

# Recommendation system and its approaches- A survey

Naresh E, Geetha LM, Vijaya Kumar BP

**Abstract**— Data size has been increasing day by day in the E commerce business. This rapid development of technology leads to overload of information. In order to get the required solution from this massive amount of data search engines are used. However this search engine doesn't provide personalized information to the user. Further, recommendation systems are introduced in order to provide the personalized information to the user. Recommendation system provides suggestions based on the user's interest. Many approaches are available for this purpose which can be used to create the recommendation list. E commerce websites uses these various approaches with different combinations in order to increase their business by attracting the users. This paper gives the overview of such approaches along with their strengths and weaknesses.

**Index Terms**— Content-based approach, Collaborative approach, Data Mining, E-commerce business, Hybrid approach, Neural network, Recommendation list, Recommendation systems, Search engines.

## 1 INTRODUCTION

Recommendation system was mostly used in the online shopping to recommend the list of product. Nowadays it is also gaining importance in different domains like news, portal and service providing web pages and so on. Recommendation system helps the user to choose the right information based on their preferences. This recommendation system builds a model using set of data and then this model is used to predict the result for recommendation [13].

As the technology advanced data generation is also increasing rapidly. Earlier, search engines were the solution to find the required data. But in the E-Commerce business, this massive generation of data leading to find the approaches that helps the customer to choose the items which are of their interest. Recommendation system is used for this purpose. Recommender system takes the input from input and gives the results or suggests services/product which is relevant to the customer by using individual or particular group details [11][15].

As the web technology developed, recommender system gained its importance in the E-commerce business. It has also become a research area on developing new approaches for the recommendation system. In order to develop the new approaches and to make the system more efficient, we need to understand the existing approaches and its limitations.

As per various survey conducted, many approaches are available for recommendation engine. These are mainly classified into following categories:

- i. **Content-based approach:** Recommendation is given based on the user's preferred items in the past;
- ii. **Collaborative approach:** Items are recommended based on the interest of similar previous people;
- iii. **Hybrid approach:** This is the combined approach of Content-based and Collaborative recommendations.

Other than these approaches many other recommendation class techniques are available. They are feature based, behavior based, citation based, context based, knowledge based, rule based and many other recommendation classes [2].

It is necessary to understand these concepts in order to extend the capabilities of recommendation engine [1]. Studies have revealed that 55% of the recommendation system uses content based filtering. 18% of recommendation system uses collaborative filtering and graph based recommendation approach is used by 16% of the system.

Though many approaches are available for recommendation, many studies have revealed that it is hard to say which concept or approach is best. Sometime Content-based approach works better compared to collaborative recommendation approach and sometime it doesn't. Hence it is difficult to choose the promising approaches in the available techniques. Sometimes it requires the combinations of available approaches. Hence, basic understandings

• Naresh E is working as Assistant professor, Dept of ISE, MSRIT and research scholar in CSE dept at Jain University Bangalore, India. E-mail: nareshkumar.e@gmail.com

• Geetha LM is currently pursuing masters degree program in software engineering in MSRIT, Bangalore, India., E-mail: geetha.lm90@gmail.com

• Vijaya Kumar BP is currently professor and head, Dept of ISE, MSRIT, Bangalore. And Senior member in IEEE.

of these concepts are necessary. This ambiguity, related to the performance of approaches, provides many opportunities for researchers [2].

As mentioned earlier it is very hard to find the promising approach in the set of available recommendation classes. However, this paper provides an overview of the approaches which have been using in the field.

## 2 CONTENT BASED APPROACH

Content based approach is the one of the most widely used recommendation technique in the business field. In this approach user interests are analyzed and based on the interaction the user recommendations are given to the customer. This process of inferring the item that users are interacted with is known as user modeling process. Most of traditional approaches use collaborative filtering and content based techniques. But sometimes context based approach works more efficiently than collaborative techniques. For instance, recommending a vacation package doesn't only requires user and items. It also requires some context to recommend the list. Contextual information can be obtained explicitly, implicitly and inferring the context using data mining techniques. Contextual information can be obtained by directly referring the relevant people. It is known as explicitly obtaining the data. By referring the location of the customer using telephone or mobile company or getting the details by referring social networks is known as implicitly obtaining data. Using the data mining techniques to get the details related to the user interacted item is known as inferring the context.

