Comparative Study on Approaches of Recommendation System

Roshni K. Sorde
Dept of CS & IT
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra

Sachin N. Deshmukh
Dept of CS & IT
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra

ABSTRACT
Recommender systems are the software engines and approaches for providing suggestion of products to the user which might be most probably matched with the user's like/choice. Generally, Recommendation system gives suggestions on content like what items to buy, which music to listen or what online news to read. Depending on the users liking the items are suggested. In this paper we will be discussing about the research done in the Recommendation System its approaches and the techniques used. At the end we will go through the main challenges and the limitation of the system.

Keywords
Recommendation System; Content based algorithm; Collaborative filtering algorithm; Hybrid approach.

1. INTRODUCTION
With the growth of Buying and selling of things Online, it became a big problem to find what the user is actually looking for. Search engines partially solved that problem, which has evolved a new branch and scope of system like Recommendation System [1].Recommender System are the software engines and approaches for providing suggestion of products to the user which might be most probably matched to the users choice. Usually the recommender system is a technology which filters out the information to envision in case a particular user will like a specific item; this is usually called as prediction problem, or to identify N set of items that will be of certain users interest called as Top N-recommendation problem [2]. From past few years the use of recommender System is being gradually increasing in various different applications, for instance application for recommending books, CDs and other products at different search engines like amazon.com , Netflix.com, ebay.com and so on. Even the Microsoft suggests many additional software's to user, to fix the bugs and so forth [3]. When a user downloads some software, a list of software is provided by the system. All the above examples would be result of diverse service, but all of them are categorized into a recommendation System, Identifying web-pages that will be of interest, or even implying backup ways of searching for information's [4]. There is huge number of algorithms used for making a personalized Recommender System, but out of them 2 algorithms became most popular they are Content Filtering from the item based filtering and Collaborative Filtering from the social community [5].They are used as the base for new era's recommender system[6].

The list of techniques which recommender system applies comes from other research domain such as Human Computer Interaction (HCI) or Information Retrieval (IR) [7]. However, most of these systems take in their core as algorithm that can be understand as a particular instance of a data mining (DM) technique. The data mining development consists of 3 steps, succession: Data Preprocessing Data Analysis and Result Interpretation [8][9].

2. BACKGROUND
The outcome of Recommender System is a list produced out of the two mainly two –collaborative filtering and Content filtering. Content based algorithm is built up solely at the time when a new profile of user is built [10]. All the information about the user's choice is stored in the users profile, from there the taste of the user is studied. In recommender system the taste is studied by combining the entire positively rated product in one group and then finding the maximum likelihood ratio of the ratio, and is suggested to the user [11].

Collaborative Filtering is one of the most research topics. The main concept is behind studying the social community and then deciding the user having the similar appreciation. If the users have similar choices and tastes then they fall in same category of choices. Even though the User has not rated the products but still it will be recommended to the user if they belong to the same category

3. TRADITIONAL RECOMMENDER APPROACHES

3.1 Content Based Filtering
This algorithm is also called as cognitive Filtering**;** this algorithm has been implemented successfully on text mining related system. Pure Collaborative Filtering Technique applies the matrix of ratings given by user; they can be in direct or indirect way, so as to generate a collaborative model [12]. These techniques treat the entire user and the item as a one unit, regardless of specifying the individual user or item [13].

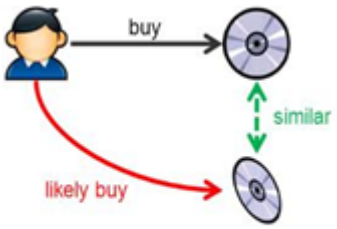

Fig. 1 Users Choice

PRES is a content-based filtering system it recommends the item by analyzing the user s profile satisfying the user's document in the archive. The contents of the archive can be

denoted wit set of terms. These terms are then extracted from the archive by running through the other algorithm [Porters 1980] (parsing algorithm). First all the HTML tags and the stop words are removed. Then the stemming is applied on the remaining words (removing the prefixes and the suffixes). For eg the word 'Plays', 'Played' and 'playing' will be constructed to the word 'Play'[14]. The Profile of User is expressed with similar terms and erected by considering the contents in which user is concerned of. The contents which are concerned to user can be an ambiguous or unambiguous data. Ambiguous data are the feedback which should be evaluated by the. User and is clear. Whereas the unambiguous data is the feedback in this the user's action is observed which is more convenient to the user but complex to implement which makes it fuzzy. Formally, an item is described as a vector (*n*) *X x* , *x* ,..., *x* 1 2 = of *n* components[15].

