A Review on Techniques of Recommendation System

Palika Jajoo[1], Dolly Mittal[2]
[1]Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur-302017 (INDIA)
[2]Department of Information Technology, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur-302017 (INDIA)
Email: **palika@skit.ac.in, dolly.mittal@skit.ac.in**
Received 16.01.2021, received in revised form 19.02.2021, accepted 19.02.2021
doi: 10.47904/IJSKIT.11.2.2021.31-36

***Abstract*- Recommendation System is a method to find the needs of the customer either it can be data or an item from the enormous amount of online data. This is technique which reforms from content to customer essential in every domain. In this paper, we present the three major techniques that are used for generating recommendation: Collaborative filtering, Content-based and Hybrid. It is a comprehensive survey of these techniques, enlighten with their advantages and drawbacks and provide a roadmap for more innovations in this area.**

Keywords– Recommendation System, Collaborative, Content, Hybrid.

1. INTRODUCTION

Recommender system came into existence in 1990s as recommendations were implemented based on rating system [1]. But in today's world, demand of online services (for e.g., online shopping, online movies, online music etc.) increasing tremendously therefore Recommendation System becomes an important part to find new services or items based on information about the user or the items.
The details of an item, content or an application can be fetched from user's tagging, reviewing, number of likes, or voting they provides. Review can also be considered in form of blog, message written or acknowledge to an online organization [1]. RS basically works on these three frameworks: Collaborative filtering, Content based and Hybrid. Other techniques that create revolution in recommender system are as follows [1]:

- GroupLens- It was the first technique based on user's ratings.
- MovieLens- It was the first Recommendation System on Movie.
- Fab– It was the first Recommendation System framed as hybrid by Marko.
- Pandora- It was the first music Recommendation System.
- Amazon received the patent rights by applying item based Collaborative filtering.

2. TECHNIQUES OF RECOMMENDATION SYSYTEM

In this section, we discuss the techniques of recommendation system in details:

2.1. Content-based Filtering

Content based recommendations are not based on similarity of other users but rather based on description of items. The user generates the data in explicit way (by giving rating) and in implicit way (by clicking on a link) [2]. User profile is generated based on the data provided by them which is further used to make recommendation to the user. For example, in a Web-based ecommerce, if some user purchased some thriller novels, then Recommendation System will likely recommend a new thriller novel that he or she has not purchased yet. It can be analysed the items which the user has purchased, frequently visited, liked, surfed the websites or rated with positive points then other similar items are also recommended. The engine become more accurate as the user keep on increasing inputs and on the recommendations able to take actions.

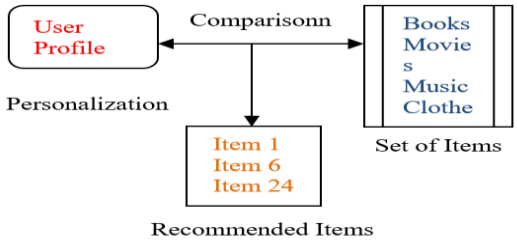

Fig. 1 Content-based Recommendation System

Here are few illustrations of Content based filtering recommendation System [3]:

- LIBRA (Learning Intelligent Book Recommending Agent): It is a content-based recommendation system on Books using learning for text categorization.
- CBMRS (Content-Based Music Recommendation System): It is a content-based recommendation System on Music.
- PRES (Personalized Recommendation System): A content-based Recommendation System on Home Improvement.
- Cobra (Content-based Filtering and Aggregation of Blogs and RSS Feeds): A content-based filtering and aggregation of Really Simple Syndication (RSS) feeds and blogs.

