

# A Review on: Designing a Recommender System Using Sequential Logs

Anupama Patel

Department of computer science  
Gyan Ganga Institute of Technology & Sciences  
Jabalpur, India

Dr. Santosh K. Vishwakarma

Department of computer science  
Gyan Ganga Institute of Technology & Science  
Jabalpur, India

**Abstract**— Recommender system has change the way of searching for information over the internet. The information over the internet is overburden. The searching for specific information sometimes becomes very time consuming. Today in the fast growing world no one wants to wait everyone wants o work fast. This in result has increased the competition for achieving success. So it is very necessary to save time in searching information. Recommender system is solution for the problem by providing recommendation to the users. This paper has discussed about the types of recommender system like collaborative filtering, content based filtering and hybrid filtering along with the challenges of recommendation system. The paper has also discussed about the sequential information of user behavior which is stored in the web servers. In this paper only the idea is being discussed.

**Keywords**— Recommender system, collaborative filtering, content based filtering, hybrid filtering, sequential information, web servers.

## I. INTRODUCTION

The internet has become a huge storage of web pages and links. Internet contains information of almost everything happening in world. The information over the internet gets modified and new information are also added every day. It is not necessary that all information that is available over the internet is useful for an individual. The user tries to search information according to its need are requirement. The massive growth in the digital information and the increase in the number of visitors over the Internet have created potential challenge of information overload which hinders timely access to items of interest on the Internet. Information retrieval systems, such as Google, Yahoo have partially solved the problem but it has also increased the demand of recommender system more than ever before.

Recommender system is a type of machine learning which has the ability to learn by itself without have any deals about the task. The recommender system predicts whether a user would prefer an item or not depending upon the user profile. Recommendation system studies the pattern of user searching over the internet and then tries to generate prediction for the new users. Recommender systems are very helpful in decision making. Recommender system are created by using data mining tools which make use of various data mining techniques such as clustering, classification, association, neural network, sequence discovery and much more. Recommendation System is used in almost every field. The most important application where recommender system is used is the e-commerce. The shopping sites gives

recommendation to the users so that they can easily sells their products and can gain more profit from it. Recommender system is categorized into three parts:

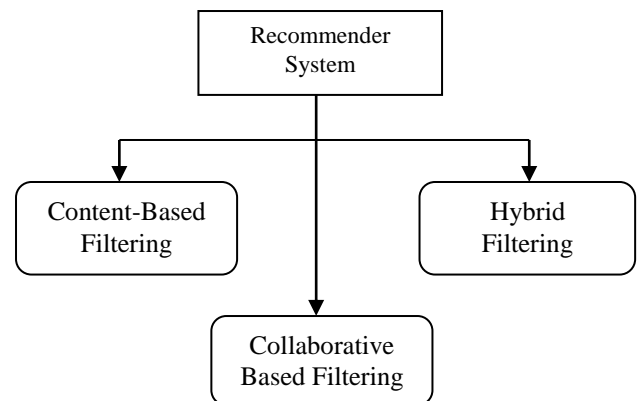


Fig: 1 Types of Recommender System

### A. Content-Based Filtering

As the name suggests the content based filtering is done on the basis of content requirement. The user provides the content or it click to the link he/she wants to see the data regarding this whole process is stored and a user profile is created. The profile contains all the information in respect to the user action. The prediction is generated based on rating given in the past. So the items are recommended to the new user based on the content matching with the previously rated content. The prediction becomes more and more accurate as the input data is increased by the user. For example if the user has affinity towards the pages having the characteristic of car or engine, then the recommendation given will be page having information about automobiles. Some time problem arises in content based filtering which is due to the content mismatch this may result in decrement of performance. Content based filtering is used by recommendation system along with information retrieval system like the search engines.

