

Report

1. Introduction

In this assignment, we focus on building a Recommender System using the MovieLens 100k dataset. The primary objective is to implement a collaborative filtering approach with Singular Value Decomposition (SVD) as the underlying model.

2. Data analysis

The MovieLens 100K dataset contains 100,000 ratings from 1000 users on 1700 movies. The dataset contains 19 genres, with each movie being assigned to one or more genres. The distribution of genres in the dataset is uneven. The most popular genres are Drama, Comedy, Action, and Thriller. The data was read into pandas dataframes and analyzed to draw some key insights. We observed that most of the users have rewarded movies they watched with a 4 star rating and followed by 3 star and 5 star. Most of the users are also students and the number of films reviewed per user is rather uneven.

3. Model implementation

The model used for this recommender system is Singular Value Decomposition (SVD). SVD is a matrix factorization technique that can be used to predict ratings for user-item interactions. It decomposes the user-item interaction matrix into three matrices: U , Σ , and V . U represents the user latent factors, Σ represents the matrix of latent factors, and V represents the item latent factors

4. Model Advantages and Disadvantages

SVD has several advantages. It can handle large datasets and can provide a good baseline for recommendation systems. It can also be used for collaborative filtering, which is a powerful technique for recommending items based on user behavior.

However, SVD also has some disadvantages. It assumes that the user-item interaction matrix is a low-rank matrix, which might not always be the case. It also requires a large amount of data to get good results.

5. Training Process

The data used were `ua.base` and `ua.test`, which split the data into exactly 10 ratings per user in the set. So, `ua.base` was used for training and `ua.test` for validation

6. Evaluation

The evaluation metric used to assess the performance of the model were the root mean squared error (RMSE), validation loss and mean absolute

error(MAE) between the predicted ratings and the actual ratings. The best result were:

val_loss: 0.91 val_rmse: 0.95 val_mae: 0.75

7. Results

The results of the SVD model on the MovieLens100k dataset showed promising results. The model was able to provide accurate recommendations for users based on their past behavior. The model's performance was evaluated using various metrics, and the results were satisfactory.