In Content based Recommendation system, Term frequency (TF) and Inverse Document frequency (IDF) are two terms that recognize the significance of a movie, an item, document or article so on [4].
Term frequency means to figure out how frequently the words occur in a document. Another term Inverse Document frequency is used to boost the learning frequency words or can say fewer common words in a document. Formula that are used as

$$TF(x,y) = \frac{\text{Word x frequency in document y}}{\text{Total words in document y}}$$

$$IDF(x) = log\left[\frac{Total\ documents}{Documents\ with\ x\ word}\right]$$
$$= log\left[\frac{z}{d_{zj}}\right]$$

Here z represents number of documents in total and dz_i are the number of documents with respect to term i.
IDF is used because of the reason that suppose we looking for "A Survey on Recommendation System" on Web search engine. It is obvious that "A" and "on" will appear more frequently than "Recommendation" but it can be observed in search query that the relative importance of word "recommendation" is more than the common words. In this scenario, weighting of TF-IDF determines the importance of an item (document) by negating the effect of frequently occurring words and in order to minimize the effect of frequently occurring words, log is used. For instance, TF=5 vs TF=6 is highly diverse from TF=20 vs TF=2000. Alternatively, the importance of a document can't be computed as normal raw count and thus equation is shown below:

$$W_{t,d} = \begin{cases} 1 + log_{10} tf_{t,d} & , \ if\ t_{f,d} > 0 \\ 0 & , \ Otherwise \end{cases}$$

Term Frequency	Weighted Term Frequency
0	0
10	2
1000	4

After computing TF-IDF scores, we use Vector space model to find what types of items are similar to one another in preference to profile of users.

2.1.1. Vector Space Model

In Vector Space model, every item is considered as its vector of its characteristics in an n-dimensional space and the angle between vectors are computed to find closeness between them [5]. Later on, user profile is generated depends upon action taken by users on earlier characteristics of items and in a similar way we can figure out the closeness between an item and a user.

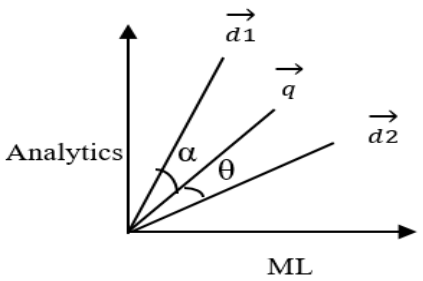

Fig. 2 Graph 1

Graph 1 determines the cosine similarity between Vectors. α and θ are two cosine angles among vectors d1, q and d2. Cosine similarity is used to determine the closeness between two vectors.
We can understand with an example, by showing 2-D representation of an attributes ML and Analytics.

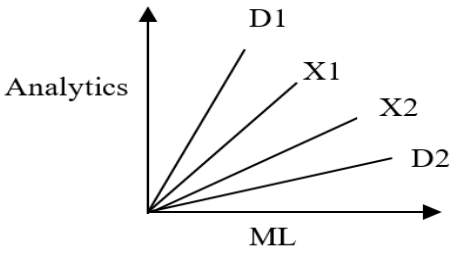

Fig. 3 Graph 2

In Graph 2 X1 and X2 are two users. D1 and D2 are two documents. The document D2 concern more on ML than Analytics and D1 concern more on Analytics than ML. User X1 is fond of articles on Analytics more than ML and conversely for user X2. The techniques to determine measures of likes and dislikes of users is computed by considering the cosine of angle between vectors of document and user profile. The reason for taking cosine is, with declined value of angle, the cosine's value will be inclined which signifies more similarity. The cosine computation is the sum-product of vectors and these vectors become of length 1 after doing the normalization.
Content Based filtering can be accomplished through Bayesian Classifier and several Machine Learning methods comprising Clustering, Decision trees and Artificial Neural Network [6][7].

2.2. ***Collaborative Filtering***

This is most common technique used for Recommendation System. It is approach towards like user recommendation. Based on evaluation of all past ratings of either implicit (it indirectly recommends user preference such as webpage views, clicks, purchase records, time spent, watched movie or not and so on) or explicit ratings (direct feedback given by user on an item in the form of rate

like 1 stars, 2 stars and so on) of all users, an item is recommended to user.

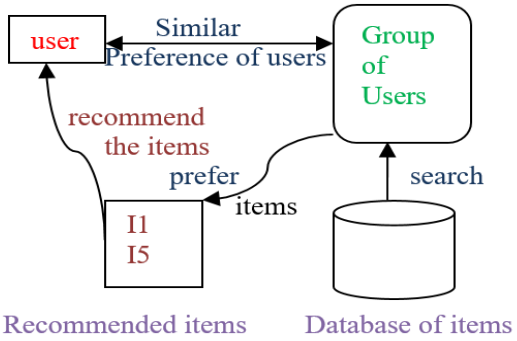

Fig. 4 Collaborative based Recommendation System

Basically CF system determines that how to recommend an item to the user by exploiting similarities in rating behaviour among various users for items which are collected through feedback [3].

