THINK INDIA JOURNAL

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

Movie Recommendation System Using Collaborative Filtering Method

Vasudha
(Assistant Professor)
Faculty of Computer Science & Technology
Lovely Professional University,Punjab, India
vasudha.23636@lpu.co.in

Acharya Thandava Krishna
(Student)
Lovely Professional University, Punjab, India
acharya.thandava@gmail.com

***Abstract*- The world is generating tons of data every second in a wide variety of forms like structured or unstructured. Every business wants to leverage the data that they produced and give a fantastic experience to its users. In particular, Recommendation systems are part of it to catch the user's pulse and increase the companies revenue. In this paper, we will discuss the power of recommendation systems and their help to businesses which rely on them. We implemented the collaborative filtering algorithm in R Programming and User Interface in Shiny R.**

Keywords: Recommender Systems, Collaborative filtering, Content based Model, K nearest neighbour

I. INTRODUCTION

Entertainment is the main factor in everyone's life to get away from the busy schedule in life. The person will look for options to release his/her stress. Movies are one of the options most people will love to choose. The film industry, on average, producing 2500+ feature films every year worldwide. It's been one of the largest entertainment industries in the world. The American media service provider Netflix alone has more than 13000+ titles in its catalog and updating over time. Apart from Netflix, we have many service providers like Prime Video, Hulu, CBS, HBO, and many more.

It's a nightmare situation for users to select a movie from a pile of collections; this is where recommender systems come into play. The system suggests the best movies to users, and these recommendations are given solely based on their past behavior. The recommender systems save a lot of time to the users for watching their favorite shows or movies rather than searching for the best choice from the vast collection.

Nowadays, the recommender system became a vital role in every industry. Many users love this feature and attracted to the website, which is offering this feature. Thus service providers taking the leverage of this feature and continually upgrading their systems to the latest technology to increase their revenue rapidly. These recommender systems are beneficial to both parties. For users, it saves a lot of time to pick a movie to watch, and for service providers, it generates income by increasing the customer base.

II. TYPES OF RECOMMENDER SYSTEMS

The main idea behind the recommender system is to recommend similar items to users who will interact with the system. To complete this task, there are two types of methods available which are content based recommender system and collaborative filtering recommender system. Let's discuss about these two recommender algorithms.[10]

A. Collaborative filtering Recommender System

Collaborative filtering recommender systems is a type of method, the result which is purely based on the past behaviour of the users to the items to generate the new suggestions of the users which will interact with the system. These interactions will be stored in the "user-item interaction matrix".[3]

Copyright © 2019Authors

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

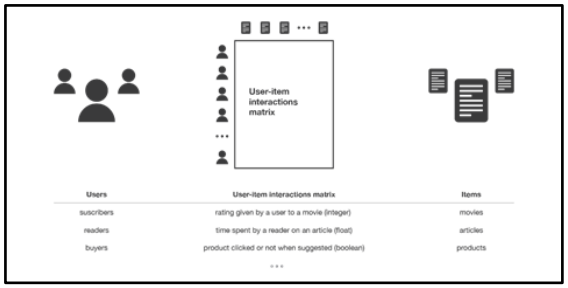


Fig.1.User-item interaction matrix illustration.

The main idea that runs behind the collaborative filtering recommender systems are the past interactions between the user and item is sufficient to generate future predictions to detect similar users and suggest recommendations.

Again the CF recommender algorithms further divided into further two subtypes. These two subtypes are called memory based and model based system. The memory based system will work with the values of past interactions, assuming that no model, and it is based on nearest neighboursfor example, to find the closer users from a user interactions and suggest the popular items among these closer neighbours.

Now Model based system assumes that an underlying suggestive model which explains the past interactions between users and items and try to find that interactions to generate new suggestions.

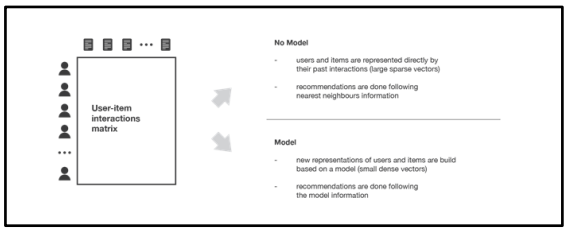


Fig.2.Collaborative filtering recommender methods

The main use of the collaborative filtering method is that the system does not require any data about the items or users so that they can be used in any conditions. We get more accurate new recommendations when more users interact with items.

As we know that it only consider the past interactions of users to items, to predict future suggestions, collaborative filtering method suffers from the issue called “cold start problem”. In matrix the users who have few interactions in past, it is not possible to recommend the new items.

