Self-Supervised Learning for Graph Data

Summary of the paper - https://arxiv.org/pdf/2103.00111.pdf

Contents

Introduction

- Deep Learning has been a subject of interest in solving a lot of complex machine learning problems, more recently on graph data
- Most solutions rely on labelled data, causing over-fitting and overall weak robustness
- Self-Supervised Learning is a promising solution which mines useful information from unlabelled data

Relevance of self-supervised learning for graph data

- Self-supervised learning (SSL) helps in understanding structural and attributive information that is present in the graph data which would otherwise be ignored when labelled data is used
- Getting labelled graph data is expensive and impractical for real world data
- SSL help learn generalized information in unlabelled graph data by performing pretext tasks.
- A **pretext task** is a combination of supplementary tasks which help in obtaining supervision signals without labelled data

Graph data and some definitions

- Graph data is a data structure which can be understood as a set of nodes and edges
- The three types of downstream graph analysis tasks are node-level tasks, link-level tasks and graph-level tasks
- **Node-level** tasks are related to nodes in the graph, for example, node classification
- **Link-level** tasks focus on edges and representation of nodes, for instance, link prediction
- **Graph-level** tasks target the representation of graphs, for example, prediction of properties of a single graph based on other graphs

Self-supervised training schemes

- Based on the relation between graph encoders, self-supervised pretext tasks and downstream tasks, self-supervised training schemes can be classified into 3 categories
  - Pre-training and fine-tuning, where the encoder is pre-trained with pretext tasks and later fine-tuned with specific downstream tasks
  - Joint learning is a scheme where the encoder is pre-trained with both pretext and downstream tasks together
  - Unsupervised representation learning, where the encoder is first pre-trained with pretext tasks and then parameters of the encoder are frozen when the model is trained with downstream tasks

Types of graph self-supervised learning

- Four different categories of self-supervised learning for graphs are
  - Masked Feature Regression (MFR)
  - Auxiliary Property Prediction (APP)
  - Same-Scale Contrasting (SSC)
  - Cross-Scale Contrasting (CSC)
  - Hybrid self-supervised learning
- In the next slides, we will look at the different approaches in detail

Masked Feature Regression (MFR)

- This technique is used in computer vision to restore damaged images by filling the masked pixels of the image
- In graph data, the features of nodes and edges are masked with zero or other tokens
- The goal of this approach is to use GNNs to recover the masked features based on the unmasked data
- Existing methods based on this approach include
    - Masked node feature regression for graph completion
    - AttributeMask
    - AttrMasking
    - Reconstruction techniques

Auxiliary Property Prediction (APP)

- This branch is used to understand the underlying graph structural and attributive information. Both classification and regression based approaches are used in this branch
- **Regression-based Approach**, where the properties of unlabelled nodes are predicted based on the representative properties of nodes learned
- **Classification-based Approach** relies on constructing pseudo labels during training and using these self-supervised labels to group the rest of the nodes

Same-scale contrasting (SSC)

- This branch of methods learn by predicting the similarity between two elements in a graph, for examples, node-node contrasting or graph-graph contrasting. The different approaches based on this method include -
- **Context-Based Approaches** (C-SSC), where the idea if to pull the contextual nodes closer in the embedding space
- **Augmentation-Based Approaches** (A-SSC), where augmented data samples are generated by this method from original data samples and samples from the same source are regarded as positive pairs, while the samples from different sources are regarded as negative pairs

Cross-Scale Contrasting (CSC) and Hybrid self-supervised learning

- In this **cross-scale contrasting** (CSC), the representations are learned by contrasting different elements in a graph, for example, node-graph contrasting, node-subgraph contrasting
- In **hybrid learning**, instead of using a single approach, different types of pretext tasks are combined for a better performance
  - For example, GPT-GNN combined MFR and C-SSC into a graph generation task to pretrain a GNN
  - Graph-Bert uses node feature reconstruction (MFR) and graph structure recovery (C-SSC) to pretrain a graph transformer model

Challenges

- All the existing methods rely on either intuition or empirical experiments. A strong theoretical foundation for graph SSL will minimize the gap between empirical SSL and graph theories
- Because there are many augmentation-based approaches for graph SSL, the data augmentation schemes should be explored further
- Existing approaches are mostly for attributed graphs, only a few of them focus on complex graphs.
- It would be promising to have more pretext tasks designed for complex graphs and more ubiquitous graphs

Conclusion

- Graph Self-supervised Learning is an interesting topic to explore as most of the data is graph structured and generally unlabelled
- Approaches like these help provide better generalization and robust models
- Using these methods, we can learn the structural and attributive information present in the graphs which would often be ignored when labelled data is used
- You can find more details about this topic in this medium article