PRIVACY PRESERVED EXPLICIT FEEDBACK SYSTEM WITH FP GROWTH METHOD

Sofia Sayed[1], Reeba R[2]

[1] *PG Scholar, Department of Computer Science & Engineering, Sree Buddha College of Engineering, Alappuzha, India*
sofivm.s@gmail.com

[2] *Assistant Professor, Department of Computer Science & Engineering, Sree Buddha College of Engineering, Alappuzha, India*
reeba.amjith@gmail.com

Abstract—*Search engines is the basic tool of the internet, from there we can collect related information and can search according to the specified query. The information on the web increasing rapidly. The users are spending lots of time on internet for finding the information they are interested in. Now a days the search engines doesn't provide enough personalized help but provide the user with huge irrelevant information. In such case, personalized web search (PWS) has emerged and demonstrated its effectiveness in improving the quality of various search services on the Internet This paper gives details about privacy protection in PWS system applications which models the user preferences in hierarchical user profiles. This paper proposes a PWS framework called UPS which will adaptively generalize profiles by queries while respecting user specified privacy requirements. It aims at providing protection against a typical model of privacy attack. To reduce the time complexity of UPS procedure FP Growth algorithm also implemented here.*

Key Words—**Privacy, Personalised web search, internal utility, incompatible metrics, closed high utility itemsets, User Privacy Customisable Search**

1. INTRODUCTION

The amount of information on the web increases continuously, it is a very difficult task for web search engines to find information which satisfies individual needs of the user. Personalized search is a technique used to improve search quality by customizing the search results for people with different information need. Many researches are focused on this area. They are categorised in to two: Re-ranking query output returned by search engines locally by using the users personal information; or by sending personal details and queries together to the search engine.

Most of the leading personalized search services online like Google Personalized Search, Yahoo etc. adopts the second approach to tailor results on the server by analysing the result of personal information.

In PWS approach the privacy issues will make on exposing personal information to a public server. It usually requires users to allow the server full access to their personal and behaviour information on the Internet. Without the user's permission, usage of personal information would violate an individual's privacy. In addition, it is very difficult to incorporate unstructured data with search engines without summarization. So, for the purpose of both web personalization and privacy preservation, it is necessary for an algorithm to collect, summarize, and organize a user's personal information into a structured user profile. Here in existing system 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. Here develop two effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. In the proposed method here implemented FP growth algorithm for avoiding the time complexity of Greedy algorithms.

.

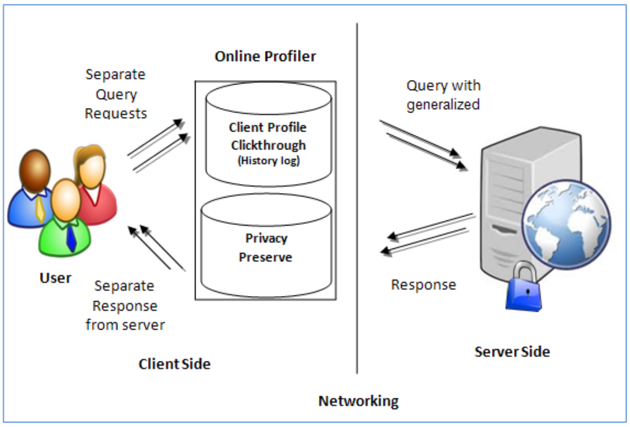

Fig. 1.1 System architecture of UPS.

2. LITERATURE REVIEW

1.1 PWS System Based on Content Analysis

Personalized web search can be achieved by checking content similarity between user profile and web pages [3]. When the user issues a query, each returned result are filtered or re-ranked and classified. Chirita et al. [3] [4] use the ODP (Open Directory Project) hierarchy structure to implement personalized search. In [5], a user profile is built as a vector of distinct terms and is constructed by aggregating past user click history [3]. Shen et al. [6] first use language modelling to mine immediate search contextual and implicit feedback information [3]. Teevan et al. [2] and Chirita et al. [12] exploit rich models of user interests, built from both search-related information, and other information about the user. [3]

1.2 PWS System Based on Hyperlink Analysis

Most generic web search approaches give rank importance to the documents based on the linkage structure. A large group of these works focuses on personalized PageRank. PageRank, proposed by Page and Z. Dou, [9], is a popular link analysis algorithm which used in web search. The fundamental motivation underlying PageRank is the recursive notion that important pages are those linked-to by many important pages [3].Qiu and Cho [8] develop a method to automatically estimate a user's topic preferences based on Topic-Sensitive WPageRank scores of the user's past clicked pages

2.3 Community-based analysis of PWS

Some approaches that personalize search results for the preferences of a community of like-minded users [3] is known as community or collaborative based search. In community-based personalized web search, when a user issues a query, search histories of users who have similar interests to the user are used to filter or re-rank search results [3]. Here] use a modified collaborative filtering algorithm to constructed user profiles to accomplish personalized search. Sun et al. [11] proposed a novel method named CubeSVD to apply personalized web search by analysing correlations among users, queries, and web pages in click through data.

