International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244
Impact Factor: 6.111

A REVIEW: COMPARATIVE STUDY OF CONTENT BASED AND COLLABORATIVE FILTERING APPROACH

Anu Taneja, Assistant Professor, BCIIT, GGSIPU, Delhi

Abstract: *As the information on the web is increasing, it is becoming difficult for users to choose the right option. So recommendation systems are developed to guide the users in selecting the best option. In this paper we did comparative analysis of recommendation approaches i.e. content based and collaborative filtering approach in which we mainly focused on which approach outperforms other and under what situations, what are the major application areas and the research challenges faced in this domain.*

Keywords*: Recommendation System, Content-Based, Collaborative Filtering, Hybrid Recommendation Systems, E-Commerce.*

I. INTRODUCTION

As we have to select options from huge set of data, so recommendation systems are becoming popular day by day and they are mainly designed to filter out the most relevant information according to user's need. This is possible by storing the user's information such as his reviews can be analysed like ratings given by user, comments given, how user interacts like clicks, time-spent on a particular page etc. By maintaining the user profile, most suitable items are recommended to the users according to user's interest. Now a day's recommendation Systems are applicable in numerous domains such as movies, books, videos, music, friends, news, and restaurants etc. For better utilization of recommendation systems, efficient recommendation approaches have been implemented. The broad categories of recommendation systems are content-based, collaborative filtering and hybrid-based approach. Collaborative Filtering Approach is further divided into three categories Memory Based/Neighbourhood Approach, Model Based and Hybrid Approach which is shown as below in fig 1:

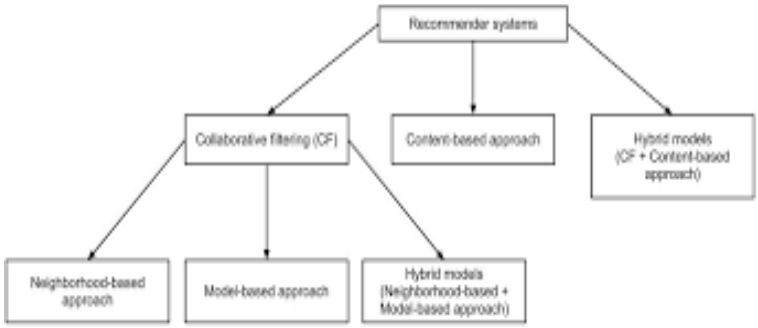

Fig1: Classification of Recommendation Systems

II. CONTENT BASED FILTERING APPROACH

The two main approaches [1] which are used in the design of the recommendation systems are:

- Content-Based Filtering.
- Collaborative Filtering.

Content Based Filtering is the technique which mainly analyses the items that are previously rated by the user and then predict the user interest based on already those rated items. So in this way system recommends the items that best fit to the user. This example can help in better understanding of content-based filtering like if one user wants to purchase some product and if that product is not available then recommendation systems will recommend similar product on the basis of items information already stored. The major advantage of this approach is that using this approach, recommendation systems start recommending as soon as the information of items is available as no user input is required to recommend in this approach. But the main limitations of this approach are firstly if a user is new to the system i.e. new user then sufficient information of user about its interests will not be available then it is difficult to recommend the items accurately. Secondly, this approach has to match up the profile which is build from user interests and items.

Work-Flow of Content-Based Filtering Approach

The flow of content-based filtering approach is as follows:

The flow of item-based collaborative filtering [2] is as follows:

- Finding the similarity among the items: The first step in this approach is to find the most similar items which can be computed using Pearson Correlation Coefficient or Cosine Distance Similarity methods.

- After finding the similarity between the items, top n items are selected that are having the highest similarity.
- Then prediction is computed as weighted average on the user's ratings on the most similar items.

III. COLLABORATIVE FILTERING APPROACH

The second approach which is used in the design of recommendation systems is collaborative filtering. Collaborative filtering [2] is the process of filtering the information from very large data-sets according to user taste. But the main assumption which collaborative filtering assumes that if one user has similar taste as some other user then it is more likely to say that these two users can have same taste on some other issue. Example of collaborative filtering include as if one user like some book of particular subject then recommendation systems can recommend the similar books or related videos too of the same subject according to the similar users interest who liked the same book in past. The major advantages of this approach are that this approach does not require any substantial information about products or users as recommendation is done on the basis of other user's experience and recommendations done by them in the past. The major limitations of collaborative filtering approach is that this in approach initially ratings of the users are required so that recommendation can be done by finding the similarity among users i.e. this approach is user-dependent. Secondly if the user is new then it is quite difficult to recommend as user has not gone through the rating process earlier, this is known as cold-start problem.

