SECURE PRIVACY ENHANCED APPROACH IN WEB PERSONALIZATION

Tejashri A. Deshmukh 1, Asst. Prof. Mansi Bhonsle 2

1 PG Scholar, Department of Computer Engineering, Savitribai Phule University, GHRCEM, Pune.
2 Asst. Professor, Department of Computer Engineering, Savitribai Phule University, GHRCEM, Pune.

ABSTRACT:

Personalized web searching is a capable to improve searching excellence by modifying searching results for people with data. Users are always uncomfortable with revealing private data to searching engines. Privacy is not compromised if there is an improvement in service to the user. So the balance must be hit between quality and protection. Our work is scalable way for the users to automatically form rich profiles. Such profiles summarize user’s interests in an organization rendering to specific interests. There are parameters which are specifying requirements for privacy are proposed to service user to select the content. We are going to present SHA algorithm and Greedy algorithm for creating Generalized Profile. We are also going to provide an online forecast mechanism so that we can decide whether a query is useful. Wide experiments prove the efficiency of framework.

Keywords: Personalized search, privacy protection, user profile, privacy risk.

[1] INTRODUCTION

The data on web continuously develops; it becomes more and more difficult for searching engines to get data that satisfies user’s different needs. Personalized searching is a way to develop searching quality by modifying searching results for public with different data objectives. Now day’s researching efforts have focused on this area. Personalization methods enable the dynamic insertion, customization or suggestion of content in any format that is relevant to the individual user.

Personalized web search is of two types, one is of clicked log based & another is profiled based method. In clicked log based method [1], it works on clicked pages in query history. It work well but, this method work on repeated no. of queries. This is the limitation for this method. The profiled based method work on all type of queries. It improved the search result quality with complicated user model. They are classified into general approaches: Re-ranking query results reverted by searching engines locally by the use of personal data; or transfer personal data and query together to searching engine [2]. A good personalization algorithm relies on rich user profiles and web corpus. As the web quantity is on server, re-ranking on client is bandwidth concentrated because it needs a huge number of searching results conveyed to the client before re-ranking. Alternatively, if the amount of data transmission is limited through filtering on the server, this results high hope on the existence of
data along filtered results. This result is not always the case. Therefore, most of personalized searching services like Google Searching [2] and Yahoo! Web [3] agree the instant approach to modify results on server by examining composed personal data, e.g. interests, searching histories. This methodology has privacy issue on revealing personal data to a public one. It needs users to allow the server full access to their personal and behaviour data on the Internet. Without the user’s permission, cleaning such data would violate an individual’s privacy. For protecting user profile in personalized web search, researcher have consider two things, first attempt to improve search quality, second is they need to hide private data which exist in user profile.

[2] RELATED WORK

This section is focus on literature of profiled based personalization & privacy protection in personalized search.

[2.1] PROFILED-BASED PERSONALIZATION

For better search results, we used profiled-based personalization.

The paper [9] focus on the different types of web mining that is web content mining, web structure mining and web usage mining. After that, they introduced the web mining techniques in the area of the Web personalization. In this paper, we have proposed a web usage mining approach for semantic web personalization. This paper focused on various Web personalization categories and their research issues.

The paper [3] described personalized web search via automated analysis of interests and activities. The paper also investigated the feasibility of personalization by constructing user profile by using ranking algorithm. Text based personalized algorithm is used, which perform significantly better than explicit relevance feedback. Such personalization algorithm improved web search.

K. Sugiyama, K. Hatano, and M. Yoshikawa[4] studied adaptive web search based on user profile. This paper provides a method to adaptive search result according to user need for relevance. User profile is constructed which based on collaborative filtering model for achieving best accuracy.

Y. Xu, K. Wang, B. Zhang, and Z. Chen [5] proposed privacy enhancing personalized web search. The paper [1] focus on the issue of privacy preservation in personalized search. they show that client-side personalization has advantages over the existing server-side personalized search services has privacy preserving, and having possible future strategies to fully protect user privacy. The paper described the method for automatic building of hierarchical profile via frequency analysis. This paper investigated feasibility in achieving a balance between users.

Charanjeet Dadiyala, Prof. Pragati Patil, Prof. Girish Agrawal [6] proposes an approach to improve the retrieval quality of web engine and refining the search results depending on the users need.

Chanchala Joshi, Teena Jaiswal, Himanshu Gaur[6] provides is an overview of a personalized Web search, it contain different search results for different users or organize search results differently for each user, based upon user interests, preferences, and
information needs. There are different personalized web search algorithms for analyzing the user interests and producing the outcome quickly. In this paper we are analyzing various issues of personalized web search.

