International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878, Impact Factor (2014): 3.05

Privacy Protection in Personalized Web Search Using Metric Prediction

Sneha M. Durugkar[1], Prof. Saudagar S. Barde[2]

[1]Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Vadgaon (BK), Pune, India
[2]Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Vadgaon (BK), Pune, India

Abstract: ***Privacy is the ability of an individual or group for seclude themselves, or information about themselves, and hereby expressing them selectively. The boundaries and content of what is considered to private differ among the cultures and individuals, but share the common themes. When something is private to a person, it usually to means that something is inherently special and sensitive to them. The domain of privacy partially it overlaps security, which can be include the concepts of appropriate use, as well as for protection of information. Privacy may also take form of bodily integrity. World Wide Web is expanding quickly throughout the many years. These days, web has ended up to tremendous wellspring of Information. In procedure of gaining data, web indexes assume an essential part. Number of query items that are acquired or indicated through different internet searchers, yet low quality and less precision for indexed lists make troublesome for client to pick up the data that is required. Personalized web Search (PWS) has shown sufficiency in upgrading the way for diverse request advantages on Internet. Regardless, to confirmations show that clients dislike to uncover the private information to the midst of request has transformed into the major prevention for the great extension of PWS. We consider security for affirmation in PWS applications that model client inclination as dynamic user profiles. Here propose a PWS framework called UPS that can be adaptively entirety up profiles by inquiries while with respect to the client defined security for requirements. Our runtime generalization goes for striking the concordance between the two farsighted estimations that survey utility of personalization and for the security danger for revealing summed up of profile. Here demonstrate two enthusiastic algorithms, to be particular Greedy DP furthermore Greedy IL, for runtime generalization.***

Keywords: Privacy protection, personalized web search, Utility, Risk, and Profile

1. Introduction

The web records have quite a while ago transformed into the most basic passage for people searching for important in formations on web. Of course, customers may experience the dissatisfaction at the point when records return insignificant happens that don't meet for the veritable desires. Such superfluity is by and large in the view of tremendous mixture of users settings and establishments, and obscurity of compositions. Personalized web search for (PWS) is general characterization of chase frameworks trying to giving good query items, which are then specially designed for the individual user needs. As the expense, user information must accumulate and examined to assess user point behind the issued query. The answers for PWS can generally be grouped into two sorts, particularly click-log-based methods and profile-based ones. The click-log based procedures are clear that they just force inclination to clicked pages in the user's history. Regardless of the way that the methodology has been indicated to perform dependably and respectably well, it can simply take problem at repeated inquiries from the same user, which is a strong containment keeping its suitability. On the other hand, profile-based procedures are upgrading that the chase contribution with jumbled user investment models that are created from customer profiling frameworks. Profile-based strategies may be perhaps effective for the essentially distinctive varieties of request.

Personalized web search (PWS) is an all inclusive class of search techniques that going for giving search results, which are altered for individual client requirements. As the consumption, client data must be made and examined to make for sense of the client that aims behind the issued query. The answers for the PWS can generally categorize into two sorts, specifically click-log-based systems and another is profile-based ones. The click-log based techniques that are clear they essentially force inclination to clicked pages in user's question history. Despite the fact that their methodology has been built to perform the reliably and extensively good [1], it can just take shot at incessant questions from same client, which is a solid restriction binding for appropriateness. In distinction, profile-based techniques gives signs of improvement for the search involvement with the complex client investment models that produced from client profiling routines. Profile based systems that it can be conceivably powerful for practically different varieties of queries, however are accounted for to be unequal under the few circumstances. The profile-based PWS has created much productivity in acculturating the nature for web search. As of late, with expanding utilization of individual and conducting data to profile its clients, who are generally, accumulated for verifiably from query history, browsing history, click-through the data, bookmarks, user archives. So, protection nerves have turned into the significant wall for wide multiplication of PWS administrations. Here author also proposed RSA algorithm that is used for encryption and decryption, as encrypted queries are sending from client side and that are received and server sending decrypted results and again this client making online profile building and also online profile generalization. On the server side two methods are used that are g click and p click.