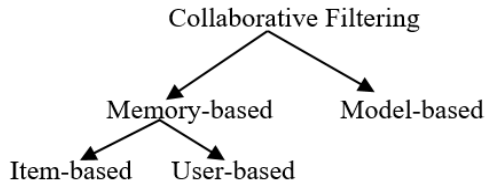

Fig. 5 Classification of Collaborative Filtering

2.2.1. Memory-based Collaborative filtering

In this approach, a user X becomes part of a group of people with analogues interests. The most popular algorithm this method uses is K-Nearest Neighbour algorithm. KNN identifies the neighbours of a new user or presently active user and recommends new items for him or her [9].

Nearest Neighbour algorithm applied in two types: User based and Item based Collaborative filtering.

A. User based Collaborative filtering

a) For any user X, it finds similar users or we can say produces set of k neighbours that are close to user X using the similarity-based methods.

b) Then apply an aggregation approach: the average, weighted sum and adjusted weighted aggregation.

c) Select top N recommendation from step 2 that provide most gratification to active user [10] [11].

There are various ways to compute closeness between users. To understand all the techniques, we consider a table which shows rating of movies given by different users [12].

Table 1: Matrix of Users & Movies

Users/Movies	AVG	SM	MAT	TT	TB
X	5		4	1	
Y	4	3			
Z			1	5	4

i. The Jaccard Similarity

To compute the similarity between X and Y using Jaccard coefficient from Table 1 as

$$sim(X,Y) = \frac{|r_X \cap r_Y|}{|r_X \cup r_Y|} = \frac{1}{4} = 0.25$$

$$sim(X,Z) = \frac{|r_X \cap r_Z|}{|r_X \cup r_Z|} = \frac{2}{4} = 0.5$$

We can conclude that sim(X, Y) < sim(X, Z) but here it ignores the rating values as we can see the difference in the ratings given by user X and Z for particular movie.

ii. Cosine Similarity

To overcome the problem of ignorance of rating, there is another technique called cosine similarity to fond the similarity between two vectors as follows:

$$sim(X,Y) = \cos(r_X, r_Y) = \frac{r_X \cdot r_Y}{||r_X \cdot r_Y||} = 0.61$$

$$sim(X,Z) = \cos(r_X, r_Z) = \frac{r_X \cdot r_Z}{||r_X \cdot r_Z||} = 0.21$$

It treats missing ratings as negative.

iii. Pearson Correlation

It is also known as Centered Cosine Similarity. It normalizes ratings by subtracting with row mean. It treats missing ratings as “average”.

After normalizing the table 1 would be as follows

Table 2: Matrix of Users & Movies

Users/Movies	AVG	SM	MAT	TT	TB
X	5/3		2/3	-7/3	
Y	1/2	-1/2			
Z			-7/3	5/3	2/3

$$sim(X,Y) = \cos(r_X, r_Y) = \frac{r_X \cdot r_Y}{||r_X \cdot r_Y||} = 0.41$$

$$sim(X,Z) = \cos(r_X, r_Z) = \frac{r_X \cdot r_Z}{||r_X \cdot r_Z||} = -0.64$$

B. Item Based Collaborative filtering:

User based CF suffers from scalability problem when number of users increases. A new method is introduced [13] called Item based CF to overcome this drawback.

- For item i, determine another homogenous items.
- Item i rating’s is estimated by considering ratings for other homogenous items.
- It computes similarity-based methods b and prediction functions by taking the weighted average or regression can be used.

User based CF similarity-based methods are used to find closeness between users and Item based CF are used to find closeness between items.

2.2.2. ***Model -Based Collaborative Filtering***

In this type of technique, a model is proposed to figure out the behaviour of user rating. It is used to make predictions based on available rating data. It learns model of user-preference from collection of ratings and probalistic approach is used based on ratings given on previously rated items and then calculates the probability of giving rating to new item by users.

