SUPPORTING PRIVACY PROTECTION IN PERSONALIZED WEB SEARCH- A REVIEW
Neha Dewangan 1, Rugraj 2

1 PG student, Department of Computer Engineering, Alard College of Engineering and Management
2 Department of Computer Engineering, Alard College of Engineering and Management,
   Savitri Bai Phule, Pune University, Pune

ABSTRACT:
Since the content in Internet is growing rapidly, the search provider users demand accurate search result as per their need. One of the options available to users is Personalized web search which presents search result based on the personal data of user provided to the search provider. However, users’ unwillingness to share their private information during search has become the major barrier for personalized web search. This paper models preference of users as hierarchical user profiles. It proposes a framework called UPS which generalizes profile at the same time maintaining privacy requirement specified by user. Two greedy algorithms namely GreedyDP and GreedyIL are used for runtime generalization. Also, an online prediction mechanism to decide whether to personalize a query or not is provided in this paper.

Keywords: Privacy Protection, profile, personalized web search, risk, UPS.

[1] INTRODUCTION

The web search engine has gained a lot of popularity and importance for users seeking information on the web. Since the contents available in web is very vast and ambiguous, users at times experience failure when an irrelevant result of user query is returned from the search engine. Therefore, in order to provide better search result a general category of search technique Personalized Web search is used. In personalized web search, user information is collected and analyzed in order to find intention behind issued query fired by user.

There are two categories of PWS, namely click-log-based and profile-based. The click-log based methods are straightforward— they simply impose bias to clicked pages in the user’s query history. This strategy has been performing well but it work on repeated queries from same user which is a strong limitation to its applicability. While profile-based methods improve the search experience generated from user profiling techniques. Profile-based methods can be potentially effective for almost all sorts of queries, but are reported to be unstable under some circumstances. There are both advantages and disadvantages for both type of PWS technique, profile based PWS is more effective for improving search result. The user profile is made from information gathered
from query history, browsing history, click-through data bookmarks, user documents and so forth. Unfortunately, such implicitly collected personal data can easily reveal a gamut of user’s private life.

[2] RELATED WORK

Previous works has focused on improving search result on profile-based PWS. Many representations for profile are available, some of them are term lists/vectors or bag of words to represent their profile while recent work create profile in hierarchical structure. The hierarchical representations are constructed with existing weighted topic hierarchy/graph, such as Wikipedia or the hierarchical profile is generated via term-frequency analysis on the user data. UPS framework can adopt any hierarchical representation.

Two classes of privacy protection problems for PWS is identified. One class treats privacy as identification of individual. Other considers data sensitivity as the privacy. Typical literature works in for class one try to solve the privacy problem on different levels, which includes the pseudoidentity, the group identity, no identity, and no personal information. The first level solution is proved to fragile and the third and fourth levels are impractical because of high cost in communication and cryptography. Therefore, the existing efforts focus on the second level. Online anonymity for PWS provide anonymity by generating a group profile of k users. Using this approach, the relation between the query and a single user is broken. The useless user profile (UUP) protocol shuffle queries among a group of users who issue them. As a result no entity can profile a certain individual. The shortcomings of class one solution is the high cost.

In Class two solutions, users only trust themselves and don’t tolerate the exposure of their complete profiles to anonymity server. Krause and Horvitz employ statistical techniques to learn a probabilistic model, and then use this model to generate the near-optimal partial profile. Privacy Enhancing personalized web search proposed a privacy protection solution for PWS based on hierarchical profiles. Using a user-specified threshold, a generalized profile is obtained in effect as a rooted subtree of the complete profile.

This paper provides personalized privacy protection in PWS. 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. Thus, this paper allows user to customize privacy requirements in hierarchical user profiles.

