Privacy Based Personalized Web Search Engine

R.Monika¹, V.Pavithra², D.PriyaDharshini³, N.Vaishnavi⁴, S.Gowthami⁵

Computer Science and Engineering, Anna University, Chennai, India / Saranathan College of Engineering

ABSTRACT:

With the growth of (WWW) World Wide Web, Search engines have more contribution in giving information from the web to the user. They help in searching information on the web robust and easy. But there is still room for improvement. Existing web search engines do not consider particular needs of user and serve each user equally. For this ambiguous query, a number of documents on distinct topics are returned by search engines. Hence it becomes difficult for the user to get the specified content. Moreover it also takes more time in searching a pertinent content. Personalized Web Search Engine is considered as a promising solution to handle these problems, since different search results can be provided depending upon the choice and information needs of users. It exploits user information and search context to learning in which sense a query refer. In order to perform Personalized Web search it is important to model User's interest. User profiles are constructed to model user's need based on his/her web usage data. This Enhanced User Profile will help the user to retrieve concentrated information. This paper proposes architecture for constructing user profile and enhances the user profile using background knowledge. It can be used for suggesting good web pages to the user based on his search query and background knowledge.

Keywords: search engine, personalized web search, background knowledge.

1. INTRODUCTION:

Searching is one of the common task performed on the Internet. Search engines are the basic tool of the internet, from where one can collect related information and searched according to the specified keyword given by the user. The information on the web is growing dramatically. The user has to spend more time in the web in order to find the particular information they are interested in.

Now, Traditional search engines produce some irrelevant information to the user. When they search a particular keyword, it gives some unrelated information. In order to solve the problem we come up with a new search engine called “PERSONALISED WEB SEARCH ENGINE”.

This search engine has demonstrated its effectiveness in improving the quality of various search services of the internet. Personalized web search is a promising way to improve the search quality by customizing search results for people with individual information goals.

However evidences show that users are not willing to disclose their private information during search that has become a major barrier for the wide use of personalized web search. Users are uncomfortable with exposing private preference information to search engines. Privacy is not absolute and it can be compromised that if there is a gain in service or profitability to the user.

Thus we need a balance that struck between search quality and privacy protection. This paper describes about the privacy protection in personalized web search applications. Moreover the personalized web search framework called “Topic Modeling” that can adaptively generalize the profiles by queries while respecting user specified privacy requirements. Ultimately our paper aims at providing protection against a typical model of privacy attack.

2. EXISTING SYSTEM:

Our existing system (i.e) web search engine is the most important thing for ordinary people looking for useful information on the web. However, users generally experience failure and get improper results. The search engines return some irrelevant results that do not match with their real intentions. A typical search engine provides similar set of results without considering of who submitted the query. This existing system was implemented by three main techniques.

I. SLDA:

SLDA (Supervised Latent Dirichlet Allocation) approach is used to display all the links. That means when a particular keyword is given by the user. It will display all the respective links associated with that keyword. This kind of
approach is also known as “BAG OF WORDS”. It means selecting a particular word from various topics

II. BERNOLLI APPROACH:

The above SLDA approach is possible by algorithm called “BERNOLLI THEOREM”. Bernoulli theorem is used to get the links for the given keyword. The particular keyword is selected from all the links and the related links will be displayed for the keyword.

3. PROPOSED SYSTEM:

   OVERVIEW:

Personalized web search is a general category of search techniques which aims to provide better search results, according to individual user needs. So, for this user information has to be collected and analyzed so that the perfect search results required for the user behind the issued query is to be given to the user. The solution to this is Personalized Web Search (PWS). It can generally be categorized into two types, first is click-log-based methods and second is profile-based ones. The click-log based methods are simple and straightforward: This method performs the search based upon clicked pages in the user’s query history.

To protect user privacy in profile-based PWS, admin have to consider two important and contradicting issues during the search process. The first issue is that, they attempt to improve the search quality with the personalization utility of the user profile. On the other hand the second issue is, they need to hide the privacy contents existing in the user profile to place the privacy risk under control. Sometimes people are willing to compromise privacy if the personalization by supplying user profile to the search engine yields better search quality.

