ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Supporting Privacy Protection in Personalized Web Search

Nitesh Chavan, Swapnil Khese, Eshwar Palve, Akshay Dongare, Asmita Mali

B.E Students, Dept. of IT, DYPIET, Pimpri, India

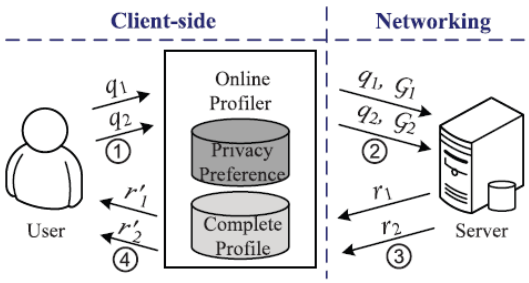
ABSTRACT: Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. The study of privacy protection in PWS applications that model user preferences as hierarchical user profiles. It propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user-specified 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. There are present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. It also provides 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.

KEYWORDS: Privacy, UPS, PWS, greedy Algorithm, Online profiling.

I. INTRODUCTION

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query. The solutions to PWS can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods are straightforward-they simply impose bias to clicked pages in the user's query history. Although this strategy has been demonstrated to perform consistently and considerably well, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. In contrast, profile-based methods improve the search experience with complicated user-interest models 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.

II. RELATED WORK



System Architecture of UPS

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

1.When a user issues a query qi on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile Gi satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.

2.Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search.
3.The search results are personalized with the profile and delivered back to the query proxy.
4. Finally, the proxy either presents the raw results to the user, or re-ranks them with the complete user profile.

UPS is distinguished from conventional PWS in that it:
1) Provides runtime profiling, which in effect optimizes the personalization utility while respecting user's privacy requirements;
2) Allows for customization of privacy needs; and
3)Does not require iterative user interaction.

III.PROPOSED SYSTEM

1.**PROPOSED SYSTEM** - It 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 its NP-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. The stability of the search quality.
2. To avoids the unnecessary exposure of the user profile.

IV.PSEUDO CODE

1.If t is a leaf node and t THðqÞ, its preferenceprefHðt; qÞ is set to the long-term user supportsupHðqÞ,3 which can be obtained directly from theuser profile.
2. If t is a leaf node and t 62 THðqÞ, prefHðt; qÞ ¼ 0.
3. Otherwise, t is not a leaf node. The preference valueof topic t is recursively aggregated from its childtopics asprefHðt; qÞ ¼
X
t02Cðt;HÞ
prefHðt0; HÞ:
Finally, it is easy to obtain the normalized preference for
each t 2 H as
Prðt j q; HÞ ¼
P prefHðt; qÞ
t02THðqÞ prefHðt0; qÞ

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Greedy Algorithm:
A greedy algorithms a mathematical process that recursively constructs a set of objects from the smallest possible constituent parts. Recursion is an approach to problem solving in which the solution to a particular problem depends on solutions to smaller instances of the same problem. Greedy algorithms look for simple, easy-to-implement solutions to complex, multi-step problems by deciding which next step will provide the most obvious benefit. Such algorithms are called greedy because while the optimal solution to each smaller instance will provide an immediate output, the algorithm doesn't consider the larger problem as a whole. Once a decision has been made, it is never reconsidered. The advantage to using a greedy algorithm is that solutions to smaller instances of the problem can be straightforward and easy to understand. The disadvantage is that it is entirely possible that the most optimal short-term solutions may lead to the worst long-term outcome.Greedy algorithms are often used in ad hoc mobile networking to efficiently route packets with the fewest number of hops and the shortest delay possible. They are also used in machine learning, business intelligence(BI),artificial intelligence(AI).

