ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

Content Based Movie Recommendation System

[1]Dr Shaifali M. Arora, [2] Dr Anshul Pareek, [3]Dr Poonam

[1]Associate Professor, ECE Department, Maharaja Surajmal Institute of Technology

[2,3]Assistant Professor, ECE Department, Maharaja Surajmal Institute of Technology

ABSTRACT

In today's digital world, there is a boundless increase in the digital content related to videos, books, movies, food etc. and it has become a tiresome task for users to find the content of their choice from this content. Therefore, there arises a need of recommendation system, that can help user to shortlist the item of their taste and liking by spending minimum time. Recommender System is basically a filtering system that help users to shortlist the content by overcoming information overload. It predicts data for users based upon the previous information/interests about the user and makes recommendation according to the interest model of users. There is also an incipient growth among the digital content providers whose motive is to engross as many operators on their service as possible and that too for maximum time. As this increases their weightage in any recommender system. A movie recommendation system plays a significant role in our social life as this system suggests a set of movies to users based upon their interests. The system is basically a filter that gathers the information from the past history and interests of users, and accordingly give outcome. With the advancement in technology and agile growth in machine learning techniques, the accuracy in the outputs of these recommendation systems is also increased. In this paper, content-based movie recommendation system has been proposed. that uses traits like director, actor, description, genre etc. for making suggestions to users. The perception behind the designed algorithm is that if a user liked a specific type of movie or show, a similar movie or a show will be liked by him/her.

Introduction:

In today's digital world, it has become an irksome task to find the content of one's liking in an endless variety of content that are being consumed like books, videos, articles, movies, etc. On the other hand, there has been an emerging growth among the digital content providers who want to engage as many users on their service as possible for the maximum time. This gave birth to the recommender system wherein the content providers recommend users' the content according to the users' taste and liking. A recommender system is basically an information filtering system that takes the information like rating, preference, past experience to predict the recommendation for the user. Various websites like Amazon, Netflix, Youtube, Tinder use prediction/recommendation engines to suggest the interest items to users based upon his/her previous history. World cinema has released more than 500000 movies, which is a number beyond one person's search control. The search engines are therefore designed to help the users to decide the movie from such a big heap.

Machine learning and AI are the approaches that help in filtering out the choices and predicting appropriate recommendations to the user based upon their past choices. The accuracy of prediction increases if sufficient data is available. But in these approaches, the challenging task is when there is no previous history of user is available. To make prediction in such case two ways has been suggested in literature, first- such users can be recommended with the highest rated/ most popular items. Such recommendations may not be accurate because

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

these are not specific to a particular user but same for all new users. Second solution is to ask for the interests of new users and based upon that recommendation is provided to them.

A Simple Recommender system offers generalized suggestions to all the users depending upon the popularity and (sometimes) genre of the movie. In this general model the most popular movies will get high probability of being recommended to the users. So this model doesn't give personalized suggestions to the users. This system can be improved by adding various recommender features. Steps to design a recommendation system are shown in Fig 1 .

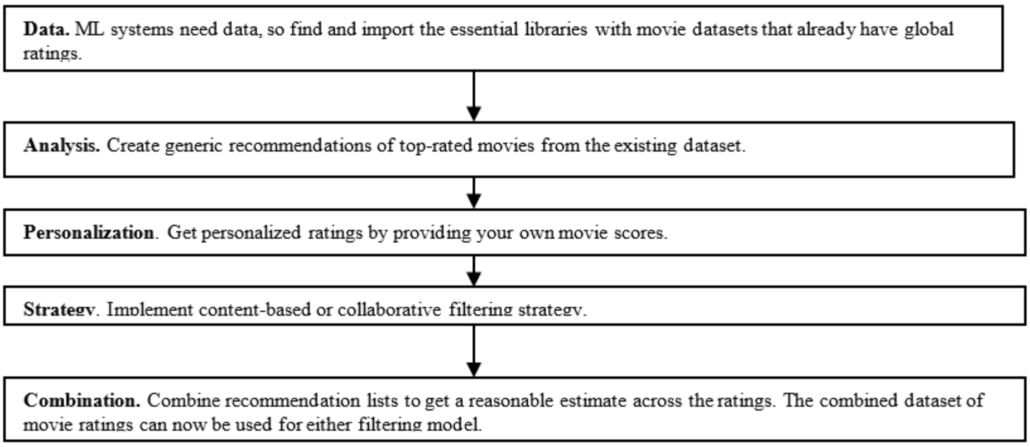

Fig 1: Steps to design a Movie Recommendation System (MRS)