3. PROPOSED SYSTEM

We propose a privacy-preserving personalized web search framework UPS, which is used to generalize profiles for each query based on user specified privacy requirements. Based on the definition of two conflicting metrics, 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 effective algorithms, GreedyDP and GreedyIL, to support runtime profiling. While), the latter attempts to minimize the information loss (IL) and the former tries to maximize the discriminating power (DP. GreedyIL outperforms GreedyDP significantly by exploiting a number of heuristics. We provide an effective mechanism for the client to decide whether there query want to personalize or not. This decision making can be done before each runtime profiling to enhance the ability of search result and to avoid the unnecessary exposure of the profile. The proposed architecture is shown in figure 3.1.

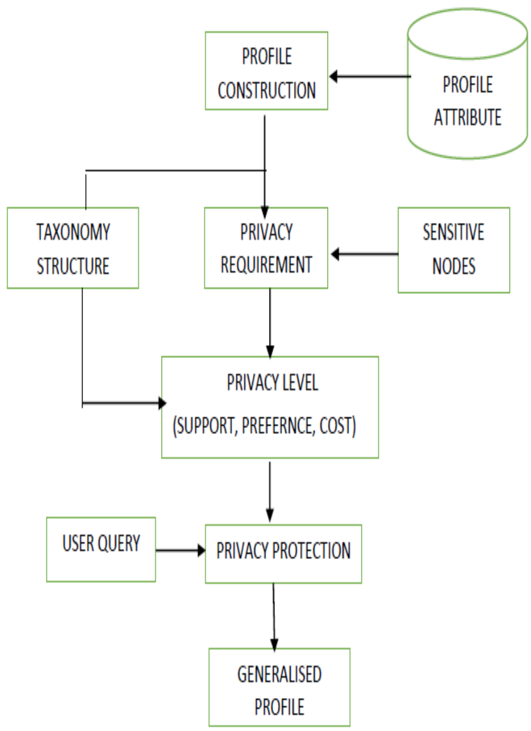

Fig. 3.1: Proposed Architecture

3.1 Data Flow Diagram

Data Flow Diagram (DFD) is a graphical representation of the flow of data through a system. It consists of data flows, processes, sources, destinations and stores. It is a control tool and tools are based upon it. Logical Data Flow Diagram shows the transformation of data from input to output through process. It shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of process or information about whether processes will operate in sequence or in parallel. The data flow diagrams is shown in figure 3.2.

.

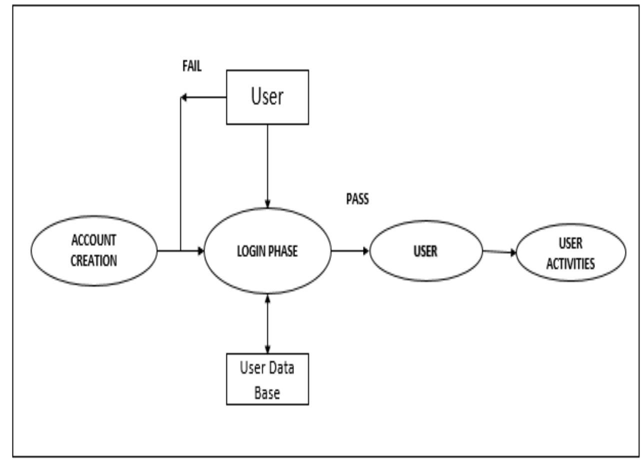

Fig 3.2 : Data Flow Diagram

3.2 Use Case Diagram

Use case diagrams are usually referred to as behaviour diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more external users of the system (actors). Each use case should provide some observable and valuable result to the actors or other stakeholders of the system. The user view represents the goal and objectives of various users and their requirements from the system. The user view represents that part of the system with which the user interacts. Actors are elements that lie outside the system. They are used to model system users. Actors interact directly with the system, exchanges messages and request for some actions to be performed. The Use Case diagram of the proposed system is shown in the figure 3.3.

