



Personalization of Web Search and its Techniques: A Survey

Raghavendra R¹, Dr. Niranjnamurthy M², Pavan Kumar M³

Assistant Professor, Department of Computer Science and IT, Jain University, Bangalore, India¹

Assistant Professor, Department of Computer Applications, MSRIT, Bangalore, India²

P.G Student, Department of Computer Science and IT, Jain University, Bangalore, India³

Abstract: The quality of information accessing from the Web (WWW) is growing rapidly per year as per the user demand, so that the quality of the results provided by the search engine will be low. This makes the retrieval of the relevant information which is extracted from the search results for the user is made more difficult. The quality of the result generated by the search engine information depends on the needs of the users and the various search techniques involved in the web search systems. Web logs keep a track of the user activities on the search engine. These web logs are mined by web usage mining which is one of the methods of web mining, where they are the rich sources for the Web Personalization. So, Web personalization becomes a basic need in web search, hence here we discuss various techniques used in the Web Personalization.

Keywords: Personalization, Web Usage Mining, Web Logs.

I. INTRODUCTION

The vast collection of documents around the world in the forms of web is known as the World Wide Web (WWW). The Data in WWW is growing exponentially due to its size and usage with respect to the time. It has proving to be a tremendous amount of data for the user access. It has become a powerful platform to retrieve, disseminate and store information as well as useful user knowledge.

To access patterns, obtain knowledge and the user behaviour's the depth study of the data source is essential. Considering the better design, structures and the content can be created for the web site and represented to the users. There are different types of data which have to be managed and it should be organized in such a way that they can be accessed by different users.

A huge amount of data is available on the web, so it makes the user to identify the most important data for particular user. Due to this Web Personalization it is required to find the relevant information hence various data mining methods are being used to deal with semi-structured or unstructured data, heterogeneous data [1].

Most of the queries given by the user in search engines are short. The needs of the every user are different with similar goals under the same query. Even the process of customizing the web search is based on the users past behaviour's which is done by the Personalization of web search. Thus the effectiveness of Personalization of web search depends upon the search contents, queries and users.

II. NEED FOR PERSONALIZATION

The results generated by the search engines based on the query of different individual user are general and not adaptable. For a particular query given in the search engine displays different results for the different users. Search results are organized for every users based on one's interest, preferences and information needs. Need for personalization is necessary when there is insufficient use of data is available in order to extract the patterns or when the website's content changes and new pages are added, but they are not included in the web logs. The user's visit usually aim at finding relevant information related to a particular subject, thus the underlying content semantics should be dominant factor in the process of Web Personalization[4].

Web Personalization Architecture



Figure 1: Web Personalization System Process



III. WEB SEARCH PERSONALIZATION APPROACHES

When it is applied to search, Web Search Personalization aims to find a subset of web data which matches the interest profile of a user or a group of users. It can be achieved by filtering web pages that are interested to the users.

Web Search Personalization Systems, it uses gathered information about the user from the profiles and cookies to conduct and also it revises the search to maximize the user's satisfactions. These user profiles are created based on the user's interest, preferences and the information needs. There are two different ways to generate the user profiles: explicit and implicit user profiling.

Implicit User Profiling: The user profile is being created from users past behaviour such as determining the documents they do select for viewing and the duration of time spent on viewing a document or a page surfing or scrolling actions. In the Explicit User Profiling: The users create their profiles manually to provide some kind of information or feedback to a search system [3].

Web Search Personalization can be done on either client or server side. But most of the problems arise at web on personalization at the server side, because server should keep track of history for each and every user. Server has to maintain the search history of different users when a user submits ambiguous query. It makes the performance of the server go down when many users submits the query at a time. So therefore most of the techniques employed in client side approach as all the histories maintained and also the query as well. This makes client side system making the faster way to access the user profile [5].

IV. WEB SEARCH PERSONALIZATION METHODS

Search engines include various personalized search strategies based on the following methods,

- User Groups
- Hyper Link Structure of the Web
- Content Analysis

A. Personalized Search Based on User Group

In this approach we take users with similar behaviour and interests. Users with similar interests with the rest of users out there are used to refine the search results [5]. So users of such kind are only responsible to provide the information which is necessary to form the user profiles. Personalized web search is done with these similarities between the users in the user group.

B. Personalized Search Based on Hyperlink Analysis
Hyperlink in web is also an important for personalization.

Thus rank algorithms are used in the web search by using personalized page rank which is to provide personalization web search. The important pages are linked to or by many important pages on the laid emphasis of Page Rank [5]. Page Rank is a modified version proposed by Personalized Page Rank to personalize web search by page. In Personalized Page Rank it is used to re-rank the search results during personalization. Different Algorithms of ranging the web pages are HITS, Sum Rank, Weighted Page Rank and Page Rank are discussed [2].

C. Personalized Search based on the Content Analysis

As the user profiles are created by the user themselves or they can be learnt using the user's past activities by implicitly. Where the user is not always ready to give their choices explicitly, so most of the work focus automatically on collecting the information from the history.

Here in the approach user profile can be built in two ways:

- Topical Categories
- Keywords

Topical Categories in user profile is framed as a hierarchy of topics or concepts.

Keywords in user profile are a list of keywords which is used to show the user preferences [5].

V. METHODS USED IN WEB SEARCH PERSONLIZATION

There have been implemented different types of methods in web search personalization. These methods have been helped to understand the user's behaviour. Such as Personalized Click Model trough Collaborative Filtering, Web Search Personalization using Fuzzy Adaptive Approach.

