

Movinder: Movie Recommendation System for a Group of People Project Summary

Using the wealth of data available on user preferences, researchers and online streaming companies have extensively researched and deployed movie recommender systems. Most recommendation algorithms process the preferences of each user, and potentially those of other similar users, to predict other movies they might like. However, watching movies is often a social activity shared by multiple actors, such as a group of friends. The social dimension of watching movies contradicts common approaches that strive to satisfy one user at a time. In this project, we propose a movie recommender system that aims to maximize the collective satisfaction of a group of users.

In order to implement our movie recommender system, we use the [MovieLens dataset](#). The data contains 100K ratings from 1K users on 1.7K movies and has been used traditionally for recommender system research. Based on the data, we construct two graphs as follows: i) a *user graph* that models the similarity of users based on their personal details (age, gender, occupation and location), and ii) a *movie graph* that models the similarity of movies based on their characteristics (release date, genre and title). Nodes have a vector signal that corresponds to the respective ratings. We plan to enrich the movie graph with features extracted from the [IMDb dataset](#) (Internet Movie Database) which contains additional information such as the cast of the movie, production companies and relationships between movies. The graph structure is used to generalize the partial information provided by the ratings.

As mentioned, the main goal of our project is to build a recommendation system using graph neural networks. Moreover, the dataset will enable us to have some insights about movie tastes in different countries and among different age groups. Using various *dimensionality reduction* and *clustering algorithms*, we may reveal any polarization in the graph.

In order to recommend the correct movies to the users, we need to define a *similarity metric* between the two movies (graph where movie is a node) or the two users (graph where user is a node). Therefore, since it is not known to be a familiar distance (e.g. Euclidean), we would need to learn this distance metric which will represent the edges of the sparsely connected similarity graph. To do that, we would need to create a dataset of ground truths containing similar movies/users from which the neural network will learn the similarity metric. Ideally, with this we would be able to learn not only the similarity metric, but also the embedding space where the movies/users will be visually close to other similar movies/users.

In addition, by determining the type of graphs, we expect to determine which graph has a better results when used as a recommender tool. If we have enough time, we are aiming to build a basic web app for movie recommendations for groups of people.