Context based approaches have gained importance in many field, however it is limited due to lack of techniques for extracting the information. And also it is hard to find that when text or context is useful. Thus many text mining technologies come into picture. Examples of context are location, identity, state, people, time, companies, activities of current user etc. For instance consider a dining event, for this event for contextual information can be considered. They are companion, time, location and occasion. Using this contextual information recommendations are given to the user. The information related to the user's review of products/review is collected using user profile from different website. From this contextual information is extracted and evaluated and finally used for recommendation.

As per studies it is revealed that context information is an important factor if contextual information is properly used. It helps to recommend the product/service efficiently. As per the studies, text mining techniques are not optimal. Contextual information doesn't help to predict the user ratings [5].

As the technology advanced, unstructured and semi structured data is increasing rapidly. These data doesn't only include textual information but also contains multimedia data such as audio, image and video etc. Filtering is a process of removal of data in the large amount of incoming data rather searching for it. This technique not only helps to predict the customer's interest but also gives in-

formation related to the details which customer don't like. With this large amount of input data, filtering is necessary for recommendation in E-commerce business. It can also be defined as the accessing information which is present in the remote system.

Initially information filtering is done on textual information. The process of filtering differs for data format such as textual and multimedia. Filtering is done based on the individual information or group information, they are called profiles. By referring these profiles, interest of the user can be predicted. Here customers see what is left after the filtrations (which are based on user's interest). As mentioned earlier, filtration can also be defined as the process of removal of data from the massive incoming. In this case customer sees the items which are extracted.

Here the concept of information retrieval and filtering is explained which is helpful to use in the recommendation system [8].

### 2.1 Framework

As stated earlier, in this approach users are recommended based on their past interest. Let  $s$  be the item and  $c$  be the customer then the utility  $u$  is calculated i.e.  $u(c,s)$ . Here utility is calculated based on the utilities which are assigned by user  $c$  to items  $si$ ,  $si$  is similar to item  $s$ . Consider an example of music album recommendation application, in order to recommend the albums the recommender system analyses user's previously rated albums. Similarly for every user  $c_i$  which belongs to  $C$ , items  $s_i$  which is belongs to  $S$  is calculated. Below figure shows the flowchart for content-based approach.

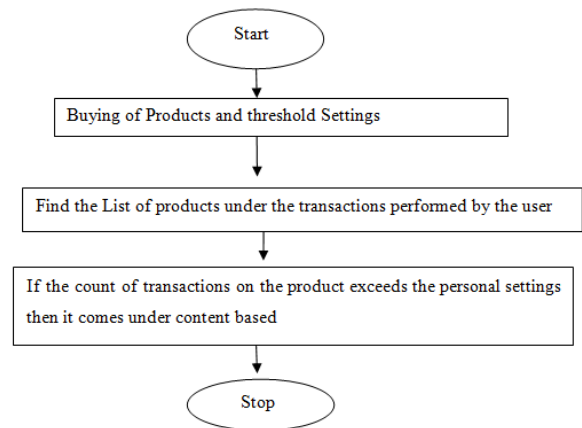


Fig. 1: Flowchart for Content based recommendation approach.

### 2.2 Advantages

- For every individual customer, content based approach uses user-based personalization technique and provides the best recommendation for each user.
- Classification overhead is very less as this approach uses user modeling concept.

### 2.3 Disadvantages

- a. Content-based filtering approach requires more computation power
- b. This approach doesn't work efficiently for new user, as it refers previously preferred items for recommendation
- c. This approach includes following steps like analysis of the features, building user models and calculation of similarity for each item that user interacted. Hence it takes much time to process.
- d. It also doesn't consider the quality aspect while recommending.

All tables and figures will be processed as images. You need to embed the images in the paper itself. Please don't send the images as separate files.

Many applications require not only items but also context related to the item. Thus we need to understand the importance of Context based approach and its application in various fields [2].