Figure 1 provides general overview of personalized web search. In this figure, an algorithm is provided for the user to automatically build a hierarchical user profile that represents the user’s implicit personal interests. The interests are put on a higher level specific interests are put on a lower level. Only portions of the user profile will be exposed to the searching engine in accordance with a user’s own privacy settings. A searching engine wrapper is developed on the server side to incorporate a partial user profile with the results returned from a searching engine. Rankings from both partial user profiles and searching engine results are combined. The customizable results are delivered to the user by the wrapper.

The solution has three parts: First, a scalable algorithm automatically builds a hierarchical user profile from available source data. Another is, privacy parameters are offered to the user to determine the content and amount of personal data that will be revealed. Third, a searching engine wrapper personalizes the searching results with the help of the partial user profile.

[2.2] PRIVACY PROTECTION IN WEB PERSONALIZATION

The privacy protection in web personalization consists of two classes. First is privacy as the identification of user and other class is the data sensitivity, user profile, exposed to personalized web search server.

The paper[5] proposed privacy enhancing personalized web search. An important difference of work is that provide privacy protection in web search. Degree of protection of privacy is specified by user, Users are allowed to customize privacy needs in their hierarchical user profiles. Queries having distinct queries generally benefit more from personalization, which is different for those with larger values. Since the latter may cause privacy disclosure, hence becomes questionable for such queries.

The paper [10] provides an article that analyzes the privacy risks associated with several current and prominent personalization trends that is social-based personalization, behavioral based, and location-based personalization. The paper surveys user attitudes towards privacy and personalization as well as technologies that can help reduces privacy risks.

The existing works of privacy preserving personalized web search are not satisfactory work. The problems for existing are as follows:

1. The old profile based PWS do not generalize user profile at runtime. A user profile is generalized only once and that is in offline mode and queries used to personalize are from the same user. One issue reported is [1] that sometimes profile based personalization method does not support ad-hoc queries. Online profiling is the better approach but no previous work has supported this.
2. The existing system does not consider modification in privacy requirements. In this, privacies are not protected properly.
3. While creating personalized search results many existing personalization techniques require repeated user interactions.

[3] PROPOSED WORK

Personal data, i.e. personal documents, browsing history and emails might be helpful to identify a user’s implicit intents. However, users have concerns about how their personal data is used. Privacy, as opposed to highly secure, highly depends on the involved person and how that person may benefit from sharing personal data. The question here is whether a solution can be found where users themselves are able to set their own privacy levels for user profiles to improve the searching quality query.

Yet, this paper is an exploratory work on the two aspects: First, we deal with unstructured data such as personal documents, for which it is still an open problem on how to define privacy. Secondly, we try to bridge the conflict needs of personalization and privacy protection by breaking the premise on privacy as an absolute standard.

[3.1] ENHANCED PRIVACY Preserving FRAMEWORK

The proposed design contains the enhanced privacy preserving personalized search framework called EPS. The entire problems are resolved in this system. The framework works with the assumption that the queries are without any sensitive information, this framework not only protect privacy of individual user but also retaining their usefulness for PWS.

The framework works in two phases, first is offline and another is online. In the offline phase, hierarchical user profiles are constructed and in online phase, it handles the different queries.
1. When a user issues a query Q on the client, a generalized user profile G is created by the proxy satisfying the privacy requirements.
2. The query Q and the user profile G are then sent together to PWS server for personalized web search.
3. Search results personalized according to the profile are sent back to the query proxy.
4. At last, the proxy either presents raw results r to user or ranks them r' with the entire user profile.

Our main contributions are as follows:

1. A Enhanced privacy-preserving personalized framework EPS, which generalizes user profiles for every query based on user specified privacy requirements.
2. The two metrics namely personalization utility and privacy risk are taken into consideration and we formulate the problem of privacy-preserving personalized search as $\delta$ risk profile generalization.
3. The simple and efficient algorithms SHA algorithm are developed to facilitate runtime profiling.
4. The client can decide to personalize a query in EPS before each runtime profiling.

[3.2] ALGORITHM FOR PROPOSED SYSTEM

In this section, we develop effective generalization algorithms, SHA stands for Secure Hashing algorithm and Greedy algorithm. The Greedy algorithm improves the efficiency of the Generalization.

Greedy algorithm is used to create generalized algorithm. This algorithm is used to maximized discriminating power and minimized information loss. In this section, we develop effective generalization algorithms, GreedyDP and GreedyIL which support runtime profiling. This algorithm tries to maximize the discriminating power (DP) and minimize the information loss (IL). The first greedy algorithm GreedyDP works in a bottom up manner. Starting from optimal profile, GreedyDP chooses a leaf topic for pruning, trying to maximize the utility of the output of the current iteration, namely optimal profile. During the iterations, we also maintain a best profile, which indicates the optimal profile having the highest discriminating power while satisfying the risk. The iterative process terminates when the profile is generalized to a root. The best profile will be the final result of the algorithm. The main problem of GreedyDP algorithm is that it requires computation of all candidate profiles (together with their discriminating power and privacy risk) generated from attempts of prune-leaf on all. This causes significant memory requirements and computational cost.