This disadvantage can be solved in a different way by recommending random items to users or items to random users, suggesting the popular items to users or items to the most regular users, suggesting a set of items to new users or new items to a set of users or, finally, using a content based method for the early stage of the user or the item. [1]

B. Content based recommender system

As we discussed earlier the collaborative filtering methods will only depend on the past behaviour between the users and items, content-based recommender methods use extra data about the users.[4]Let's take an example of a movie recommendation system, this extra data can be, gender,age, location, occupation etc for users data as well as the movies data, the lead cast, the duration of the movie or release country, are the characteristics for the movies which are items in our case.

The main theme of the content based algorithm is to create a model, which is created on the available attributes in the data of user-item interaction matrix. Taking into account users and items, we will try, for example, to mock-up the fact that some movies rate better by young women, and some other movies rate better by young men. If we managed to create such type of model, then, generating new suggestions for a user will be easy for us. We need to look at the data of this users and from this data. We need to find similar movies to recommend for users.

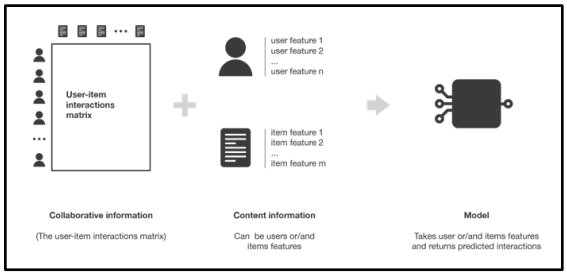


Fig.3.Content based recommender methods

Content based recommender system will not suffer from the cold start problem like collaborative filtering methods.The new users or new things can be portrayed by their content, thus important recommendations will make for new elements. the new users or items which the model does not seen features in the past will suffer from this problem, but when our system is old enough, this problem will not occur.

Copyright © 2019Authors

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

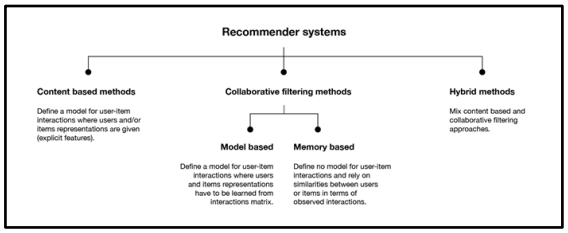


Fig.4.Overview of types of recommender systems

III. IMPLEMENTATION & RESULT

A. DATASET

The MovieLens Full Dataset was readily available at the GroupLens Website. This dataset contained 100k ratings from 610 users on 9742 movies. The ratings for each movie are in the scale of 1 to 5. One file in this dataset, is links.csv, contained the TMDB and IMDB IDs for all the movies. Movies dataset consists of all the details of movies like release year, movie id and movie genre. Ratings dataset consists of user ratings given to the movies.[2]

B. DATA WRANGLING

Overview

This section describes the various data cleaning and data wrangling methods applied on the datasets to make it more suitable for further analysis.

Importing the Dataset into the Environment

The Datasets are imported to the R Studio for further analysis of the data. Two datasets were imported movies and ratings, movies dataset consists of all the details of movies like release year, movie id and movie genre. Ratings dataset consists of user ratings given to the movies.

Data Cleaning

- The Datasets are pre-processed after importing for further analysis like in movies dataset the genres column is delimited.
- Extracted release year from movie title column.
- Creating new column year which is extracted from above.
- Created search data frame to extract movies released in particular year and genre.
- Created binary ratings to the user ratings i.e. “1” for rating greater than 3 and “-1” for less than 3.
- If the Rating is 1 then the movie is liked by the user and if it is -1 then the movie is not liked by the user.
- Removed movies which are not rated by any user.

C. MODEL CREATION

Item-Based Method

Collaborative filtering method runs on the information of different users. The Collaborative word refers to the fact that users collaborate to recommend items. The algorithm takes into consideration of user ratings and their preferences.
The rating matrix which we will create, the rows correspond to the users and columns correspond to items in our case movies. This Item based algorithm work on the following steps below

1. For every item, it identifies the k, where k is the most similar items
2. For every both items, it will measure how similar they are in terms of similar ratings by similar users.
3. It finds the items that are most likely to the users interest for every user.

USER-BASED CF MODEL

According to the user-based CF model, when a new user is given, first it identifies the similar users. After that, the top rated items which are rated by the users are suggested.[5]
For every new user, below are the steps followed

1. We have to measure how much similarity is there for each user to a new user.
2. We have to find the similar users.
    a. Take into consideration of the top users for k nearest neighbours
    b. Consider a significant level and only take into consideration of the users whose similarity is above that level.

3. Rate the movies which are rated by the similar users. We will consider the rating is the mean rating of similar users, and methods are the following:
    a. Mean ratings
    b. Weighted mean rating,
    c. Select the movies which are top rated.

