IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11618

Movie Recommendation System

Kassetty Krishna[1], Arnav Srivastava[2], Manash Protim Dutta Phukan[3], Deeksha Lokesh[4], Dr. S Vijay Kumar[5]

[1-5]Department of Computer Science and Engineering, Jain University, Bangalore, India

Abstract: Recommender systems are software tools used to leverage different strategies to generate suggestions for movies and other entities and make them available to users. Hybrid recommender systems combine two or more recommendation systems in different ways to take advantage of their complementary benefits. This systematic literature review reveals the latest technology in hybrid recommender systems over the last decade. An overview of relevant data mining and recommended techniques used to address and overcome the most relevant issues under consideration. It also considers the hybridization class to which each hybrid recommender belongs, the application domain, the evaluation process, and the proposed future research direction. Based on our results, most studies combine collaborative filtering with other techniques and are often weighted.

The Hybrid Recommender System is a hot topic and provides a good foundation for responding to new opportunities by exploring new opportunities, Contextualization recommendations, embedding parallel hybrid algorithms, and processing large datasets.

Keywords: Movie Recommender System, Hybrid Recommender System, Content – Based System, Collaborative based System.

I. INTRODUCTION

Hybrid recommender systems are a unique type of recommender system that combines content with collaborative filtering. The merge of collaborative filtering and content-based filtering can help you overcome the limitations you face when used alone and can be more effective in some situations. The hybrid recommender system approach can be implemented in a variety of ways. Individual predictions has been generated using content and collaboration methods and merged them, or add the functionality of collaboration techniques to a content-based approach (or vice versa).
Some studies have compared the performance of traditional and hybrid approaches and concluded that hybrid approaches provide more accurate recommendations.

Despite substantial study over the last few years, both collaborative and content-based approaches to content filtering and recommendation still have some flaws. One of the most noticeable problems is what is known as a cold start issue. This means that a system can't offer accurate recommendations for new users who haven't seen or rated any material yet. Similarly, a system cannot propose a new item that is notably different from previous things reviewed by the user. Due to these drawbacks, a hybrid recommender system may be able to enhance recommendation outcomes. We did not attempt to solve the cold start problem because our study concentrated on hybrid recommender systems, but we are aware of it and intend to address it.

The major goal is to develop a parallel hybrid recommender that does not limit the number of possible recommendations and allows for adjustable weighting of both recommender types. This means that CBR and CF recommenders' classification judgements for potentially recommendable content items can be freely blended based on a variety of decision characteristics. These characteristics include the predictability of both recommenders' predictions, the availability of content recommendations from one or the other technique, and so on. Furthermore, unlike any other hybrid recommender, the parallel hybrid recommender is simpler to implement.

II. LITERATURE SURVEY

K-NN (K-Nearest Neighbours) is a widely known class set of rules with numerous variations and implementations, extensively applied in severe statistics mining and different applications. This approach is famous amongst collaborative filtering RSs which constitute the maximum not unusual place in the circle of relatives of recommenders. It is commonly applied to investigate communities and discover customers of comparable profiles or examine objects` catalogue and discover objects with comparable characteristics. K-NN became determined in a complete of fifty nine studies.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022
DOI: 10.17148/IJARCCE.2022.11618

Clustering There are numerous clustering algorithms utilized in RSs and different statistics mining applications. They normally attempt to place up a fixed number of classes with which statistics may be identified. The most famous is the K-method which walls the complete statistics into K clusters. In RSs clustering is commonly implemented to pre-process the statistics. In the researcher's test with K-way (much like K-means) clustering and Bisecting K-method for grouping distinctive styles of getting to know objects. They additionally use CBF to create learners` profiles and construct an e-getting to know recommendation with advanced accuracy. A different instance is
in which websites are clustered the use of co-occurrence of pages and the content material statistics of pages. The consequences are aggregated to get the very last guidelines and conquer statistics sparsity. In general clustering algorithms have been utilized in 34 studies.

Dr V Subedha, Sandhya M, Shree Lakshmi, Swathi A, April 2021, proposed a CNN approach for Sign Language Recognition and it became capable of classifying Sign Languages. The writer makes use of simply the Alphabets while in actual existence the speech impaired cannot usually use Alphabets to deliver his/her message.

Keerthana P, Nishanth M, Karpaga Vinayagam D, Alfred Daniel J, Sangeetha K, March 2021, proposed the approach of Pattern Recognition the use of Computer Vision Techniques for Recognition and the approach became capable of comprehend the Sign Languages.