MRS use different filtering techniques and filtering algorithms to find out the best possible result. There are two popular Machine Learning algorithms 1) content-based recommendation system 2) collaborative filtering-based recommendation system for designing MRS. In this paper, the content-based filtering approach is used.

Content-based recommender system:

Content-based recommender system shown in Fig 2, is one of the filtering techniques in which similarity among different items is evaluated depending on the data collected for the items that are selected by users, and then according to that suggest the recommendation. In this users' and products' profiles are built and these profiles are continuously updated when the user watch/ buy a new item. So, these recommendation systems will compare the profile of user and items and then items that are most similar are recommended.

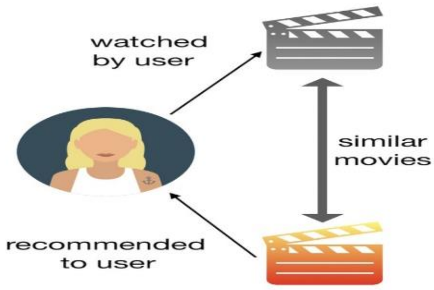

Fig 2: Content based Recommendation System

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

This can be explained with an example, if a movie like 'Mission Impossible' with 'Tom Cruise' as actor, is liked by the user, then some other action movies of same actor can be recommended to him/her so here the two features are worked upon in the recommendation system are actor and the other is 'Action'.

In the above example while filtering, two types of data could be used. First, the interests of the user, their personal information such as their gender, age or, sometimes the history of user can also be used. This collected data is stored in a vector named user. Second vector is formed by collecting the information related to similar items i.e. item vector. U