A. Personalized Click Model trough Collaborative Filtering

Click Model is used in the search engine, where it predicts the user's clicking behaviour. These search click data are stored in click-through logs. This Click-through log learns the user preferences for the search results. This approach was proposed by Shen et .al. The documents or the pages showed at the top on the web pages will be accessed by many numbers of user's than the pages below and having more relevance. By putting forward Personalized Click Model it employs a collaborative filtering method in it. Two such models are Matrix Factorization Click Model and Tensor Factorization Click Model.

1. Matrix Factorization Click Model

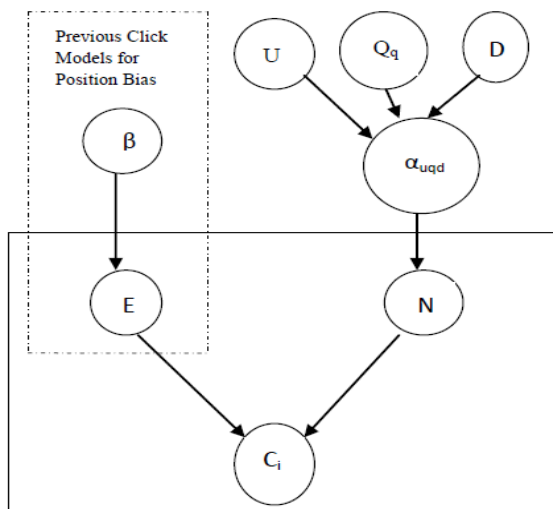
In the previous models individual's queries and documents were not thoroughly considered. Matrix Factorization Click Model is proposed to focus on the query and documents interactions through their latent feature vector [6].



Consider a query q which is submitted in a session and N documents which are fetched, the i^{th} document to being d_i . If there are in total M_q queries and M_d documents then let $Q \subset \mathbb{R}^{F \times M_q}$, where the set consists of all possible real-valued matrices with size $F \times M_q$, and $D \subset \mathbb{R}^{F \times M_d}$ represents the latent factors of queries and documents, respectively. F is the number of factors. Let $N(\mu, \sigma^2)$ be the probability density functions of Gaussian distribution with mean μ and variance σ^2 .

2. TENSOR Factorization Click Model

In this click model it introduces Personalized Click Model which includes personalization in our click model to extend the Matrix Factorization Click Model to use in the user domain. Consider there are M_u users. Let $U \subset \mathbb{R}^{F \times M_u}$ denote the latent factors of user domain. The event of user being personally interested in i^{th} document is indicated by N_i . Let α_{uqdi} denote the probability of event N_i [7].



In figure 1 it gives a graphic representation of Personalized Click Model. It shows the connections of parameters in the graph model. As the user examination various assumptions of previous click events can be made, consider the event of click made at the position i which is dependent on the user examination and personal document interest is modelled as based on the latent factors of users, queries and documents themselves.

B. Web Search Personalization using Fuzzy Adaptive Approach

As we know data in the web is growing rapidly in size. The retrieval of relevant information from the web is a big task. This makes a challenge to the web search engines. Search results give number of information to the users which are not relevant too. Fuzzy Logic is a tool which is used in personalization of web search. Fuzzy logic selects the input given by the user in the query and clusters the similar information. The selected cluster which is similar is used to recommend the Fuzzy ranked set of documents

[8]. A Fuzzy variable is given to clarify the relevance of each document and then calculate the relevance of each document the abstract terms are been used. Each abstract has a common term with query vector, these common terms are counted. The number of common terms used in the query vector is calculated by the values of Fuzzy Variables [9].

VI. CONCLUSION

As the tremendous growth in the data in web, WWW is growing rapidly. This makes the user very difficult to retrieve the relevant information from the web. Personalization of web search makes the user to extract the relevant information. Here in this paper, we have focused on a survey of Personalization of web search. Firstly we have discussed about the need of Personalization, Secondly on the approaches of Personalization of web search and later on with the methods of Personalized Web Search.

REFERENCES

- [1] Eirinaki, M., Mavroudis, D., Tsatsaronis, G. and Vazirgiannis, M. 2006. Introduction to Semantics in Web Personalization: the Role of Ontologies. EWMF/KDO5.
- [2] Charanjit Singh, S.K. Kautish "Page Ranking Algorithm for Web Mining:A Review" Published in International Conference on Advancements in Engineering and Technology(ICAET 2015).
- [3] Pooja Mehtaa, Brinda Parekh, Kirit Modi, and Paresh Solanki : Web Personalization Using Web Mining: Concept and Research Issue. International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012 .
- [4] Himani Arya , Jaytrilok Choudhary, Deepak Singh Tomar : A Survey on Techniques for Personalization of Web Search. International Journal of Computer Applications (0975 – 8887) Volume 94 – No. 18, May 2014.
- [5] K. Sridevi and Dr. R. Umarani : Web Personalization Approaches: A Survey. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, March 2013.FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
- [6] Mohammad. S Norouzzadeh, Ayoub Bagheri, Mohammad. H Saraae. 2009. Web Search Personalization: A Fuzzy Adaptive Approach. IEEE.
- [7] Si Shen, Botao Hu, Weizhu Chen, Qiang Yang, "Personalized Click Model through Collaborative Filtering".
- [8] Mariya Savova Evtimova and Ivan Momtchilov Momtchev: Personalize Web Searching Strategies Classification and Comparison. Journal of Communication and Computer 13.
- [9] Dr. Suruchi Chawla: Effective Personalization of web search based on Fuzzy Information Retrieval. Suruchi Chawla / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2831-2837.