[3] EXISTING SYSTEM

The existing profile-based Personalized Web Search does not support runtime profiling. User profile is generalized only once offline, and used to personalize all queries from a same user. Such “one profile fits all” strategy has drawbacks for the variety of queries. Also, the existing profile-based personalization does not even help to improve the search quality for some ad hoc queries. The existing methods do not take into account the customization of privacy requirements. In existing system, all the sensitive topics are detected using an absolute metric called surprisal.
based on the information theory which assumes that the interests with less user document support are more sensitive. However, this assumption can be doubted with a simple example: If a user has a large number of documents about “sex,” the surprisal of this topic may lead to a conclusion that “sex” is very general and not sensitive, despite the truth which is opposite.

Iterative user interactions are required in many personalization techniques for creating personalized search results. Search results are refined with some metrics such as rank scoring, average rank, and so on. This is infeasible for runtime profiling, since it pose too much risk of privacy breach, and also require processing time for profiling. Therefore, we need predictive metrics to measure the search quality without iterative interaction of user.

**Disadvantage:**
1. All the sensitive topics are detected using an absolute metric called surprisal based on the information theory.
2. The existing profile-based PWS do not support runtime profiling.
3. The existing methods do not take into account the customization of privacy requirements.
4. Personalization techniques require iterative user interactions when creating personalized search results.

**[4] PROPOSED SYSTEM**

This paper proposes a privacy-preserving personalized web search framework called UPS i.e User customizable Privacy-preserving Search, that generalize profile for every query as per user privacy specification. Based on personalization and privacy risk metric, this paper formulate Risk Profile Generation, with its NP-hardness proved. It develops two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. GreedyDP maximize the discriminating power (DP) while GreedyIL minimize the information loss (IL). This paper also provide a mechanism for the client to decide whether or not to personalize a query in UPS. This decision is made before each runtime profiling to enhance the stability of the search results.

**Advantages:**
1. It enhances the stability of the search quality.
2. It avoids the unnecessary exposure of the user profile.
3. The framework allowed users to specify customized privacy requirements via the hierarchical profiles. In addition, UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality.
[5] METHODOLOGY

As shown in [Figure-1] UPS consists of number of clients/users and a server for fulfilling clients request. In clients machine, the online profiler is implemented as search proxy which maintains users profile in hierarchy of nodes and also maintain the user specified privacy requirement as a set of sensitive nodes. There are two phase, namely Offline and Online phase for the framework. During Offline, a hierarchical user profile is created and user specified privacy requirement is marked on it. The query fired by user is handled in the online phase as:

When user fires a query on the client, proxy generates user profile in run time. The output is generalized user profile considering the privacy requirements. Then, the query along with generalized profile of user is sent to PWS server for personalized web search. The search result is personalized and the response is sent back to query proxy. Finally, the proxy presents the raw result or reranks them with user profile.

[6] GREEDY ALGORITHM

A greedy algorithm is an algorithm that follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum. Greedy algorithm considers easy to implement and simple approach and decides next step that provide beneficial result. In many problems, a greedy strategy does not produce an optimal solution, but a greedy heuristic yields locally optimal solutions that approximate a global optimal solution in a reasonable time.

[6.1] GREEDYDP ALGORITHM

It works in a bottom up manner. Starting with the leaf node, for every iteration, it chooses leaf topic for pruning thus trying to maximize utility of output. During iteration a best profile-so-
far is maintained satisfying the Risk constraint. The iteration stops when the root topic is reached. The best profile-so-far is the final result. GreedyDp algorithms require recomputation of profiles which adds up to computational cost and memory requirement.

[6.2] GREEDYIL ALGORITHM

GreedyIL algorithm improves generalization efficiency. GreedyIL maintains priority queue for candidate prune leaf operator in descending order. This decreases the computational cost. GreedyIL states to terminate the iteration when Risk is satisfied or when there is a single leaf left. Since, there is less computational cost compared to GreedyDP, GreedyIL outperforms GreedyDP.

[7] CONCLUSION

A client side privacy protection framework called UPS i.e User customizable Privacy preserving Search is presented in the paper. Any PWS can adapt UPS for creating user profile in hierarchical taxonomy. UPS allows user to specify the privacy requirement and thus the personal information of user profile is kept private without compromising the search quality. UPS framework implements two greedy algorithms for this purpose, namely GreedyDP and GreedyIL.
REFERENCES