Cosine similarity is evaluated to suggest the recommendations is given by

$$\cos(\theta) = \frac{\sum_i Ui\ Ii}{\sqrt{\sum_i Ui^2}\sqrt{\sum_i Ii^2}}$$

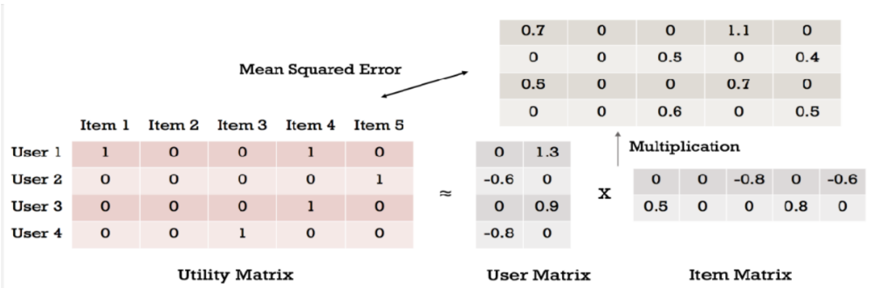

Fig 3: Similarity Calculation Matrix

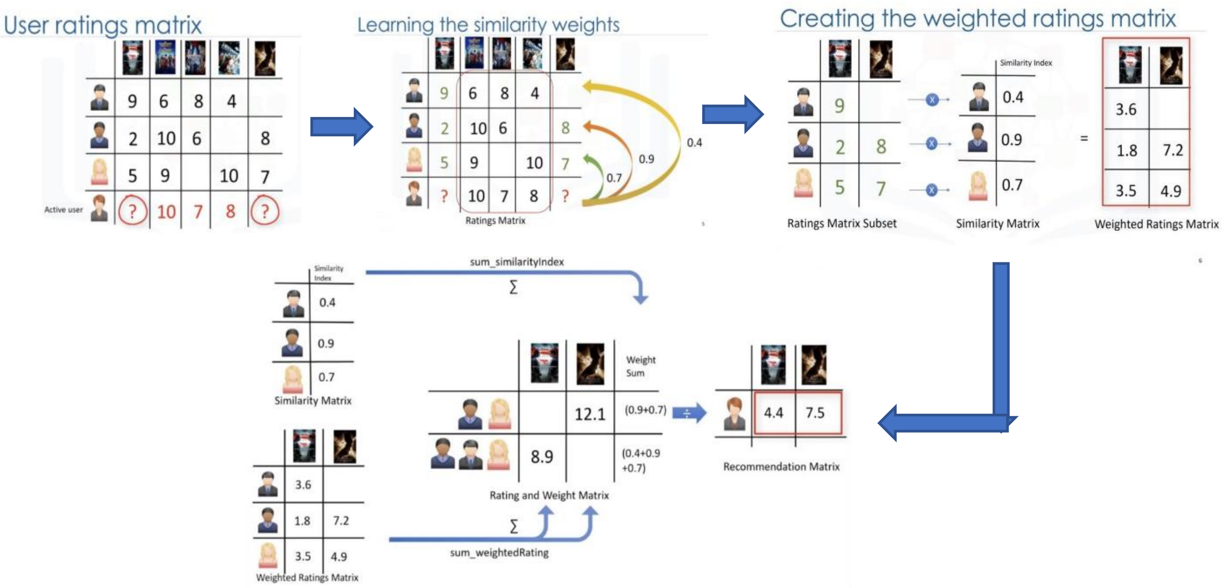

Fig 4: Steps followed in Content Recommendation System

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

Where user vector is given by U and item vector is given by I. Values are calculated using cosine similarity and then the matrix is sorted in the descending order. The items that are on the top are recommended to the user. The main advantage of this system is that the users get suggestions and they can select by sending small time. The user is satisfied by the type of recommendation. The only disadvantage of this technique is to new business providers because the users don't try a different type of new products. And also, if the user information matrix is changed then the cosine similarity matrix is required to be calculated again.

2. Methodology:

To design content based recommendation system with ML approach, it is required to have a big data set so as to master the algorithm. The right movie datasets that is most valuable for ML must contain information on cast, screen time, script, reviews, plot etc. In this work, TMDB dataset is taken. Following steps are followed while designing the algorithm:

1. First the data is preprocessed so as to filter out the desired information. Next from filtered data, further filtering is done based upon the votes/ ratings given to a movie. 95th percentile cutoff is chosen and the movies below that are rejected from the recommendation system. Also data is filtered on the basis that a movie has to have at least 434 votes on TMDB or the other condition is that the average movie rating is 5.244 out of 10. After these filterations only 2274 movies are qualified. This is depicted in the outcome shown in fig 5.

```
m = vote_counts.quantile(0.95)
m

434.0

# Pre-processing step for getting year from date by splliting it using '-'

md['year'] = pd.to_datetime(md['release_date'], errors='coerce').apply(
    lambda x: str(x).split('-')[0] if x != np.nan else np.nan)

qualified = md[(md['vote_count'] >= m) &
               (md['vote_count'].notnull()) &
               (md['vote_average'].notnull())][['title',
                                                'year',
                                                'vote_count',
                                                'vote_average',
                                                'popularity',
                                                'genres']]

qualified['vote_count'] = qualified['vote_count'].astype('int')
qualified['vote_average'] = qualified['vote_average'].astype('int')
qualified.shape

(2274, 6)
```

Fig 5: 95th percentile Outcome from chart

2. First fifteen movies has been selected. Further, this filtered data recommendation is made based on description and tagline.