In an identical situation, significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) portion of the user profile, namely a generalized profile. Thus, user privacy can be protected without compromising the personalized search quality. In general, there is a compromise between the search quality and the level of privacy protection achieved from generalization. To prevent the information loss while performing runtime generalization, a greedy algorithm is used here. This work can be enhanced for complex query also.

II. SSLDA ALGORITHM:

This system feature propose selective supervised Latent Dirichlet Allocation (ssLDA), where the discrimination power of words specified to topics (the topic-specific discrimination power of words) is taken as a model parameter to be estimated and each word in a given document is selected as strongly or weakly discriminative one according to a per-word binary indicator variable (the value 1 or 0 corresponds to strongly or weakly discriminative), which is referred as “probabilistic word selection”.

Normally, SSLDA Algorithm is categorized into General and Personalised.

General is nothing but the SLDA Algorithm. The user is giving a keyword and the related links for the particular keyword is displayed. Moreover, simply query is taken into our database.

But in Personalised Web Search, Query along with history of that user is taken into our database. If the user want to search anything, he/she must login into our site by registering before itself. Once the user is sign into our site, the particular users downloads, browsing details and bookmarks will be recollected by database. Again the user want to search the same keyword, then the users expected results will be displayed by the browser.

III. GLM APPROACH:

GLM – GAUSSIAN LINEAR MODEL

If the user is new user (i.e) first time the user is registering and login into our website. He/she search a particular keyword at first, the respective link will be stored in the database. Then the user searches a same keyword second time, the user predicted results will be displayed.

In case the user want to search a new keyword, then the keyword searched by the user will be saved in database and the respective links will be displayed.

The Predicted results (i.e) user expected results will be displayed by the approach called “GAUSSIAN LINEAR MODEL”. Instead of Bernoulli approach in existing
system, we are using GLM approach. In GLM approach, the related links are displayed for given keyword with the help of “BAG OF TOPICS”.

IV. GREEDY INFORMATION LOSS ALGORITHM:

The Greedy IL algorithm improves the efficiency of the generalization using heuristics based on several findings. One important finding is that any prune-leaf operation reduces the discriminating power of the profile. Greedy IL algorithm is used to provide pruning approach to eliminate the user details from server histories.

The most important feature in our paper is “PRIVACY PROTECTION”, providing security to the user who logged into our browser. We are providing the privacy for the user profiles, because users are logged in by using their personnel details. Privacy is provided by what is called “GREEDY INFORMATION LOSS ALGORITHM”. Greedy IL algorithm is used to prune the leaf at top level. It will hide the details of the user by pruning and added to that it provides the related links quickly by using shortest path.

In our browser, user id along with the browsed links are stored in the server. But this user id can be hacked by user and hijacking is also possible. So in our Personalised Web Search, privacy is giving to user id by hiding the value of user id (i.e) user id is hided in the server. It means admin alone is allowable to see the user id. In addition to this, greedy information loss analyze the history for giving results.

V. ACCURACY RATE:

Classification Accuracy is calculated for our system, in order to know about the true result and false result. True result is nothing but the number of results matches with the given keyword. And false result is nothing but the number of results that does not match.

Accuracy rate is equal to the ratio of successful results to the total number of successful results.

VI. DATABASE:

In our paper we maintain local database for storing the user details and browsed information when user search the query for first time results will be displayed by fetching from google server and this will be stored in our local database. Again if user searches for same keyword, results will be taken from local database and then it displays the preferred links by analyzing the history of that user. The common results are fetched from the google server by using the query called “googlequery.java”.

4. CONCLUSION:

In this paper we provide a security to the user browsed details in our browser. This is provided by using SSLDA (Selective Supervised Latent Dirichlet Allocation) approach which is the extension of SLDA (Supervised Latent Dirichlet Allocation) approach. we also use “Bag Of Terms” concept using LDA and GLM approach to extract the relevant and exact results for query given. In this module we proposed a new profile generalization algorithm called greedyIL. Improves the efficiency of the generalization using heuristics based on several findings.

GreedyIL algorithm is used to provide pruning approach to eliminate the user details from server histories. we also maintain local database to store the user details in order to provide privacy to the user details and browsed informations.

5. REFERENCES: