International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 4 Issue 10 || December. 2016 || PP-09-13

Enhanced Web Usage Mining Using Fuzzy Clustering and Collaborative Filtering Recommendation Algorithms

A.Sangeetha[1], C.Nalini[2]
[1]*Ph.D Scholor, Department of Computer Science & Engg. ,Bharat University*
[2]*Professor, Department of Computer Science & Engg. ,Bharat University*

ABSTRACT: *Information is overloaded in the Internet due to the unstable growth of information and it makes information search as complicate process. Recommendation System (RS) is the tool and largely used nowadays in many areas to generate interest items to users. With the development of e-commerce and information access, recommender systems have become a popular technique to prune large information spaces so that users are directed toward those items that best meet their needs and preferences. As the exponential explosion of various contents generated on the Web, Recommendation techniques have become increasingly indispensable. Web recommendation systems assist the users to get the exact information and facilitate the information search easier. Web recommendation is one of the techniques of web personalization, which recommends web pages or items to the user based on the previous browsing history. But the tremendous growth in the amount of the available information and the number of visitors to web sites in recent years places some key challenges for recommender system. The recent recommender systems stuck with producing high quality recommendation with large information, resulting unwanted item instead of targeted item or product, and performing many recommendations per second for millions of user and items. To avoid these challenges a new recommender system technologies are needed that can quickly produce high quality recommendation, even for a very large scale problems. To address these issues we use two recommender system process using fuzzy clustering and collaborative filtering algorithms. Fuzzy clustering is used to predict the items or product that will be accessed in the future based on the previous action of user browsers behavior. Collaborative filtering recommendation process is used to produce the user expects result from the result of fuzzy clustering and collection of Web Database data items. Using this new recommendation system, it results the user expected product or item with minimum time. This system reduces the result of unrelated and unwanted item to user and provides the results with user interested domain.*

KEYWORDS: *fuzzy clustering, collaborative filtering, recommender*

I. INTRODUCTION

A web search engine is a software system that is planned to seek for information on the World Wide Web (WWW). The search consequences are commonly presented in a stroke of results regularly called to as search engine results pages (SERPs). The information may be a blend of web page, pictures, and other types of files. Some engines dig for data available in databases or open directory. Nothing like web directories, which are maintain only by human editors, search engines also maintain real-time information by executing an algorithm on a network crawler.

A search engine maintains the following processes in near real time:

1. Network crawling
2. Indexing
3. Searching

Web search engines obtain their information by network crawling from one site to other site. The "spider" also called crawler checks for the customary filename robots.txt, addressed to it, before sending that information back to be indexed depends on many factors, such as the titles, JavaScript ,page content, Cascading Style Sheets (CSS), headings, as proof by the standard HTML markup of the informational content, or its metadata in HTML Meta tags.

Indexing means associating words and other definable tokens initiate on web pages to their domain names and HTML-based fields. The associations are done in a public database, made available for web search queries. A query from a user can be a solitary word or set of words. The index helps find information linking to the query as rapidly as possible. Some of the methods for indexing and caching are secrets for the trade, while network crawling is a basic process of visiting all sites on an orderly basis.

Naturally when a user enters an inquiry into a search engine it is a few words. The index previously has the names of the websites containing the words, and these are directly obtained from the index. The actual processing load is in generating the web pages that are the search consequences list: Every page in the entire list

must be prejudiced according to information in the indexes. Then the peak search result item requires the lookup, reconstruction, and markup of the snippets showing the situation of the words matched. These are simply part of the processing each search results requires, and further pages (next to the top) require extra of this post processing.

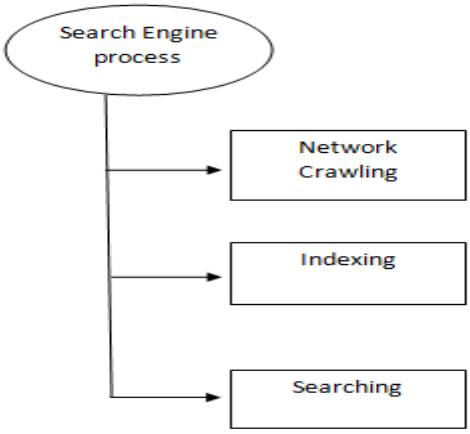

Fig.1 Search Engine Process

In a search engine, the user choices will be determined based on the previous histories, which is helpful to build a recommendation system. Recommendation system is a subdivision of information filtering system that search to forecast the ranking or fondness that a user would give to an item. These systems have become tremendously frequent in latest years, and are applied in a diversity of applications. The most admired ones are movies, social tags, news, books, research articles, search queries, music, and products in general. Other than this, there are also recommender systems for experts, collaborators, jokes, hotels, financial services, life insurance, people (online dating), and Twitter followers. This system follows two approaches:

- Collaborative Filtering
- Content based Filtering

Collaborative filtering approaches structuring a model from a consumer's precedent behavior (stuff previously purchased or chosen and/or numerical ratings given to those items) as well as alike decisions made by other users. This model is used to forecast items that the consumer may have a curiosity in. Content-based filtering approaches make use of a sequence of discrete features of an item in order to suggest added items with alike properties. These two approaches are joined to form a Hybrid Recommender Systems.