3.2 Collaborative Based Filtering

3.2.1 Memory Based System (User Based) uses the ratings to make the Prognosis.

Table 1: Example of a Rating Matrix

Person/ Subject	Sonu	Amit	Monu	Ram	Adi
Mathematics	5		4	2	4
Computer	2			5	1
English	4	3			
Social Science	4		4	5	1

This Algorithm directly uses the system similar like the polling for the judgment of the rating of the active user. One possible way is to assume all the users but it usually accepts the neighborhoods algorithm [16]. This algorithm is called as the "User-Based Nearest Neighbor Algorithm" and is commonly used.

To find out the poll of the acquaintance match to the user, so as to give the weights and ratings to the acquaintance [17]. The commonly used method to solve such problem is done by the Pearson correlation formula. The below is the formula of the Pearson correlation.

$$userSim(u,n) = \frac{\sum_{i \subset CR_{u,n}} (r_{ui} - \overline{r}_u)\,(r_{ni} - \overline{r}_n)}{\sqrt{\sum_{i \subset CR_{u,n}} ((r_{ui} - \overline{r}_u)^2)} \sqrt{\sum_{i \subset CR_{u,n}} ((r_{ni} - \overline{r}_n)^2)}}$$

The above formula is used for the finding user similarity formula using Pearson correlation, Where $CR_{u,n}$ denotes the set of co-rated items between u and n. In the Table I show the rating based user item choices and likings, it has the person and subjects as the prime factors for discussion.

3.2.2 Model Based systems (item Based) make its own model for Prognosis

In contrast to the above algorithm, this algorithm try to model the user based on their past ratings for predicting the unseen ratings of the products. It is type of probabilistic approach and analyzes the collaborative Filtering process as the computing the expected value of the user forecast, given his/her rating on the products[18]. This algorithm has been performed on several different Machines learning Algorithm like Bayesian Network, Rule-based approach. A Probabilistic Model is defined by the Bayesian Network for the Collaborative Filtering problems [19]. The Clustering Model treats this algorithm as a classification problem and performs by clustering user in same class and evaluating the probability that a particular user is in particular class C[20][21], and from that it computes the probability rating. Similarly the association rule is applied to find the relationship between the purchased items and generate the association rating between items.

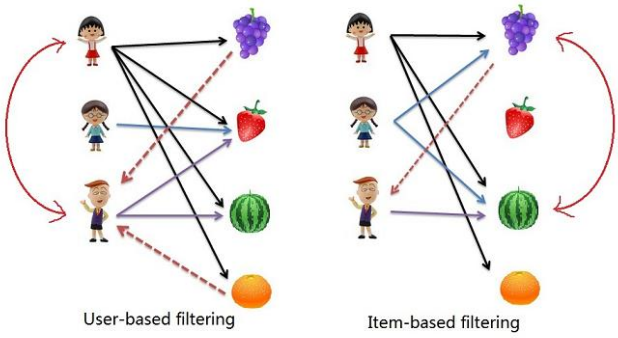

Fig 2: Difference between User and Item Based Filtering Techniques

There are many model-based CF algorithms. These include Bayesian networks, clustering models, Multiple Multiplicative Factor, latent semantic models such as singular value decomposition, Latent Dirichlet allocation, markov decision process based models, and probabilistic latent semantic analysis [22].

3.3 Hybrid Based Filtering

Both of the above mentioned approaches have certain strengths and drawbacks. In order to achieve better recommendation results, both the techniques are combined to take over the main advantages and removing the drawbacks of the approach [23]. In general, hybrid recommenders are systems that combine multiple recommendation techniques together to achieve a synergy between them. Although there exist a number of recommendation approaches that are practical to merge (i.e. Collaborative, Content-based, Demographic and Knowledge-based Recommender), here we have considered only on the combination of CF and CBF.The table 2 shows the difference between the three approaches.