MODULES DESCRIPTION :

1. Profile-Based Personalization
2. Generalizing User Profile
3. Online Decision
4. Privacy Protection in PWS System

V.SIMULATION RESULT

In this section, we present the experimental results of UPS.We conduct four experiments on UPS. In the firstexperiment, we study the detailed results of the metricsin each iteration of the proposed algorithms. Second, welook at the effectiveness of the proposed query-topicmapping. Third, we study the scalability of the proposedalgorithms in terms of response time. In the fourthexperiment, we study the effectiveness of clarity predictionand the search quality of UPS.

5.1Experimental Setup

The UPS framework is implemented on a PC with aPentium Dual-Core 2.50-GHz CPU and 2-GB main memory,running Microsoft Windows XP. All the algorithms are implemented in Java.The topic repository uses the ODP web Directory. To focus on the pure English categories, we filter out taxonomies "Top/World" and "Top/Adult/World." The click logs are downloaded from the online AOL query log, which is the most recently published data we could find. The AOL query data contain over 20 million queries and 30 million clicks of 650k users over 3 months (March 1, 2006 to May 31, 2006). The data format of each record is as follows: huid; query; time½; rank; url_i; where the first three fields indicate user uid issued query attimestamp time, and the last two optional fields appear when the user further clicks the url ranked at position rank in the returned results.The profiles used in our experiment can be either synthetic or generated from real query logs: .Synthetic. We cluster all AOL queries by their DP into three groups using the 1-dimensional k-means algorithm. These three groups, namely Distinct Queries, Medium Queries, and Ambiguous Queries,can be specified according to the following empirical rules obtained by splitting the boundaries betweentwo neighboring clusters.Distinct Queries for DPðq; RÞ 2 ð0:82; 1Medium Queries for DPðq; RÞ 2 ð0:44; 0:82Þ. Ambiguous Queries for DPðq; RÞ 2 ð0; 0:44Þ. Each synthetic profile is built from the click log ofthree queries, with one from each group. The forbidden node set S is selected randomly from the topics associated with the clicked documents.Real. The real user profiles are extracted from50 distinct user click logs (with #clicks _ 2;000) from AOL. For each user, the user profile is built with the documents dumped from all urls in his/her log.6 The sensitive nodes are randomly chosen from no more than five topics (with depth _ 3Þ.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

VI.CONCLUSION ANDFUTURE WORK

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.

REFERENCES

[1] Z.Dou, R.Song, and J.-R. Wen, A Large-Scale Evaluation and Analysis of PersonalizedSearch Strategies, Proc. Intl Conf. World Wide Web (WWW), pp. 581-590, 2007.
[2] J.Teevan,,S.T.Dumais, and E.Horvitz, Personalizing Search via AutomatedAnalysis of Interests and Activities, Proc. 28th Ann. Intl ACM SIGIR Conf. Researchand 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 Intl Conf. Web Intelligence (WI), 2005.
[4] B. Tan, X. Shen, and C. Zhai, Mining Long-Term Search History to ImproveSearch Accuracy, Proc. ACM SIGKDD Intl Conf. Knowledge Discovery and DataMining (KDD), 2006.
[5] K. Sugiyama, K. Hatano, and M. Yoshikawa, Adaptive Web Search Based onUser Profile Constructed without any Effort from Users, Proc. 13th Intl Conf. WorldWide Web (WWW), 2004
[6] X. Shen, B. Tan, and C. Zhai, Implicit User Modeling for Personalized Search,Proc. 14th ACM Intl Conf. Information and Knowledge Management (CIKM),2005.
[7] X. Shen, B. Tan, and C. Zhai, Context-Sensitive Information Retrieval Using ImplicitFeedback, Proc. 28th Ann. Intl ACM SIGIR Conf. Research and DevelopmentInformation Retrieval (SIGIR), 2005.
[8] F. Qiu and J. Cho, Automatic Identification of User Interest for Personalized.

BIOGRAPHY

Nitesh Chavan, Swapnil Khese, Eshwar Palave and Akshay Dongare pursuing B.E. degree in Pune University. They are working on project named Privacy Preservation In Personolized web Search.