1.**Collaborative Filtering:**

Collaborative Filtering strategies make pointers for a consumer primarily based totally on rankings and alternatives statistics of many customers. The foremost underlying concept is if customers have each favoured sure not unusual place objects, then the objects that one consumer has favoured that the alternative consumer. has now no longer but attempted may be advocated to him. We see collaborative filtering strategies in movement on diverse Internet structures inclusive of Amazon.com, Netflix, Facebook. We advocate objects primarily based totally at the rankings and buy statistics that those structures acquire from their consumer base. We additionally explored one set of rules for Collaborative Filtering called the Nearest Neighbour's Algorithm. This method is predicated on the concept that customers who've comparable score behaviours so far, percentage the identical tastes and could in all likelihood show off comparable score behaviours going forward. The set of rules first computes the similarity among customers through the use of the row vector withinside the rankings matrix similar to a consumer as an illustration for that consumer. The similarity is computed through the use of both cosine similarity or Pearson Correlation. In order to expect the score for a specific consumer for a given film j, we discover the pinnacle of comparable customers to this precise consumer after which take a weighted common of the rankings of the ok comparable customers with the weights being the similarity values. Now we flow on in addition to the Second set of rules called Content Based Filtering Algorithm.

2.**Content Based Filtering:**

Content Based Filtering set of rules takes under consideration the likes and dislikes of the consumer and generates a User Profile. For producing a consumer profile, we don't forget the object profiles(vector describing an object) and their corresponding consumer score. The consumer profile is the weighted sum of the object profiles with weights being the rankings consumer rated. Once the consumer profile is generated, we calculate the similarity of the consumer profile with all of the objects withinside the dataset, that is calculated the use of cosine similarity among the consumer profile and object profile. Advantages of Content Based method is that statistics of different customers isn't always required and the recommender engine can endorse new objects which aren't rated currently .

III. METHODOLOGY

We have referred tow datasets for this entire project:

1. Movie Lens Data Set
2. TMDB Dataset

We are considering the Movie Lens dataset which contains approximately 100k ratings distributed across 943 users and 1682 movies. Additionally, Movie Lens also provides user information such as gender, age, occupation, etc. For retrieving information related to movies, our project makes use of IMDB/TMDB which is available online. IMDB contains item features such as movie title, cast, release date, imdb/tmdb url, etc.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

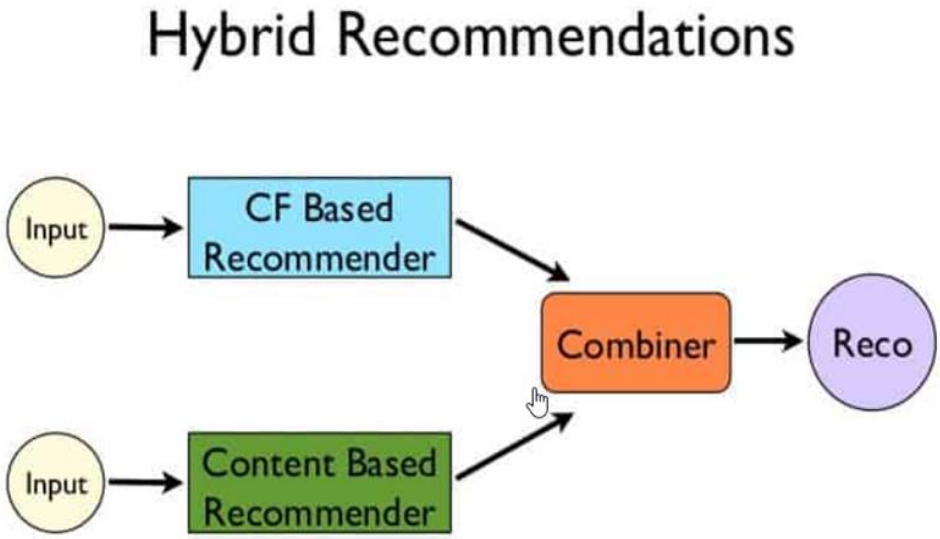
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11618

Process:

1. Input will be made to the Collaborative Filtering based recommender system as well as content based recommender system.
2. Then the combiner will take the recommendation from both the recommendation engine and combine it.
3. Then the combiner will recommend.



IV. IMPLEMENTATION

- Data Analysis & ML Model

➔ Cosine similarity:

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space. It is defined to equal the cosine angle between them, which is also the same as inner product of same vectors normalized to both have length 1.

- Cosine similarity is bounded in the interval [-1,1] for any angle.
- -1 means exactly opposite
- 1 means exactly same
- 0 means either completely opposite or completely same
- Formulae: A*B = ||A||*||B||*COS(angle)

➔ Singular Value Decomposition (SVD):

The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices. It has some interesting algebraic properties and conveys important geometrical and theoretical insights about linear transformations

- In singular value decomposition the estimator does not center the data before computing SVD, which means it can work for sparse matrices efficiently

Steps:

1. Similarity computation will be done first
2. Then selection of clusters will be made using SVD.
3. Then the prediction will be made using the combination.
4. Then in the real time based system the user ratings will be updated in the databases.
5. Then the whole process will be continued again for to make it better.

