Genetic Algorithm for Privacy Protected Personalized Web Search

Shrikant A. Shinde[1], Abhilasha V. Biradar[2]
[1]*Department of Computer Engg, SPPU, Pune, India*

[2]*Department of Information Technology, SPPU, Pune, India*

Abstract— The search engine becomes the most important gateway for ordinary people who are looking for useful information on the web. In spite of, users might sense failure when search engines return incorrect results that do not meet their real objectives. Such inappropriateness is due in large part to the variety of users' contexts and backgrounds, as well as the uncertainty of texts. Personalized web search (PWS) is a type of search techniques which aims at providing better search results, which restricts to individual user needs. In this paper, UPS (User customizable Privacy- preserving Search) framework is proposed which generalize user profiles for each query given to user-specified privacy specification. This generalization has aims of keeping a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the user generalized profile. When considering bi-objectives problem through greedy search, it does not adequately handle the objectives dynamically or adaptively. So, to handle these goals genetic algorithm developed in this paper. The derived results will be compared based on precision and time.

Keywords- Personalized web search, user profile, utility, risk, genetic algorithm.

1. INTRODUCTION

Searching is one of the standard technique to find out the information from the internet with the help of the Web search engine (WSE). It is necessary for every retailer who is providing services to provide coherent and easy to use of information. The user gets a variety of related information for their queries. To render more proper and useful results to the user, Personalization technique used. Web search engines make finding information on the Internet quick and easy. For a given query, a personalized Web search (PWS) can provide different search results for various users or organize search results differently for each user, based upon their interests, preferences, and information needs. Personalized web search varies from generic web search, which returns identical research results to all users for identical queries, regardless of varied user interests and information needs. For example, In Google's beta version, for instance, users are asked to select the categories of topics which they are interested in, and the search engine applies this information during the retrieval process.

The solutions to PWS can be characterized into two types, namely click-log-based methods and profile-based ones. The click-log based methods merely impose bias to clicked pages in the user's query history. It can only work on repeated inquiries from the same user, which is a substantial limitation confining its applicability. In contrast, profile-based approaches improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be potentially useful for almost all sorts of queries. During the search process, two contradicting effects [1] It analyze, improve the search attribute with the personalization utility of the user profile and the necessary to hide the privacy contents being in the user profile to place the privacy risk under control.

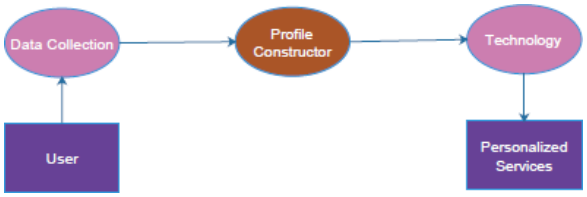

Fig1. Overview of Profile based personalization

As shown in Figure 1, the user profiling process consists of three main phases. First, an information collection process is used to gather raw information about the user. The second stage focuses on user profile construction from the user data. The final stage, in which technology or application exploits information in the user profile to provide personalized services.

Besides the personalized results, security is required in the personalized web search. Users are not interested in exposing their information during a web search. It becomes a significant concern in profiling the user in personalized web search. There should be a technique that recognizes profiles according to information was given by the user. If the search engine knows more about the user, more concrete results will be gained by search provider. Search engines can deliver accurate and exact data if users think search engine and provide more information. Hence, search engines should give security mechanism such that user will be assured of its privacy and its information should be kept secure. In personalized web search, user information is gathered and evaluated to discover the intention of issuing query fired by the user. The search completed by providing queries to retrieval system in the

form of a set of words. If remote users fire the same query, the system will yield the same results without considering the user. But search results should be produced by considering the user so that specific users can get various search results for the same query by keeping the record of user's personal information and interests.

The residue of this paper organized as follows: Section 2 discusses some related work. Section 3 presents implementation details of proposed system. Chapter 4 presents expected results. Section 5 concludes the paper and discusses future work.

2. RELATED WORK

Susan T. Dumais et al. [2] proposes a search algorithm that considers user's prior interactions with a wide variation of content, to personalize their current web search. Rather than relying on the unrealistic presumption that people will precisely specify their intent when searching, it seeks techniques that leverage accurate information about the user's interests. This information is used to re-rank web search results within a relevance feedback framework. It explores expensive models of user interests, built from both search-related information such as previously issued queries and earlier visited web pages and other information about the user such as documents and email the user has read and created. The research suggests that rich illustrations of the user and the corpus are necessary for personalization but that it is possible to approximate these representations.

Y. Xu, K. Wang, G. Yang introduced the notion of online anonymity [3] to enable users to issue personalized queries to an un-trusted web service while with their anonymity preserved. The challenge for providing online anonymity is administering with anonymous and dynamic web users who can get online and offline at any time. Proposes the notion of online anonymity to ensure that each query entry in the query log cannot link to its sender and an algorithm that achieves online anonymity through the user pool is proposed. This approach can be continued to deal with personally identifying information that may contain in the query. The method is also suitable for general web services where there is a need to anonymize the query, with or without personalization.

