भारतीय प्रौद्योगिकी संस्थान जोधपुर

Indian Institute of Technology Jodhpur

PRML Project

Movie recommendation system

Authors: Vatsal Dadhaniya, Prakash Nandaniya, Vekariya Sagar, Gadhiya Ronak, Pareen Shah
Guided By: Anand Mishra

April 21, 2024

Abstract

Aim of this is project is to develop a personalised movie recommendation system using traditional ML techniques. Project revolves around 4 models based on collaborative filtering using various ML techniques. In real life scenario this can be used on streaming platforms to provide recommendations.

Keywords: Collaborative Filtering, KNN, PCA, SVD, Hybrid Model

Contents

1 Introduction **3**
1.1 Background . . . 3
1.2 Major Findings and Overview of our work . . . 3

2 Approaches Tried **3**
2.1 Collaborative filtering using KNN . . . 3
2.2 Collaborative filtering using PCA . . . 4
2.3 Collaborative filtering using SVD . . . 4
2.4 Hybrid Model . . . 4

3 Experiments and Results **4**
3.1 About Dataset . . . 4
3.2 Results . . . 5

4 Summary **7**

A Contribution of each member **7**

1 Introduction

1.1 Background

Our project centers on exploring traditional machine learning (ML) models commonly used in movie recommendation systems. These models include collaborative filtering[2] techniques such as user-based and item-based filtering, which leverage similarities between users or items to generate recommendations. User-based filtering identifies users with similar preferences and recommends movies based on their collective behavior, on the other hand item-based filtering suggests similar movies given one movie as an input.

By focusing on these established ML models, our objective is to develop a reliable movie recommendation system that provides personalized suggestions based on user behavior and preferences, without relying on deep learning architectures.

1.2 Major Findings and Overview of our work

We explored many different techniques to be used in the movie recommendation system. Majorly two types of methods are used, user-based collaborative filtering and item based collaborative filtering. User based collaborative filtering is implemented using 3 different models which uses KNN, PCA and SVD [1], similarly item based collaborative filtering is also implemented using the same models. After analysing the results of this model we were able to conclude it practically that PCA and SVD are similar techniques and hence they also gives similar results.

We have also created a hybrid model which uses K-means clustering and Random forest to generate recommendations based on user's activity as well as analysing various features of movies. By creating this model we were able to link different ML techniques in one model. This Model is discussed in details in following parts of this report.

2 Approaches Tried

‘ Following are the approaches used:

1. Collaborative filtering using KNN
2. Collaborative filtering using PCA
3. Collaborative filtering using SVD
4. Hybrid Model

2.1 Collaborative filtering using KNN

Using KNN we have implemented both user and item based collaborative filtering. Firstly, movie-user matrix is created to represent the ratings given by each user for a particular movie. By identifying the k nearest neighbors to a user's vector, we personalize movie recommendations to align with similar users' tastes. This aspect is particularly beneficial for streaming platforms, where tailored content suggestions on a user's homepage to recommend the movies on a general basis without any genre bias.

Now, this is further extended to give genre wise recommendations, for this we added only movies of particular genre in movie-user matrix and implemented KNN similarly.

On the other hand, our model gives item-based recommendations by identifying similar movies using KNN, instead of focusing solely on user preferences. This functionality proves invaluable when users search for specific movies or upon completing a movie, where recommending related content enhances the overall user experience.

2.2 Collaborative filtering using PCA

Similarly using this model also we have implemented three types, general and genre wise in user based filtering and item based filtering. To implement item based filtering, similar movie-user matrix is created. Next, we compute the covariance matrix from the normalized movie matrix to determine eigenvalues and eigenvectors. These eigenvalues and eigenvectors are pivotal in calculating cosine similarity scores. By leveraging cosine similarity, we identify the top "n" movies closely related to a given movie, thus providing tailored recommendations based on item similarities.

To implement user based filtering, for all the movies that were rated by user, item based filtering was applied and from each recommendations two were selected to give final recommendation. It was ensured that the movies rated by users were selected in descending order of rating to ensure that the given recommendations corresponds to some of the top rated movies by user. For genre specific recommendations movies were selected from the recommendations only if they belongs to that particular genre.

2.3 Collaborative filtering using SVD

SVD splits normalised movie-user matrix into three simpler matrices: one representing the relationship between users and latent factors, another representing the importance of each latent factor, and the third representing the relationship between items and latent factors. Then using the cosine similarities recommendations were given for item based filtering.

