DA_REPORT_MARSHMALLOW

by Shreesha Sk

Submission date: 07-Dec-2021 06:09PM (UTC+0530)
Submission ID: 1723320931
File name: DA_REPORT_MARSHMALLOW.docx (1.07M)
Word count: 1718
Character count: 10118

Netflix Recommendation System

Shreesh Ravindra Devi
Computer Science and Engineering PES University
Bangalore, India
shreeshrdevi2@gmail.com

Shreesha Seerya Krishna
Computer Science and Engineering PES University
Bangalore, India
skshreesha@gmail.com

Aditya Vijay Kavatage
Computer Science and Engineering PES University
Bangalore, India
write2adityavk@gmail.com

Aryan Kumar
Computer Science and Engineering PES University
Bangalore, India
aryan.kumar.052001@gmail.com

ABSTRACT- With the increase in growth of web usage and streaming in the recent years, recommender systems have become a very important domain to address in different domains such as business , Netflix , amazon or youtube streaming etc. The purpose of the project is to analyze the various approaches inherited to solve the problem of providing a movie recommendation on Netflix and formulate a model that would provide the best recommendation for the user. Content based filtering and dimensionality reduction help to cluster similar movies to users by generating precise and highly efficient robust recommendations. The results of our model have shown that this new method significantly improves the recommendation system's performance.

KEYWORDS-Vectorization(CountVector,TfidVector),Content based Filtering ,IMDB Classification, collaborative filtering

INTRODUCTION

Recommender systems nowadays have become heart of business techniques for business companies or service/product based companies examples can be companies like amazon, youtube, Flipkart etc. The algorithms used in these companies help predict user's choices or preferences based on their search or watching acivities or behavior and help improve the users satisfaction towards the items.

Many e-business Companies of various domains ranging from movies, books, and news to research articles now consider recommendation system a very vital part of their business. These systems try to acquire feel/sentiment about particular products or services from communities of users to promote these products for other customers with similar interests. For example, recommender systems can find the existing connections or similarities between users and their friends and recommend new tags for auser in a social network context

Usually recommender systems are sub-divided into three categories: content-based, collaborative or hybrid filtering methods. Content-based focuses on users preference and description provided for the item in general to be able to recommend items that are similar to the items that were highly rated earlier. Collaborative method takes a lot more attributes than content based for providing the recommendation such as activities, behavior, users preference into consideration and recommend .Hybrid methods are a combination of content-based filtering and collaborative filtering which adds up on the advantages of both to provide more suitable items.

Our approach introduces a new content-based recommendation algorithm considering dimensionality reduction as well as clustering. Goal is to improve the performance of recommender systems and to overcome their flaws such as cold-start as well as the scalability issues. Results have shown that our method improves the performance of recommendation systems.

LITERATURE REVIEW

1) An Approach for Netflix Recommendation System using Singular Value Decomposition

Results show an improvement in overall performance for all algorithms and improvement in performance as the data size grow[1]. Individual increase in the scalable performance of algorithms have implicated a majestic difference in their effectiveness, especially having considered between the matrix factorization algorithms SVD and ALS. Singular Value Decomposition gave the best ratings with RMSE (root mean square error) of 0.8190 considering 1M data set and an increase in the performance to 0.09936% from 100K to 1M database. Root Mean Square Error show how well these algorithms have predicted ratings for non-rated items. Given these results, it is suggested to make use of SVD out of the four considered algorithms for a better recommender system

in production. Paper focuses on comparing four unique collaborative filtering methods, considering which the goal was to figure out which gave the best prediction rate. The algorithms taken were K-Nearest Neighbours, Singular Value Decomposition, Slope One and the ALS Also this paper considered two means that is arithmetic and weighted means to determine if math models could provide better prediction ratings than the considered algorithms. SVD had the best prediction rate and the ALS had the worst rate. But, ALS had highest performance improvement .Arithmetic mean had slightly better rate of prediction than SVD, Weighted Arithmetic Mean had the overall worst prediction rate.