2. Literature Survey

Huge amount of the information gets added for Web every day. Nowadays visible text creation is of order of suppose some GB per day and private text creation including users email, IM messages, tags, reviews etc is of the order of 3 terabytes per day [5]. This rapidly increasing for scale of the web is in many ways limiting the utility of the web.

There is high level of noise beginning from the spam and ending with lot of uninteresting, irrelevant and duplicated content. Search engines and other forms of the ranking are unable to keep up with this. Recently, search engines have started showing Wikipedia links as top search result because ranking has become very hard. Personalization [6] is a playing an increasingly important role for creating good Internet experiences. Recent applications of personalization have focused on improving the search experience [9]. An important view of personalization is creation of user profile. The user profile [3] could be created on client PC or on Internet server. Client side profiles offer better privacy, a more complete view of user data. Server side profiles enable collaborative filtering and profile portability.

Although personalized search that has been proposed for many years and many personalization strategies have been investigated, it is unclear that personalization is consistently effective on different queries for different users, and under different search contexts. So this problem and provide some preliminary conclusions, also present a large-scale evaluation framework for personalized search based on query logs, and then evaluate for five personalized search strategies using 12-day MSN query logs. By analyzing the results, here reveal that personalized search has significant improvement over common web search on some queries but it has little effect on queries e. g. , queries with small click entropy. It even harms search accuracy under some situations. Furthermore, we show that straightforward click-based personalization strategies perform consistently and considerably well, while profile based ones are unstable in our experiments. Also reveal that both long term and short-term contexts are very important for improving search performance for profile based personalized search strategies.

Online offerings such as different web search, news portals, and e-commerce applications face that challenge of providing high-quality service to a large, heterogeneous user base. Recent efforts have highlighted potential to improve performance by the introducing methods to personalize services based on special knowledge about users. For example, a user's demographics, location, and past search and browsing may be useful in enhancing the results offered in response to web search queries. However, reasonable concerns about privacy by users, providers, and government agencies acting on behalf of citizens, may limit to access by services to such information. Here also introduce and explore an economics of privacy in personalization, where people can opt to share personal information, in a standing or on-demand manner, in return for expected enhancements in the quality of an online service and focus on the example of web search and formulate realistic objective functions for search efficacy and privacy. Author demonstrates how it can find a provably near-optimal optimization of the utility-privacy tradeoff in efficient manner. Also evaluate our methodology on data drawn from a log of the search activity of volunteer participants. Author separately assess users preferences about privacy and utility via a large-scale survey, aimed at eliciting preferences about peoples willingness to trade the sharing of personal data in returns for gains in search efficiency. Here shows that a significant level of personalization can be achieved using a relatively small amount of information about users.

3. Proposed System

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 figure, UPS consists of a non trusty search engine server and a number of Clients. Each client (user) accessing 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. The framework works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchicaluserprofileisconstructedandcustomizedwiththeuser-specifiedprivacy requirements.

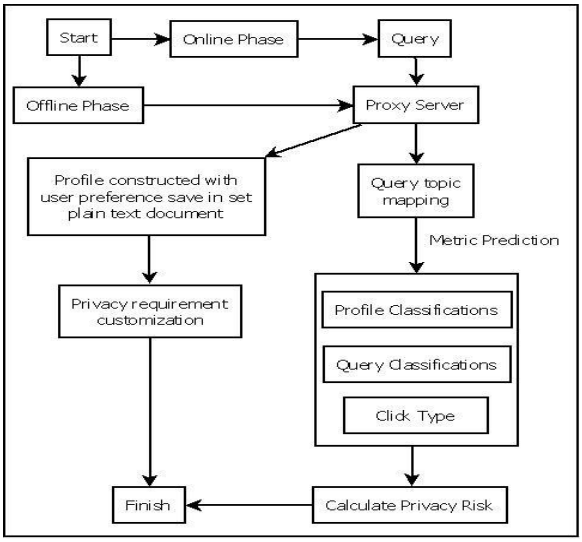

Figure1: System Architecture of UPS

The online phase handles queries as follows:

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 users' privacy requirements;
2. Allows for customization of privacy needs; and
3. Does not require iterative user interaction.