User based filtering was implemented similarly as in PCA, by applying item based filtering on the movies rated by user and giving recommendations based on its output. As known in theory, it was also practically observed that PCA and SVD are similar technique as they produced similar results.

2.4 Hybrid Model

In this model, we initially employed the K-means clustering algorithm to group users based on the ratings given by them, this ensures that user with similar test are grouped in same clusters. Then training dataset was made using the movies rated by users in that particular cluster. Random forest regressor was trained on this dataset, it was chosen as classifier because of the availability of many features in the dataset. All the extra features extracted from various sites were also used as features here along with genres.

Then for all the movies rating were predicted and recommendations were given accordingly. This two-step approach ensures that recommendations are personalized within user clusters while also leveraging the predictive power of Random Forest to enhance recommendation accuracy by taking various features in account.

3 Experiments and Results

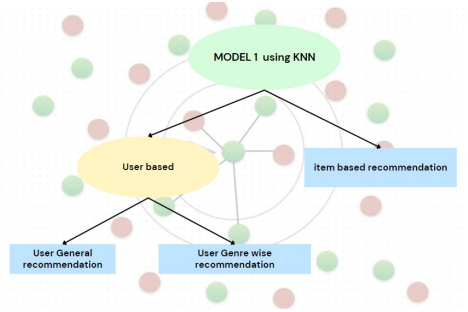
3.1 About Dataset

The dataset used in this project describes 5-star rating and free-text tagging activity from MovieLens, a movie recommendation service. It contains 100836 ratings and 3683 tag applications across 9742 movies. All selected users had rated at least 20 movies. It has 4 files links.csv, movies.csv, ratings.csv and tags.csv. Detailed description can be found in the readme file.

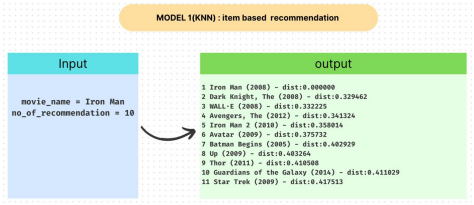
Two dataset files imdb.csv and tmdb.csv were genrated by scrapping data for all the movies from the corresponding sites to use some important features like rating, user and critic reviews etc.

3.2 Results

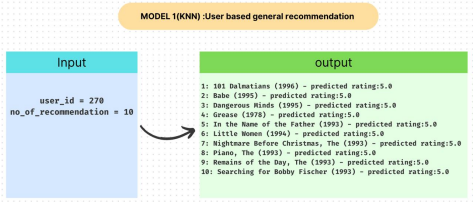
1. Collaborative filtering using KNN
    - Results from this model are shown in 1



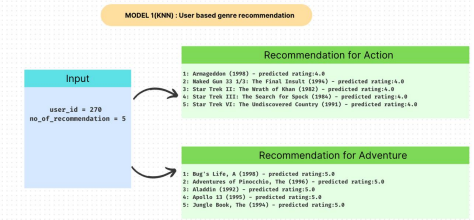
(a) model



(b) Movie(item) based using KNN



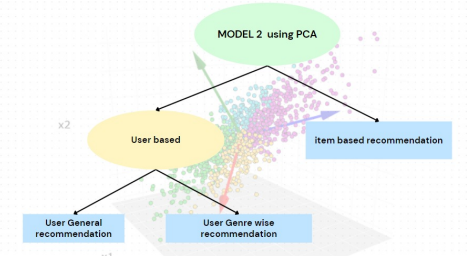
(c) User based general Recommendations



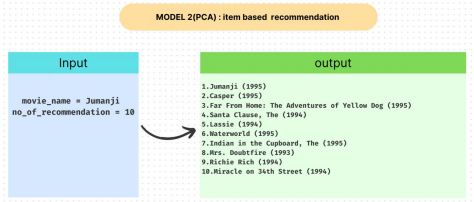
(d) User based genre wise recommendations

Figure 1: Recommendations using KNN

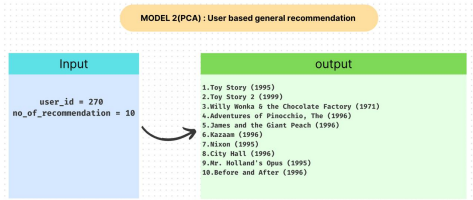
2. Collaborative filtering using PCA
    - Recommendations by this model for are shown in 2



