IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10448

Hybrid Movie Recommendation System Based on Collaborative Filtering & Content Based Filtering

ShivendraYadav[1], Sameera Khan[2]

Student, Amity School of Engineering & Technology, Amity University Chhattisgarh[1]

Assistant Professor, Amity School of Engineering & Technology, Amity University Chhattisgarh[2]

Abstract: In a growing world, where internet has everything, literally everything which makes tedious to search each and everything over and over. Like movies, series, songs and much more. A recommendation system is a specialized system, made for a specific domain, which shows similar searches, result based on our search and query. We live in an era of OTT, which made it possible that we don't need theaters and TV for movies and series altogether. The need of recommendation rose up quickly. To give recommendation, once a series or a movie is finished, based on what is just finished. Write a name of a series of movies in Google, scroll down, and in the right you will see "people also search for". That shows how recommendation systems are important. In this project, an attempt is made to make a hybrid movie recommendation system, by combining two techniques, collaborative filtering and content-based filtering.

Keyword: Content-based filtering, collaborative filtering, nearest neighbors, user-item interaction matrix.

I. INTRODUCTION

For a user to search things one after another is quite tedious. Instead how it would be like, you search for one and similar things show up as well. Because that will not only save time for you to search the next thing, but will also get to know what might not know about. And that's what a recommendation system does.

For example, when you will type Avengers in Google, and hit enter. Scroll down and you will find in "people also search for" movies like Guardian of the Galaxy, WandaVision and more. Now instead of searching specifically, these suggestions help us to reduce our search time and ease our work.

In this project, am attempt to make a hybrid recommendation is done. Two filtering techniques are together being used, at first collaborative filtering and then content-based filtering.

II. LITERATURE SURVEY

In Content-Based Movie Recommendation System Using Genre Correlation by SRS Reddy, Sravani Nalluri, Subramanyam Kunisetti, S. Ashok and B. Venkatesh [1] and A movie recommendation algorithm based on genre correlations by Sang-Min Choi, Sang-Ki Ko, Yo-Sub Han [2] talked about content-based approach in movie recommendation field. While A Collaborative Filtering Algorithm and Evaluation Metric that Accurately Model the User Experience Matthew R. McLaughlin and Jonathan L. Herlocker [3] talked about using collaborative filtering for movie recoomendation system. Many more works has been done in the field of recommendation systems which has been taken into consideration.

III. DATASET

For making this recommendation system, dataset from grouplens site is taken, which contains many dataset both for educational and research process.

The name of the dataset used if called MovieLens](http://movielens.org)and in that MovieLens 25M dataset is used to make this project. The dataset is in the form of csv files.

IV. MOVIE RECOMMENDATION SYSTEM

A movie recommendation system is a system in which recommends movies based on your given input. There are many filtering ways so as to implement a recommendation system: content-based filtering, collaborative filtering, and hybrid model. In collaborative filtering, there are two types of model, memory based and model based. A hybrid model is a model, in which combination of all models is used so as to achieve a unique model.

V. CONTENT BASED FILTERING

A content-based filtering relies on similarity on items (in our case, similarity in movies). The basic idea of content-based filtering is that, if you like an item, most probably you will like an item similar to that item [1].

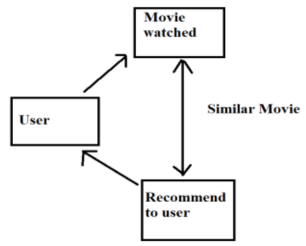


Fig 1. Content Based Filtering

As shown in diagram, if a user has watched a movie, and there is a similar movie, then recommending it to the user is beneficial as it is of similar type of what the user has just finished.

In content-based filtering, the emphasis on item taken into consideration and then similar items are searched [2]. The similarity can be based on item to item. Like in movies, if we want to get similarity, we will go for genre. If we talking about sings too, to find similar songs, we can either go for the same singer's other songs, or same genre songs.

VI. COLLABORATIVE FILTERING

Collaborative Filtering is a method of making automatic predictions about the interest of a user by collecting preferences from many users [5]. The base of collaborative filtering is that if A and B buy similar products, A is more like to buy the product B has bought, compared to any other random person [9]. There are two types of collaborative filtering, memory based and model based.

