), \quad \mathbf{z}_\mathcal{G} = \text{MLP}(f(\mathcal{G})), \tag{30}$$

where $\mathbf{h}_n^k$ is the embedding of the vertex v_n at the kth layer with $\mathbf{h}_n^0 = \mathbf{x}_n$, $\mathcal{N}(n)$ is the neighborhood set of v_n. After k layer propagation, the output embedding for $\mathcal{G}$ is summarized via any permutation-invariant READOUT function over the final layer embeddings followed by a multi-layer perceptron (MLP).

Graph contrastive learning is performed by maximizing the agreement between

two augmented views of the same graph via a contrastive loss in the latent space. First two views i, i' are generated for each graph in the sampled batch graphs. Then graph representations $\mathbf{z}_i, \mathbf{z}_{i'}$ are extracted by applying a GNN-based encoder and a readout function for each graph. The normalized temperature-scaled cross entrophy loss is applied then:

$$\mathcal{L} = -\frac{1}{n} \log \frac{\exp(\text{sim}(\mathbf{z}_{n,i}, \mathbf{z}_{n,j})/\tau)}{\sum_{n'=1,n' \neq n}^{N} \exp(\text{sim}(\mathbf{z}_{n,i}, \mathbf{z}_{n',i})/\tau)} \tag{31}$$

where τ is the temperature parameter.

Four augmentation tasks include node dropping which assumes that missing part of vertices does not affect the semantic meaning of $\mathcal{G}$; edge perturbation which assumes that the semantic meaning of the graph has certain robustness to the edge connectivity pattern variances; randomly masking attributes which prompts models to recover masked vertex attributes using their context information; randomly samples a subgraph which assumes that the semantics of the global graph can be preserved much in its local structure.

Numerical results demonstrate several meaning and important insights. First data augmentations are crucial in graph contrastive learning. Composing different augmentations avoids the learned features trivially overfitting low-level shortcuts, making features more generalizable. Edge perturbations benefits social networks but hurts biochemical molecules. Applying attribute masking achives better performance in denser graphs. Node dropping and subgraph are generally beneficial across datasets.

Bibtex:

@articleyou2020graph, title=Graph contrastive learning with augmentations, author=You, Yuning and Chen, Tianlong and Sui, Yongduo and Chen, Ting and Wang, Zhangyang and Shen, Yang, journal=Advances in Neural Information Processing Systems, volume=33, year=2020