

Comparison of Collaborative Filtering, Content-based Filtering, and Hybrid Recommendation System

¹Aman Kandari, ²Satnam Kaur

¹Student, ²Assistant professor

^{1,2}Computer Science and Engineering Department

^{1,2}SGT University, Gurugram, Haryana India

Abstract: In the current scenario of the internet which is filled to brim with content. Recommendation systems are of extreme importance to filter and recommend data and, content based on user preference, profile, previously preferred content, and similar relevant areas of interest. Providing the user with content that they would like is extremely important to keep their attention on platforms like Netflix and also it is equally important to recommend similar [6]items on ecommerce websites like amazon.To make this possible it is important recommendations are made on shows, clothes to purchase, shows to watch. A recommendation system should make it easy for user to find their preferred items and also it is important for it not to ruin their experience by recommending too much of same content day in day out. To better the user recommendation different platforms are using user data and different machine learning algorithms to properly recommend content to the user.

This research paper aims to compare various recommendation techniques for content recommendation. Techniques such as content-based filtering, Collaborative filtering, and hybrid filtering are used for such tasks, and an effort is made to try to find best suitable content filtering and recommendation technique by comparing their pros and cons.

Keywords - Ease of use, Filter data, Internet, Machine learning, OTT platforms, Preferences of user, Recommendation system, Unique content.

I. INTRODUCTION

Internet with its new way to interact with world has changed the way we humans live our lives. In 21st-century digital platforms such as Amazon and Netflix makes it convenient to order things online and consume digital content. According to a recent report Netflix have over 203 million paid customers for their streaming platform, whereas Amazon has more than 200 million paid subscribers for its prime membership which is fast shipping, streaming, and multi-utility platform [7].

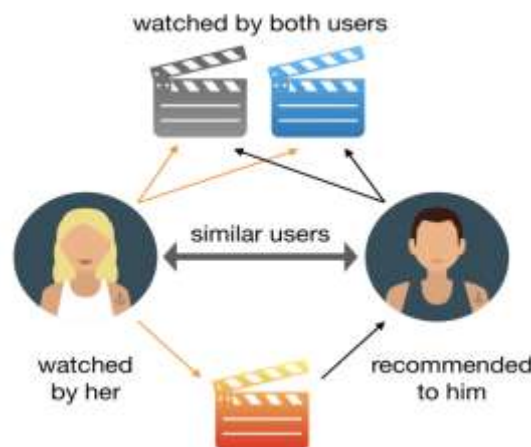


Fig. 1 Content recommendation in day to day life

Now the problem is quantity which is both a boon and bane in digital world has made it extremely difficult to choose. As of February 2018, Netflix US has 1,569 TV shows on offer and 4,010 movies with such huge volume it becomes extremely hard to choose new content, apart from viral shows. This, is where Recommendation systems (RSs) help net users, they suggest content and items based on user preferences, past and personal information of the user. Items of user's interest are recommended by such systems. All major platforms use recommendation systems, such as Amazon for products, myntra for clothes, Netflix for shows/movies and so and so forth. Currently there is also Scienstein which is recommendation system for research papers [5]

For recommendation systems to make an as accurate prediction as possible RSs collect user information.

II. LITERATURE SURVEY

Recommendations are necessary to provide the user with preferred data. [1][2]Bobadilla divides the recommendation system into four categories:

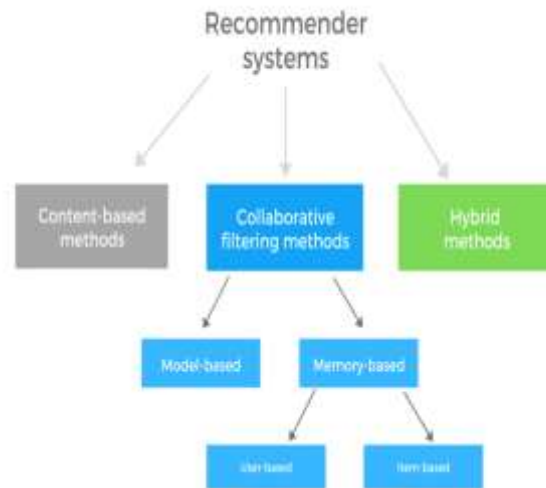


Fig 2. Different types of recommendation systems

1. Content-based recommendation system
2. Collaborative recommendation system
3. Demographic Filtering
4. Hybrid recommendation system.