CONTENT-BASED FILTERING

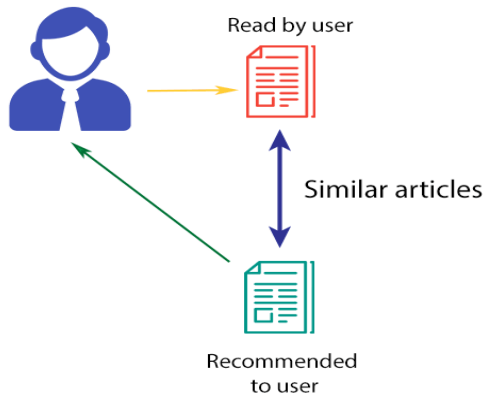


Fig. 1 Example of content-based filtering

**B. Collaborative-Based Filtering**

Collaborative filtering provides result on the basis of prediction. Collaborative filtering tries to predict the options on the behalf of user. Collaborative filtering is a type of machine learning, which has the ability to learn automatically without having any guidance about the process. According to collaborative filtering the users having similar preference will also like the same item which the other user liked in the past. Suppose if A is drinking a diet coke and if B has the same taste as A then, it is possible that B will also like the same coke which A was drinking, so the recommendation will be given to the user B. This can also be termed as learning from the experience. The closely related items are being predicted and are given as recommendation to the users.

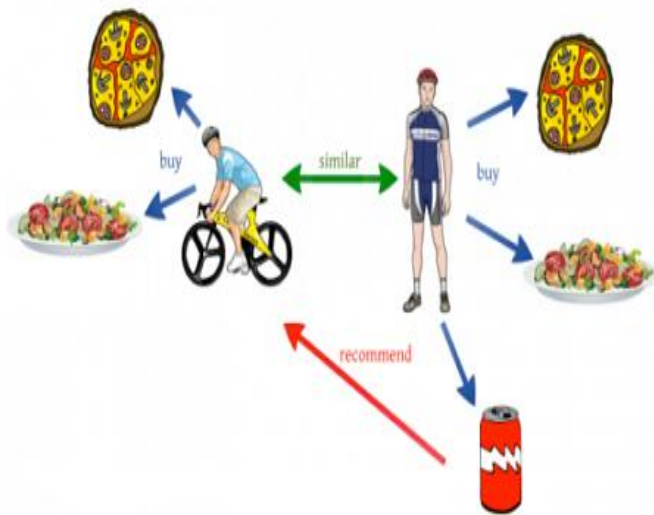


Figure 1 Example of Collaborative filtering based recommendation

Collaborative filtering has two types of approaches:

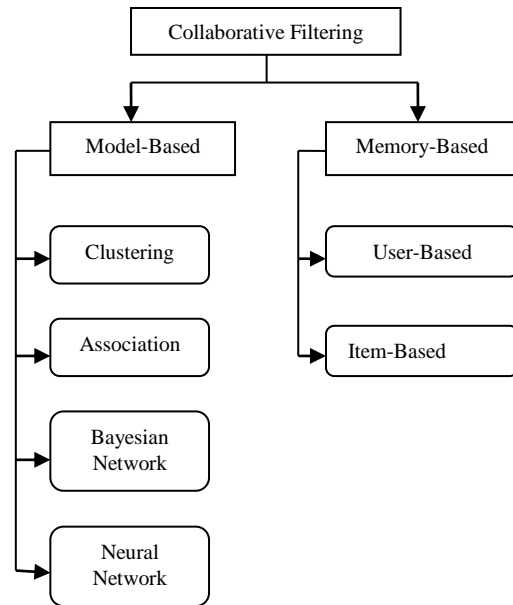


Figure 2 Types of collaborative filtering

**1) Model-Based**

In model based collaborative filtering the structures are created. There are many data mining algorithms which help in creating the structure. Every algorithm have their own concept so, different structures are created by using different algorithms. The main use of creating models is to generate prediction. The prediction is made based on the habits over the real data. Techniques used by model based filtering are

- **Clustering:** Clustering is a technique of arranging similar types of data in one cluster. It is also used for detecting the outliers. The clusters are arranged in such a manner that the inter cluster similarity must be low and the intra cluster similarity should be high.
- **Association:** Association shows the probability of occurrence of one item with respect to the other item. It shows the relationship between the user and the items. Association is a type of machine learning which learns from the users searching pattern and thus generate prediction. Example if user A has selected milk then the probability of selecting bread will increase with respect to milk. The main working is based on support and confidence.