Work-Flow of Collaborative Filtering Approach

The flow of collaborative filtering approach is as follows:

- Firstly the user rates the items i.e. shows his preferences in the specific domain.
- Then the system matches the user's ratings with other users which help in finding the similar users.
- Now the products are recommended to that specific user according to the ratings done by the similar users.

IV. COMPARISON OF CONTENT BASED AND COLLABORATIVE FILTERING APPROACH

The major difference between content-based filtering and collaborative filtering is that collaborative filtering approach do not use any actual information of item rather this approach is based on preference patterns of other users whereas content based filtering approach uses the actual information of the item. In content based approach, the major drawback is that too much similar items are recommended to the user, so he never gets a chance to explore other options whereas in collaborative filtering approach, those items are recommended which are similar to other user's interests so user gets chances of exploring different products. The cold-start problem i.e. new user problem [4] exists in both approaches as if user is new to the system then in case of content-based approach, his past choices will be unknown to the system, so it will be difficult to recommend the items accurately and similarly in collaborative filtering approach, if user is new to the system then it is difficult to find similar users as his interests are unknown and finding the similar users is the basic principle of collaborative filtering. The major advantage of content based filtering is that there is no new item problem as if this item is of interest to any user, then it will be in consideration for recommendation but new item problem exists in collaborative filtering approach as if this new item is not rated by large number of users, then this new item has less chances for being in consideration for recommendation.

V. RELATED WORK

Now a day, recommendation systems have become so popular that it is applicable in numerous domains such as movies, music, news, friends, books etc. In this section related work in this domain has been elaborated.

Music Recommendations

Gideon Dror [5] in his work on music recommendations uses Yahoo's data as music dataset which contains million of users, thousands of musical items and millions of ratings. All musical categories of items are linked together within a defined taxonomy. A mechanism of timestamp is involved due to which session analysis of user activities is possible which helps in determining the exact order in which ratings were given. As collaborative technique suffers from cold-start and long tail problem so a novel method is designed by using music taxonomy information. The model used is matrix factorization which involves temporal

analysis of user ratings and item popularity trends. As user preferences and popularity of music will change with time so modelling of temporal dynamics with collaborative filtering (CF) models helps in getting better results.

Seok Kee Lee [6] in his work on recommending music for mobile devices confesses that explicit ratings as done in traditional collaborative techniques to collect user preferences is a limitation as compared to implicit ratings because of poor interfaces and high telecommunication costs in mobile web environment. So a new collaborative filtering method with ordinal scale based implicit ratings is proposed. To keep record of implicit preferences information mobile web usage mining technique is used by using web log data. But this information can be insufficient so to handle this consensus model can be the solution which addresses the problem of aggregating preferences of users into a compromised preference. This methodology is implemented in real mobile web environment and results show that the performance of this methodology is better than traditional methodology of collaborative filtering approach.

Products Recommendation

Hui Li [7] in his work focus on recommending the products the user by mitigating the sparsity problem generally encountered in collaborative filtering approach. The traditional collaborative filtering approach uses user-item matrix as the base which consists of ratings done by various users on different items to predict the products. But as there are large numbers of users, and we need to find the most similar users so there is extreme sparsity of users rating data. So in this a new approach has been developed to find the similar users by first refining the user ratings data first so that sparsity problem can be reduced to certain extent. Though various approaches have been developed to overcome sparsity problem such as clustering technique, weighted co-clustering technique multilayered semantic social network model, bi-clustering collaborative filtering technique. But this proposes a new technique that introduces the concept of attribute deduction and neighbour domain i.e. computing the distance to all other users in the rating matrix. So this approach is more scalable as compared to existing approaches. But the further work can be carried out by taking into consideration the factors that impact the customer feelings as customer satisfaction is the major requisite for recommendation systems.

Gal Oestreicher [8] in his work mainly focus on the long-tail problem of collaborative filtering approach i.e. there might be some items which are not rated by the users so they remain unnoticed and user is unaware of such items. To solve this problem a new approach is designed for linking recommendation networks to the long tail by connecting the position of products to the relative demand and revenue within their respective categories. In such a network having large number of links, firstly the position of the product is measured using Google's Page Rank. Then average influence of the network is associated with the relative demand and revenue using Gini's coefficient derived from the Lorentz's curve.