## 3 COLLABORATIVE APPROACH

Collaborative filtering approach works based on the preferences and ratings given by the like-minded people. It implies that the current user is recommended based on the ratings given by the previous like-minded people [1]. "Collaborative filtering" word was coined by *Goldberg et al.*, who proposed a theory that "information filtering technique can be more effective if the human involvement is high in the filtering process". Here the recommendations are given based on the user similarity but not based on the item similarity. When like-minded users are identified, items which are rated by one user are recommended to the other similar users [2].

Prediction of items based on the other user's rating is the common approach used for recommendation in E-commerce business. However these approaches don't use the capabilities of machine learning algorithm. Collaborative filtering techniques are gaining importance in online shopping. Like minded preferences towards items are analyzed and then neural network technology is used to recommend the list for the user.

It works in two steps. First, Similarities between users are computed. Once these similarities are recorded, it can be used to prepare the recommendation list. Collaborative filtering can be used along with the machine learning concept in order to improve its prediction accuracy [6]. The most popular collaborative approach is Nearest Neighborhood (NNH) algorithm. There are two types of NNH approaches, they are

- i. **User-based nearest neighborhood:** Here Users profiles, who share the same ratings, are first identified. Then likeminded user's ratings are used for

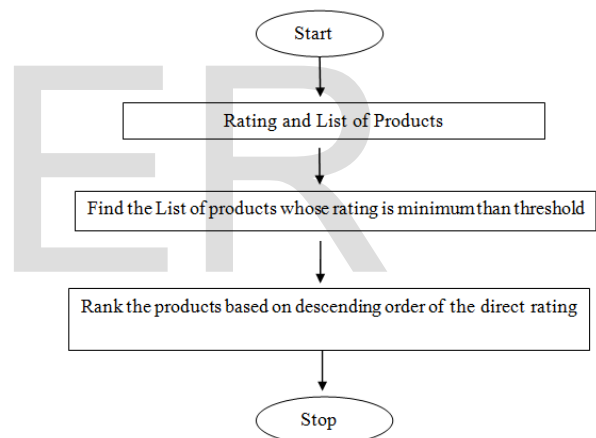
the prediction. In order to generate the more accurate results, neighbors are assigned the weight. Based on this prediction is done.

- ii. **Item-based nearest neighbor algorithm:** Here the prediction is done based on the similarities of the items. This method is the transpose of the User-based nearest neighborhood algorithm [10].

By referring this we can understand the difference between traditional collaborative approaches. It also helps to improve the performance of collaborative filtering approach [6].

### 3.1 Framework

Unlike Content-based approach, in this approach items are predicted which have been rated by the users who are similar to the current user. Let  $s$  be the item and  $c$  be the customer then the utility  $u$  is calculated i.e.  $u(c,s)$ . Here utility is calculated based on the utilities which are rated by the user  $ci$  to items  $s$ ,  $ci$  is similar user. The flowchart is as shown in below.



### 3.2 Advantages

- a. Collaborative filtering is content independent hence it doesn't requires error prone item processing.
- b. Since the Collaborative filtering includes human involvement, it considers the quality aspects.

### 3.3 Disadvantages

- a. Collaborative filtering approach requires user participation to predict the recommendation. However, the motivation for this participation of user is not available. This is known as cold-start problem.

- b. If the item is new and it is not rated by any customer then it can't be recommended to any user.
- c. Further, if the new user doesn't rate any item then it will be not possible to find the like-minded people for further recommendation [2].

## 4 HYBRID APPROACH

Many recommendation approaches such as content based technique, collaborative approach and many other approaches are available in survey. Content based approach works on the concept that item with similar feature gets the same ratings. But this is not always true and difficult for the prediction.

Each recommendation systems have both strength as well as weakness. Hence these approaches solely can't achieve the better performance. Thus hybrid approach is developed to enhance the performance of recommendation system. In this approach collaborative approach is usually combined with other approaches. Few hybridized methods are mentioned and defined as below.

- i. **Weighted method:** Hybrid recommendation approach is developed based on the scores of other recommendation techniques.
- ii. **Switching method:** Here multiple techniques are used. Based on the requirements system switches between the recommendation techniques.
- iii. **Mixed method:** Different recommend techniques are used to recommend the item at the same time.
- iv. **Feature combination method:** Here the features of the multiple recommendation approaches are used together based on the requirements.
- v. **Cascade method:** Multiple recommendation systems are used and here one approach refines the results which are given by the other approaches.
- vi. **Feature augmentation method:** Here output of one approach is used as input to the other technique.