Author proposes 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. Also use RSA algorithm for encryption and decryption the data, here client sends encrypted queries for the server and that of server sends decrypted results for the client and clients forms outline profile building as g0, and also outline profile generalization.

4. Algorithm

1) Proposed Algorithm
Let,
The system Sis represented as: S=G, S, R, Rr (1)
•Generation of User Profile: G=Generating user profile
Here, user issues query q, proxy generates user profile P, output of user profile Gi. Q=issues Query on client
Gi=Output of profile
•Query and User Profile Sent to PWS:
PWS=Personalized Web Search=PWS1 request=r1, r2. . . rn.
•Personalized Search Result with profile and sent to proxy:
R=result set
Pr=Proxy=pr1
•Present Search result or re-rank: Rr=Re-ranking
D=display search result

2) RSA Algorithm
For Efficient encryption and decryption operations:

In RSA states that "computing M^e (mod n) requires at most $2 \cdot log_e$(e) multiplication and $2 \cdot log_e$(e) divisions" if we use their procedure below. It is important for us to know the amount of steps it would take a computer to encrypt the message so we can see if a method is fast and efficient, or not. We now "exponentiations by repeated squaring and multiplication":

Step1. Let $e_k e_{k-1} \ldots \ldots e_1 e_0$. . . be the binary representation of e.
Step2. Set the variable C to 1.
Step3. Repeat steps 3a and 3b for i = k, k – 1, …. , 0:
Step3a. Set C to the remainder of C^2 when divided by n.
Step3b. If e_i = 1 then set C to the remainder of C · M when divided by n.
Step4. Halt. Now C is the encrypted form of M.

There are more efficient procedures out there, but this one is good too. Also, since decryption follows the same identical procedure as encryption, we can implement the whole operation on a few integrated chips. According to the authors of RSA, "the encryption time per block increases no faster than the cube of the number of digits in n.

5. Results

Table 1: Time per User Profile

Query Type	*Existing System*	*Propose system*
Distinct Query	689	600
Medium Query	756	702
Ambiguous Query	1232	1123

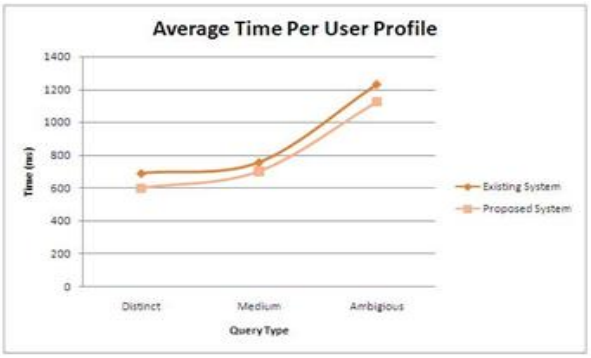

Figure 2: Average Time per User Profile

In our graph, our system take less response time in user profile compare to existing system time.

Table 2: Average Accuracy of a Query Result

Query Type	*Existing System*	*Propose system*
Distinct Query	93%	96%
Medium Query	85%	90%
Ambiguous Query	76%	82%