These could be defined as:

1. Content based based recommendation system:

In this method attributes of content are used to recommend similar content

2. Collaborative recommendation system:

This method observes how similar one user's profile is to that of other user, by this way it finds people similar to user and recommend content they like.

3. Demographic Filtering:

This technique decides which items could be ideal for recommendation based on demographic data (which includes age, race, ethnicity, gender, income, education etc).

4. Hybrid recommendation system:

It is a recommender system which is combination of multiple recommendation methods.

According to Bobadilla Recommendation Systems collect information on users preference based on certain areas of interest such as songs liked by users, Job searches, Web series watched, items purchased, etc.

Information collected could be done explicitly like for example using users rating, or score or it could be done implicitly which is indirect way of collecting information, like monitoring of user behavior.

RS uses different sources of information and use them to predict the kind of content that user might like, and trying to balance various factors like accuracy, pleasure, novelty, etc. This is important as user might like novelty of multiple recommended shows, but also want the recommendation to be accurate, to get things right all areas have to be well balanced.

There is a growing trend of using information from the Internet of things.

The use of RS has been increasing which has resulted from its use in diverse areas. In the most recent research paper on RS, the focus has been on Netflix and movie recommendations. As RS evolved, research has shown that hybrid RS, which merges different techniques to get advantages of each, has been the most advantageous. Hybrid RS is explained in more detail ahead in this research [3]

III. TECHNIQUES AND APPROACHES

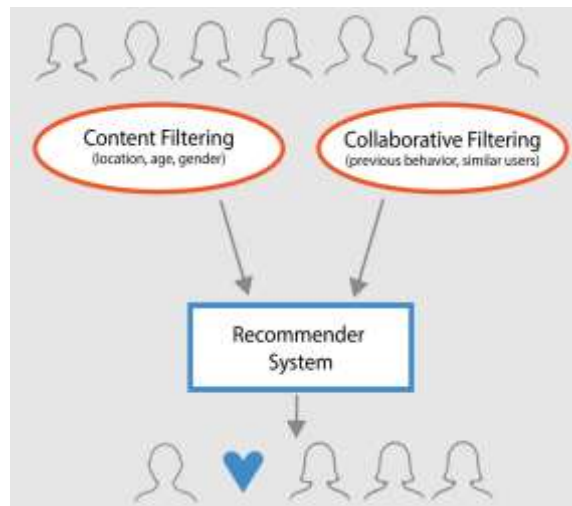


Fig 3. Parameters used by content-based and collaborative filtering

3.1 Collaborative Filtering:

First, we will review what happens when we use collaborative filtering as a recommendation system.[4]

Of all recommendation methods, collaborative filtering is most used. In collaborative filtering, content is recommended, depending on how similar the profile of one user is to that of another, it finds people similar to you and recommend content they have liked, note this type of recommendation system is based on past evaluation of a large group of individuals.

Now we would try to understand two types of collaborative Filtering. These are 1. Memory-based collaborative filtering and 2. Model-based collaborative filtering

3.1.1 Memory based collaborative filtering

In this type similar users are found based on cosine similarity or person correlation, and weighted avg. of ratings is taken.[9]

Memory based filtering can be further divided into two categories: User-item filtering and item-item filtering

a. User-item:

This type of filtering takes a user and find users similar to that user and then recommend items that those similar users liked.

b. Item-item:

Item-item collaborative filtering looks for similar items, based on items users have already liked. It takes items and outputs other items as recommendations.

