# A Survey: Collaborative Filtering, Content-based Filtering, Hybrid Recommendation Approach

**Tarang Raval[1], Yask Patel[2]**
[1]M.Tech, Department of Information Technology, PIET-Parul university, Vadodara, India
[2]Asst.Professor, Department of Information Technology, PIET-Parul university, Vadodara, India
Email - TarangRaval1992@gmail.com Yask.patel@paruluniversity.ac.in

***Abstract:*** *The World Wide Web have presented to us overabundance information in shifted fields and of the information or data over-burdening, it is extremely exhausting to discover related information. So, Recommendation System comes into existence. Recommended systems or recommendation systems are a subset of information filtering system that used to expect the evaluation or preference that user would receive to an item. In recent few years E-commerce applications are widely using recommended system. Generally the about popular E-commerce sites are likely music, news, books, research articles, and products. Recommended systems are further available for business experts, jokes, restaurants, financial services, life insurance and twitter followers. Recommended systems method based on demographic, content-based filtering and collaborative filtering. Currently, these systems are making use of social implicit information for boosting of suggestion process, recommended systems handle personal, implicit and local information from the World Wide Web. This paper has a summary of recommended systems which include collaborative filtering, content-based filtering and hybrid approach of recommended system.*

***Keywords:*** *recommendation system, content-based, filtering, collaborative filtering, Hybrid Recommendation Approach.*

## 1. INTRODUCTION:

Recommended systems have become indeed popular in crisp years and are used in various World Wide Web applications. Recommended Systems (RSs) are software tools that are used to give suggestions to user according to their need. The suggestions associate mutually various decision-making processes, one as which items to bought for a song, what music to listen, Item is the general term used to denote what the system recommends to users.

A recommender system is the information filtering that applies data analysis techniques to the problem of helping customers find the products they would like to purchase by producing a predicted likeness score or a list of recommended products for a given customer[10]. Recommender systems work from a specific type of information filtering system technique that attempts to recommend information items (TV program/show/episode, Movie, books, news, images, music, web pages, scientific literature etc.) and social elements (e.g. events or groups, people) that are likely to be of interest to the user[8]. Broadly speaking, a RS suggests to a user those items that might be of user's interest.

Recommendation systems method are classified into 3 approaches i.e. collaborative, content based or knowledge-based, and hybrid approach. Each of them Explain in below section.

## 2. COLLABORATIVE FILTERING:

Collaborative filtering is practically extensively used approach to design recommended system. Collaborative Filtering (CF) methods spring an important role in the advice process, during the time Collaborative filtering is regularly used along mutually other filtering techniques like content-based, knowledge-based [1]. Basically Collaborative filtering methods are established on gathering and examining a large amount of information which based on users perspective, activities or preferences and anticipating taste of that at variance user by using their similarity with other users [5]. In collaborative filtering recommendation system recommended objects are occupied on the basis of previous evaluations of a large collection of users. The primary assumption in Collaborative filtering would be that the individuals who have preferences before may also like to have same kind of personal preferences in the future[6].

Collaborative filtering techniques can be classified memory-based and model-based collaborative filtering[7].

### *2.1 Memory Based approach*

The memory base collaborative selection technique is use complete dataset related to user-item dataset. Memory-based CF methods generally use rating matrix to maintain user-item repository to generate recommendation. Generally in memory base collaborative filtering technique use neighborhood item datasets to obtain the interest of

user, designed to use in future for all the ratings by referring to users or items whose ratings are similar to the other user or items[7].

In memory based methods commonalities of user and item are calculated depending on whether it is item or user based techniques. They rely upon their neighbors. So fortunately they are called as "k Nearest Neighborhood method (KNN)". KNN filtering suggests items to user, established on the similarity procedures [7]. It is the most popular approach and it performs the following three tasks to produce recommendations [7,8].

(1) Find similar users (neighbors) for the user a,
(2) Apply an aggregation approach,
(3) From step 2 select the top N recommendations

Memory based technique is classified into User based and Item Based approach.

1. User Based Approach

In the User-based approach the user plays an significant role. If certain majority of the client have matching taste formerly they join into the one collection .Recommendations are supposing to user based on evaluation of items by other users form the same group, with whom he/she shares commonplace preferences. If the item was actually rated by the person in the street, it will be preferred to the user[4].

2. Item Based Approach

The item-based approach investigates the set of items graded by target user and calculates their similarity with the prospective item and then chooses most similar items. Their representing similarities are also computed at the same time. Formerly the most similar items are discovered, after that by taking a weighted mean of the target customer's ratings on these similar items the prediction is calculated[4].similarity computation basically different types of similarity measures are used and weighted sum and regression used for prediction computation[4].

### *2.2 Model Based approach*

The main drawback of memory base collaborative filtering technique is it manage complete dataset related to user item Datasets and because of that this system is not work as fast as other collaborative system and besides occurs scalability Problem when generate real-time entries in recommendations program database[7]. To overcome those problems, model-based recommendation systems are introduced by researchers. In Model-based recommendation systems use some small datasets called model. This model is design using extracting some information from the huge database related to particular
Parameter/attribute and uses this model every time without using full database, because of that models increases both speed and scalability of recommendation system[6,7].