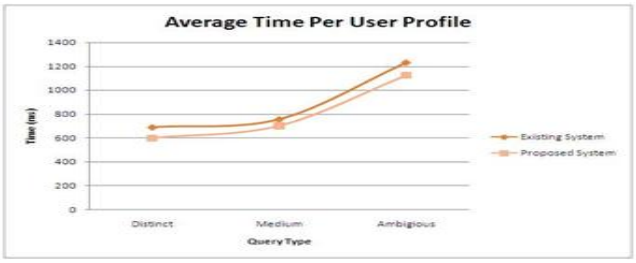

Figure 3: Average Accuracy of Query Result

6. Conclusion

The framework displayed a client side protection insurance structure called UPS for personalized web search. UPS could conceivably be embraced by any PWS that catches client profiles in a progressive scientific categorization. The system permitted clients to detail modified security necessities by means of various leveled profiles and UPS likewise performed online generalization on client profiles to ensure that individual protection without compromising the search quality. Author proposed two ravenous algorithms, in particular Greedy DP and Greedy IL, for the online generalization. Also propose metric prediction using ranking search techniques which include p-click and g-click, here RSA algorithm is used for the encryption and decryption and also we used query classification and profile classification. A metric prediction algorithm is for predict the performance of UPS framework. This metric predicts the search quality of the query on a generalized profile. We

transform the utility prediction problem to the analysis of distinguishing power of a given query on a generalized profile. Similar assumption has been made in to model utility, but this metric cannot be used in our problem settings, as have a profile with hierarchical structure instead of flat one.

References

[1] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search, " Proc. 14th ACM Intl Conf. Information and Knowledge Management (CIKM), 2005.
[2] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia, " HP Labs, 2008.
[3] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy, " Proc. ACM SIGKDD Intl Conf. Knowledge Discovery and Data Mining (KDD), 2006.
[4] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories, " Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
[5] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search, " Proc. 16th Intl Conf. World Wide Web (WWW), pp. 591-600, 2007.
[6] Z. Dou, R. Song, and J. -R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies, " Proc. Intl Conf. World Wide Web (WWW), pp. 581-590, 2007.
[7] F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search, " Proc. 15th Intl Conf. World Wide Web (WWW), pp. 727-736, 2006.
[8] P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschu tter, "Using ODP Metadata to Personalize Search, " Proc. 28th Ann. Intl ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.
[9] D. Xing, G. -R. Xue, Q. Yang, and Y. Yu, "Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies, " Proc. Intl Conf. Web Search and Data Mining (WSDM), pp. 139-148, 2008.
[10] C. C. Chen, M. C. Chen, PVA: A self-adaptive personal view agent, Journal of Intelligent Information systems, 18:2/3, pp 173-194, 2002.
[11] H. R. Kim and P. K. Chan, Learning implicit user interest hierarchy for context in personalization, Proc. of International conference on Intelligent User Interfaces (IUI), Miami, Florida, 2003.
[12] D. Godoy and A. Amandi, User profiling for web page filtering, IEEE Internet computing, July-August 2005.
[13] K. Sugiyama, K. Hatano, M. Yoshikawa, Adaptive web search based on user profile constructed without any effort from users, WWW 2004.
[14] Y. Xu, B. Zhang, Z. Chen and K. Wang, Privacy enhancing personalized web search, WWW 2007.
[15] H. Dai et. al, Detecting online commercial intention, WWW 2006.
[16] N. Nanas, V. Uren and A. D. Roeck, Building and applying a concept hierarchy representation of a user profile, SIGIR 2003.
[17] Trajkova J. and Susan Gauch, Improving Ontology based user profiles, Proceedings of RIAO 2004, University of Avignon, France.
[18] P. A. Chirita , W. Nejdl, R. Paiu, C. Kohlschutter, Using ODP data to personalize search, SIGIR 2005.
[19] E. Gabrilovich and S. Markovich, Overcoming the brittleness bottleneck with Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge, Proc. of the AAAI conference, 2006.
[20] Z. Ma, G. Pant and O. R. L. Sheng, Interest based personalized search, ACM transactions on Information systems, Vol. 25, no. 1, February 2007.
[21] Susan Gauch et. al, User profiles for personalized information access, Ch. 2, "The adaptive web", Springer LNCS 4321.
[22] Ahu Sieg, Bamshad Mobasher, Robin Burke, Web search personalization with ontological user profiles, CIKM 2007