Among various model-based methods, some of in-demand methods are used like matrix factorization, fuzzy systems, Bayesian Classifier, Neural Networks, Markov Decision Process (MDP). The details of the technique are mentioned [14].

In Table 3 we listed the benefits and drawbacks of Content-based, Collaborative filtering and Hybrid based techniques [1] [3] [4] [7]:

2.3. **Hybrid based Recommendation System**

Hybrid means to aggregate so here Hybrid based RS join different recommendation system so as to build more robust framework. As we know Content based and Collaborative filtering has its own benefits and drawbacks. Therefore, in this, we basically reduce the demerits of any technique by using the merits of other techniques and construct a robust system [15].

2.3.1. ***Hybrid Recommender combining Collaborative filtering algorithms***

Instead of using pure memory based or model-based CF, it's better to use both memory based and model-based CF techniques as a combination to make prediction performance better.

2.3.2. ***Hybrid Recommender incorporate Content based and Collaborative filtering feature***

It is approach of combining Collaborative filtering techniques and Content based methods so as to overcome the weakness of each other. It is main source of hybrid technique. For instance, the models fails if there are no ratings for new items in CF but in Content based methods anywhere information about the item is available, new items can be recommended more accurately and precisely. There are various ways of doing it but the major steps involved [10]:

- Implement them independently and then integrate their predictions.
- Integrate some of content-based features within Collaborative features.
- Integrate some of collaborative based features within Content-based features.
- It forms a unified model after integrating content and collaborative methods.

Hybrid based RS is useful as it overcomes the limitations of Cold start problem, Sparsity etc. [10].

3. LITERATURE REVIEW

In this section, we have done some literature review on various recommendation system on various items using different techniques so as to overcome the limitation of existing methods as shown in Table 4.

Table 3: Advantages ad Limitations

Technique	Algorithms	Advantages	Limitations
I.Content-based filtering	Bayesian Classifier, Clustering, Decision Trees, Artiifical Neural Network	1. Not required data of other users 2. No cold start problem 3. No data sparsity	1. To define item features it is important to do content analysis. 2. Can't estimate about the quality of products
II.Collaborative filtering a. Memory-based CF b. Model- based CF	 K-Nearest Neighbour Matrix factorization, Bayesian Classifier, Markov Decision process.	1. Memory based CF is easy to implement. 2. The content of an item to be recommended is not considered here. 1. Model based CF better solves the problems of sparsity & scalability. 2. It improves the prediction performance	1. Cold start issue for new products and for different users. 2. Dependent on user's ratings. 1.Trade-off between scalability and prediction performance 2. It can be expensive as model is built.
III. Hybrid - based filtering	Combination of Collaborative and Content based filtering	1. Overcomes the limitation of Cold-Start Problem. 2. Reduces Sparsity. 3. Improve Prediction Performance	1. Complexity is increased. 2. Implementation is expensive.

Table 4: Literature Review

Recommendation System	Author	Work
Movies RS	Yukun Cao et.al [16]	Introduced a model combining both content-based and collaborative approach which suggest movies depends on user's interest using ML techniques.
Fashion products RS	Hyunwoo et.al. [17]	Proposed a method in which they extend the existing CF methods to recommend and reflect the characteristics of fashion products.
Consumer Electronics RS	Yukun Cao et.al. [18]	Proposed an intelligent fuzzy based recommendation system for consumer to retrieve best favoured electronic products. This system builds the personalized recommendation system to find the best-favoured products by considering the domain expert knowledge and information offered by consumer.
Citation RS	Wenyi Huang et.al. [19]	Introduced citation RS called RefSeer. It can be used to check completeness of references based on content of a paper manuscript. The complexities of training and recommending show that system is efficient and scalable.
MOOC Courses RS	Tao Huang et.al. [20]	MCRS has made great improvement in course recommendation model algorithm. MCRS algorithm is more efficient than Hadoop Apriori algorithm recommend apriori course to user.
Personalized RS	Robin van et.al. [5]	Described a recommender system PRES that recommends small articles about home improvement by using content-based filtering techniques.