This approach reduces the dimension. So, this approach reduces the memory and reduces the processing time. Through this approach system can visualized more accurately and reduce the error also. Different methods are available to find hidden (latent) features. Most commonly used methods are MF (Matrix factorization), SVD (Singular value decomposition)[7].

### *2.3 Similarity Measures*

In user based CF similarity between user based and in Item based similarity is calculated. There are many functions to believe similarities, a well-known as the Pearson correlation, cosine based function, Jaccard coefficient and Manhattan distance. Any of these functions can be used as long as they have the related input format within the related range and get back on one feet a ratio which shows a high degree of similarity for higher values[6,7,1].

Similarity is define by data analysis in term of distance function. The distance function can be calculated using Euclidean distance[] or Manhattan Distance[].

$$Dist_{XY} = \sqrt{\sum_{k=1}^{m} (X_{ik} - X_{jk})^2}$$ [4]

$$Dist_{XY} = |X_{ik} - X_{jk}|$$ [4]

Cosine Based Similarity

It can be described as measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0 is 1, in simple fact it is less for any other angle [].

$$\cos(\theta) = \frac{x \cdot y}{\|x\|\|y\|} = \frac{x1 * x2 + y1 * y2}{\sqrt{x1^2 + y1^2}\sqrt{x2^2 + y2^2}} \quad [4]$$

Pearson Correlation

The similarity between any two vectors can even be calculated using Pearson correlation[]. The outcome ranges from -1 to 1. Here 1 and -1 represents that they are strongly correlated. If perhaps 1, then they are positively related. If -1 they are negatively related. In negative correlation the value of one vector diminishes as other vector raises. Towards 0 the level of correlation is null[].

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad [4]$$

Prediction Formula

Prediction is computed as the weighted average of deviations from the neighbor's mean, we can predict the rating for entire user-item pair by using the idea of weighted sum. First we require all the items similar to our target item, and from those similar items, we retain items which the active user has rated. We weight the user's rating for each of these items by the similarity between that and the target item. Finally, we scale the prediction by the sum of similarities to get a reasonable value for the predicted rating[].

$$\kappa_{x,y} = sim(u_x, u_y) = \frac{\sum_{h=1}^{n'}(r_{u_x,i_h} - \overline{r_{u_x}})(r_{u_y,i_h} - \overline{r_{u_y}})}{\sqrt{\sum_{h=1}^{n'}(r_{u_x,i_h} - \overline{r_{u_x}})^2}\sqrt{\sum_{h=1}^{m'}(r_{u_y,i_h} - \overline{r_{u_y}})^2}} \quad [4]$$

### *2.4 Limitation of Collaborative filtering*

*Cold-start problem*: There needs to be stuffing other users already in the system to find a match. Collaborative filtering is entirely depends on similar neighbor in the system, yet if these similar neighbors are not available in the system in the arch phase that is known as ―cold-start problem. This problem can be avoided by the hybrid approach[6,7].

*Data sparsity:* If there are large amount items to be recommended, even if there are many users, the user/ratings matrix is sparse, and it is hard to find users that have rated the related items[7]. In online shops that have a massive amount of users and items there are almost always users that have rated rarely a few items. Using collaborative and distinctive approaches recommender systems generally construct neighborhoods of users using their profiles. If a user has evaluated rarely few items then it's quite difficult to determine his taste and he/she could be related to the wrong neighborhood. Sparsity is the problem of lack of information[6,7].

*Scalability:* recommendations for at variance environments to what place of users and products exist. Therefore, a massive amount of computation power is periodic essential to compute recommendations.

## 3. CONTENT BASED FILTERING:

Content-based filtering (CBF) tries to recommend items to the perceptive user based on similarity count which is rated by that user surely in the past []. For example, if a user likes a web page with the racket ,mobile, pen drive ,and ,RAM, the CBF will promote pages related to the electronics world. Item recognition and a profile of the user`s orientation play an significant role in Content-based filtering. Content-based filtering algorithms offer to urge items based on similarity count. The best-matching items are chosen by comparing various candidate items with items previously rated by the user.

Content-based Recommender system work with user profiles which consists of user preferences for an peripheral []. It compares the peripheral lists and attempts to recommend items which are similar to that of user's chosen items in the past. It mostly uses keywords, tags and weights to characterize the items. These keywords are secondhand in information filtering and information retrieval. These results summary or abstract features of an peripheral in the system. For an instance Tf– idf representation (also called vector space representation) is a mostly used algorithm to obtain abstract features of the item[10].

### *3.1 Limitation of Content Based Filtering*

*Limited Content Analysis*: It's tough to analyze sufficient set of features that would represent the complete content. It's very difficult to analyze whether it is a good or bad document if they contain same terms and phrases. It doesn't guarantee the quality of the document or item. For multimedia data, audio streams, video streams and graphical images it is tough to extract the feature factors [10].