### 4.1 Framework

Content based and collaborative techniques are combined in different ways which are discussed below.

- i. Content based and collaborative techniques are implemented separately then these techniques are

combined to produce the hybrid approach. Here output of the each technique is combined or the better result in these two techniques can be chosen.

- ii. Content based characteristics are added to traditional collaborative approach. This combination avoids the weakness of collaborative techniques.

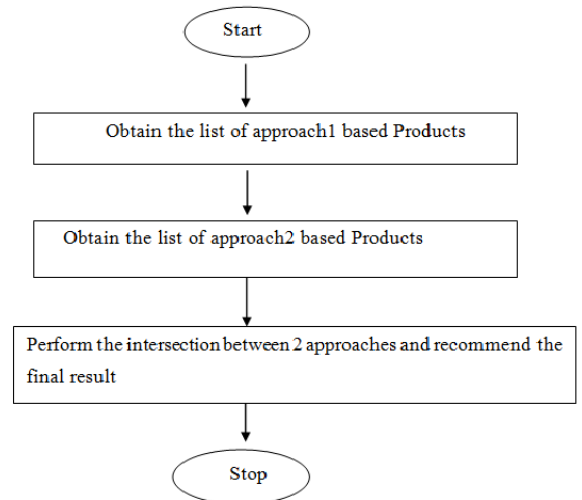


Figure 3: Flowchart for hybrid recommendation

- iii. Collaborative characteristics are added to the content based approach. This technique improves the performance of pure content based approach.
- iv. Unified model is developed using probabilistic method in order to combine Content based and collaborative [1].

This combination of approach is done based on the requirements and to achieve the peak performance [9]. The figure shows the flowchart for hybrid recommendation approach.

### 4.2 Advantages

- a. Hybrid approach increases the benefits and reduces the drawbacks of approaches which are used in the combination [11].

### 4.3 Disadvantages

- a. Different approaches can be combined for hybrid approach but the most promising approach is not found yet.

## 5 DATA MINING (DM) TECHNIQUES FOR

## RECOMMENDATION SYSTEM

Many traditional approaches have been used to improve the efficiency of recommendation system. Data mining technique is used to form a data model from the large set of data, further this model can be used by the recommendation system [16].

Data mining technique consists of three steps they are preprocessing of data, analysis of data and interception of result. In the first step distance measure, sampling and dimensionality reduction techniques can be used. In the Analysis step classification, association rule mining and clustering technique can be used

Large set of real time data should be preprocessed in order to use it for further analysis steps. In data preprocessing step data is defined as object along with its attributes. Here attributes are the characteristics of object. These defined objects are used in the next step.

In the second step, numbers of objects are reduced from large amount of data set by preserving its main features. In this step training and testing data sets are identified which can be used in the analysis step [17].

### 5.1 Neural network for an intelligent Recommendation system

An intelligent recommendation system can be developed using a hybrid approach, this is the combination of neural network and data mining technologies. One of the approaches in neural network technology is back propagation learning algorithm. In this approach the networks are trained by propagating set of data into the network [12]. Its output is compared with the input and then error is calculated. Then this error is adjusted for better result.

### 5.2 Framework

Basically neural network is developed using the concept of brain. The brain's smallest unit is neurons, it reacts when it is stimulated. Similarly artificial neural networks are developed, here the smallest unit is known as perception (similar to neurons). These perceptions are connected with each other using synapses. Each connection has weight, based on this weight action is triggered. This artificial network consists of three layers. First layer is known as input layer which consists of nodes each node in turn consist of values. The second layer is known as middle layer which also consist of nodes. The last layer is known output layer from where the output is produced.

A set of training data is required to train the neural networks. Networks are trained using different data patterns which are present in the training data set. Error back-propagation methodology is used to train the artificial neural networks. In this approach the training data propagates to the layers which are present in the neural networks. The nodes present in the neural networks produces the output for this training data set. Difference between actual output values and current output values is calculated, this is

known as error. This error is corrected by adjusting the weights of the network connections. Here training the networks means that finding the correct weights for the connections in the network. Thus the intelligence of the system is based on the weights in the network [13]. This neural network concept can be used as an alternative for context based approach and it is combined with collaborative approach for personalization and to recommend the items [14].