Copyright © 2019Authors

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

D. CROSS-VALIDATION TO VALIDATE MODELS

Cross Validation is a technique which is the most accurate one, even though its computationally large. Using this k-fold cross-validation approach, we will split the given data into some parts, and then take a subset out as the test set, and then we will evaluate the accuracy of the model. Then, we will do the same with other parts of the data and calculate the mean efficiency.[6]

IV. USER INTERFACE

The user interface for our system was developed with R Shiny package and it is deployed into cloud with R Shiny Server.

The web application which is created using R shiny is hosted on shinyapps and you can find the web application from the following link "https://joeysplpizza.shinyapps.io/recommend"

A. RECOMMENDED MOVIES PAGE

After entering the web application, you will land on to the "Recommended Movies" page. The page consists of a Text Box where the user has to enter their id and click on the GO button to generate recommended movies.[7]

You can see the image below for reference.

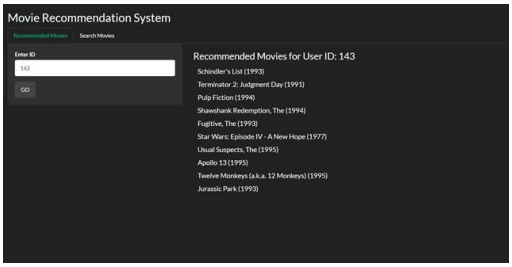


Fig.5.Recommended movies page

This is the main page for the entire project, this page is integrated with machine learning model which we created in R Programming.

B. SEARCH PAGE

The second page is "Search Movies", this page consists of a slider to choose the year and a drop-down menu to choose the genre. After selecting the two fields the user has to click on the GO button the get the results.[8]

The results contain movies released in that particular year and genre which was the user selected.

You can see the image below for reference

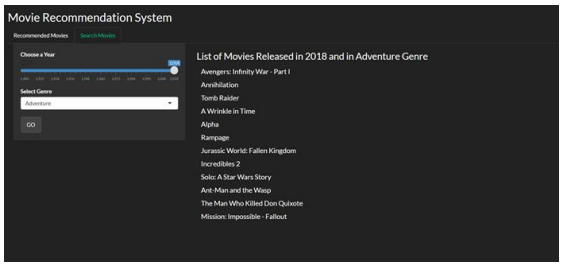


Fig.4.Search movies page

The entire UI is developed by only using R Shiny Package which is available in R Programming language. It is very easy to implement the web application in R shiny when compared to other frameworks.[9]

We can easily integrate the machine learning model with the web applications without any hassle. The Web application can be developed within minutes rather than hours or days.

After creating the application in R Shiny, we can even deploy the application the cloud using Shiny server for free.

V. CONCLUSION

In this paper, we have developed user-based collaborative filtering recommendation system. It allows user to enter his id in the web application and then user gets the recommended movies based on their past interactions to items. We are constantly developing the current system for better experience. In future, we would like to implement voice based recommendations and even better user interface for fantastic experience to the users.

VI. REFERENCES

[1] Baptiste Rocca. Introduction to Recommender Systems [Internet]. Medium. Towards Data Science; 2019. Available from: https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada

Copyright © 2019Authors

ISSN: 0971-1260
Vol-22-Issue-3-July-2019

[2] F. Maxwell Harper, Joseph A. Konstan. The MovieLens Datasets. History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19.; Available from: https://doi.org/10.1145/2827872

[3]Schafer, J.B., Frankowski, D., Herlocker, J. and Sen, S., 2007. Collaborative filtering recommender systems. In *The adaptive web* (pp. 291-324). Springer, Berlin, Heidelberg.

[4] Basu, C., Hirsh, H., & Cohen, W. (1998, July). Recommendation as classification: Using social and content-based information in recommendation.

[5] Gong S. A collaborative filtering recommendation algorithm based on user clustering and item clustering. JSW. 2010 Jul;5(7):745-52.

[6] Le Gratiet, Loic, and Claire Cannamela. "Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes." *Technometrics* 57.3 (2015): 418-427.

[7] Cremonesi, P., Elahi, M., &Garzotto, F. (2017). User interface patterns in recommendation-empowered content intensive multimedia applications. *Multimedia Tools and Applications*, *76*(4), 5275-5309.

[8] Roßner, Daniel, Claus Atzenbeck, and Tom Gross. "Visualization of the relevance: using physics simulations for encoding context." In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, pp. 67-76. ACM, 2019.

[9] Di Noia, Tommaso, et al. "Linked open data to support content-based recommender systems." *Proceedings of the 8th international conference on semantic systems*. ACM, 2012.

[10] Chen, Y. L., Cheng, L. C., & Chuang, C. N. (2008). A group recommendation system with consideration of interactions among group members. *Expert systems with applications*, *34*(3), 2082-2090.

Copyright © 2019Authors