*Over-Specialization: Overspecialization* restricts users to items similar to the ones defined in their respective profiles and thus new items and other options are not discovered [9].

## 4. HYBRID RECOMMENDATION APPROACH:

Recent research has demonstrable that a hybrid approach could be preferably effective in some cases. Basically Collaborative filtering and Content-based filtering approaches approaching extensively used in information filtering application. As we visualize that every coin has two sides equally each approach has its own reward and weaknesses. Basically the main target of hybrid approach is to aggregate collaborative filtering and content-based filtering to improve recommendation accuracy[10].

Hybrid approaches can be implemented in various ways [10]:

1. Implement collaborative and content-based methods individually and aggregate their predictions.
2. Integrate some content-based characteristics into a collaborative approach,
3. Comprise some collaborative characteristics into a content-based approach, and
4. Construct a general consolidative model that integrates both content-based and collaborative characteristics.

Empirically, several studies compared the performance and quality of hybrid approach and proved that the hybrid approaches provide recommendations more accurately than original recommender approaches. The very common issues cold start and sparsity are also eliminated [8].

- **Weighted Method**

Different recommendation components scores are mutually statistically. This class aggregates scores individually factor by ad additive formula.

- **Switching**

From available recommendation components system chooses particular element and applies the picked out one. Switch one technique to another based on situations.

- **Mixed**

Different recommender provides their recommendation that will be received together. This share is based on merging and presentation of multiple rated lists into base hit rated list. First of all content based method is used for textual description and use of collaborative method for finding the preferences of the user and Recommendations from the two techniques lead to suggest a final program. With the help of this mixed method new item -start up problem can be overcome.

- **Feature combination**

In Feature combination the features from different recommendation data sources are used together into a single recommendation algorithm. The working of actual recommender is depends on the data modified by the contributing one. The contributing one throws features of one source on to the other components source.

- **Feature Augmentation**

This class is similar to the feature combination hybrids but only difference is that the contributor gives novel characteristic. It is more elastic than feature combination method.

- **Cascade**

This class plays a role of tie breaker. Here for all recommender associate some priority and through that assign priority, lower priority recommenders play an tie breakers role over higher priority.

- **Meta-level**

Here two recommendation techniques can be merged by using the model generated by one as the input for another.

## 5. CONCLUSION:

This kind of paper presents an summary of Recommender systems, techniques or methods of Really simply syndication based on the knowledge prepared, and their advantages and disadvantages. We certainly have also talked about the Evaluation measures; these can be used to measure the overall performance and quality of the system. Currently, the hybrid algorithms are being used. To increase the quality of recommender systems anticipations future research will concentrate on progressing the current methods and algorithms, using hybridization methods like measured method, which is employed to overcome the certain limits Moreover, paper also details about the various methods about the hybridization and it will help to know whether which approach to be applied for hybridization

## REFERENCES:

1. Lops, Pasquale, Marco De Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State of the art and trends." Recommender systems handbook. Springer US, 2011. 73-105.
2. X. Luo, Y. Xia, Q. Zhu, “Applying the learning rate adaptation to the matrix factorization based collaborative filtering”, Knowledge Based Systems 37 (2013) 154–164.

3. J. Bobadilla, A. Hernando, F. Ortega, J. Bernal, "A framework for collaborative filtering recommender systems", Expert Systems with Applications 38 (12) (2011) 14609–14623.
4. J. B. Schafer, D. Frankowski, J. Herlocker, S. Sen,"Collaborative filtering recommender systems", in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, 2007, pp. 291–324 .
5. Sánchez Sánchez, José Luis. "Improving Collaborative Filtering Based Recommender Systems Using Pareto Dominance". Diss. E_Informatica, 2013.
6. Paritosh Nagarnaik, Prof. A.Thomas, "Survey on Recommendation System Methods", International Conference On Eletronics And Communication System (ICECS), 2015 IEEE International, 26-27 Feb. 2015.
7. Poonam B. Thorat, R. M. Goudar, Sunita Barve, "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System ", International Journal of Computer Applications,Volume 110 – No. 4, January 2015
8. G. Suganeshwari and S.P. Syed Ibrahim, "A Survey on Collaborative Filtering Based Recommendation System", Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges. 2016.
9. J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, "Recommender systems survey", Knowledge-Based Systems. 109–132, July 2013.
10. Mrs.M.Sridevi, Dr.R.Rajeshwara Rao, Dr.M.Varaprasad Rao, "A Survey on Recommender System", (IJCSIS) International Journal of Computer Science and Information Security, *5, May 2016* .
11. Haibo Liu, Hongjie Xing, Fang Zhang. 2012. "Web Personalized Recommendation Algorithm Incorporated with User Interest Change". Journal of Computational Information Systems 8(4), 1383-1390
12. C.C. Aggarwal, Recommender Systems the textbook
13. M. Pazzani, "A framework for collaborative, content-based, and demographic filtering", Artificial Intelligence Review-Special Issue on Data Mining on the Internet 13 (5-6) (1999) 393–408.