Table 2: Overview of Collaborative Filtering Techniques

CF Categories	Representative techniques	Main advantages	Main short comings
Memory based CF	• Neighbor based usually top n recommendations are considered • Algorithms like Pearson /Vector cosine is used	• Implementation is easy • Easily new data can be added • Not necessarily considers the recommended items • Match well with co-rated items	• Depends on users ratings • Poor performance for sparse data • Herd to recommend for new users and items • Limited scalability for large datasets
Model based CF	• Bayesian belief nets CF • Clustering CF • MDP base CF • Latent semantic CF • Sparse factor analysis • CF using dimensionality reduction using SVD,PCA	• Understands the sparsity , scalability and other problems • Predictions performance is improved • Easy to understand for recommendation	• Expensive model building • For dimensionality reduction technique useful information is lost • Have to deal with others factors
Hybrid	• Content based CF • Content–boosted CF • Hybrid CF combining memory based and model based CF algorithms	• Prevail over limitations other recommenders • Better forecast performance • Prevail over CF such as sparsity and gray sheep problems	• Implementation is costly and increased complexity • External information are needed which are unavailable

4. EXISTING SYSTEMS

There is several of existing recommendation system and several research works is going on to make that system robust as well adding new features each system are made with different techniques and algorithms. Each system are have been made for different purposes. In some systems ratings take the main role for finding the recommendation items. These ratings could be taken directly by users surfing history and other sources it depends on the system what type of data is required for what type of system. The table 3 shows the name of the system and their application along with the technique and algorithms used for making that system.

Table 3: Existing Recommendation System

	Techniques	Domain
Grouplens	CF(Pearson)	Usenet Net news
Ringo	CF(Constrained Pearson)	Audios CD
Krakatoa	CBF(TF-IDF)	Newspaper
CiteSeer	Citation indexing	Research paper
Fab	CB,CBF	Web Pages
CBCF	CF, CBC	Movies
P-Tango	CF, CBF	Newspaper

Fab is a hybrid based Recommendation system which recommends the web pages to the specific user by analyzing the users profile with the help of Collaborative Filtering Technique. This system works on 2 principles they are information gathering and Collecting and Selecting.

P-Tango is hybrid Recommender system which uses the average on the basis on content by using the collaborative Filtering and then does the prediction. In this system the user can rate to the items explicitly.

Krakatoa chronicle is a system used to clarify newspaper and creates a business like realistic newspaper with multi column format. In this system the user profile is built with the e help of the keyword search d by the user. The archive is constructed on by three frameworks they are score given by user (TF-IDF), average score received by all the users and the size of the article.

Grouplens system is a collaborative system to recommend Usenet net Users based on the rating score. Although this system is not currently been used but it was said to be the most successful system which had employed the CF technique.

Ringo system the domain of this system is similar to that of Grouplens but instead of having 5 rating scale it has used the scale as 7. The constrained Pearson was used to implement this system but it reduced the analysis.

CiteSeer it is the largest storehouse for the research paper related to Computer Science on web. it also supports automated citation indexing for processing Documents with the help of Machine Learning Techniques.

Content–boosted Collaborative Filtering (CBCF) is a system which is used to recommend movies to the user. It uses the Collaborative Filtering technique for comparing the profiles of different users and then compares the result and according to that returns the recommended output.

5. CHALLENGES AND ISSUES

5.1 Cold-start

For the new user it is very difficult to recommend them any items as their profile would be empty and they would have not yet rated any item. So it becomes difficult for the system to recommend any product this issue is called as cold start problem. In some system this problem is overcome by having a short survey from the user to know the taste of the user, whereas some systems use the hybrid model to solve such problems

5.2 Trust

Sometimes the reviews and rates given by user who rarely uses his profile is irrelevant comparing to the profile having a great history. This problem can be solved by giving the priorities to the user by noticing and then evaluating the things they generally buy and surf for.

5.3 Scalability

The number of user and items are increasing day by day so much that more resources are required for processing those information to get a perfect and efficient recommendation system. Most of the system use collaborative filtering to know the tastes of the user. This problem can be solved by using a hybrid approach.

5.4 Sparsity

In online shops that have a huge amount of users and items there are almost always users that have rated just a few items. Using collaborative and other approaches recommender systems generally create neighborhoods of users using their profiles. If a user has evaluated just few items then it's pretty difficult to determine his taste and he/she could be related to the wrong neighborhood. Sparsity is the problem of lack of information.

