Graph-Based Node Classification using GNNs

Mohini Jasthi, Sanjeev Durge, Sivagugan Jayachandran, Prasad Upasani, Siddish Nirgude

Abstract

This project focuses on the problem of node classification in a graph using machine learning techniques. We explore multiple approaches, including Multi-Layer Perceptron (MLP) as a baseline and Graph Neural Networks (GNNs) such as Graph Convolutional Network (GCN), Graph Attention Network (GAT), and GraphSAGE. The objective is to predict node labels based on feature vectors and structural relationships.

Introduction

Graphs are an essential data structure for modeling relationships in social networks, citation networks, and recommendation systems. In this project, we aim to classify nodes in a given graph dataset, which consists of 2480 nodes, 10100 edges, and 7 classes. Each node is represented by a feature vector, and our goal is to predict the labels for test nodes using machine learning techniques that incorporate both node attributes and graph topology.

Proposed Methodology

We plan to implement and compare the following models for node classification:

Baseline: Multi-Layer Perceptron (MLP)

A simple MLP that only uses node features and disregards the adjacency matrix will serve as our baseline. This provides a benchmark for evaluating the effectiveness of graph-based models.

Graph Neural Networks (GNNs)

To leverage graph structure, we will explore the following GNN architectures:

- **Graph Convolutional Network (GCN)**: Uses convolutional operations on graphs to aggregate information from neighboring nodes.
- **Graph Attention Network (GAT)**: Employs attention mechanisms to assign different importance levels to neighboring nodes.
- **GraphSAGE**: Samples and aggregates node neighborhood information to create embeddings.

We anticipate that GNN-based models will outperform MLP due to their ability to capture graph structure and relational dependencies.

Expected Outcome

We expect GNN-based models to achieve higher accuracy than the MLP baseline by incorporating graph structure. Our final submission will include predicted labels for the test nodes, aiming for a high classification accuracy.