A. Organization

The enduring of the paper is explained as follows: In Part II, related work is clearly explained .In Part III, system architecture is clearly explained with the techniques used in the work. The techniques explained are fuzzy clustering and collaborative filtering. In Part IV, conclusion is provided.

II. RELATED WORK

In recent years, there have been numerous works based on the recommendation. Few of the works are to be discussed as follows: In Improving efficiency of personalized web search [1], the user search is analyzed using content and keyword extraction technique. The main aim of that work is to progress the search engine quality. Depend on the user query, the search results are obtained. The content and keywords of results are analyzed. The query is preprocessed and root words are found out. Based on the words, the dictionary is constructed. The query is compared with the dictionary and the words are weighted and ranked. This work was implemented in client-side. In survey on web search engines [2], the search engine basic working is explained. This work explains the working and ranking concept of familiar search engines. Each search engine has its own searching methodology. Some of the search engines which are explained in that work are as follows: Archie, Gopher, Google, Bing, Yahoo and Ask. Archie uses File Transfer protocol concept to list all the search files.

Gopher user gopher protocol. It is an internet protocol to carry out search in the internet. Google is the popular search engine and it uses the page ranking algorithm to rank the web pages. Whereas Bing uses the number of back links to rank the results. Ask is an answering engine and it is not widely used now. Yahoo's ranking algorithm is based more on heading of the websites.

In Winsome [3], a search engine is designed based on the concept of Entity ranking and time cache results. This search engine provides the relevant results using the process of entity ranking. Entity ranking is applied to filter the irrelevant entities for the result and groups the relevant entities to provide the relevant results. But this engine does not concentrate on the user activities so the user query is not analyzed for providing recommendation scheme.

III. ARCHITECTURE

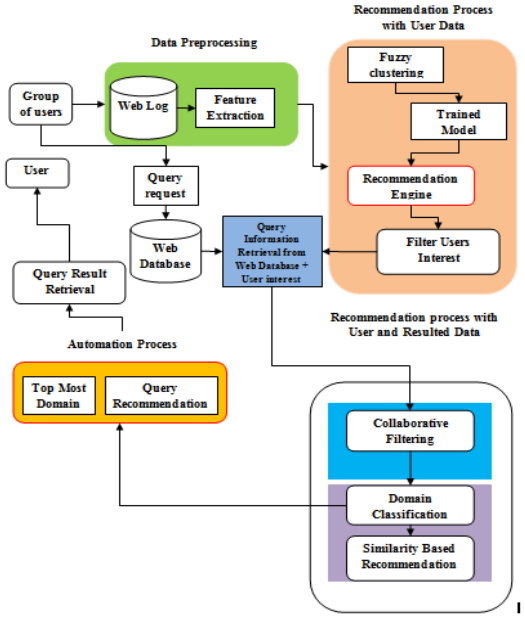

Fig.2 System Architecture

The system architecture modules are given as follows:

- Data Preprocessing
- Recommendation process with user data
- Recommendation process with user data and Resulted data
- Automation Process

Data preprocessing is a key part in the process of data mining. The saying "garbage in, garbage out" is mainly pertinent to data mining and learning projects. Data-gathering methods are regularly loosely controlled, follow-on in out-of-range values (e.g., Weight: −120), impossible data combinations (e.g., Education: Degree, Illiterate: Yes), missing values, etc. Analyzing data that has not been cautiously screened for such problems can create deceptive results. Thus, the illustration and quality of data is primary and leading ahead of running an analysis.

Data preprocessing is sub divided into following:

- Data Cleaning
- Data Integration
- Data Transformation
- Data Reduction

The system architecture of this recommendation system is shown in the above figure 2.This tool uses two main methodologies to provide an efficient recommendation system. The methodologies are:

- Fuzzy Clustering
- Collaborative Filtering

Fuzzy clustering (soft clustering) is the process of clustering in which every data point can belong to more than one cluster or partition. It was developed by J.C. Dunn in 1973, and enhanced by J.C. Bezdek in 1981. Clustering is the process of assigning data points to clusters or same classes. These are identified through similarity contains intensity, distance and connectivity. Distinct similarity may be selected depending on the application or data. The fuzzy algorithm is called Fuzzy C-means (FCM) algorithm. It is similar to k-means algorithm. This algorithm is used in Bioinformatics, Marketing and Image Analysis.