K.W.T. Leung et.al. [4] have proposed a web search personalization method that captures the user's interests and preferences in the form of concepts by mining search results and their clickthroughs. They depart concepts into content concepts and location ideas, organize them into ontology for the creation of an ontology-based, multi-facet (OMF) profile to precisely capture the user's content and place interests. The Experimental results prove that OMF improves the precision significantly as compared to the baseline.

B. Smyth [5] introduced a community-based approach to provide efficient personalizing Web search. At the community level, knowledge reflects within search communities by collecting user search query and result chosen by the user. All gathered data is used to prepare a significance model that gives promotive community-related results for all web search. Collaborative web search approach is used here that recommend valuable and shareable knowledge.

P. Palleti et al., [7] By using probabilistic query expansion author developed personalized web search. In this approach, the authors developed a personalized Web search system applied at proxy which changes to user interests perfectly by generating user profile with the use of collaborative filtering. A user profile primarily consists of probabilistic correlations among query terms and document terms which utilized for providing personalized search results. Experimental outcomes prove that this proposed personalized Web search system is very effective and efficient.

In this paper [6] the author studied the existing generalization methods are insufficient because they cannot provide assurance privacy protection in all cases, and frequently acquire redundant information loss by performing too much generalization. In this paper, the author suggests the idea of personalized secrecy and develops a new generalization structure that takes into account customized privacy necessities. A person can specify the degree of privacy protection for her/his sensitive values by specifying "guarding nodes" in the taxonomy of the sensitive attribute. The drawback of this paper is, the greedy algorithm presented in this document was not optimal and also did not support runtime profiling.

Xu et al. in [8] proposed a privacy protection solution for PWS based on hierarchical profiles. Using a user-specified threshold, a generalized pattern achieved in effect as a rooted subtree of the complete profile. These profiles review a user's interests into a hierarchical organization according to particular interests. Two parameters for specifying privacy conditions are proposed to help the user to choose the content and degree of detail of the profile information that exposed to the search engine. The main drawback of this strategy was that this work does not address the query utility, which is crucial for the service quality of PWS.

In [9], the prototype of UPS is proposed, together with a greedy algorithm GreedyDP to support online profiling based on predictive metrics of personalization utility and privacy risk. The limitation of this paper, it requires more computational cost & recomputation of all queries.

3. PROPOSED SYSTEM

The search result of different search engines depends only on the text entered as query and not depends on user interest. Therefore, sometimes the user may get fail to retrieve the exact result. However, some questions are ambiguous in nature, and when it is submitted to search engines, it returns result regardless of who submitted the question for what requirement. So, search results have to rearrange according to user's intention which can get by user's general information; we called as a user profile.

A personalized web search gives different search results for different users based upon their interests, preferences, and information needs. The privacy issues in personalized web search are the main drawback. The proposed UPS will give the generalized profiles for each query according to user-specified privacy needs.

This framework assumes that questions do not contain any sensitive information. A Trade of between search quality and level of privacy protection achieved from generalization. Generalization algorithm is used namely GeneticIL to find out and utilization of user profile and improving performance.

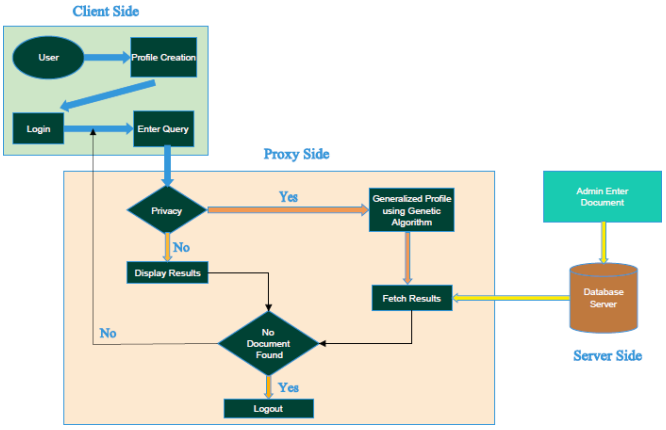

Fig2. System architecture

As shown in figure 2, the user fires a query 'q' through a proxy refer as an on-line profiler to the server. Then generalized profile is created by a proxy, and both generalized profile and query passed to the server. The server gives response 'r' back to the proxy, then it decides either to re-ranked the search or provide as it results to the client as per the query.

3.1 Proposed Modules

1. User Login

   End user gives query to the search engine. Initially, the user has to login to the search engine. Users have to provide username and password for login. If the user is new, then registration should be done. After successful login, the user will go to Personalized Web Search (PWS) System.

2. Profile Construction

   In this module, the user profile is created by collecting information about a user (e.g., Name, age, demographic data, interest, etc.). When the user enters a query, depending upon the refer to the attributes of the user profile are taken and shown in the hierarchical form.

