Movie Recommendation

Pankaj Katkar(160010002)
Aman Kumar(160010030)
Vaishnav Tammadwar(160010022)
Team 9

November 25, 2018

1 Introduction and Overview

In this project we worked on how to recommend a user new movie based on its previously watched movies and ratings it had given to watched movies. Movie recommendation is used in many areas like Netflix(https://www.netflix.com/in/) uses it to recommend movies to its user, also Amazon Prime(https://www.primevideo.com/) for movie recommendation. It gives user a complete new experience of watching movies. Thus it motivated us to do project on this topic.

Section 2 covers methods which are used in this project and in that we have explained one method of how to do movie recommendation. Section 3 covers the analysis of our experiments and also contains dataset link and results. Section 4 covers discussion and future directions.

Related work

Many similar work are done in these field.

- Related work using content-based filtering is done on (https://bit.ly/2Anc0E4). But it uses Term-frequency matrix rather than term frequency inverse document frequency matrix.

- Also related work using Singular Value Decomposition is done on (https://bit.ly/2nuQTJv). In that it explains how to apply Singular Value Decomposition method.

2 Methods

In these project we have used two methods namely

- Content-based Filtering

- Singular Value decomposition

The link for it is given in Related work section.So in Content based filtering approach is based on previously watched movie of user. It relies on the similarity of the movies being recommended. The basic idea is that if you are in some mood and watched a movie, then you will also watch a "similar" movie in that your mood.For example if you are today in mood of watching a action movie and you watched one action movie then you will again want to watch action movie similar to it as you are in mood of watching action movies so we will recommend action movies similar to previously watched movie to you using content based approach. It generally works on the basis of genre of movie.

We used concept of Term Frequency Inverse Document Frequency(TFIDF) in these method.We will calculate TFIDF matrix where each entry of matrix is found using below formula.

$$\mathbf{tfidf}_{i,j} = \mathbf{tf}_{i,j} \times \log\left(\frac{\mathbf{N}}{\mathbf{df}_i}\right)$$

$\mathbf{tf}_{i,j}$ = total number of occurences of i in j
$\mathbf{df}_i$ = total number of documents (speeches) containing i
$\mathbf{N}$ = total number of documents (speeches)

In these matrix columns correspond to words and rows correspond to documents where there is document for each movie's genre and document consist movie's genre. By calculating TFIDF matrix we have found TFIDF vector for each movie. Now we have to find similarity between two movies based on their genres. So for that we will used cosine value to find similarity between two TFIDF vector.Here cosine value is just the cos of angle between two TFIDF vectors. So more the cosine values, more the two TFIDF vectors will be similar to each other and more two movies will be similar to each other. Below image gives idea of content-based approach.
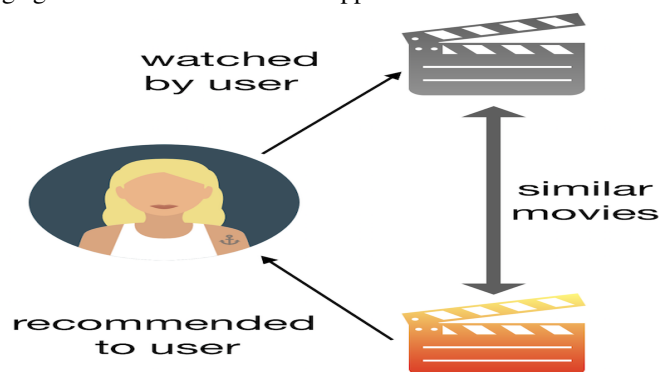
3 Experimental Analyses

In these project we have worked on movie dataset (https://bit.ly/2SdrTUD) which has around 1 million ratings for different movies and also has around 6000 movies with their genre. In these project we have apply two methods. Description corresponding to each method is given below in next paragraph.

In Content-based Filtering, we have found cosine matrix in which (i, j)th entry corresponds to cosine value of ith and jth movie TFIDF vector. Thus when user gives name input of previously watched movie, our code will recommend list of 10 movies which are similar to previously watched movie. Suppose user watch previously "Jumanji (1995)" so it will recommend next 10 movies of which five names are:

- Kids of the Round Table (1995)

- Indian in the Cupboard, The (1995)

- NeverEnding Story III, The (1994)

- Escape to Witch Mountain (1975)

- Labyrinth (1986)

The results seems to be accurate as the genre for "Jumanji" is adventure, children and fantasy. And the genre of all five recommended movies have the same genre. The movies are recommended based on cosine values,one with more cosine values will be recommended first. Two or more movies having same cosine values will be recommended in order.

In Singular Value Decomposition approach, movies are recommended based on the ratings the user given to the movies he/she watched. In it SVD decomposes user-movie rating matrix A into two unitary matrices and a diagonal matrix as shown. $A = U. E .VT$ where A (user's ratings) is input matrix, U is left singular matrix (user - features matrix), E is diagonal matrix of singular values (essentially weights of each concept), VT is right singular matrix (movie - features matrix).

- Roger Me (1989)

- Boat, The (Das Boot) (1981)

- To Kill a Mockingbird (1962)

- Sex, Lies, and Videotape (1989)

- Boys Don't Cry (1999)

The results seems to be accurate as user with id '6040' has highly rated to movies with genre drama, war, action, comedy, documentary and recommended movies has these genre in common. So recommendation seems to be accurate for user with id '6040'.

Datasets

Dataset that we have used is given on link (https://bit.ly/2SdrTUD). It consists data on movies, users and rating.

Results

Using two methods we got the 10 movies recommended for the user. First method gives recommendation based on user's previously watched movie and second method gives recommendation based on user's rating. Top five Recommendation based on first method when user previously watched "Tom and Huck (1995)" is

- Amazing Panda Adventure, The (1995)

- Casper (1995)

- Far From Home: The Adventures of Yellow Dog (1995)

- Secret Adventures of Tom Thumb, The (1993)

- Black Beauty (1994)

Top five Recommendation based on second method where user id is "2"

- Men in Black (1997)

- Rock, The (1996)

- Schindler's List (1993)

- Rain Man (1988)

- Good Will Hunting (1997)

4   Discussion and Future Directions

Our approach helps to recommend movies to user to enhance it's experience. Our approach is good because in content based filtering recommendation are done based on the taste of user in movies like user wants to watch romantic movies it will recommend user more romantic movies and in second approach recommendation are done based on user's rating to movies. However there are cons for content-based approach which is it does not recommend items outside a user's content profile and also it is unable to exploit quality judgments of other users. Thus it is need to further develop these model in that direction to improve it. It can be used to recommend movies in Netflix(https://www.netflix.com/in/) and Amazon prime(https://www.primevideo.com/). It can be further extended for product recommendation in Amazon(https://www.amazon.in/) and Flipkart(https://www.flipkart.com/).

References

[1] https://bit.ly/2Anc0E4

[2] https://bit.ly/2nuQTJv

[3] https://bit.ly/2SdrTUD