

INFORMATION FILTERING & IT'S IMPACT IN RETRIEVING INFORMATION

Ruban S¹, Athul C² & Rakshak³

Abstract-Information filtering can refer to a wide range of strategies or solutions for refining data sets. This means the data sets are refined into simply what a user needs, without including other data that can be repetitive, irrelevant or even sensitive. The information filtering will involve taking out information that is useless to a reader or information that can be confusing. Generated reports and query results from database tools often result in large and complex data sets. Redundant or impartial pieces of data can confuse or disorient a user. Filtering information can also make results more efficient. Some examples of Information filtering are: filters for search results on the internet that are employed in the Internet software, personal e-mail filters based on personal profiles, newsgroups filters for groups or individuals, browser filters that block non-valuable information, filters for e-commerce applications that address products and promotions to potential customers only, and many more. The paper defines a framework to classify Information filtering systems according to several parameters, and illustrates the approach with commercial and academic systems. The paper describes the underlying concepts of Information filtering systems and the analysis of the filters based on few features.

Keywords: Information filtering, Types of information filtering, Information Retrieval.

1. INTRODUCTION

Now a day's more than 3 billion people are using the Internet, according to the United Nations agency that oversees international communications. So whenever user try to search information or data from internet user will get more unwanted data rather than important. This is due to lack of information on internet. This scenario is commonly referred to as the problem of information overload. Information overload has become a popular buzzword of our times and people feel overwhelmed when navigating through today's information and media landscape. The information filtering (IF) can be defined as process of monitoring large amounts of dynamically generated information and pushing to a user the subset of information likely to be of her/his interest. In other words delivery of information that the user is likely to find interesting or useful.

The information filtering will involve taking out information that is useless to a reader or information that can be confusing. This information filtering process is used everywhere on the internet for instance, online shopping (Amazon, flip kart, Snap deal etc.). An information filtering system can be designed as being an automatic mechanism with the capacity of monitoring a continuous flow of document and ability to select documents considering its relevance for a certain user or users groups, according to its needs.

We use various types of filtering methods such as content-based filters, Collaborative filters and Hybrid filters. This type of system is designed for unstructured or semi structured data. Information filtering systems deals with textual information.

The content based filtering can be also called as cognitive filtering the information source that content-based filtering systems are mostly used with are text documents. Collaborative filtering, also known as social filtering, filters information by using the recommendations of other people. The goal of hybrid filters is to take the best features of each and minimize the impact of their weakness with the goal of outperforming each individually. Basically the information filtering is divided into content-based and collaborative filtering the two approaches can also be used together i.e. Hybrid system.

2. INFORMATION FILTERING

Now a days humans are becoming dependent on the internet. Every individual work can be done by online. No work has to be done manually due to advanced technology. The work can be done by using internet or any system. Due to this, the information on internet becoming more. So whenever user try to search on internet a particular information user will get unwanted information rather than important. This leads to confusion for users to select information of his/her needs. This scenario is commonly referred to as the problem of information overload. To solve this problem we have information filtering which is introduced by Luhn in 1958 as "Selective Dissemination of Information" and later named "Filtering" by Denning in 1975. At present, these techniques are used in different applications, not only in the web context, but in thematic issues as varied as voice recognition, classification of telescopic astronomy or evaluation of financial risk. An information filtering

¹ Asst. Professor, Department of IT, St. Aloysius College (AIMIT), Mangalore.

² Aloysius Institute of Management and Information Technology (AIMIT), Beeri, Mangalore, Karnataka, India

³ Aloysius Institute of Management and Information Technology (AIMIT), Beeri, Mangalore, Karnataka, India

system assists users by filtering the information source and deliver relevant information to the users. The information filtering system can be called as recommender system if the delivered information comes in the form of suggestions.

3. TYPES OF INFORMATION FILTERS

Basically there are three types of content filtering or information filtering. Which are as follows,

- Content based Filters.
- Collaborative Filters.
- Hybrid Filters.

