A Review on Privacy Preservation in Personalized Web Search

Najneen Tamboli, Sathish Kumar Penchala

M.E Student, Dept. C.E., Dr. D. Y. Patil School of Engineering, Savitribai Phule Pune University, Pune, India

Assistant Professor, Dept. C.E., Dr. D. Y. Patil School of Engineering, Savitribai Phule Pune University, Pune, India

ABSTRACT: World Wide Web (WWW) expanding its information very fast. Search engines (e.g. Google, Yahoo, etc.) is capable to retrieve the information according to user's need. Users are always curious in getting the accurate search result. The role of search engine is to serve all users without depending the unique needs of any individual user. One of the best methods for users is Personalized Web Search (PWS) which helps to search the required result, according to the users demand by supplying the personal data of the user to the search provider. In this paper PWS application model user preference as stratified user profiles is proposed using a PWS framework called UPS. It generalizes the profiles also parallel and manage   the privacy requirement which is stated by user. In this paper two algorithms are stated, namely GreedyDP and GreedyIL for runtime association. The online prediction mechanism is also stated to distinguish whether the provided query is personalized or not.

KEYWORDS: Privacy preservation, personalized web search, GreedyIL, GreedyDP

I. INTRODUCTION

In web life web search engines play a very vital role. The web is made up of 60 trillion individual pages and it is constantly expanding, to find documents of our need we follow links from page to page. People are becoming more dependent on the web for the needed information[7]. There are many reasons for people for searching daily needs as well as getting different articles, web search engine sort tremendous amount of data and send response to the users. As information is growing rapidly on the web, users may get relevant or irrelevant results i.e., queries submitted by various users with various requirement may get same results. This irrelevant results is mainly because of different variety of users submitted the same query. Let's consider, for a query "orange" one user want to get the information about benefits of orange fruit while other user may interested in getting the information about orange color by submitting the same query. So universal search engines are not able to discriminate such type of cases and return the same results to all users. To overcome these limitations it is important to personalize the web search.

Personalized search refers to search participation that are altered specifically to an individual's concern by incorporating knowledge about the individual beyond particular query provided. There are two general approaches to personalizing search results, one involving modifying the query of the user and the other re-ranking search results. Personalization is the process of determining the largest set of possible choices[10]. Personalized web search (PWS) alters the search experience specifically to fulfill the user interest by consolidating the information about the individual beyond the specific query. The explanations to PWS can be divided into two categories: click-log-based methods and profile-based methods. In this paper a framework called UPS (User customizable Privacy-preserving Search) which gives the search result by generating the information of user's requirement. An online prediction mechanism is also provided to differentiate whether to personalize the query or not.

II. LITERATURE REVIEW

In[1] G. Chen, H. Bai, L. Shou, personalized web search (PWS) has demonstrated the efficiency in the development of quality of search service. In this paper privacy protection in PWS engines capture personalities of user profiles. They propose a PWS framework called UPS which can create profiles for each query according to user's privacy requirements. This paper also provides an online prediction mechanism which is based on query utility for thinking whether the query is personalize in UPS or not. Expanded experiments exposes the efficiency and effectiveness of the

framework.In[2] A. Viejo and J. Castell a-Roca, a new scheme designed to preserve the protection of the clients from a web crawler which tries to profile them. A distorted user profile is provided by using social networks to the web search engine. The recommended convention submits standard inquiries to the web crawler; therefore it doesn't require any adjustment in the server side. Protocol enhances the current arrangements regarding query delay. In addition, the contorted profiles still permit the clients to get a proper administration from the web search engines.In[3] Y. Zhu, L. Xiong, and C. Verdery a novel bundling technique that chunks client profiles into gatherings by considering the semantic connections between the terms while refreshing the privacy restraint. In this paper, a Bayes-optimal privacy notion for profiles of users are represented as set-valued data. In[4] A. Krause and E. Horvitz they propose an economics of privacy in personalization. User preferences are separately assessed about privacy and utility. The important level of personalization can be conclude using user's information.In[5] Y. Xu, K. Wang, G. Yang an online anonymity approach to enable users to subject personalized queries to an entrusted web services. The threat for providing online anonymity is dealing with unfamiliar and random users of web those who get online and offline at any moment. They also suggested an approach to keep online anonymity by time. This approach uses a third party which is called as user pool and there is no need of user pool to be trusted.

III. PRIVACY PRESERVATION IN PWS

Two classes are there for a solution to the privacy preservation in PWS. One class consist of that work, treat privacy as the recognization of an individual. The other class consist of the data sensitivity, specifically the profiles of users disclosed to PWS server.