The GreedyIL algorithm improves the efficiency of the generalization using heuristics based on several findings. One important search is that any prune-leaf operation reduces the discriminating power of the profile. The benefits of making are, it enhances the stability of the search quality and it avoids the unnecessary exposure of profile of user. Therefore, GreedyIL is accepted significantly.
Secure Privacy Enhanced Approach In Web Personalization

www.ijcea.com ISSN 2321-3469

Inputs: Query Q, User profile G, Privacy threshold δ.
Output: Generalized Profile G*

Step 1: If G’ is user profile obtained by pruning on G,
    Then
    \[ \text{DP}(Q,G) \geq \text{DP}(Q,G’). \]

Step 2: Each candidate in the queue having tuple like
    \[ C = (t, \text{IL}(t,Gi)), \text{where } t \text{ is leaf to be pruned by } C \text{ and } \text{IL}(t,Gi) \text{ indicates the IL incurred by pruning } t \text{ from } Gi. \]

Step 3: The iterative process can be terminated when risk satisfied.

Step 4: While risk (Q,Gi) > δ then pop the prune leaf operation on t from q.

Step 5: In C1 case, if t is a node with no sibling, then insert s to Q. The case C2 requires introducing shadow sibling of t.

Step 6: If we attempt to prune t, we merge t into shadow to obtain new shadow leaf.

Step 7: Prune-leaf only operates on single topic t. While in case C2, pruning t incurs re-computation of the preference values of its sibling nodes.

Step 8: Once a leaf t is pruned, only the candidate operators pruning t’s sibling topics need to be updated. Thus, it does not impact the IL of other candidate in Q. In case C2, pruning t incurs re-computation of the preference values of its sibling nodes.

GreedyIL traces the information loss instead of the discriminating power. This enhanced approach saves a lot of computational cost.

[3.3] SHA ALGORITHM

SHA1 is used SHA hash function, in this it will be replaced by newer and more secure than others.SHA1 outputs is of 160 bits including all of any size of file or input. It uses a 512 bit block size and has a maximum message size of \(2^{64}-1\) bits.

- In SHA algorithm, first we extend message until the final block has 448 bits. Then initialized five hash block to specific constant. Allocate 80 words array for message schedule. the words are generated using the following algorithm

  - word[i-3] XOR word[i-8] XOR word[i-14] XOR word[i-16] then rotated 1 bit to the left. Calculate SHA function().Then concatenate output whis is message digest.

The advantages of Enhanced Privacy Preserving Framework are as follows:

- It enhances the stability of the search quality.
- It is fast retrieval process with secure data.
- It improves the privacy protection against different type of attacks.
- It avoids the unnecessary exposure of the user Profile.
- It provides runtime profiling.
[4] RESULT

In this section, we present the experimental results of SHA1 algorithm. And also generate graph for SHA1 algorithm.

In this experiment, users are required to try different privacy thresholds to explore the relationship between privacy preservation and search quality. Uploaded file is converted into encrypted format. It is opposed to decrypting the text to obtain original version. Hacker can’t use that file.

Supposed we upload some that is downloaded by only that user. if hackers try to download that file, there was problem to download that file. To download that uploaded file, we provide better security, that is we enter security question, authentication key that is only known by user itself.

The SHA1 algorithm is simply that takes a string of any length reduce it to a unique fixed length string. By using SHA1 algorithm, we generate encrypted key and authentication delay graph as shown below.

The proposed graph also provides more security than previous method. This graph is as shown in follows.
[6] CONCLUSION

The personalized searching is a promising way to improve search quality. However, this approach requires users to grant the server full access to personal data on Internet, which violates users’ privacy. We propose EPS framework for personalized web search which do client side privacy protection. Any personalized web search having user profiles in hierarchical structure can use this framework. This framework provides some functions to the user like to specify their privacy requirement, online decision making. EPS framework protects user’s personal privacy without compromising the search result. We use greedy algorithm for creating generalized profile and SHA1 algorithm for online generalization. In this paper, we investigated the feasibility of achieving a balance between users’ privacy and searching quality. There are a few of promising directions for future work. In particular, we are considering ways of quantifying the utility that we gain from personalization, thus users can have clear incentive to comprise their privacy. Also, we suspect that an improved balance between privacy protection and searching quality can be achieved if web searching is personalized by considering only exposing those data related to a specific query. We will try to find valuable relationship. We will try to find more appropriate solution to predict the performance and find better method to construct user profile.

[7] ACKNOWLEDGMENT

We would like to thanks all the authors from which we referred for giving their suggestions and making the material available for us to refer.
REFERENCES


T. A. Deshmukh and M. Bhonsle