3. In the above output, improvement is done in this by taking into consideration the cast, crew, keywords and genre. Current dataset is merged with the crew and the keyword datasets. Different weights for the features like directors, actors, genres has been used by limiting the number of keywords and gives the output shown in fig:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

```
get_recommendations('The Ugly Truth').head(10)

5127    Win a Date with Tad Hamilton!
3532                   Legally Blonde
7611                          Killers
6201                   Monster-in-Law
6986                               21
7522                     When in Rome
9112                           Plan B
5030                     I'm No Angel
7711                Life As We Know It
1951                   You've Got Mail
Name: title, dtype: object
```

```
] get_recommendations('Mean Girls').head(10)

3319              Head Over Heels
4763                Freaky Friday
1329             The House of Yes
6277             Just Like Heaven
7905          Mr. Popper's Penguins
7332     Ghosts of Girlfriends Past
6959      The Spiderwick Chronicles
8883                     The DUFF
6698          It's a Boy Girl Thing
7377        I Love You, Beth Cooper
Name: title, dtype: object
```

Fig 6: Recommendation Result with description and tagline

4. Another step towards the improvement is done by taking into consideration the rating/popularity. Therefore, these features have been further added to improve the recommendation.
5. Also, a mechanism has been added to remove bad/low rated movies of users favorite actor/director etc. Now the system will return movies which are popular and have had a good critical response.

RESULTS AND DISCUSSION

1. Content based Recommender model using Similarity Score

```
get_recommendations('The Dark Knight').head(10)

7931                         The Dark Knight Rises
132                                 Batman Forever
1113                                Batman Returns
8227      Batman: The Dark Knight Returns, Part 2
7565                   Batman: Under the Red Hood
524                                         Batman
7901                              Batman: Year One
2579                 Batman: Mask of the Phantasm
2696                                           JFK
8165      Batman: The Dark Knight Returns, Part 1
Name: title, dtype: object
```

Fig 7: Outcome of Content based Recommender model using Similarity Score

- It has been observed that when the system is identifying Batman movies, ‘The Dark Knight Rises’ movie has been recommended as only genre features were added. But this movie has very bad reviews and it should not be recommended.
- This recommendation system does not take into considerations some of the important features like crew,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

director of the movie, cast etc. which also are the factors while rating a movie.

- User who liked The Dark Knight movie may have high chances of liking movie like NOLAN hate other Batman movies.

2. Content-based Recommender system is based on the matching of keywords like directors and genre.

```
get_recommendations('Fingers').head(10)

2819          Black and White
7323                    Tyson
1429       Two Girls and a Guy
5100        The Pick-up Artist
4021      The Long Good Friday
3768              Dinner Rush
217           The Glass Shield
2272               In Too Deep
3664               Next of Kin
4868             Time and Tide
Name: title, dtype: object
```

Fig 8: Outcome of Content based Recommender model using Similarity Score based on keyword, director and cast

- The second system took genre, keywords, cast and crew into consideration and have recognized other James To back movies because of high score given to director ,therefore these movies got highest recommendations.

3. Content based Recommender model using Similarity Score and filtering of data using minimum number of votes.

- This system recommends movies without considering the ratings of the movies. Although there are lot of common characteristics in Batman and Robin in comparison to The Dark Knight movie but this movie shouldn't be recommended as it has lot of terrible content.
- Therefore, while designing this system a mechanism has been developed to remove bad movies and only recommend movies that are popular and high rating.

```
improved_recommendations('The Dark Knight')
```

	title	vote_count	vote_average	year	wr
7648	Inception	14075	8	2010	7.917588
8613	Interstellar	11187	8	2014	7.897107
6623	The Prestige	4510	8	2006	7.758148
3381	Memento	4168	8	2000	7.740175
8031	The Dark Knight Rises	9263	7	2012	6.921448
6218	Batman Begins	7511	7	2005	6.904127
1134	Batman Returns	1706	6	1992	5.846862
132	Batman Forever	1529	5	1995	5.054144
9024	Batman v Superman: Dawn of Justice	7189	5	2016	5.013943
1260	Batman & Robin	1447	4	1997	4.287233

Fig 9: Outcome of Content based Recommender model using Similarity Score and filtering of data using minimum number of votes.

