www.ioirp.com
International Journal of Innovative Research in Computer Science and Engineering (IJIRCSE)
ISSN: 2394-6364, Volume – 1, Issue – 2. January 2015

Proficient Search Results for Users with Generalization Algorithm

Anas V A
ME CSE, Maharaja Prithvi Engineering College, Avinashi, India
anasva77@gmail.com

***Abstract*— Web search engines (e.g. Google, Yahoo, Microsoft Live Search, etc.) are widely used to find certain data among a huge amount of information in a minimal amount of time. However, these useful tools also pose a privacy threat to the users: web search engines profile their users by storing and analyzing past searches submitted by them. In the proposed system, we can implement the clustering algorithms for improving the better search quality results. To address this privacy threat, current solutions propose new mechanisms that introduce a high cost in terms of computation and communication. Personalized search is a promising way to improve the accuracy of web search, and has been attracting much attention recently. However, effective personalized search requires collecting and aggregating user information, which often raises serious concerns of privacy infringement for many users. Indeed, these concerns have become one of the main barriers for deploying personalized search applications, and how to do privacy-preserving personalization is a great challenge. In this we propose and try to resist adversaries with broader background knowledge, such as richer relationship among topics. Richer relationship means we generalize the user profile results by using the background knowledge which is going to store in history. Through this we can hide the user search results. By using this mechanism, we can achieve the privacy.**

I. INTRODUCTION

The search engines are attempts to satisfy the user's needs by ranking the web pages with respect to queries. The recent improvement and extensive use of high-throughput technology are producing day by day in using query and click log data in web search. The explosion of the information offered on the internet and its heterogeneity present a challenge for keyboard based search technologies to find useful information for users. Those technologies have a systematic behavior of showing the same result for all users query at a certain time. Same time there is ambiguous on user's query; it reduces the performance of technologies. This investigation shows that the main reason is that they do not considering the user perspective in the retrieval process.

As search engines reach the limits inherent in selecting data and are hungry for more data in a competitive environment, mining cursor movements, hovering, and scrolling becomes important. A series of interactions of a process of querying, learning, and reformulating between users and search engine have to satisfy a single info need. In the proposed system, they have better method to convert user query to user search goal that is known to call as feedback session, pseudo-document, and clustering technique. And they have using criteria to evaluate the performance and number of clustering using the technique called Classified Average Precision (CAP).When compared to existing methods of extracting the information about the user click through data's as from direct click-through logs, in the proposed system they are using feedback session which will reduce the noise by using the information from directly click throughs. The pseudo-documents are mapped from the feedback session in existing paper they use binary vector method which won't provide the exact details to extract the keywords.

Web search query is a query that a user enters into a web search engine to satisfy his or her information needs. Web search queries are distinctive in that they are often plain text or hypertext with optional search-directives (such as "and"/"or" with "-" to exclude). They vary greatly from standard query languages, which are governed by strict syntax rules as command languages with keyword or positional parameters.

There are four broad categories that cover most web search queries. Informational queries – Queries that cover a broad topic (e.g., Colorado or trucks) for which there may be thousands of relevant results. Innavigational queries, queries that seek a single website or web page of a single entity (e.g., YouTube or delta air lines). Intransactional queries, queries that reflect the intent of the user to perform a particular action, like purchasing a car or downloading a screen saver. Search engines often support a fourth type of query that is used far less frequently: Connectivity queries,queries that report on the connectivity of the indexed web graph

UPS is distinguished from conventional PWS in that it; 1)provides runtime profiling, which in effect optimizes th personalization utility while respecting user's privacy requirements; 2) allows for customization of privacy needs; and 3) does not

require iterative user interaction. Our main contributions are summarized as following: 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. Our extensive experiments demonstrate the efficiency and effectiveness of our UPS framework.

A. Dataset preprocessing

Here, choose input dataset. Choosen dataset has been loaded into the database. After loading the dataset into the database, we can view the dataset. By using the string matching algorithm we filter out unwanted values in the dataset and it has been preprocessed and store into the database. a client-side privacy protection framework called UPS for personalized web search was proposed. UPS could theoretically be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The context allowed users to stipulate 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. In this they proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. In this for query mapping process it has various steps to compute the relevant items.

Most works on anonymization focus on relational data where every record has the same number of sensitive attributes. There are a few works taking the first step towards anonymizing set-valued or transactional data where sensitive items or values are not clearly defined. While they could be potentially applied to user profiles, one main limitation is that they either assume a predefined set of sensitive items that need to be protected, which are hard to done in the web context in practice, or only guarantee the anonymity of a user but do not prevent the linking attack between a user and a potentially sensitive item.

B. User Login

Here, users are entered by using the unique id and password. In this module, users are entered after registering. After registering each user has unique id. After login, user posts some queries which is based on our dataset which is loaded into the database.

C. Query Submission and Query Retrieval Based on User

Here, user submits query. Based on the query, some relevant results has been displayed and also based on the submitted query some history results are displayed. Based on the query and already posted queries, we can calculate the similarity values between them. Based on the similarity values, we can estimate some results.

Using the information of user browsing history and domain knowledge, we create an Enhanced User Profile. Once the Enhanced User Profile is created, we take the user query and suggest the relevant web pages with respect the query. In our Experiment, we have used User Profile as a base case for suggesting the relevant pages and compared the results with the pages suggested from Enhanced User Profile. For each query, we suggest top 20 relevant documents from User Profile and for the same query we also suggest top 20 relevant documents from Enhanced User Profile. In order to compare the efficiency of the result, we compared the similarity of suggested documents with the user query.

D. Estimate Relevant Results

Here, user posts sub query also. Based on the query and sub query, we can estimate some results. Based on the relevant results and total number of datas in the dataset, we estimate the support values. In this, some history results are also found. Web search engines (e.g. Google, Yahoo, Microsoft Live Search, etc.) are w widely used to find certain data among a huge amount of information in a minimal amount of time. However, these useful tools also pose a privacy threat to the users: web search engines profile their users by storing and analyzing past searches submitted by them. In the proposed system, it implement the clustering algorithms for improving the better search quality results. It is retrieved by using the Correl (Correlated Item) clustering algorithm. To address this privacy threat, current solutions propose new mechanisms that introduce a low cost in terms of computation and communication. In this paper we present a novel protocol specially designed to protect the users' privacy in front of web search profiling. In this we propose and try to resist adversaries with broader background knowledge, such as richer relationship among topics. Richer relationship means we generalize the user profile results by using the background knowledge which is going to store in history. Through this we can hide the user search results. In the Existing System, Greedy IL and Greedy DP algorithm, it takes large computational and communication time.

II. CONCLUSION

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. In this paper, we provide better efficiency results when compared with existing system. It provides privacy mechanism when adversaries retrieve the results by using background knowledge. In this similarities are calculated based on the similarity algorithm.As a future work; implement the clustering algorithms for improving the better search quality results.

References

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

[2] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search viaAutomated Analysis of Interests and Activities," Proc. 28th Ann.Int'l ACM SIGIR Conf. Research and Development in InformationRetrieval (SIGIR), pp. 449-456, 2005.

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

[4] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search Historyto 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 WebSearch Based on User Profile Constructed without any Effortfrom Users," Proc. 13th Int'l Conf. World Wide Web (WWW),2004.

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

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

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

[8] 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.

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