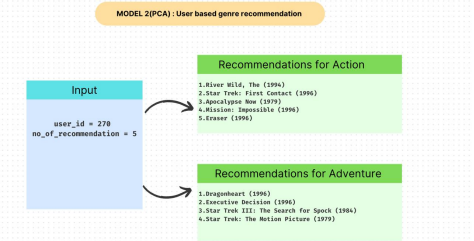
(a) model



(b) Movie(item) based using PCA



(c) User based general Recommendations

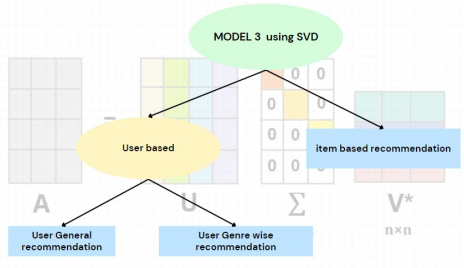


(d) User based genre wise recommendations

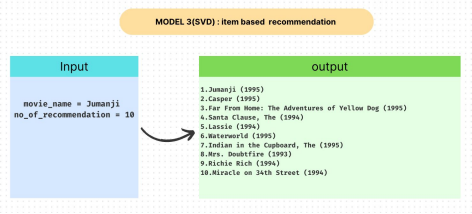
Figure 2: Recommendations using PCA

3. Collaborative filtering using SVD.

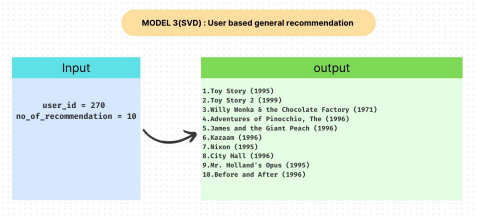
- Recommendations by this model are shown in 3



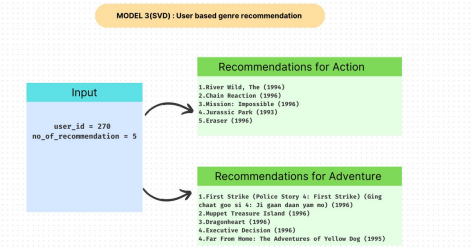
(a) model



(b) Movie(item) based using SVD



(c) User based general Recommendations

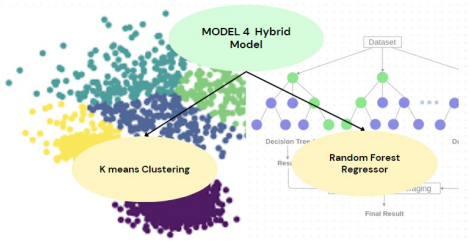


(d) User based genre wise recommendations

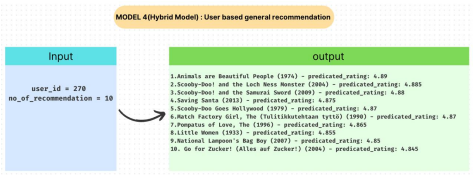
Figure 3: Recommendations using SVD

4. Hybrid Model

- Recommendation by this model for are shown in 4



(a) model



(b) Recommendation by Hybrid Model

Figure 4: Recommendations using Hybrid Model

4 Summary

The project outlines a movie recommendation system based on traditional machine learning models, specifically focusing on collaborative filtering techniques like KNN, PCA, and SVD, as well as a hybrid model combining K-means clustering and Random Forest. The system aims to provide personalized movie suggestions by analyzing user behavior and preferences. Further developments can be done by using advanced ML techniques like neural network to increase the scope and accuracy of model.

References

[1] Hervé Abdi. Singular value decomposition (svd) and generalized singular value decomposition (gsvd). URL https://www.cimat.mx/~alram/met_num/clases/Abdi-SVD2007-pretty.pdf.

[2] Yuliia Kniazieva. Introduction to the movie recommendation system architecture. URL https://labelyourdata.com/articles/movie-recommendation-with-machine-learning.

A Contribution of each member

1. Ronak Gadhiya: Data scrapping from web, main part of report and portion of project page.
2. Prakash Nandaniya: Data scrapping from web, main part of report and portion of project page.
3. Pareen Shah: Main work of project page.
4. Sagar Vekariya: Designing of models, source code writing and image designing on canva.
5. Vatsal Dadhaniya: Designing of models, source code writing and portion of report.