VOL. 10, NO. 7, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

PRIVACY PROTECTION IN PERSONALIZED WEB SEARCH USING GENERALIZED PROFILE

M. Radhika and V. Vijaya chamundeeswari
Computer Science and Engineering, Velammal Engineering College, Chennai, India
E-Mail: radikamoorthi@gmail.com

ABSTRACT

Personalized web search is effective way of improving the quality search result based on the use profile. But people who want to search in internet do not want to reveal his identity or profile to the outside world. In this paper, the privacy of the user profiles is analyzed as a hierarchical data structure. The framework captures the user profiles in a hierarchical taxonomy. The framework adaptively generalize a user profile for a query, keeping the preferences of the user privacy. Since the hierarchical taxonomy which maintain in the server will take more time to search if the number of users increases. So, a greedy algorithm in map reduce paradigm is evolved to process the hierarchical taxonomy tree structure in parallel.

Keywords: privacy protection, personalized web search, utility, risk, profile.

1. INTRODUCTION

As the measure of data on the web persistently extends, it has developed incrementally hard for web search engines to discover data that fulfills clients' individual needs. Personalized web search 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 profile based approach get better search observation of events with entangled user-interest representation generated from user profiling method. Profile based mode can be likely effective for about all category of queries, but are described to be changeable under some conditions [1]. The profile based personalized web search has illustrated more powerful in enriching the quality of web search recently, with improving usage of personal and behaviour information to profiles its users, which is usually accumulated implicitly from query history [2],browsing history [4], bookmark [5] and so on.

Unfortunately, such indirectly collected personal data can easily disclose a complete range of user's private activity. Secrecy issues inclining from the absence of security for such data, for the sample America Online Incorporation (AOL) query logs scandal [7], not only increase panic among distinct users but also dampen the information-publisher's interest in offering personalized service. In particulars, privacy task have become the important obstacle for large increase of personalized web search services.

2. RELATED WORK

Earlier tasks on profile based personalized web search especially focus on enhancing the search utility. The principle idea of profile based personalization is tailored the searched results by referring indirectly often a client profiles that uncovers an individual report objective.

An alternate work in [6] builds the sequential profile naturally by means of term-recurrence investigation on the client information. The user customizable privacy preserving search structure does not concentrate on the execution of the client profiles. Normally, our system can possibly receive any hierarchical representation focused on a classification of data. Normalized Discounted Cumulative Gain [10] is a typical measure of the adequacy of a data retrieval system. It is focused on a human graded significance positions in the result list. Therefore, it known for is high cost in unambiguous feedback accumulation.

To decrease the human contribution in execution measuring, specialists also propose different measurements of personalized web search that depend on clicking choices, including Average Precision (AP) [11] Rank Scoring [9], Average Rank [3] .The Average Precision metric utilization is proposed by Dou et al. [1], to measure the ability of the personalization in user customizable privacy preserving search. The two forecasting measurements are personalization utility and security hazard, on a profile occurrence without asking for client comments.

In [8], Krause and Horvitz enlist information systems to learn a probabilistic model and then use this paradigm to produce the near-optimal incomplete profile. One important impediment in this work is that it builds the client profile as a limited set of properties, and the probabilistic model is prepared through predefined regular queries. These expectation are unreasonable in the environment of personalized web search. Wang *et al.* [6] proposed a privacy protection solution for personalized web search based on taxonomy. Using a client defined edge, a small portion of user profile is captured in effect as a rooted sub tree of the complete profile.

3. FRAMEWORK

The overview of User Customizable Privacy-preserving Search (UPS) framework

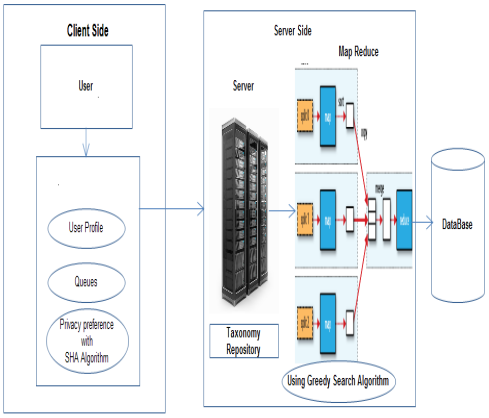


Figure-1. System architecture.

When a user issues a query on the client the online profiler generate a user profile in runtime. The outcome of this process called generalized profile. The generalized profile satisfy the privacy requirement. The privacy risk is protected using secure hash algorithm. The generalization process consider the privacy preferences and personalization utility.

The query and the generalized profile are sent together to the server for personalization. The taxonomy repository structure is maintained in the server. For the scalability and retrieval effectiveness the map reduce is used. Map reduce supports the parallel computing for large data sets. So, the map reduce function splits, sorts and merges the information. The searched results are personalized and delivered back client.

4. IMPLEMENTATION

The proposed approach consists of three modules: web crawler, creation of taxonomy, generalization using greedy algorithm in map reduce paradigm.

a) Web crawler

When a user issues a query the search result is displayed. The web crawler collects information about the browsed website. The crawler reads the website uniform resource locator, web page title, meta tag contains web page descriptions, keywords of the page. The crawler read the contents of the browsed web pages and find which user is viewing for a specific web query. The web text is fixed by clustering the client web query data. The crawled data's are stored in the radial tree structure.

b) Creation of taxonomy

A taxonomy data structure is created for all the information queried by the user. The hierarchical taxonomy is maintain in the server. The tree structure contains all the browsed information. The browsed information are displayed in radial tree structure because the large information sets stored in server.