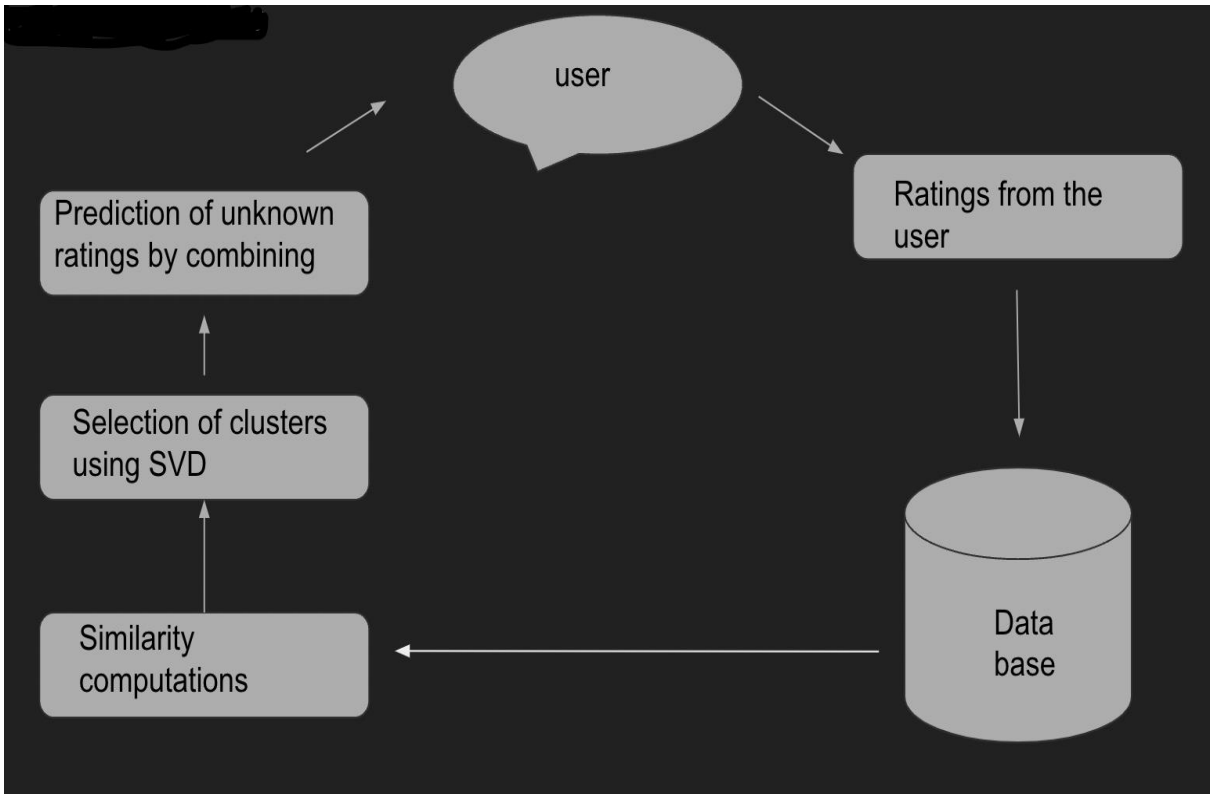
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

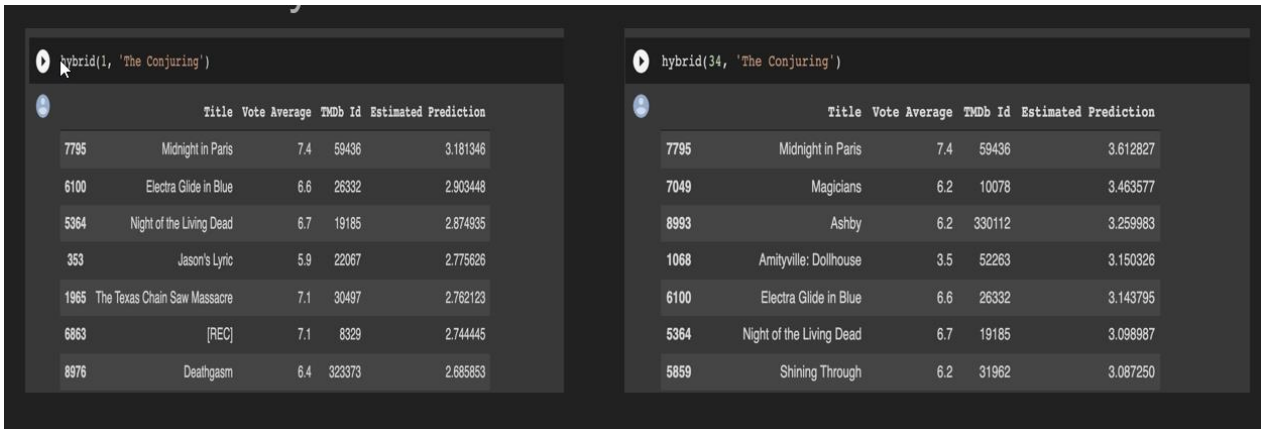
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11618

