Learning-With-Subgraphs

Skand Raturi

`GitHub Repository Link`

Introduction

Traditional Graph Neural Networks (GNNs) often process entire graphs, which can limit scalability and obscure interpretability. This project introduces a framework for **weakly supervised graph classification** that leverages **subgraph-based learning**. The key idea is to break large graphs into smaller, meaningful substructures, enabling better learning with limited supervision.

Methodology

1. BFS-Based Subgraph Extraction

- Performs breadth-first traversal from unvisited nodes.
- Controls depth and minimum size constraints.
- Preserves graph connectivity within subgraphs.

2. Sliding Window Subgraph Extraction

- Creates fixed-size subgraphs from node sequences.
- Ensures consistent subgraph dimensions.
- Captures localized neighborhood information.

3. Graph Attention Networks (GAT)

- Multi-head attention for node aggregation.
- Subgraph-level attention to rank importance.
- Global pooling for graph-level classification.

4. Attention-Based Subgraph Selection

- Top-k most informative subgraphs selected via GAT attention scores.
- Aggregated predictions improve classification robustness.

Dataset

- **MSRC_21** from TU Dataset Collection.
- Contains multiple node features, variable graph sizes, and 20 object classes.
- Designed for graph-level classification tasks.

Results and Evaluation

- **Accuracy**: Improved classification performance compared to whole-graph baselines.
- **Interpretability**: Attention maps highlight critical subgraphs influencing predictions.
- **Scalability**: Subgraph extraction reduces complexity for large graphs.
- **Visualization**: Training curves, confusion matrices, and attention-based subgraph plots provided.

Key Contributions

1. Dual subgraph extraction strategies (BFS and sliding window).
2. Integration of GAT for attention-based subgraph ranking.
3. Modular, reusable codebase with visualization support.
4. Framework optimized for **weakly supervised settings**.

Future Work

- Incorporating hierarchical subgraph structures.
- Extending to larger real-world datasets.
- Exploring alternative GNN architectures beyond GAT.
- Investigating a **hybrid BFS + Sliding Window strategy**: extracting windows within the n-depth area of a BFS-selected center node, then shifting the center to the next least-visited node (acting as stride). This aims to balance connectivity (BFS) with coverage and consistency (sliding window).