2)A New Collaborative Filtering Recommendation Algorithm Based on Dimensionality Reduction and Clustering Techniques

The approach proposed a new technique for the building of recommender systems that can benefit from the potentials provided in by clustering algorithms such as SVD (Singular Value Decomposition) and the k-means technique. Initially, K-means clustering was taken to cluster users/customers in the similar partition according to their choices/preferences, and then the SVD (Singular Value Decomposition) was used for dimensionality reduction but this this technique was also a powerful mechanism, that could efficiently find most similar users. For evaluating the performance of the considered method, experimentations were conducted on datasets for the recommendation called MovieLens 1million (1M) and MovieLens 10 million (10M), these ratings were made by users anonymously.RMSE (root mean square error) metric was also taken to evaluate the accuracy of the considered method in comparison to k-nearest neighbor (KNN) based and also the k-means recommendation methods. Results show that our method/approach improved the performance significantly.

3) Movie Recommender System Based on Percentage of View

This paper proposed an approach on Implicit Opinion Measure to improve on the performance of recommender systems to recommend movies on implicit feedback from the users. Dataset by Namava, [3].
The authors of this paper propose or consider a percentage view approach to find relevant movies for recommending to the users. In order to obtain this prediction, collaborative based filtering method, content-based filtering method and a new method called residual method was taken.
Correlation between the like probability and percentage view have shown there is a positive correlation which exists between like probability and percentage view. For evaluating the model of recommendation system, the average of the top 20 recommendations precisely for all the customers was obtained.
The results for the content-based filtering, collaborative Filtering and also the memory-based techniques are reported.
For improving on the accuracy of prediction, sign and the error value was calculated independently in each particular level. Also for comparison, a random recommender model was considered as baseline and accuracy of the recommendation model was compared with it.

Hence the recommendation system developed from using these different prediction models, works 5 much better than the average recommendation systems.

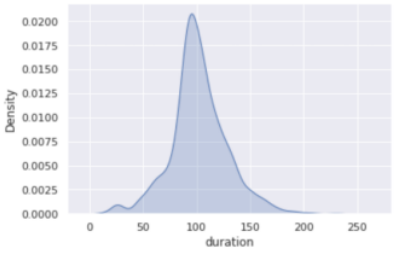
4)Predictive self-learning content recommendation system for multimedia contents

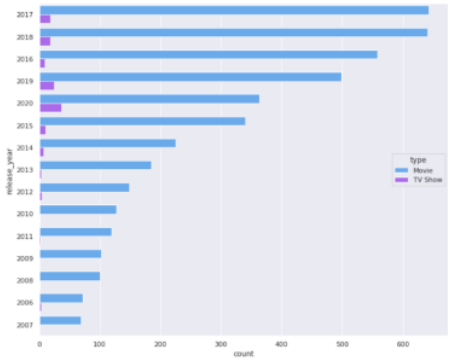
Goal or approach of this paper was to create a model that can predict what user google or search for by using prior information collected and using ML (machine learning) to analyze the user behavior.
Recommendation models have several issues such as incomplete data, outliers, noisy data etc because the users don't usually cooperate. Hence recommender models are general or mild and usually cannot be changed according to certain user characteristics given the suggested data or database. This given predictive self-learning recommender model uses the filtering technique such as Collaborative Filtering as well as criterias such as similarity , popularity, currency, safety, importance etc in addition to users' profiles to make the recommendations. This model is different from general recommender systems as it allows for diverse suggestions while at the same time does not compromise on the performance of the system or the response time.

DATASET and INITIAL INSIGHTS

For the purpose of providing a recommendation to the user we have taken the Netflix movies and TV-shows dataset. This dataset consisted of 12 columns and 8807 rows. Since we have worked on another aspect or another factor of recommendation of movies considering the IMDb rating and rotten tomatoes we have chosen to merge these two columns from another dataset and joined them over a common attribute or column of Movie Titles.

Initial insights have shown most people prefer to watch shows or movies on Netflix rather than other available platforms and the insights have also shown most movies average around 92 minutes and the United States produces the most number of movies and tv shows. Also the the number of movies and tv shows shown an increase in trend in the recent years.





PROBLEM STATEMENT AND PROPOSED SOLUTION

The aim of this project was to develop a profound recommender model or system for providing recommendations to the user based on his likes of a movie or given a movie provide recommendations for it. Our approach differ from other general trends as we use both content based filtering and tfid vectorization. Also used a soup(bag of words/attributes) considering the CountVetorizer to provide the results .

```
get_recommendations_tf('peaky blinders')
```

```
Out[4]: 7683                    our godfather
        2646                  my stupid boss
        3133                             don
        8293                        the fear
        7140    jonathan strange & mr norrell
        7785                power rangers zeo
        8467                        the prison
        8539                        the tudors
        1510                    the con is on
        8391     the legend of michael mishra
        Name: Title, dtype: object
```

1) Using content based filtering and tfid vectorization

2)Using cosine similarity and CountVector
Our approach also considers the use of cosine similarity which helps in overcoming the flaw such as 'count-the-common-words' or also the Euclidian distance approach which are usually fundamental flaws. It in terms of mathematical value measures cosine of angle between two vectors projected in n-dimensional space vector.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

```
In [8]: get_recommendations_multiple('peaky blinders', cosine_sim2)
```

```
Out[8]: 4809                  kiss me first
        5032    the frankenstein chronicles
        6922                   happy valley
        3034                    giri / haji
        2184                       get even
        5716                       paranoid
        7333                    london spy
        3789                  killer ratings
        5278                         apaches
        1991                   criminal: uk
        Name: Title, dtype: object
```

2) Using cosine similarity and CountVector

We have also attempted to implement a type of collaborative filtering technique. Since the dataset doesn't actually contain other user ratings, we have instead used IMDb ratings as a form of external rating/collaborative results.
This approach provides us with recommendations as well as a predicted rating, which helps us with the cold start problem i.e. when a new TV Show/Movie has unknown rating.

EXPERIMENT RESULTS AND CONCLUSION

Our approach has also considered the given IMDb ratings and worked on generating IMDb ratings for new movies and have been successful with an accuracy of 0.96 or 96%.

```
In [21]: prediction = actual = 0
         for i in range(100):
             ratings, IMDb = get_similar(i)
             prediction += max(ratings)
             actual += IMDb
         print("Accuracy: ", prediction/actual)
```

```
Accuracy:  0.9612554112554114
```

ACKNOWLEDGEMENT

We would like to express our profound gratitude to Prof Bharathi, for encouraging and providing us with this opportunity to get hands-on experience in the field, and guiding us along the way. We would also like to thank the Computer Science and Engineering department at PES University, for always inspiring us to conduct frequent research and inculcating a problem-solving discipline in us.

REFERENCES

[1] Hafed Zarzour, Ziad Al-Sharif, Mahmoud Al-Ayyoub , Yaser Jararweh.” An Approach for Netflix Recommendation System using Singular Value Decomposition”, 2018.

[2] Ankur A. Ranjan1*, Amod Rai2, Saiful Haque3, Bhanu P. Lohani4 and Pradeep K. Kushwaha “A New Collaborative Filtering Recommendation Algorithm Based on Dimensionality Reduction and Clustering Techniques”, 2019

[3] Ramin Ebrahim Nakhli, Hadi Moradi, Mohammad Amin Sadeghi “Movie Recommender System Based on Percentage of View”, 2019

[4] Ibrahim Hussein Mwinyi, Husnu S. Narman, Kuo-Chi Fang, and Wook-Sung Yoo “Predictive self-learning content recommendation system for multimedia contents”,2018

[5] https://www.sciencedirect.com/topics/computer-science/cosine-similarity

DA_REPORT_MARSHMALLOW

ORIGINALITY REPORT

24%	4%	23%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Hafed Zarzour, Ziad Al-Sharif, Mahmoud Al-Ayyoub, Yaser Jararweh. "A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques", 2018 9th International Conference on Information and Communication Systems (ICICS), 2018 Publication	10%
2	Ramin Ebrahim Nakhli, Hadi Moradi, Mohammad Amin Sadeghi. "Movie Recommender System Based on Percentage of View", 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), 2019 Publication	5%
3	Ibrahim Hussein Mwinyi, Husnu S. Narman, Kuo-Chi Fang, Wook-Sung Yoo. "Predictive self-learning content recommendation system for multimedia contents", 2018 Wireless Telecommunications Symposium (WTS), 2018 Publication	3%

4	Submitted to Birla Institute of Technology and Science Pilani Student Paper	1 %
5	www.univ-soukahras.dz Internet Source	1 %
6	Manoj S Hegde, Ganesh Krishna, Ramamoorthy Srinath. "An Ensemble Stock Predictor and Recommender System", 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018 Publication	1 %
7	towardsdatascience.com Internet Source	1 %
8	mdpi-res.com Internet Source	1 %
9	K. V. Sandeep, Shailja Agarwala, R. Sharath, Viraj Kumar. "DSAdvisor: Facilitating Deeper Inquiry in the Data Structures Course", 2016 International Conference on Learning and Teaching in Computing and Engineering (LaTICE), 2016 Publication	<1 %
10	sersc.org Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		