3.1.2 Model based collaborative filtering

In this approach we use machine learning to find user ratings of unrated items [10] e.g.: PCA, neural networks, matrix factorization [11] etc.

Advantages of collaborative filtering

- No domain knowledge necessary
- Domain knowledge isn't needed as embeddings are learned automatically.
- Serendipity
- The model can help users discover new interests.
- Businesses can be expanded and can popularize new products.

Disadvantages of collaborative filtering

- Cannot handle fresh items (cold-start problem) i.e. items that have not been purchased or rated.
- Hard to include side features for query/item Side features are features beyond the current query.
- User's previous history is required or data for products is required based on the type of collaborative method used.

3.2 Content-based filtering

In this method attributes of the content are used to recommend similar content.[8] This type of recommendation techniques are free from problem such as that of Cold start, as it works through attributes such as artist and genre in case of song database, so that it is easy to recommend new content.

Process of content-based filtering:

In content-based filtering, the most used algorithm is tf-idf representation [12]. To create a user profile mostly two areas are focused on:

- User preference model
- User's interaction log with recommender system;

The creation of content-based profiles of the user is done with the help of a weighted vector of item features.

Advantages:

- The data about other users isn't required as recommendations are specific to user. This makes the model scalable.
- Specific requirements of user could be captured and then recommendation of niche content could be made.
- User is usually satisfied as they get a recommendation based on their preference.

Disadvantages:

- Domain knowledge required in this method is quite high as certain feature representations are hand engineered.
- Recommendation could be made only on existing interests of user. New items are difficult to recommend.
- It would be extremely difficult to expand the business as the user does not try a different type of product.

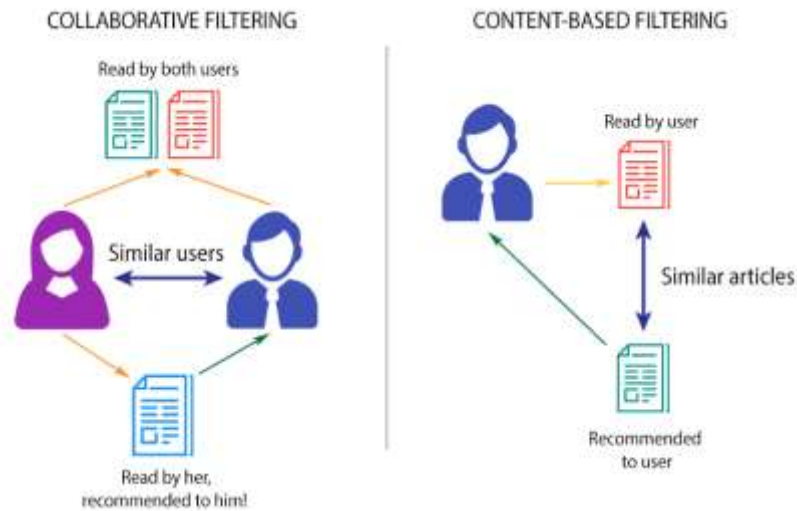


Fig 4. Head-to-head comparing Collaborative Filtering and content-based filtering

3.3 Hybrid Recommender system

It is a recommender system which is combination of multiple recommendation methods. [13] Collaborative filtering is used most frequently used in combination with some other technique in an attempt to avoid the ramp-up problem.

The hybrid recommendation could be more effective than both content-based and collaborative recommenders, as it minimizes downsides and improves recommendation accuracy. Hybrid Recommender is a common combination of Collaborative filtering with demographic filtering or Collaborative filtering with content-based filtering to maximize the benefits of these techniques.

Some ways in which hybrid approaches could be implemented are:

- Add some of content based characteristics into collaborative approach.
- Individually work on collaborative and content based systems and choose best from pool of predictions by both.
- Add some collaborative characteristics into a content based approach.[14]

Note: In any recommendation system, biggest problems are those of cold start and sparse data. Both of these could be solved using a hybrid recommender.