Conclusion:

Content based filtering give the recommendation based on the features of the given item and compare it with other items features. In this paper, a content-based recommendation system is suggested different weight to different feature. Also the movies which are low rated are removed from the recommendation system so that

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

these low rated movies with similar features as high rated movies are not recommended to the user.

References:

[1] Ahuja R, Solanki A, Nayyar A. Movie recommender system using K-Means clustering and K-Nearest Neighbor. In: Proceedings of the IEEE 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2019. https://doi.org/10.1109/CONFLUENCE.2019.8776969.

[2] Wu, C.-S. M., Garg, D., & Bhandary, U. (2018). Movie Recommendation SystemUsingCollaborativeFiltering,23-25Nov, 2018, IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018. doi:10.1109/icsess.2018.8663822

[3] M.M. Reddy, R.S. Kanmani and B. Surendiran, "Analysis of Movie Recommendation Systems; with and without considering the low rated movies, " 24-25 Feb , *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*,Vellore,India,2020,pp.1-4,doi:10.1109/ic-ETITE47903.2020.453.

[4] Singh S.P., Solanki S.,"A Movie Recommender System Using Modified Cuckoo Search," 24April, 2019 Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 545. Springer, Singapore, 2019.doi:10.1007/978-981-13-5802-9_43

[5] H. W. Chen, Y. L. Wu, M. K. Hor and C. Y. Tang, "Fully content-based movie recommender system with feature extraction using neural network," 9-12 July 2017 International Conference on Machine Learning and Cybernetics (ICMLC), Ningbo, China, 2017, pp. 504- 509. doi: 10.1109/ICMLC.2017.8108968

[6] Rajarajeswari S., Naik S., Srikant S., Sai Prakash M.K., Uday P.," Movie Recommendation System, " 03 May 2019 Emerging Research in Computing, Information, Communication and Applications. Advances in Intelligent Systems and Computing, vol882. Springer, Singapore, 2019. doi: 10.1007/978-981-13-5953-8_28

[7] J.Lund and Y.Ng, "Movie Recommendations Using the Deep Learning Approach," 6-9 July 2018 IEEE International Conference on Information Reuse and Integration (IRI),Salt Lake City, UT,2018,pp.47-54,doi:10.1109/IRI.2018.00015.

[8] M. K. Kharita, A. Kumar and P. Singh, "Item-Based Collaborative Filtering in Movie Recommendation in Realtime," 15-17 Dec,2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp.340-342, doi: 10.1109/ICSCCC.2018.8703362.

[9] L.A. Hassanieh, C.A. Jaoudeh, J.B. Abdo and J. Demerjian, " Similarity measures for collaborative filtering recommender systems,"18-20 April, 2018 IEEE Middle East and North Africa Communications Conference (MENACOMM), Jounieh, 2018, pp.1-5, doi:10.1109/MENACOMM.2018.8371003.

[10] N. Ifada, E. H. Prasetyo and Mula'ab, "Employing Sparsity Removal Approach and Fuzzy C-Means Clustering Technique on a Movie Recommendation System," 26 -27 Nov, 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM), Surabaya, Indonesia, 2018 ,pp.329-334, doi:10.1109/CENIM.2018.8711270

[11] Y. Song, S. Liu and W. Ji, "Research on personalized hybrid recommendation system," 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), Dalian, 2017, pp.133-137, doi:10.1109/CITS.2017.8035321.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,Iss 7, Nov 2022

[12] A. Pal, P. Parhi and M. Aggarwal, "An improved content based collaborative filtering algorithm for movie recommendations," 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, 2017,pp.1-3,doi:10.1109/IC3.2017.8284357.

[13] Li, Jing & Xu, Wentao & Wan, Wenbo & Sun, Jiande. (2018). Movie Recommendation Based on Bridging Movie Feature and User Interest. Journal of Computational Science. 26. 10.1016/j.jocs.