3. Generalization of user profile

   This procedure generalizes profile in an iterative manner relying on privacy. Also, this procedure computes the discriminating power of online decision on whether personalization should be employed.

4. Personalized Results

   After generalization of a user profile, a generalized profile, and a query will be given to server by proxy for further processing. In this module, server will give personalized search results based on generalized profile and query

3.2 Pseudocode

Input: Seed profile G_0; query q; privacy threshold δ

Output: Generalized profile G_* satisfying δ-Risk

// Let Q be the IL priority queue of prune-leaf decisions; i be the iteration index, initialized to 0.

if $DP(q,R) < \mu$, then
obtain the seed profile G_0

insert $\langle t, IL(t) \rangle$ in to q for all $t \in T_H(q)$,

while $risk(q,G_i) > \delta$ do

Apply genetic cross over operator
Apply genetic mutation operator

Find the prune-leaf $G_i \rightarrow G_{i+1}$
if t has no siblings, then
insert $\langle s, IL(s) \rangle$ to Q

else if t has siblings then
merge t into shadow-sibling;
if no operations on t's siblings in Q then
insert $\langle s, IL(s) \rangle$ to Q
else
update the IL values for all operations on t's siblings in Q
update $i \leftarrow i + 1$

return G_i as G_*;
return $root(R)$ as G_*;

5. EXPECTED RESULT

Currently, search engines are trying to do a personalized web search. But these search engines uses a greedy algorithm for the generalization of a profile which requires more computation time. In comparison with

greedy, the genetic algorithm requires less amount of time. Based on this comparison we can expect our result as shown in figure 3.

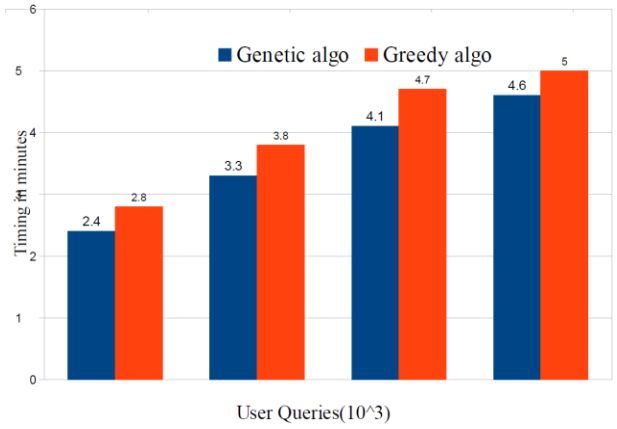

Fig3. Comparison graph

The main limitation of proposed system is it depends on the proxy server. If a proxy server fails, the whole system fails down.

6. CONCLUSION AND FUTURE SCOPE

A client-side privacy protection framework called UPS for personalized web search is presented. UPS could be adopted by any PWS that captures user profiles in a hierarchical manner. The structure allowed users to specify customized privacy requirements via the hierarchical profiles. Proposed system improves web search by placing the privacy under control and also improves the time required for generalization of the user profile by using a Genetic Algorithm. In future, this proposed system can be used for implementation of recommendation system.

REFERENCES

[1] Lidan Shou, He Bai, Ke Chen, and Gang Chen "Supporting privacy protection in personalized web search" IEEE Transactions on Knowledge and data engineering, Vol. 26, No. 2, February 2014.

[2] J. Teevan, S.T. Dumais, and E. Horvitz, “Personalizing Search via Automated Analysis of Interests and Activities,” Proc. 28th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449- 456, 2005.

[3] Y. Xu, K. Wang, G. Yang, and A.W.C Fu, “Online anonymity for personalized web services” Proc.18th ACM conformation and knowledge management (CIKM), pp 1497-1500, 2009.

[4]] K.W.T. Leung, D.L. Lee and Wang-Chien Lee, "Personalized Web search with location preferences", IEEE 26th International Conference on Data Engineering, pp. 70 I - 712, 2010.

[5] B. Smyth, “A Community-Based Approach to Personalizing Web Search,” IEEE Journals, Computer, Vol. 40, No. 8, Pp. 42 –50, 2007.

[6] P. Palleti, H. Karnick and P. Mitra, "Personalized Web Search Using Probabilistic Query Expansion," International Conferences on Web Intelligence and Intelligent Agent Technology Workshops (IEEE/WIC/ACM), Pp. 83 – 86, 2007.

[7] X. Xiao and Y. Tao, “Personalized Privacy Preservation,” Proc.ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2006.

[8] Y. Xu, K. Wang, B. Zhang, and Z. Chen, “Privacy-Enhancing Personalized Web Search,” Proc. 16th Int’l Conf. World Wide Web (WWW), pp. 591-600, 2007.

[9] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, “Ups: Efficient Privacy Protection in Personalized Web Search,” Proc. 34th Int’l ACM SIGIR Conf. Research and Development in Information, pp. 615-624, 2011.