Project Proposal Group 26

Name: DENG, King Ho (20960866)

YUEN, Tsz Cheuk (20957326)

Topic: Movie Recommender system

Objective:

The objective of developing a movie recommendation system using a Graph Neural Network is to enhance the accuracy and personalization of recommendations by effectively capturing complex relationships between users, movies, genres, and metadata. GNNs can represent these interactions as nodes and edges in a graph structure, allowing the model to learn from both direct and indirect relationships, such as shared genres, favorite actors, or similar viewing patterns. Unlike traditional recommendation systems, GNNs excel at incorporating multi-hop connections, meaning they can discover deeper associations between movies and users, improving recommendations even when data is sparse, or user-movie interactions are limited. This approach ensures that recommendations are not only based on explicit preferences but also on implicit connections between users and movies, leading to more accurate and meaningful suggestions. This system should be able to drive higher engagement, satisfaction, and discoverability of movies by better matching users with content they might enjoy.

Background:

Recommender systems are essential for personalizing user experiences on platforms like Netflix, Amazon, and YouTube. Traditional recommendation algorithms, such as collaborative filtering, suffer from limitations in effectively capturing the relationships between users and items. A more recent and powerful approach is to represent users, items, and their interactions as a graph, where machine learning models like GNNs can be used to model these complex relationships.

In this project, we will focus on creating a **movie recommendation system** where:

- **Nodes**: represent users and movies.
- **Edges**: represent user-movie interactions (e.g., ratings given by users).

Using this graph structure, we will address the following key task:

- **Recommending new movies**: Predicting which movies a user is likely to watch based on their past interactions.

Experiment Plan

Dataset:

The MovieLens dataset is a widely used benchmark dataset in the field of recommender systems and machine learning research. It was developed by GroupLens Research at the University of Minnesota and contains various versions with different sizes, offering movie ratings, metadata, and user information. The primary purpose of the dataset is to enable the development and testing of movie recommendation algorithms.

Experimental Tasks: Recommending New Movies

- **Objective**: Recommend movies to users that they are likely to enjoy based on their past interactions.
- **Method**: This task will be framed as a link prediction problem, where the goal is to predict the probability of new edges (i.e., user-movie interactions) forming between users and movies in the graph. Graph Attention Network (GAT) will be used to predict the likelihood of new interactions.
- **Evaluation Metric**: The model will be evaluated using Bayesian Personalized Ranking (BPR) to optimize ranking of relevant vs. irrelevant items for each user. BPR loss ensures relevant items rank higher, improving personalized recommendations. Precision and recall from the confusion matrix will also be considered for comparison.

Experiment Design

The following steps will be carried out:

1. **Step 1: Data Preprocessing**: Convert the MovieLens dataset into a graph representation, where users and movies are nodes, and interactions (ratings) are weighted edges.
2. **Step 2: Model Training**: Train GCN, GAT, and LightGCN models using the training set. We will also tune hyperparameters based on the validation set.
3. **Step 3: Evaluation**: Evaluate the models on the test set using Root Mean Square Error (RMSE) for rating prediction and BPR for ranking movie recommendations. The performance of the graph-based models will be compared to traditional matrix factorization methods to highlight potential improvements in recommendation accuracy and user preference ranking.
4. **Step 4: Model Comparison**: Compare the graph-based models with traditional approaches to understand the added benefits of graph-based learning in movie recommendation.

Feasibility Analysis

The project is feasible within the available time and resources. The **MovieLens** dataset is well-documented and frequently used for benchmarking movie recommendation systems. With popular graph-based libraries like **PyTorch Geometric** and **DGL**, implementing and experimenting with GNNs is straightforward. Additionally, by leveraging pre-built models and publicly available datasets, the project is highly practical and achievable.