ARCHITECTURE:



RESULT:



hybrid(1, 'The Conjuring')

	Title	Vote Average	TMDb Id	Estimated Prediction
7795	Midnight in Paris	7.4	59436	3.181346
6100	Electra Glide in Blue	6.6	26332	2.903448
5364	Night of the Living Dead	6.7	19185	2.874935
353	Jason's Lyric	5.9	22067	2.775626
1965	The Texas Chain Saw Massacre	7.1	30497	2.762123
6863	[REC]	7.1	8329	2.744445
8976	Deathgasm	6.4	323373	2.685853

hybrid(34, 'The Conjuring')

	Title	Vote Average	TMDb Id	Estimated Prediction
7795	Midnight in Paris	7.4	59436	3.612827
7049	Magicians	6.2	10078	3.463577
8993	Ashby	6.2	330112	3.259983
1068	Amityville: Dollhouse	3.5	52263	3.150326
6100	Electra Glide in Blue	6.6	26332	3.143795
5364	Night of the Living Dead	6.7	19185	3.098987
5859	Shining Through	6.2	31962	3.087250

Model-based approaches including SVD algorithms tackle the problem with the help of Dimensionality reduction and matrix Factorisation. To better the performance and accuracy of these models, by stacking SVD and content-based model we built a hybrid model, which gives an improved performance and accuracy of 20 percent more than the independent models.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11618

```
Evaluating RMSE, MAE of algorithm SVD on 5 split(s).

                  Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Mean    Std
RMSE (testset)    0.8767  0.8713  0.8763  0.8729  0.8702  0.8735  0.0026
MAE (testset)     0.6739  0.6720  0.6719  0.6719  0.6673  0.6714  0.0022
Fit time          4.74    5.26    4.75    4.65    4.63    4.80    0.23
Test time         0.15    0.24    0.15    0.23    0.15    0.18    0.04
```

V. CONCLUSION

This paper introduces a hybrid method of recommender system where the combination of collaborative filtering as well as the content-based filtering will be used for the recommendation. This also minimizes the problems associated with the system by this combination. It used the cosine similarity and SVD method of machine learning for the process and also the movie lens dataset.SVD and SVD++ are model based filtering approaches implemented to overcome the limitations of memory- based models. We can still further increase the efficacy of this model by increasing the datasets to 10M or more and by using different types of ML based models and by adding more features also.

REFERENCES

[1] Movie recommender systems using hybrid models based on graphs with co-rated, genre, and closed caption features. By Putra Pandu Adikara , Yuita Arum Sari , Sigit Adinugroho and Budi Darma Setiawan Department of Information Engineering, Universitas Brawijaya, Malang, Indonesia Graduate School of Information Science and Engineering, Ritsumeikan University, Japan

[2] The evaluation of a hybrid recommender system for recommendation of movies By Tomaz Po and Matevz Kunaver and Matevz Poga and Jurij F. Tasic.

[3] Hybrid Movie Recommendation System Using Machine Learning IEEE International Conference on Communication information and Computing Technology (ICCICT), June 25-27, 2021, Mumbai, India By Sarvesh Kulkarni & Sakina Salmani

[4] Hybrid Recommendation System For Movies International Research Journal of Engineering and Technology (IRJET) By Rohan Nayak, Aniket Mirajkar, Jeetesh Rokade, Prof. Girish Wadhwa B.E. Student, Department of Information Technology Engineering , Vidyalankar Institute of Technology, Maharashtra, India

[5] S. Bhat and K. Aishwarya, "Item-based Hybrid Recommender System for newly marketed pharmaceutical drugs," 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013.

[6] J. Chen, K. Chao and N. Shah, " Hybrid Recommendation System for Tourism," 2013 IEEE 10th International Conference on e-Business Engineering, 2013

[7] P. Darshna, "Music recommendation based on content and collaborative approach and reducing cold start problem," 2018 2nd International Conference on Inventive Systems and Control (ICISC), 2018

[8] V. Powar, S. Girase, D. Mukhopadhyay, A. Jadhav, S. Khude and S. Mandlik, "Analysing recommendation of colleges for students using data mining techniques," 2017 International Conference on Advances in Computing, Communication and Control (ICAC3), 2017