A. *Individual Recognization:*

Classic works in the preserving user's identification literature (class one) we try to determine the privacy problem on different level. It consist of the group identity, no identity, pseudo integrity and no personal information. It is proved that the solution to the first level is weak[8]. Because of high cost in communication and cryptography the level third and level forth are impractical. So the existing efforts focus on the level second.

The class two solutions do not require collaborations or third party assistance between social network entries. In these explanations, users only trust themselves and do not allow the disclosure of their complete profiles to an uncertaintyserver[7].

B. *Online Anonymity:*

Online anonymity based on user profiles by creating a group profile of users. This approach is used, to break down the link between the query and a single user.

C. *Useless User Profile(UUP):*

This approach is recommended to mix the queries among the user's group who issue them. So that certain individual cannot profiled by any entity. These works assume the existence of an authentic third-party an anonymizer.

D. *Legacy Social Network:*

In this approach every user perform as a search agency as his nearby companion instead of a third party which supply a bended client profile to the web search engine.

Data Sensitivity:

The class two solutions do not require collaborations or third party assistance between social network entries. In these explanations, users only trust themselves and do not allow the disclosure of their complete profiles to an uncertaintyserver[9].

- *Statistical Technique*:

    This approach states that a probabilistic model, and then to generate the near-optimal partial profile by using this model. The drawback of this work is that the user profiles are created as a limited arrangement of characteristics, and the probabilistic model is prepared by predefined frequent queries. These predictions are impractical in the context of PWS.
- *Generalized Profiles*:

This approach recommends a solution for PWS based on hierarchical profiles. A generalized profiles are obtained in effect as a rooted sub tree of the complete profile by using a user specified threshold.

IV. ALGORITHMS

Greedy Algorithms:

A greedyalgorithm is an algorithm which follows the solving approach that making the optimal choice locally at each stage with the hope of finding a global optimum[9]. In many problems, a greedy strategy does not in general produce an excellent solution, but nevertheless a greedy heuristic may return locally optimal solutions that approximate a global excellent solution in a reasonable time.

- *GreedyDP Algorithm:*

Greedy DP supports online profiles based on foreboding metrics of personalization utility and privacy risk. Greedy DP algorithm follows in a bottom up manner approach. The main problem of Greedy DP is that it requires re-commendation of all candidate profiles created from the attempts of prune-leaf manner.

- *GreedyIL Algorithm:*

GreedyIL algorithm is called as a new profile generation algorithm. The efficiency of the generalization is improved by using Greedy IL algorithm which is based on various conclusions. One significant result is that any prune-leaf operation diminishes the discriminating power of the profile.

V. CONCLUSION

 Personalized search is a bright way to enhance the search quality. This survey paper represents the various approaches which helps to implement the privacy in personalized web search. The algorithm used are very helpful for finding the solutions for privacy preservation. The future work will help the researchers for promoting a suitable solution for personalized web search (PWS) technique.

REFERENCES

1. LidanShou, H. Bai, Ke Chen, and G. Chen, ''Supporting Privacy Protection in Personalized Web Search'', IEEE Transactions on Knowledge and Data Engineering, volume. 26, no. 2, Feb 2014.
2. L. Shou, G. Chen, Y. Gao, K. Chen, and H. Bai''Ups: Efficient Privacy Protection in Personalized Web Search'', Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information, pp. 615-624, 2011.
3. A. Viejo and J. Castell_a-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", Computer Networks, volume. 54, no. 9, pp. 1343-1357, 2010.
4. Y. Zhu, C. Verdery and L. Xiong, "Anonymizing User Profiles for Personalized Web Search",Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.
5. A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services", J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.
6. K. Wang, Y. Xu, A.W.-C. Fu and G. Yang,"Online Anonymity for Personalized Web Services",Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.
7. Greeshma A S., Lekshmy P. L, "Privacy Protection in Personalized Web Search- A Survey", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 6, Vol 2 (June 2015).
8. PriyankaDeulkar, Dr.A.D.Gawande, "A Study on Securing Privacy In Personalized Web Search",International Journal Of Engineering And Computer Science ISSN:2319-7242 Vol 4 Issue 3 March 2015, Page No. 11100-11103.
9. B.Upender, BathulaRevathi, "Analysis on Supporting Privacy Protection in Personalized Web Search",International Journal of Research and Computational Technology, Vol.7 Issue.2.
10. KhwajaAamer, Dr. A. S. Hiwale, "A Survey on Privacy Protection in Personalized Web Search", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.