The FCM algorithm attempts to divide a finite collection of n elements $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ into a compilation of c fuzzy clusters according to some given condition.

Given a limited set of data, the FCM returns a list of c cluster centres $C = \{\mathbf{c}_1, ..., \mathbf{c}_c\}$ and a partition matrix

$W = w_{i,j} \in [0, 1],\ i = 1, ..., n,\ j = 1, ..., c$, where each element, w_{ij}, tells the extent to which element, $\mathbf{x}_i$, belongs to cluster $\mathbf{c}_j$.

The FCM aims to minimize an objective function:

$$\arg\min_{C} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| \mathbf{x}_i - \mathbf{c}_j \right\|^2,$$

where:

$$w_{ij}^{m} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\|\mathbf{x}_i - \mathbf{c}_k\|} \right)^{\frac{2}{m-1}}}.$$

Collaborative filtering (CF) is a method used by several recommender systems. This filtering has two senses, a narrow one and a more general one. In general, the process of filtering for information or patterns using methods involving collaboration among multiple agents, viewpoints, data sources, etc. It involves very large data sets applications. It has been applied to many diverse kinds of data such as: sensing and monitoring data, such as in mineral exploration, environmental sensing over huge areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or in e-commerce and web applications where the focus is on user data, etc. In the narrower one, collaborative filtering is a method of building automatic predictions (filtering) about the interests of a user by collecting preferences or tang information from several users (collaborating).

The collaborative filtering system process is given as follows:

1. A consumer expresses his or her opinions by ranking items (e.g. images, movies or CDs) of the system. These rankings can be seen as a rough illustration of the consumer's interest in the consequent domain.
2. The system matches this consumer’s ratings against other consumers’ and finds the people with most "related" tastes.
3. With related consumers, the system recommends stuff that the alike consumers have ranked highly but not yet being ranked by this consumer (most probably the nonexistence of ranking is often referred as the unfamiliarity of an item).

The architecture is explained as follows: From the group of users, the user’s interest is stored in web log. The web log is used to extract the features based on that the recommendation process can be provided. The user’s query request is stored in the web database. Then the recommendation process is carried out .it includes fuzzy clustering, trained model, recommendation engine, filter user’s interest. Fuzzy clustering is applied to get a trained model. The trained model data is used for the recommendation scheme. Upon that, list of recommendation can be obtained. The list is filtered based on the user’s interest to get the suitable recommendations. These recommendations are applied collaborative filtering, domain classification, similarity based recommendation. Collaborative filtering is normally applied to get the best recommendation approach. From this approach top most domain and recommended query is suggested to user .This process is fully automated. From the recommended the user may select a particular query based on their choice so that the time to search new query get reduced.

IV. CONCLUSION

This recommendation tool is designed with two key concepts fuzzy clustering and collaborative filtering .This work is to provide an efficient recommendation scheme for the user. This tool provides excellent recommendations for the user due to the usage of two techniques.

REFERENCES

[1] Improved Algorithm For Inferring User Search Goals Withfeedback Sessions" in International Journal of Researchin Computer Applications and R obotics, A.Sangeetha,C.Nalini,Bharath University Department of Computer Science and Engineering, Bharath University, Tamil Nadu, India

[2] Retrieving Relevant Links from the Web Documents through Web Content Outlier Mining From Web Clusters , **Volume 5, Issue 2, Februar y 2015 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering A. Sangeetha , T. Nalini** Department of Computer Science and Engineering, Bharath University, Tamil Nadu, India

[3] D.Keerthika and G.Sangeetha, "Winsome – A Search Engine", IJETCSE, Vol.21, Issue-3, April 2016.

[4] A Survey of Information Retrieval in Web MiningT. Nalini and G. Sangeetha Department of Computer Science and Engineering, Bharath University, Chennai, India Middle-East Journal of Scientific Research 19 (8): 1047-1052, 2014 ISSN 1990-9233 © IDOSI Publications, 2014 DOI: 10.5829/idosi.mejsr.2014.19.8.1506

[5] Web Search Personalization with Ontological User Pro f ile Ahu Sieg,Bamshad Mobasher,Robin Burke

[6] J. Srivastva, P. Desikan, and V. Kumar, *Web mining – Concepts,Application and Research direction*, pp. 51, 2009.

[7] O. Etzioni, "The World-Wide Web, Quagmire or Gold Mine?" *Communications of the ACM*, vol. 39, no. 11, pp. 65–68, 1996.