4. CONCLUSIONS

Numerous research work has been done on the approaches and techniques of Recommendation System. In this paper, an overview of various kinds of recommendation system approaches such as Content-based, Collaborative filtering, Hybrid has been described. We also mentioned advantages and limitations of these approaches so. Also outlined few of the recommendation system on Movies, Electronic products, Fashion products, MOOC and so on by various researchers. From this we just set the outset for any researchers to select recommendation system approach based on the requirements.

This research doesn't limit here. Work on recommendation system with Machine Learning and Deep Learning has been done and in progressive to provide better recommendations in aspect of quality.

5. REFERENCES

[1] G. Suganeshwari and S.P. Syed Ibrahim" A Survey on Collaborative Filtering Based Recommendation System", 3rd International Symposium on Big Data and Cloud Computing Challenges (2016) Pages: 503- 517.

[2] Content Based Recommender Engines https://www.analyticvidhya.com

[3] L. Sharma and Anju Gera "A Survey of Recommendation System: Research Challenges" International Journal of Engineering Trends and Technology (2013) Volume 4 Issue 5, Pages 1989-1992.

[4] M.Sharma and S. Mann "A Survey of Recommender Systems: Approaches and Limitations", ICAECE (2013), ISSN: 2319-1058.

[5] R.V Meterena and M. V.Someren – "Using Content-Based Filtering for Recommendation", NetlinQ Group (2000).

[6] A. Stefanos, D.Georgios, Z. Xarilaos, K. Savvas "Content based Recommendation System" (2008).

[7] I.Portugal, P.Alencar and D.Cowan "The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review", Expert Systems with Applications (2018) Volume 97, Pages 205-227

[8] A. Hameed and O.A Jadaan "Collaborative Filtering Based Recommendation System: A survey" International Journal on Computer Science and Engineering (2012), Vol. 4 No. 05, Pages 859-876.

[9] B.Bhatt, P. J Patel and H. Gaudani "A Review Paper on Machine Learning Based Recommendation System International Journal of Engineering Development and Research (2014), Volume 2, Issue 4, Pages 3955-3961.

[10] P. B. Thorat, R. M. Goudar and S. Barve "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System" International Journal of Computer Applications (2015), Volume 110 – No.4, Pages 31-37.

[11] J. Bobadilla, F. Ortega and A. Hernando, A. Arroyo "A Balanced Memory-Based Collaborative Filtering Similarity Measure" International Journal of Intelligent Systems (2012) VOL. 00, Pages 1–8.

[12] V. Verma and R.K. Aggarwal "Accuracy Assessment of Similarity Measures in Collaborative Recommendations Using CF4J Framework", I.J. Modern Education and Computer Science (2019), Volume 5, Pages 41-53. Volume 2, Issue 4, Pages 3955-3961.

[13] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl "Item- Based Collaborative Filtering Recommendation Algorithms, ACM (2005), Pages 285-295.

[14] M. P. Thi Do, D.V Nguyen, and L.Nguyen "Model based Approach for Collaborative Filtering", The 6th International Conference on Information Technology for Education, Vietnam (2010).

[15] D.Das, L.Sahoo and S.Dutta "A Survey on Recommendation System" International Journal of Computer Applications (2017), Volume 160 – No 7, Pages 6-10.

[16] F. Furtado and A, Singh "Movie Recommendation System Using Machine Learning" International Journal of Research in Industrial Engineering, Vol. 9, No. 1 (2020), Pages 84–98.

[17] Hyunwoo Hwangbo, Yang Sok Kim and Kyung Jin Cha

"Recommendation system development for fashion retail e-commerce", Electronic Commerce Research and Applications 28 (2018) Pages 94–101.

[18] Y. Cao and Y. Li "An intelligent fuzzy-based Recommendation system for consumer electronic products" Expert Systems with Applications 33 (2007), Pages 230–240.

[19] W.Huang, Z.Wu, P.Mitra and C.L. Giles "RefSeer: A Citation Recommendation System", IEEE./ACM Joint Conferences on Dn Digital Libraries (2014), London.

[20] Tao Huang, Gaoqiang Zhan, Hao Zhang and Heng Yang "MCRS: A Course Recommendation System for MOOCs", International Conference on Advanced Technologies Enhancing Education (2017), Volume 68, Pages 82-85.