3.1 Content based Filters-

Content based filters can also be referred as cognitive filtering. It works based on comparison of item features or user features. The recommender system suggests other objects that are similar, or in other words liked by other users. The opinion of the users can be identified using two methods Explicit or Implicit. Or we can say that by keeping track of user's rate items or looking at likes or clicks the system can acquire user's opinion. For e.g. If the user is trying to search for a particular book, the search engine will recommend some of the similar titles from their past likes. This technology is used by some of the major companies like Netflix, Pandora's search engines. Such type of systems are mostly used with text documents.

3.2 Collaborative Filters-

Collaborative Filters can also be referred as Social filters. This is commonly used in Recommender Systems. This system uses database of users likes to find users with similar interests. In this system filters will predict the titles or information based on the other user's interests. The information need not be seen it can be unseen. For e.g. If the person wants to watch a movie, he/she might ask other users opinion or friends about the particular movie. Because different peoples have different opinions. So that in this case only those people can see a movie who has similar interests.

3.3 Hybrid Filters

The hybrid filters is a combination of both content based filters and collaborative filters. Generally the process of filtering starts with any one of the filtering method (Content based or Collaborative) and later includes features from another type of filter. By doing this the performance of filter will be good. This method can also be done by making content-based and collaborative-based predictions separately and then combining them. One of the major American entertainment company named as Netflix is example for hybrid filters. The recommendations can be made by comparing the searching and watching habits of the users.

4. INFORMATION FILTERING IN E-COMMERCE

Now a days online shopping is on trend. We get almost everything from online shopping and because of this the information on internet server is also high. Due to this the information filtering technique is also available here to overcome this problem. The companies like Amazon, Flip kart recommends their customers based on the interest of the user and also by checking the ratings. Similarly this can be seen in Google's play store where the recommendation is based on the ratings by means of stars (*).

5. RESULTS & DISCUSSION

Features	Content-Based	Collaborative	Hybrid
1. Works with existing profiles of users	✓		✓
2. Compares the items that are already positively rated by the user	✓		✓
3. Recommendation based on rating.	✓		✓
4. Efficient and personalized	✓	✓	✓
5. User behaviour is more important		✓	✓
6. Adaptive: Quality improves over time.	✓	✓	✓

Table 1: Comparative Study of different Filters

As discussed in the previous table, with respect to the features like working with the existing profile of users, comparing the items that are already positively rated by the user, recommendation based on rating, efficient and personalized behaviour, Analyzing user behaviour and the improvement of quality over time, we found hybrid filters perform better than the other filters.

6. CONCLUSION

At last we conclude that information Filtering is a system where only important information is selected and shown to users and unwanted information is discarded. In this paper we discuss about IF, types of filtering and techniques used in information retrieval. This paper also discussed about Information filtering and its impact in informational retrieval.

7. REFERENCES

- [1] Josep Silva "Information Filtering and Information Retrieval with the Web Filtering Toolbar", Electronic Notes in Theoretical Computer Science 235 (2009) 125–136.
- [2] URI HANANI, BRACHA SHAPIRA and PERETZ SHOVAL, "Information Filtering: Overview of Issues, Research and Systems", Department of Information Systems Engineering, Ben-Gurion University, Beer-Sheva 841055.
- [3] Colm O' Riordan IT Centre, NUI, Galway, Ireland. c.oriordan@ucg.ie, Humphrey Sorensen Department of Computer Science, University College Cork, Ireland. h.sorensen@cs.ucc.ie, "Information Filtering and Retrieval: An Overview".
- [4] Sharifa Rania Mahmud, "A Simple Information Retrieval Technique", Int. J. on Recent Trends in Engineering and Technology, Vol. 8, No. 1, Jan 2013.
- [5] Niclas Eberhagen, "Information filtering", Box 5053, S-351 95 Växjö June 1993
- [6] Nicholas J. Belkin and W. Bruce Croft, "Information filtering and information retrieval: Two sides of the same coin?", Dec 1992 v35 n12 p29(10)
- [7] Daniel Billsus and Michael J. Pazzani, "Learning Collaborative Information Filters".