**2) Memory-Based**

Memory based collaborative filtering is different from the model based filtering. In memory based filtering the prediction is done in a different manner. The rating given to the items by the user are used to generate prediction for the new users. The prediction is generated for every user whether the user is been registered or unregistered with the website. Memory based filtering have two techniques:

- **User-Based:** The rating given by the users in the past are stored in the database and are fetched when it is needed. The user based filtering technique is based on the ratings which are given in past by the uses. The working behind this is that if taste of user A and B matches then the item

when user A has rated will also be recommended to user B but these items are not rated by user B.

- **Item-Based:** This technique was developed by Amazon. This technique is used to draw interpretation among the different items. The pattern for the items which are commonly purchased is stored in the database, when every one of the items among the stored pattern items is selected the remaining other items are automatically recommended

### C. Hybrid-Based Filtering

Hybrid based filtering technique is the combine form of different techniques. This includes collaborative-based filtering, content-based filtering and others. The results integrated and then the predictions are done improving the performance of the model.

### D. Challenges of Recommendation System:

a) *Sparsity:* This is considered a serious problem in the case of collaborative filtering. The problem occurs when there is less number of ratings available over the items. The users provide ratings to the item based on their experiences over the items. If there is only one ratings then to generate prediction for recommendation is nearly impossible. The problem arises due to three main reasons firstly if there is new item available then the rating for the new item will only be given when it is being tested by the users. Secondly if there are new users then also the rating is effected because of the lack of trust. This can be sloved by inceasing the domains of information or by making assumptions on behalf of several other data having less ratings.

b) *Cold-start Problem:* This is bit similar to the sparsity problem but the difference arises when there is no rating then there is recommendation. The recommendation is not being generated until there are sufficient amount of ratings given by sufficient number of users. It mainly focuses on the counting of ratings given to an individual item. This problem has firstly arise on collaborative filtering.

c) *Scalability:* The problem of scalability takes place when there is massive increase in items and in users. The increase in user will also increase in demand of items; this is a big problem especially when prediction is done on the basis of real time. To solve this nearest neighbor algorithm is used. The item which is closely related to the item request is given as recommendation to the user. it also help in increasing the performance.

d) *Synonyms:* Synonyms mainly refers to the problem where the same item is stored with different name. This create confusion during the selection and prediction proces by the recommendation system. For examole there is a film which is based on childrens but is stored by two different name at two different place such as "kid film" and "children film". The matter is same but the name is different. This kind of problem lead to different other problems like the Gray sheep problem and Black sheep problem. To cope up with these problems Support vector machine and latent semantic indexing is being used.

## II. LITERATURE REVIEW

Literature review is the base for the researchers including any field. The literature review contains the summarized form of any type of reach. So far many research works have been done in the field of web recommender system. Some of the studies related to web recommender system are discussed below

Rajhans Mishra et al. [1] have proposed a model to predict a user's next page visit. The next page visit of a user will provide the useful information about their likes and interest. For this, a rough set based similarity upper approximation concept is used during clustering to generate soft clusters. The  $S^3M$  similarity measures have been utilized which is a hybrid of content and sequential similarity measures. The generated soft clusters have been utilized to create a response matrix which is used by SVD to generate predictions. The result of the model is then compared with the random prediction and first order Markov based models. The datasets used are MSNBC web navigation dataset, simulated dataset and CTI data set.

Pradeep Kumar et al. [3] have proposed a clustering algorithm which is based on the similarity upper approximation of a rough set cluster using  $S^3M$  similarity measures.