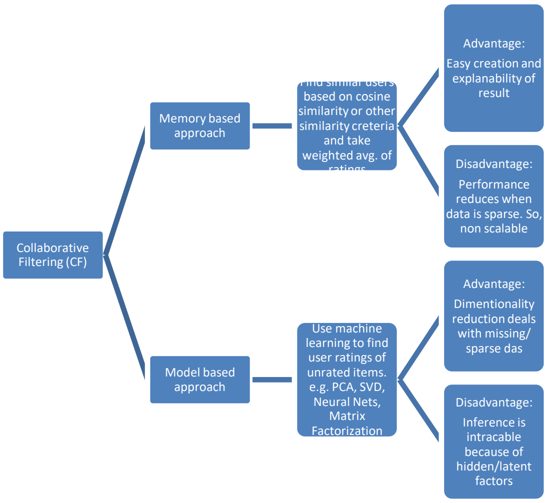
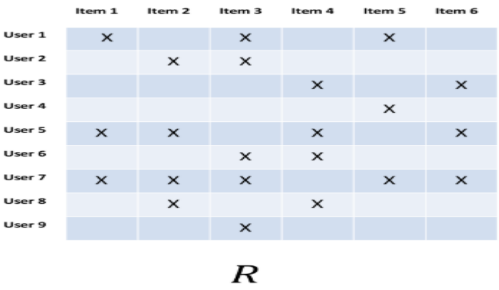

Collaborative Filtering (CF)
Memory based approach
Find similar users based on cosine similarity or other similarity creteria and take weighted avg. of ratings
Advantage: Easy creation and explanability of result
Disadvantage: Performance reduces when data is sparse. So, non scalable
Model based approach
Use machine learning to find user ratings of unrated items. e.g. PCA, SVD, Neural Nets, Matrix Factorization
Advantage: Dimentionality reduction deals with missing/ sparse das
Disadvantage: Inference is intracable because of hidden/latent factors

Fig 2. Collaborative filtering and its types

Our model is a memory based one and would be proceeding with user-based collaborative filtering approach. Collaborative filtering systems focus on the relationship between users and items.

VII. USER-ITEM INTERACTION MATRIX

The user-item interaction matrix defines the interaction between the users to the item. Interaction matrices are usually sparse since users only interact with some items. (Considering the size of the matrix is too big, which in reality usually found).



	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	×		×		×	
User 2		×	×			
User 3				×		×
User 4					×	
User 5	×	×		×		×
User 6			×	×		
User 7	×	×	×		×	×
User 8		×		×		
User 9			×			

R

Fig 3. Interaction Matrix

The image shows a user-item interaction matrix which is a part of user-based collaborative filtering. In user-based collaborative filtering, it is considered that a user will like the items that are liked by the users with whom it shares similar taste [8].

VIII. NEAREST NEIGHBORS

Nearest Neighbor is a supervised algorithm. The principle behind nearest neighbor method is to find a predefined number of training samples closest in distance to the new point, and predict the label from these. It is useful to take more than one neighbor into account so the technique is more commonly referred as k-nearest neighbors [4].

Despite its simplicity, nearest neighbors has been seen successful in a large number of classification and regression problems. The below figure depicts a 3-Nearest Neighbor classifier on a two-class problem in a two-dimensional feature space [4].

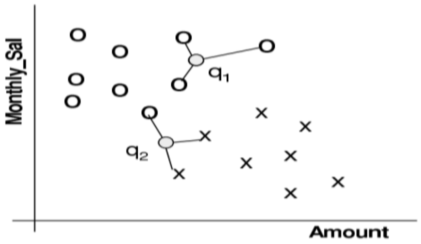


Fig 4. Nearest Neighbor Classifier

IX. PRE- PROCEDURE

So as to proceed, we narrowed down our dataset by implying the following constraints

- Only those movies are considered, who have been rated at least 500 times
- To narrower it down, users have been filtered in the way that, only those users are taken into consideration, who have rated at least 50 movies

DOI 10.17148/IJARCCE.2021.10448

All these constraints were done so as to narrow down our space search to avoid as many outliers as we can which might disrupt our final output.