News Recommendation

Jiahui Lui [9] in his work on news recommendations analyses the user's log and based on this developed a Bayesian Framework and predict the news which user would like to read. The major challenges in existing algorithms are firstly as in collaborative filtering approach items are predicted on user's past behaviour but news are generated within short span of time so they must be reached to users timely. But using existing algorithm they will not reach until rated by users. Secondly, all users cannot be treated equal as most popular news can be flashed to maximum users even he is not interested in watching that news. So solution to these problems is that user's profile should be generated on the basis of his reading interest. The main issue is that user interest may change over time so system should be able to keep the updated profile to reflect the changes in interest. In future all the factors such as user profile, time spent, short term interest according to event or long term interest like in which kind of news the user is generally more interested can be combined to improve the news recommendations.

VI. COMMON RESEARCH CHALLENGES

- **Data Sparsity** [3]: As dataset is very large and user-item matrices are sparse in nature then it is difficult to predict the recommendations accurately.
- **Scalability:** As number of users and items are growing day by day then traditional collaborative filtering systems will suffer from scalability problems. As in some systems it is needed to make recommendations for users immediately then in such cases highly scalable collaborative filtering systems are required.

- **Synonyms:** This refers to that similar items may have different names so this problem is not being resolved by recommendation systems i.e. it treats the products differently which reduces the performance of recommendation systems.
- **Gray Sheep:** This refers to those users whose opinions or interests do not agree with other users. So such types of users are not benefitted from the design of recommendation systems.
- **Shilling Attacks:** This problem arises among competitors when they just rate their own items and give negative ratings for the competitors so such type of practice should be prevented while designing the recommendation systems.
- **Cross-Domain:** As majority of the recommendation systems are designed to recommend the items in a single domain like last.fm suggests only music compositions, it would be better if recommendation systems can recommend in multiple domains like in addition to music compositions according to user's interest, relevant movies can also be recommended.
- **Group Recommendations:** When a number of users are involved then it is difficult to recommend the items in such a way that it can satisfy all the users to the maximum extent.
- **Long Tail Recommendations:** There might be some products which may remain unrated and due to this, users will never be aware of such products so design of recommendation systems must be done in such a way that this problem can be handled efficiently.
- **Less Reliable:** It can be the case that some unauthorized users rate the items which just reduces the performance of the recommendation systems.

VII. CONCLUSION

In this paper we discussed about how recommendation systems have gained popularity, comparative study of recommendation approaches have been done with their major advantages and limitations, application areas of this domain have been elaborated and research challenges in which further work can be carried out to better utilize the efficiency of recommendation systems have been discussed. The further work can be carried out in these aspects of implementation of recommendation approach in different domains.

REFERENCES

[1] Xiaoyuan Su and Taghi M. Khoshgaftaar, "A Survey of Collaborative Filtering Techniques", "Published in ACM Journal of Advances in Artificial Intelligence", Volume 2009, Article No. 4, January 2009.

[2] Yue Shi, Martha Larson and Alan Hanjatic," Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges", "Published in ACM Journal of Computing Surveys", Volume 47, Issue 1, Article No. 3, July 2014.

[3] https://en.wikipedia.org/wiki/Collaborative_filtering

[4] G Adomavicius, A Tuzhilin, Towards the next generation of recommender systems, a survey of the state-of-art and possible extensions, "Published in IEEE transactions on Knowledge and Data Engineering, 2005.

[5] Noam Koenigstein, Gideon Dror, Yehuda Koren, "Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy", Published in Proceedings of the fifth ACM conference on Recommender systems, Pages 165-172, 2011.

[6] Seok Kee Lee and Soung Hie Kim, "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations", Published in Elsevier Journal of Information Sciences, Volume 180, Issue 11, Pages 2142–215, 1 June 2010.

[7] Hui Li, Shu Zhang and Xia Wang, "A Personalization Recommendation Algorithm for E-Commerce", Journal of software, Academy Publishers, Vol. 8, No. 1, January 2013.

[8] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324064

[9] Jiahui Liu, Peter Dolan and Elin Rønby Pedersen, "Personalized news recommendation based on click behaviour", Published in Proceedings of the 15th international conference of ACM on Intelligent user interfaces, Pages 31-40.