Movie Recommender System

Aditya Kumar 210070003
Azeem Motiwala 210070018
Sahil Dharod 210070026
Chanakya Varude 210070092

April 28, 2023

1 Introduction

With the increasing number of movies available on various streaming services, users often find it challenging to decide which movie to watch next. Movie recommender systems have been developed to provide personalized recommendations based on users' preferences and viewing history. This paper presents a movie recommender system that utilizes collaborative filtering and matrix factorization techniques to offer accurate and relevant recommendations to users.

1.1 Motivation

Recommender systems have become increasingly popular and important in recent years, particularly in the e-commerce and entertainment industries. In the movie industry, recommender systems have been shown to increase user engagement and satisfaction, as well as revenue for movie streaming services. The goal of our movie recommender system is to help users find movies they will enjoy, while also improving the overall user experience and loyalty towards the streaming service.

1.2 Objectives

The main objectives of our movie recommender system are to:

- Provide accurate and relevant movie recommendations to users based on their preferences and viewing history
- Improve user engagement and satisfaction with the streaming service

- Increase revenue for the streaming service by encouraging users to watch more movies

2 Related Work

Collaborative filtering is a popular technique that generates recommendations based on users' behavior and preferences. This technique assumes that users who have similar viewing histories and preferences will have similar opinions on movies. Collaborative filtering can be further divided into two categories: user-based and item-based. User-based collaborative filtering recommends movies to a user based on the preferences of other users who have similar tastes. Item-based collaborative filtering recommends movies based on the similarities between the movies themselves.

Content-based filtering suggests movies based on the characteristics of the movies themselves, such as genre, director, or actors. This technique assumes that users will prefer movies that are similar to the ones they have enjoyed in the past.

Hybrid filtering combines both collaborative and content-based filtering to provide recommendations. This technique has been shown to outperform either technique used alone, as it can overcome some of the limitations of each individual technique.

3 Proposed Method

Our movie recommender system combines collaborative filtering techniques with content based filtering to provide accurate and relevant recommendations to users. The system takes into account both user preferences and similarities between movies to generate recommendations.

3.1 Data Collection and Preprocessing

We implemented our system using the MovieLens dataset, which contains over 100,000 ratings of movies by users. The dataset includes information about the movies, such as the title, year of release, and genre, as well as the ratings given by users. We preprocessed the data by removing rows with lot of NaN values, rows with duplicate movie ID's and movies with too few ratings (¡20).

3.2 Content-Based Filtering

We first used content-based filtering to provide addition. We extracted features from the movie metadata, such as genre, director, and overview using basic EDA,we then combined these columns by assigning them proper weights in order to get the best output. We used the TfidfVectorizer to convert this column into a vector.

The TF-IDFVectorizer takes the TF-IDF values for each term in each document and creates a sparse matrix, where each row represents a document and each column represents a term. The value in each cell of the matrix represents the TF-IDF value of the corresponding term in the corresponding document. We then used the cosine similarity function to create a similarity matrix between all the movies ,this matrix provided us with the required measure of similarity between the movies. We selected the top i.e. the most similar 30 movies and then ordered them according them to their respective ratings to provide us with the desired output.

3.3 Collaborative Filtering

We used user-based collaborative filtering to generate recommendations for user. We created a user-item matrix that represents the ratings given by users for each movie. The matrix has dimensions of $m \times n$, where m is the number of users and n is the number of movies. The matrix is sparse, as most users have not rated most movies. The ratings in the matrix range from 1 to 5, with 0 representing a missing rating. For each user, we identified the most similar users based on their ratings and preferences. We then fill the user-movie matrix entries which were not filled by the target user with average rating of all users similar to the target user.

3.4 Hybrid Filtering

To improve the accuracy and relevance of our recommendations, we combined both collaborative and content-based filtering using a hybrid filtering approach. We then took the recommendations based on a movie search and then sorted them as per the ratings given by the target user and finally recommend them, if the user/similar users haven't watched any of them we end up recommending movies only via content based filtering.

4 Conclusion

Thus, a movie recommender system that utilizes collaborative filtering techniques to generate accurate and relevant recommendations for users. We combined both user-based and content-based filtering using a hybrid approach to improve the accuracy and relevance of our recommendations. Our system achieved competitive performance on the MovieLens dataset, suggesting that it could be an effective tool for movie streaming services to improve user engagement and satisfaction.

Future work could involve exploring other techniques like matrix factorization, as well as incorporating more features from the movie metadata. Additionally, testing the system on a larger and more diverse dataset could provide more insights into the effectiveness of the system for different types of users and movies.