X. PROCEDURE

In procedure first collaborative filtering is used and then content-based genre score.

Our procedure starts with taking two inputs from the user, one of them being the names of the movies and second, the ratings of those movies. This step all along is a process to know the user, know its preferences. From our dataset, we will find those movies and see what genre they belong to. We make our nearest neighbor model with 11, brute and Euclidean as parameters. Where 11 represents number of neighbors, brute represents algorithm and Euclidean as metric.

Now we find nearest neighbors with the help of our model. We gather 10 similar users and do the following:

- We remove the movies names already watched by our input users.
- Remove all those movies whose ratings are less than 1.

Now the work of collaborative filtering is over. We have now four things right now,

- Users' input movie names
- User's input movie ratings
- Unwatched movies names for recommending to user
- Ratings of all those movies.

Now we take into consideration of users input movies genre scores and multiply it by the ratings. Now we sum column wise so as to get which genre is more watched by our user. And now we normalize the rating.

Now in the next step, we take into consideration of unwatched movies genre scores and multiply it with our normalized ratings. And then sum all those ratings and assign them to the respective unwatched movies.

Now these ratings are the ratings that we got taking into consideration of both genre and the movie's ratings itself. And with this we recommend ten movies with the highest ratings to our user.

XI. ADVANTAGES

As this system takes both collaborative filtering and genre score into consideration, we try to get more accurate result [10].

As collaborative filtering helps in find similar users and we extract unwatched movies from it. Using genre score on unwatched movies helps in suggesting best movie genre wise in the list.

So, the suggestions aren't only based upon ratings, but also the genre of the movies.

XII. DRAWBACKS

As using collaborative filtering, we encounter the problem of cold-start. Cold-start problem defines the difficulty to recommend things when the user is new [2]. Collaborative filtering needs some amount of information so as to find neighbors and recommend. But if a user is new, as there's no data of past behavior or interest, it becomes hard to recommend something. If there are not enough data, then the system becomes very unreliable because of the cold start problem.

Next problem is the sparsity; this occurs when optimal amount of data isn't available about the user. As this leads to many entries to be 0 in user-item similarity matrix and this will lead to less accurate recommendation compared to when optimal amount of data is made available.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10448

XIII. CONCLUSION

A hybrid recommendation system is a system which is made by combining of two or more methods. Hybrid systems are advantageous as it uses multiple techniques from all the base methods so as to produce optimal output and too can overcome individual methods drawbacks. In this attempt, a hybrid system using collaborative filtering and content-based filtering (we used genre score) is made. In future, more new techniques can be added to replace with new techniques or inevitably remove one of the techniques.

XIV. REFERENCES

1- Content-Based Movie Recommendation System Using Genre Correlation by SRS Reddy, Sravani Nalluri, Subramanyam Kunisetti, S. Ashok and B. Venkatesh
2-A movie recommendation algorithm based on genre correlations by Sang-Min Choi, Sang-Ki Ko, Yo-Sub Han
3-A Collaborative Filtering Algorithm and Evaluation Metric that Accurately Model the User Experience Matthew R. McLaughlin and Jonathan L. Herlocker
4-k-Nearest Neighbour Classifiers P´adraig Cunningham and Sarah Jane Delany
5-A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research M Viswa Murali, Vishnu T G, Nancy Victor.
6-Application of Content-Based Approach in Research Paper Recommendation System for a Digital Library by Simon Philip, P.B. Shola, Abari Ovye John.
7-Recommender system techniques applied to Netflix movie data, Research Paper Business Analytics by Steven Postmus
8-Content-based Recommender System for Movie Website by KE MA
9-A Review Paper On Collaborative Filtering Based Moive Recommedation System by Nirav Raval, Vijayshri Khedkar
10-Movie Recommender engine using Collaborative filtering by Howal Sadanand, Desai Vrushali, Nerlekar Rohan , Mote Avadhut , Vanjari Rushikesh, Rananaware Harshada