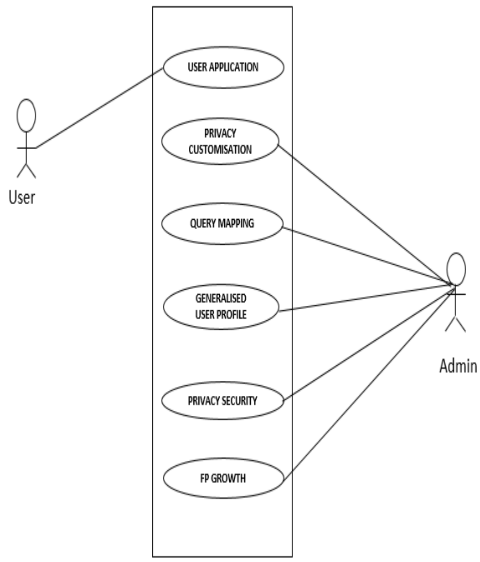

Fig 3.3: Use Case Diagram

3.3 Algorithm

A greedy algorithm is 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.

3.3.1 Greedy Discriminating Power (Dp) Algorithm

This algorithm works in a bottom up manner. Starting with the leaf node, for every repetition, it chooses leaf topic for clipping thus trying to maximize convenience of output. During repetition a best profile-so- far is maintained satisfying the Risk constriction. The repetition stops when the root topic is reached. The best profile-so-far is the final result. GreedyDP algorithms require recompilation of profiles which adds up to computational cost and memory requirement**.**

3.3.2 Greedy Information Loss (IL) Algorithm

GreedyIL algorithm is advances simplification productivity. GreedyIL continues importance queue for candidate clip leaf operator in descending order. This decreases the computational cost. GreedyIL states to dismiss the repetition when Risk is satisfied or when there is a single leaf left. Since, there is less computational cost compared to GreedyDP, GreedyIL out performs GreedyDP.

3.3.3. FP Growth

The FP-Growth Algorithm, proposed by Han in [1], is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth, using an extended prefix-tree structure for storing compressed and crucial information about frequent patterns named frequent-pattern tree. Here FP-Growth is used to develop frequent patterns of query and used to identify its risk factor. So it will improve the protection and reduces the time complexity.

4. Conclusion

A client-side privacy protection framework called UPS for personalized web search. It can be potentially adopted by any PWS that captures user profiles in a hierarchical taxonomy. And framework allows users to specify customized privacy requirements via the hierarchical profiles. It also performs online generalization on user profiles to protect the personal privacy without compromising the search quality. Two greedy algorithms, namely Greedy DP and Greedy IL are used for the online generalization.

ACKNOWLEDGMENT

First of all, we thank ALMIGHTY for giving us strength and courage for taking a relevant area and to do research on that. We thank everyone, who directly or indirectly helped us for doing the thesis.

REFERENCES

[1] Z. Dou, R. Song, and J.-R. Wen, “A Large-Scale Evaluation and Analysis 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 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, 2005.
[3] M. Spertta and S. Gach, “Personalizing Search Based on User Search Histories,” Proc. IEEE/WIC/ACM Int’l Conf. Web Intelligence (WI), 2005.
[4] B. Tan, X. Shen, and C. Zhai, “Mining Long-Term Search History to Improve Search Accuracy,” Proc. ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD), 2006.
[5] K. Sugiyama, S.T. Dumais, and E. Horvitz, “Adaptive Web Search Based on User Profile Constructed without

any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW), 2004.

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

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

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

[9] 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.

[10] A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services," J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.

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

[12] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschu¨tter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.

[13] A. Pretschner and S. Gauch, "Ontology-Based Personalized Search and Browsing," Proc. IEEE 11th Int'l Conf. Tools with Artificial Intelligence (ICTAI '99), 1999.

[14] E. Gabrilovich and S. Markovich, "Overcoming the Brittleness Bottleneck Using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge," Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI), 2006.

[15] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia," HP Labs, 2008.

[16] K. Ja¨rvelin and J. Keka¨la¨inen, "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, 2000.

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

[18] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.

[19] Y. Xu, K. Wang, G. Yang, and A.W.-C. Fu, "Online Anonymity for Personalized Web Services," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.

[20] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing User Profiles for Personalized Web Search," Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.

[21] J. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32 , no. 13/14, pp. 1541-1551, 2009.

[22] A. Viejo and J. Castella-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.

[23] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.

[24] R. Chan, Q. Yang, and Y. Shen, "Mining high utility itemsets," Proc. of IEEE Int'l Conf. on Data Mining (ICDM), pp. 19-26, Nov., 2003.