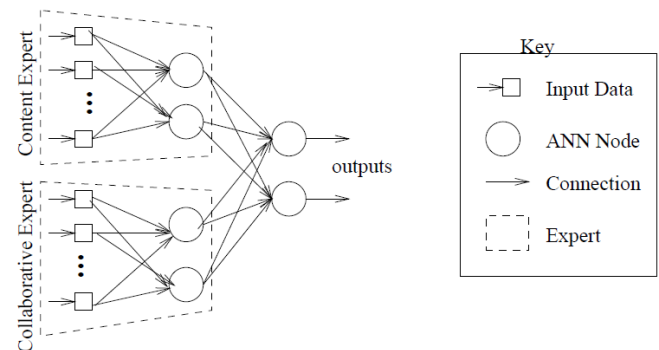


Figure 4: Architecture of Artificial Neural Networks

### 5.3 Advantages

- Artificial neural network can work efficiently even if the part of the network fails due to its parallel working system.

### 5.4 Disdvantages

- Time required to train the artificial neural network using training data set is high.
- Time required to train the artificial neural network using training data set is high.

## 6 RECOMMENDATION APPROACH COMPARISION

Research has been going from past several years on recommendation approaches that use various technologies like machine learning, information retrieval, and other advanced technologies. The techniques used in collaborative, content-based and hybrid methods are shown in the table below along with its definition. The table also shows the percentage of approach used in the recommendation system as per the literature [1].

A survey is conducted on the restaurants recommendation system with 44 users along with three different approaches. Comparison is done between precisions with different approaches like content based, collaborative based and hybrid based. Below graph depicts the comparison.

The graph shows that the precision is high when the approaches are combined i.e hybrid approaches.

- Content-based recommendation with Bayesian classifier
- Collaborative is standard using Pearson correlation
- Collaboration via content uses the content-based user profiles

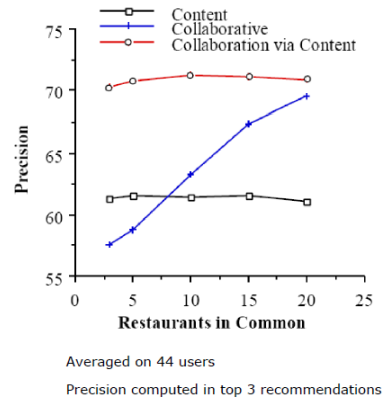


Figure 5: Comparison of three approaches

Here Content based and the collaborative based approach is combined. This combined approach gives better result compared to individual approach.

Recommendation approach	Definition	Techniques used	% of approach used in recommendation system
1.Content based	Recommendation is given based on the user's preferred items in the past	1. Information retrieval 2. Clustering 3. Artificial networks 4. Bayesian classifiers	55%
2. Collaborative	Recommendation is given based on the interest of similar previous people	1. Clustering 2. Artificial networks 3. Bayesian classifiers 4. Linear aggression 5. Probabilistic model	18%
3.Hybrid	This is the combined approach of Content-based and Collaborative recommendations	1. Incorporating one component as a part of the model 2. Building unifying model	16%

## 7 PERFORMANCE EVALUATION

In the early days accuracy factor is used as a factor in order to evaluate the value of recommendation techniques. However as more and more approaches are introduced, the research has revealed that the accuracy factor is not the only metric to evaluate the values of recommendation approaches. Recent studies have suggested metrics such as coverage, diversity, scalability, user preferences reliability and similarity also plays important role in the evaluation of recommendation system. These different parameters are mentioned and discussed below.

- Accuracy:** This metric is used to calculate the simi-

ilarity between users, which can be used in the collaborative filtering process. This metric can be formulated by the linear combination of values and weights. Values are calculated based on the similarity between the users. Weights are based on the data of recommendation system. Collaborative filtering technique uses similarity between users similarly trust is introduced to Collaborative filtering. This factor improves the Prediction of quality, recommendation of quality and performance.