5.5 Privacy:

Privacy is the prime factor for all the system, to give the user the most accurate recommendation, the system has to use the history of the user , sometimes its locations too in such case the eyes is raised on the reliability of confidential data of the user. Many of the online shops have used certain algorithms to solve such problems so that the privacy of user is protected.

6. CONCLUSION

Recommender systems help people find items of interest from large information spaces. This direct interaction with end users creates new and difficult challenges than has been explored by machine learning and AI research in the past. These difficulties become more serious as recommenders move into information spaces where users may bring a wide variety of information needs. We claim that by understanding the user's information seeking task, we can generate a more useful recommendation list. To do this, not only do we need to understand user tasks, we need to rethink about how we evaluate recommender algorithms, possibly using a wider.

7. ACKNOWLEDGMENT

The review has been partially supported by Department of Computer Science and IT Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. The views expressed here are those of the authors only

8. REFERENCES

[1] Karypis, George. "Evaluation of item-based top-n recommendation algorithms." Proceedings of the tenth international conference on Information and knowledge management. ACM, 2001.

[2] Asanov, Daniar. "Algorithms and methods in recommender systems." Berlin Institute of Technology, Berlin, Germany (2011)

[3] Survey Paper on Recommendation System Mukta kohar,Chhavi Rana Department of computer science and Engg U.I.E.T, MDU,Rohtak .

[4] Gunawardana, Asela, and Guy Shani. "A survey of accuracy evaluation metrics of recommendation tasks." The Journal of Machine Learning Research 10 (2009): 2935-2962.

[5] Cover, T.,and Hart, P., Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1):21–27, 1967

[6] Miranda, Tim, et al. "Combining content-based and collaborative filters in an online newspaper." In Proceedings of ACM SIGIR Workshop on Recommender Systems. 1999.

[7] Almazro, Dhoha, et al. "A survey paper on recommender systems." arXiv preprint arXiv:1006.5278 (2010).

[8] Pazzani, Michael J., and Daniel Billsus. "Content-based recommendation systems." The adaptive web. Springer Berlin Heidelberg, 2007. 325-341.

[9] Sarwar, Badrul, et al. "Item-based collaborative filtering recommendation algorithms." Proceedings of the 10th international conference on World Wide Web. ACM, 2001.

[10] Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007, May). Google news personalization: scalable online collaborative filtering. In Proceedings of the 16th international conference on World Wide Web (pp. 271-280). ACM.

[11] Van Meteren, Robin, and Maarten Van Someren. "Using content-based filtering for recommendation." Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop. 2000.

[12] Pazzani, Michael J. "A framework for collaborative, content-based and demographic filtering." Artificial Intelligence Review 13.5-6 (1999): 393-408.

[13] Melville, Prem, and Vikas Sindhwani. "Recommender systems." Encyclopedia of machine learning. Springer US, 2010. 829-838.

[14] Van Meteren, Robin, and Maarten Van Someren. "Using content-based filtering for recommendation." Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop. 2000.

[15]] Kuppusamy, K. S., and G. Aghila. "A personalized web page content filtering model based on segmentation." arXiv preprint arXiv:1202.1881 (2012).

[16] Golbeck, Jennifer. "Combining provenance with trust in social networks for semantic web content filtering."

Provenance and Annotation of Data. Springer Berlin Heidelberg, 2006. 101-108.

[17] Basilico, Justin, and Thomas Hofmann. "A joint framework for collaborative and content filtering." Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2004.

[18] Massarani, Leonardo C. "Content-indexing search system and method providing search results consistent with content filtering and blocking policies implemented in a blocking engine." U.S. Patent No. 6,336,117. 1 Jan. 2002.

[19] Breese, J. S., Heckerman, D., and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52.

[20] Basu, C., Hirsh, H., and Cohen, W. (1998). Recommendation as Classification: Using Social and Content-based Information in Recommendation. In Recommender System Workshop '98. pp. 11-15.

[21] Ungar, L. H., and Foster, D. P. (1998) Clustering Methods for Collaborative Filtering. In Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence.

[22] Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." Advances in artificial intelligence 2009 (2009): 4.

[23] David H. Stern, Ralf Herbrich, and Thore Graepel. Proceedings of the 18th international conference on World Wide Web, page 111--120. New York, NY, USA, ACM, (2009)