The queried information is protected by secure hash algorithm (SHA-256). The SHA-256 is the best public cryptanalysis algorithm. The algorithm structure consists of padding bits, appending length as 64 bit unsigned, Buffer initiation, processing of message. The protection is given for each user profile.

c) Greedy IL algorithm in map reduce

Greedy IL gives the optimal solution from the wide range of applications. It solves the problem heuristic of making the locally optimal choice with the hope of finding a global optimum. To construct the solution in an optimal way. The greedy algorithm consists of functions. A function that checks whether chosen set of items provide a solution. A function that checks the feasibility of a set. The selection function tells which of the candidate is the most promising function. An objective function which does not appear explicitly, it gives the value of a solution.

The map reduce model supports parallel computing on large data sets. The map reduce consists of map, shuffle , reduce phases. The map function input data is in the form of (key,value) pair. The output data to the map function is structured as (key,value) pairs called intermediate (key,value) pairs. In turn reduce function receives the intermediate (key,value) pairs in the form of group of intermediate values associated with one intermediate key (key, [set of values]) pairs as output. The map reduce is used for parallel computation and scalability.

Algorithm

Input: keyword to search
Output: personalized search results

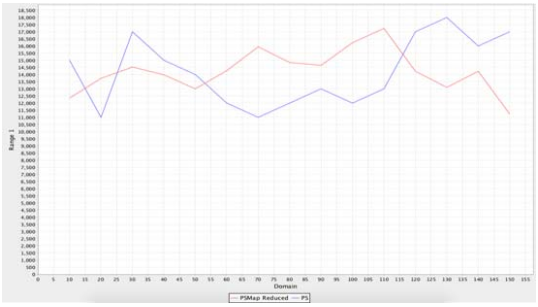
1. **Get** user search keyword and make a crawl and display results for that keyword
a. Restrict results to avoid data population
b. Record user click events
2. Recording user choice. Let **N** be the user choice and this choice should be either placed in session or as cookie. And the **no of hits 'f'** should be noted.
// Record user choice and store in database
3. For any **N -> f >0** (pre condition)
a. Invariant
Record user choice and this choice should not be 0 or less than 0
b. Post condition
For any **N** display records which have more **hits**
//Map reduced Personalized search
4. Declare a function with Map - > **M**
5. Loop each word in database which are equal to user search query
6. Increment **counter** (I)
7. Solution **keyword** => **counter**
8. Get values from Map
9. For Each key-value pair – add value to **counter**
10. Return keywords => counter

5. EXPERIMENTAL RESULTS

The experimental result of user customizable privacy preserving search framework.

a) Scalability of MapReduce algorithms

The scalability of the proposed algorithms by varying data set size and time taken to achieve results.



Red line will be personalized search results with map reduced and blue line will be normal search results (personalized). To achieve the results data length is referred as y axis and time is referred as x axis.

6. CONCLUSIONS AND FUTURE WORK

The Customized pursuit is making a guarantee way to seek the quality. The possibility of accomplishing conformity between clients' security and pursuit quality. The users are allowed to specify customized privacy requirements which captures the users profiles in hierarchical structure.

The user customizable privacy preserving search structure perform online generalization on user profiles. The utility and privacy risk are the two predictive metrics to satisfy. The greedy IL algorithm in map reduce enhances the scalability. For future work, the personalized search can perform on dynamic user profiles and non-text data. Dynamic user profiles allow a more up to date representation of users. Changes in their interests, their learning progress or interactions with the system are noticed and influence the user models. The models can thus be updated and take the current needs and goals of the users into account.

REFERENCES

[1] Z. Dou, R. Song and J.-R. Wen. 2007. “A Large-Scale Evaluation and Analysis of Personalized Search Strategies,” Proc. Int’l Conf. World Wide Web (WWW), pp.581-590.

[2] J. Teevan, S.T. Dumais and E.Horvitz. 2005. “Personalizing Search via Automated Analysis of Interests and Activities,” Proc. 28th Ann.Int’l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456.

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

[4] K. Sugiyama, K. Hatano and M. Yoshikawa. 2004. “Adaptive Web Search Based on User Profile Constructed without any Effort from Users,” Proc. 13th Int’l Conf. World Wide Web (www).

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

[6] Y. Xu, K. Wang, B. Zhang and Z. Chen. 2007. “Privacy-Enhancing Personalized Web Search,” Proc. 16th Int’l Conf. World Wide Web (WWW), pp. 591-600.

[7] K. Hafner, Researchers Yearn to Use AOL Logs, but They Hesitate, New York Times, August. 2006.

[8] A. Krause and E. Horvitz. 2010. “A Utility-Theoretic Approach to Privacy in Online Services,” J. Artificial Intelligence Research. Vol. 39, pp. 633-662.

[9] J.S. Breese, D. Heckerman and C.M. Kadie. 1998. “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI), pp. 43-52.

[10]K. Ja¨rvelin and J. Keka¨la¨inen. 2000. “IR Evaluation Methods for Retrieving Highly Relevant Documents,” Proc. 23rd Ann. Int’l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), pp. 41-48.

[11]R. Baeza-Yates and B. Ribeiro-Neto. 1999. Modern Information Retrieval. Addison Wesley Longman.

[12]M. Radhika, Dr. V. Vijaya chamundeeswari, R. Ramya Devi. 2015. “A survey On Personalization Of Web Search Using Generalised Profile” 14 Th International Conference & International Journal of Advanced Computational Engineering and Networking, Vol. 3, Issue-2, February.