ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Supporting Privacy Protection in Personalized Web Search

I.Katyani
M.Tech Student,
Dept of CSE,
KITS for Women's, Kodad, T.S, India.

Mr.K.Laxmaiah
Associate Professor,
Dept of CSE,
KITS for Women's, Kodad, T.S, India.

ABSTRACT :

Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting userspecified privacy requirements. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework. The experimental results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.

[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. Privacy issues rising from the lack of protection for such data, for instance the AOL query logs scandal, not only raise panic among individual users, but also dampen the data-publisher's enthusiasm in offering personalized service. In fact, privacy concerns have become the major barrier for wide proliferation of PWS services.

1.1 Motivations:

To protect user privacy in profile-based PWS, researchers have to consider two contradicting effects during the search process. On the one hand, they attempt to improve the search quality with the personalization utility of the user profile. On the other hand, they need to hide the privacy contents existing in the user profile to place the privacy risk under control.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A few previous studies [10], [12] suggest that people are willing to compromise privacy if the personalization by supplying user profile to the search engine yields better search quality. In an ideal case, 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 tradeoff between the search quality and the level of privacy protection achieved from generalization.

1.2 Contributions:

The above problems are addressed in our UPS (literally for User customizable Privacy-preserving Search) framework. The framework assumes that the queries do not contain any sensitive information, and aims at protecting the privacy in individual user profiles while retaining their usefulness for PWS. As illustrated in Fig. 1, UPS consists of a nontrusty search engine server and a number of clients. Each client (user) ccessing the search service trusts no one but himself/ herself. The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself. The proxy maintains both the complete user profile, in a hierarchy of nodes with semantics,and the user-specified (customized) privacy requirements represented as a set of sensitive-nodes.

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. GENERALIZATION TECHNIQUES:

In this section, we first introduce the two critical metrics for our generalization problem. Then, we present our method of online decision on personalization. Finally, we propose the generalization algorithms.

3.1 Metrics:

3.1.1 Metric of Utility:

The purpose of the utility metric is to predict the search quality (in revealing the user's intention) of the query q on a generalized profile G. The reason for not measuring the search quality directly is because search quality depends largely on the implementation of PWS search engine, which is hard to predict. In addition, it is too expensive to solicit user feedback on search results. Alternatively, we transform the utility prediction problem to the estimation of the discriminating power of a given query q on a profile G under the following assumption. Assumption 3. When a PWS search engine is given, the search quality is only determined by the discriminating power of the exposed query-profile pair hq; Gi.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

4.EXISTING SYSTEM :

The existing profile-based Personalized Web Search do not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Such "one profile fits all" strategy certainly has drawbacks given the variety of queries. One evidence reported in is that profile-based personalization may not even help to improve the search quality for some ad hoc queries, though exposing user profile to a server has put the user's privacy at risk. The existing methods do not take into account the customization of privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. For example, in, all the sensitive topics are detected using an absolute metric called surprisal based on the information theory, assuming that the interests with less user document support are more sensitive. However, this assumption can be doubted with a simple counterexample:

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. Unfortunately, few prior work can effectively address individual privacy needs during the generalization. Many personalization techniques require iterative user interactions when creating personalized search results. They usually refine the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on. This paradigm is, however, infeasible for runtime profiling, as it will not only pose too much risk of privacy breach, but also demand prohibitive processing time for profiling. Thus, we need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

Disadvantage:

All the sensitive topics are detected using an absolute metric called surprisal based on the information theory.

5.PROPOSED SYSTEM :

We propose a privacy-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified privacy requirements.

Relying on the definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as Risk Profile Generalization, with itsNP-hardness proved. We develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). By exploiting a number of heuristics, GreedyIL outperforms GreedyDP significantly. We provide an inexpensive mechanism for the client to decide whether to personalize a query in UPS. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile.

Advantages:

1. It enhances the stability of the search quality.
2. It avoids the unnecessary exposure of the user profile.

6.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.

7. 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.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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.

7.1 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.

7.2 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.

7.4 Result :

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements. 2. Select methods for presenting information. 3. Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.

Convey information about past activities, current status or projections of the Future. Signal important events, opportunities, problems, or warnings. Trigger an action. Confirm an action.

8 CONCLUSIONS:

This paper presented a client-side privacy protection framework called UPS for personalized web search. UPS could potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. 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. We proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution. For future work, we will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentiality, and so on), or capability to capture a series of queries (relaxing the second constraint of the adversary in Section 3.3) from the victim. We will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

9.ACKNOWLEDGMENTS:

I am I.Katyani and would like to thank the publishers, researchers for making their resources material available. I am greatly thankful to Associate Prof N.Lakshmi Priya for their guidance. We also thank the college authorities, PG coordinator and Principal for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

10.REFERENCES:

[1] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation andAnalysis of Personalized Search Strategies," Proc. Int'l Conf. WorldWide Web (WWW), pp. 581-590, 2007.

[2] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search viaAutomated Analysis of Interests and

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.

[3] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.

[4] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy," Proc. ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining (KDD), 2006.

[5] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW),2004.

[6] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search," Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM), 2005.

[7] X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," Proc. 28th Ann. Int'l ACMSIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.

[8] F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.

[9] J. Pitkow, H. Schu¨ tze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel, "Personalized Search," Comm. ACM, vol. 45, no. 9, pp. 50-55, 2002.

Author's Details:

Ms.I.Katyani. MTech student, in M.Tech Student, Dept of CSE in KITS for women's,kodad, T.S, India

Mr.k.Laxmaiah working as a Associate at CSE in KITS for women's,kodad, T.S, IndiaJNTUH Hyderabad. He has 2 years of UG/PG Teaching Experience