IV. CONCLUSION

The recommendation system is used by nearly every social media website(YouTube, Facebook, Instagram),e-commerce websites like Amazon, and streaming platforms like Netflix, things are recommended based on similarity of the user which is collaborative filtering and based on attributes, where content which have similar attributes to one liked by user is recommended which is content based filtering, both these methods are highly used and have their unique benefits, but they also have downsides which drag them down, one that sticks out is cold start problem for collaborative filtering, where recommender can not recommend fresh items, items which have not been purchased or rated. In the case of content-based filtering, as content is highly personalized, low-quality items which are similar to user's preference might be recommended. To overcome such problems, Hybrid recommendation, which has properties of different recommendation systems, is used. Hybrid recommendation system is more effective than both content-based and collaborative recommenders as it minimizes problems such as that of cold start and sparse data and it increases the accuracy of recommendations.

In the end, we observe that different recommendation systems are used by different companies but, hybrid recommendation systems are the most efficient with the least downside.[15]

V. REFERENCES

- [1] Poonam, Goudar, Barve.(2015) "Survey on Collaborative filtering, Content based Filtering , and Hybrid Recommendation System".
- [2] J. Bobadilla, A. Hernando, F. Ortega, J. Bernal, "A framework for collaborative filtering recommender systems", *Expert Systems with Applications* 38 (12) (2011) 14609–14623.
- [3] R. Burke, Hybrid recommender systems: survey and experiments.
- [4] Singh, Pradeep & Dutta Pramanik, Pijush & Dey, Avick & Choudhury, Prasenjit. (2021). Recommender Systems: An Overview, Research Trends, and Future Directions. *International Journal of Business and Systems Research*. 15. 14–52.
- [5] Gipp, Bela & Beel, Joeran & Hentschel, Christian. (2009). *Scienstein: A Research Paper Recommender System*.
- [6] Gómez-Uribe, Carlos & Hunt, Neil. (2015). The Netflix Recommender System. *ACM Transactions on Management Information Systems*. 6. 1-19. 10.1145/2843948.
- [7] Lad, Advait & Butala, Shivani & Bide, Pramod. (2020). A Comparative Analysis of Over-the-Top Platforms: Amazon Prime Video and Netflix. 10.1007/978-981-15-3325-9_22.
- [8] Zisopoulos, Charilaos & Karagiannidis, Savvas & Demirtsoglou, Georgios & Antaris, Stefanos. (2008). Content-Based Recommendation Systems.
- [9] M. R. Zarei and M. R. Moosavi, "A Memory-Based Collaborative Filtering Recommender System Using Social Ties," *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, 2019, pp. 263-267, doi: 10.1109/PRIA.2019.8786023.
- [10] Do, Minh-Phung & Nguyen, Dung & of Loc Nguyen, Academic Network. (2010). Model-based approach for Collaborative Filtering.
- [11] Dziugaite, Gintare & Roy, Daniel. (2015). Neural Network Matrix Factorization.
- [12] Kim, SW., Gil, JM. Research paper classification systems based on TF-IDF and LDA schemes. *Hum. Cent. Comput. Inf. Sci.* **9**, 30 (2019).
- [13] Çano, Erion. (2017). Hybrid Recommender Systems: A Systematic Literature Review. *Intelligent Data Analysis*. 21. 1487-1524. 10.3233/IDA-163209.
- [14] Xiao, Yao & Shi, Quan. (2015). Research and Implementation of Hybrid Recommendation Algorithm Based on Collaborative Filtering and Word2Vec. 172-175. 10.1109/ISCID.2015.211.
- [15] Turnip, Rudolf & Nurjanah, Dade & Kusumo, Dana. (2017). Hybrid recommender system for learning material using content-based filtering and collaborative filtering with good learners' rating. 61-66. 10.1109/IC3e.2017.8409239.