- Coverage:** Item proportion that the recommendation system can recommend is known as coverage. Collaborative filtering process calculates similarity between users. This obtained similarity provides better quality and faster results and in the coverage.
- Quality:** Hybrid technique uses user's previous history and interest of similar user to provide the recommendation. Combining the information filtering and collaborative filtering gives the qualitative result to the user. Further Pre-filtering concept improves the quality measures.
- User Preferences:** Most of the traditional approaches like collaborative filtering provide more importance to accuracy than user preference. User profiles needs to be studied in order to find the user preferences.

Metric	Recommendation approach used to enhance the metric	Related criteria
Accuracy	Collaborative filtering process	Prediction of quality, recommendation of quality and performance.
Coverage	collaborative filtering technique	Provides better quality and faster results
Quality	Hybrid approach technique	Increases online performances, quicker result.
Scalability	Clustering technique	Improves online performance
Reliability	Collaborative filtering	Accuracy and reliability

- Scalability:** Since recommendation system work on large number of data item, it should be designed in such a way that it can scale up to large data set. Scalability is measured in terms of growing data set and resource consumption. Clustering techniques are used to improve this factor.

- vi. **Reliability:** Collaborative approach provides the recommendation by analyzing user's interest. Experience and trust models are included to this Collaborative approach which in turn increases accuracy and reliability [18].

## 8 CONCLUSION

Nowadays as the technology developed data size also increasing rapidly. With this rapid development we require different recommendation approaches to find people interests and to recommend them. In this paper various recommendation techniques are discussed along with their strength and weakness. And also how other technologies can be combined to traditional approaches in order to improve the performance is also discussed which in turn helps the user to find items of their interest with less effort and time. Finally different techniques that can be used in each approach is showed in the table.

## ACKNOWLEDGMENT

We indebted to the management of MSRIT Bangalore, for excellent support in completing this work at right time, and also for providing the academic and research atmosphere at the institute. A special thanks to the authors mentioned in the references.

## REFERENCES

- [1] Gediminas Adomaviciu and Alexander Tuzhilin. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", 2005.
- [2] Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitingner. "Research-Paper Recommender Systems: A Literature Survey", 2015.
- [3] Greg Linden, Brent Smith, and Jeremy York. "Amazon.com recommendations, item-item collaborative filtering", 2003.
- [4] Gediminas Adomavicius, Alexander Tuzhilin, "Context-Aware Recommender Systems".
- [5] Yize Li, Jiazhong Nie and Yi Zhang. "Contextual Recommendation based on Text Mining", 2010.
- [6] Daniel Billsus and Michael J. Pazzani. "Learning Collaborative Information Filters".
- [7] S.Ganesh Kumar, P.Hari Krishna. "Ranking Based Approach To Maximize Utility Of Recommender Systems ", 2013.
- [8] Nicholas J. Belkin and W. Bruce Croft. "Information filtering and information retrieval: Two sides of the same coin?", 1995.
- [9] Robin Burke. "Hybrid Recommender Systems: Survey and Experiments."
- [10] Stephan Spiegel, "A Hybrid Approach to Recommender Systems based on Matrix Factorization."
- [11] Amir Hossein Nabizadeh Rafsanjani, Naomie Salim, Atae Rezaei Aghdam, Karamollah Bagheri Fard. "Recommendation Systems: a review", 2013.
- [12] KanokwanKongsakun, Chun Che Fung. "Neural Network Modelin'g for an Intelligent Recommendation System Supporting SRM for Universities in Thailand".
- [13] Wendy Kogel. "Faster Training of Neural Networks for Recommender Systems", 2002
- [14] Charalampos vassiliou, Dimitris Stamoulis, Drakoulis Martakos. "A Recommender System Framework combining Neural Networks &

Collaborative Filtering", 2006.

- [15] J. Ben Schafer, Joseph Konstan, John Riedl. "Recommender Systems in E-Commerce"
- [16] J. Ben Schafer, "The Application of Data-Mining to Recommender Systems", University of Northern Iowa.
- [17] Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol, "Data Mining Methods for Recommender Systems".
- [18] Sonali Karmalkar, Dr. Shanta Sondur, "A Survey on Evaluation and Improvement Techniques for Recommendation Systems", 2015.