Rajhans Mishra et al. [4] have implemented the algorithm proposed by Pradeep Kumar having similarity measures for finding the clusters and outliers. Four similarity measures have been used that are Jaccard, Dice,  $S^3M$  and Levenshtein. The algorithm uses a rough set upper approximation for finding the clusters of the users and outliers.  $S^3M$  measures show the better result among all four similarity measures because it measures the similarity based on content as well as sequence.

Prajyoti Lopes et al. [5] this paper focuses on providing real-time recommender to all the visitors of the website irrespective of been registered and unregistered. Action based rational technique is used which construct lexical patterns in order to generate item recommender. This proposed system dynamically provides a recommender as per changing user behavior and traversal patterns. The system also minimizes the false positive error that occurs frequently in traditional recommender system.

R. Suguna et al. [6] has proposed an efficient Web Recommender System by using Collaborative filtering and Pattern Discovery algorithm. In this paper an improved Apriori algorithm is used to find the frequently visited web pages by the user. The improved Apriori algorithm is modified by adding the time duration spent on each web page. Along with this Markov model is also used to recommend pages to the new user on the bases of the previous browsing history of the user.

The next paper entitled Applying Web Usage Mining Techniques to Design Effective Web Recommender System. A case study [7] has discussed about the concepts and techniques of recommender system. In addition to this the paper has also discussed about how to apply web usage mining over the web logs for discovering access patterns. Lastly an analysis is done over the problem causing during the deployment of recommender system and has also proposed solutions which address the problems.

### III. METHODOLOGY

The aim is to study the working of the recommendation system and to generate a real time recommendation for the new web user. The prediction should be generated on the basis of the previous rating given by the users over the items. The idea for creating a recommendation system model is by using the k-NN operator; this will help in predicting the user rating for whose rating is not given. Then on the basis on this predicted rating the recommendation will generated according to the highest ranking given to the predicted rating. The ranking will show the highest number rating given to an individual item. The performance of the model will also be measured. Description for the operator is given below:

K-NN: K-Nearest Neighbor is also considered as the lazy learning. It is a type of machine learning method which does not require learning process. K- Nearest Neighbor is used for two types of task firstly for classification and for regression. In classification is used for arranging the input data in there relevant group. And regression is used for generating the bonding between the input and the remaining data.

### iv CONCLUSION AND FUTURE WORK

This paper has discussed about the recommendation system and its techniques. In future a recommendation system can be proposed by using the user k-NN method which will help in generating predicting for the new users. The web log files can

be used as database or the system. This will give us the real time prediction.

### REFERENCES

- [1] Rajhans Mishra, Pradeep Kumar and Bharat Bhasker, "A Web Recommender System Considering sequential information", Decision Support Systems 75(2015) 1-10.
- [2] P. Kumar, P.R. Krishna, R. S. Bapi and S.K. De, "Clustering using Similarity Upper Approximation", IEEE International Conference on Fuzzy Systems, Vancouver, 2006, pp. 893-844.
- [3] Rajhans Mishra, Pradeep Kumar, "Clustering Web Logs Using Similarity Upper Approximation with Different Similarity Measures", International Journal of Machine Learning and Computing, Vol. 2, No. 3, June 2012.
- [4] Prajyoti Lopes, Bidisha Roy, "Dynamic Recommender System using Web Usage Mining for E-commerce Users", International Conference on Advanced Computing Technologies and applications, Procedia Computer Science 45(2015) 60-69.
- [5] R. Suguna, D. Sharmila, "An Efficient Web Recommender System using collaborative Filtering and Pattern Discovery Algorithms", International Journal of Computer Applications (0975-8887), Vol. 70-No. 3, May 2013.
- [6] Maryam Jafari, Farzad soleymani Sabzchi and Amir Jalili Irani, "Applying web usage mining Techniques to design effective web Recommender systems: A case study". ACSIJ Advances in Computer Science: an International Journal, Vol. 3, Issue 2, No. 8, March 2014.
- [7] B. Sarwar, G. Karypis, J. Kostan, and J. Riedl. Analysis of recommendation algorithms for e-commerce. In EC, pages 158-167, 2000.
- [8] [www.google.com](http://www.google.com)