© 2015 IJSRSET | Volume 1 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Personalizing Search Based on user Search Histories

Thenmozhi. M, Swathishri. J, Nivedha. A, Kalaiselvi. A
Dhanalakshmi College of Engineering, Chennai, Tamilnadu, India

ABSTRACT

In improving the quality of various search services on the Internet, Individualized web search (IWS) has demonstrated its effectiveness. User preferences are modelled as hierarchical user profiles in IWS applications. We propose an IWS framework called UPS that can adaptively generalize profiles by queries. Our runtime generalization evaluates the utility of personalization and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. For deciding whether personalizing a query is beneficial, we also provide an online prediction mechanism. The experimental results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.

Keywords: GreedyIL, GreedyDP, Individualized web search, profile based methods, log based, UPS, IWS

I. INTRODUCTION

The search engine used in the web is one of the most important tools to gather information on the internet. But most of the time, users' may experience inconvenience due to irrelevant result shown to them. This is due to users' backgrounds and context, and also typing errors made by the user. To provide better result to the individual users, we have created the Individualized web search (IWS) technique. The main idea behind this technique is that information of the user is collected and analyzed to find out the intended user query.

IWS consists of two types of solutions. They are click-log-based and profile-based methods.

Due to increasing usage of personal and behaviour to profile its users, information are collected from click-through data[1],7],[8], query history[2],[3],[4],browsing history[5],[6], bookmarks[9], user documents[10],[2], etc.,

These implicitly collected personal data leads to high risk of revealing the users' personal information. It increases the fear of users in the protection of data. Two things are considered to preserve the user's privacy in profile-based IWS. Scientist states that in these two things, one of the most important factors is that the quality of search process should be improved with personalization utility of the user profile. And also the second one states that users' personal information should be hidden to keep the risk under control. For this, the research of the studies tells that people are ready to compromise on their privacy to yield better search quality. This is achieved by generalization of the user profile.

II. METHODS AND MATERIAL

A. Existing System

Individualized web search (IWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. User information has to be collected and analysed to figure out the user intention behind the query. Techniques used are click log based and profile based methods. Click-log based simply imposes bias to clicked pages in the user's query. Profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. The drawbacks in this are no customization allowed and only work on repeated queries from the same user which is a strong limitation.

B. Proposed System

We propose an IWS framework called UPS (User customizable Privacy-preserving Search) that can adaptively generalize profiles by queries while respecting user specified privacy requirements. We also present two greedy algorithms, namely Greedy DP and Greedy IL, for runtime generalization. The technique used is privacy-preserving individualized web search framework called UPS. It generalize profiles by queries while respecting user specified privacy requirements and also aims at protecting the privacy in individual user profiles while retaining their usefulness for IWS.

It works on different types of queries from user and also customization of privacy requirements is available. The framework works in two phases, namely the offline and online phase, for each user. Queries with smaller click-entropies, namely distinct queries, are expected to benefit more from personalization, while those with larger values (ambiguous ones) are not. To overcome this we go for UPS based user customization. The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself.

III. RESULTS AND DISCUSSION

To connect with server user must give their username and password then only they can able to connect the server. If the user already exits directly can login into the server else user must register their details such as username, password and Email id, into the server. Server will create the account for the entire user to maintain upload and download rate.

In query processing, the data is given by customer requests goes to server, when a user issues a query 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 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 were administrator contains all files and responsible for storing that files into cloud.

In combining user profile and query module, user given query and the generalized user profile are sent together to the IWS server for individualized search. Query with related user preferences stored in a user profile with the aim of providing better search results.
In online generalization, user given query based on privacy requirements and cost of profiling search results are checked whether to personalize or not. In search personalization, user given query search results are individualized according to user profile and delivered back to the query proxy. After that results are shown to user.

ALGORITHM USED

A. GreedyDP Algorithm

The first greedy algorithm GreedyDP works in a bottom up manner. We introduce an operator called prune-leaf, which indicates the removal of a leaf topic t from a profile. Obviously, the optimal profile can be generated with a finite-length transitive closure of prune-leaf. Starting in every ith iteration, GreedyDP chooses a leaf topic t for pruning, trying to maximize the utility of the output of the current iteration. During the iterations, we also maintain a best profile-so-far, having the highest discriminating power while satisfying the risk constraint. The iterative process terminates when the profile is generalized to a root-topic. The best-profile-so-far will be the final result of the algorithm.
The main problem of GreedyDP is that it requires recompilation of all candidate profiles. This causes significant memory requirements and computational cost.

B. GreedyIL Algorithm

This algorithm improves the efficiency of the generalization using heuristics based on several findings. One important finding is that any prune-leaf operation reduces the discriminating power of the profile. Algorithm 1 shows the pseudo-code of the GreedyIL algorithm. In general, GreedyIL traces the information loss instead of the discriminating power. This saves a lot of computational cost.

GreedyIL has computational complexity. However, this is extremely rare in practice. Therefore, GreedyIL is expected to significantly outperform GreedyDP.

significantly outperform GreedyDP.

Algorithm 1: GreedyIL($\mathcal{H}$, q, δ)

Input : Seed Profile $\mathcal{G}_0$; Query q; Privacy threshold δ
Output: Generalized profile $\mathcal{G}*$ satisfying δ-Risk

```
let Q be the IL-priority queue of prune-leaf decisions;
    i be the iteration index, initialized to 0;
// Online decision whether personalize q or not
if DP(q, R) < μ then
    Obtain the seed profile G_0 from Online-1;
    Insert ⟨t, IL(t)⟩ into Q for all t ∈ T_H(q);
    while risk(q, G_i) > δ do
        Pop a prune-leaf operation on t from Q;
        Set s ← par(t, G_i);
        Process prune-leaf G_i --t--> G_{i+1};
        if t has no siblings then                    // Case C1
            Insert ⟨s, IL(s)⟩ to Q;
        else if t has siblings then                  // Case C2
            Merge t into shadow-sibling;
            if No operations on t's siblings in Q then
                Insert ⟨s, IL(s)⟩ to Q;
            else
                Update the IL-values for all operations on
                t's siblings in Q;
        Update i ← i + 1;
    return G_i as G*;
return root(R) as G*;
```

IV. CONCLUSION

This paper presented a privacy protection framework called UPS for individualized web search which could provide customized privacy requirements. UPS also performed online generalization on user profiles to protect the privacy without compromising the search quality. We proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. UPS could achieve quality search results while preserving user's customized privacy requirements. