A Synopsis of major project on

Movie Recommender System

National Institute of Technology Hamirpur

Guided By:

Dr. Pardeep Singh

Submitted By:

Mukesh Kumar Kharita (14MI542)
Atul Kumar (14MI534)
Aarti Vaid (14MI557)
Mayank (14MI502)
Arun Sharma (14MI504)
Anil Sai. D (14MI552)

Indexes:

Abstract

A recommendation system is a type of information filtering system which attempts to predict the preferences of a user, and make suggests based on these preferences.

Movie recommendation systems provide a mechanism to assist users in classifying users with similar interests. This makes recommender systems essentially a central part of websites and e-commerce applications. Movie recommendation system's primary objective is to suggest a recommender system through data clustering and computational intelligence.Basically there will be two types of techniques that will be used in movie recommended system, Content based filtering and collaborative filtering. Collaborative filtering produces recommendations based on the knowledge of users' attitude to items, that is it uses the "wisdom of the crowd" to recommend items. In contrast, content-based recommender systems focus on the attributes of the items and give you recommendations based on the similarity between them.Basically these are information filtering tools that aspire to predict the rating for users and items, predominantly from big data to recommend their likes.

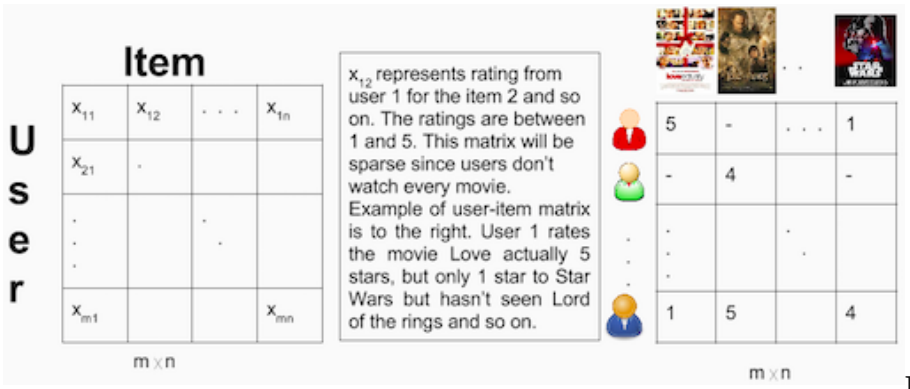
Introduction:

Now a days, *recommender systems* are used to personalize your experience on the web, telling you what to buy, what to watch, where to eat or even who you should be friends with. Recommender systems try to capture these patterns to help predict what else you might like. E-commerce, social media, video and online news platforms have been actively deploying their own recommender systems to help their customers to choose products.

Collaborative filtering:

Collaborative Filtering techniques make recommendations for a user based on ratings and preferences data of many users. The main underlying idea is that if two users have both liked certain common items, then the items that one user has liked that the other user has not yet tried can be recommended to him.
There are two algorithms for Collaborative filtering, the Nearest Neighbors Algorithm and the Latent Factors Algorithm.

Nearest Neighbors Collaborative Filtering: This approach relies on the idea that users who have similar rating behaviors so far, share the same tastes and will likely exhibit similar rating behaviors going forward. The algorithm first computes the similarity between users by using the row vector in the ratings matrix corresponding to a user as a representation for that user. The similarity is computed by using either cosine similarity or Pearson Correlation.



Latent Factor Methods: The latent factor algorithm looks to decompose the ratings matrix R into two tall and thin matrices Q and P, with matrix Q having dimensions num_users × k and P having the dimensions num i tems × k where k is the number of
latent factors. The decomposition of R into Q and P is such that

$$R = Q.P \text{ ^T}$$

Any rating *Rij* in the ratings matrix can be computed by taking the dot product of row q i of matrix Q and p j of matrix P. The matrices Q and P are initialized randomly or by performing SVD on the ratings matrix.

Content Based Recommendations:

Content Based Recommendation algorithm takes into account the likes and dislikes of the user and generates a User Profile. For generating a user profile, we take into account the item profiles (vector describing an item) and their corresponding user rating. The user profile is the weighted sum of the item profiles with weights being the ratings user rated. Advantages of Content Based approach is that data of other users is not required and the recommender engine can recommend new items which are not rated currently, but the recommender algorithm doesn't recommend the items outside the category of items the user has rated.

Dataset:

We will use MovieLens dataset, which is one of the most common datasets used when implementing and testing recommender engines. It contains 100k movie ratings from 943 users and a selection of 1682 movies.
For the content filtering portion, information was scrapped from the IMDB website for the corresponding movie.

Implementation

Baseline methods :
There are some following simple baseline methods that give the idea of the performance to expect from the:-
Global Average.
User average.
Movie average.

Collaborative Filtering : We will implement the nearest neighbors and latent factor methods for Collaborative filtering. However, if we try to use the simple implementations on the larger 20 million dataset, the code breaks. If we store the ratings matrix for the 20 million dataset in a dense format, it would take up around $140000 \times 27000 \times 8 = 28GB$ of memory, assuming 8 bytes per matrix entry.

Movie Plot keyword search :
We will implement the additional feature ,that is seraching movies based on keyword searches. The IMDB movie id is provided along with the MovieLens dataset. After accessing the webpage, the metadata from the movie is collected and then pickled and stored in a dump.

Person name search :
In addition to searching the movie plot, we will implement a separate search for names of people associated with the movie. These could be the actors, the directors, or the writers.

Technologies:
We will use HTMl5, CSS, ReactJS, for front end, In the backend we will use python ,some of python's libraries and maching learning for recommendations.

Expected Result

The expected result of our project will be to provide the movie information of the best rating movies and movie information according to the keyword entered in the search bar. A unique user can give rating to the particular movie, According to the provide rating and searched movies, our system will recommend the user the movies similar to the ratings provided by the user, director, actors.

References

[1] A Survey of Collaborative Filtering Techniques; Su et al;
https://www.hindawi.com/journals/aai/2009/421425/
[2] Google News Personalization: Scalable Online Collaborative Filtering; Das et al;
https://www2007.org/papers/paper570.pdf
[3] Intro to Recommender Systems: Collaborative Filtering;
http://blog.ethanrosenthal.com/2015/11/02/intro-to-collaborative -filtering/
[4]:http://www.sciencedirect.com/science/article/pii/S1110866516300470#f0005
[5]:https://pdfs.semanticscholar.org/767e/ed55d61e3aba4e1d0e175d61f65ec0dd6c08.pdf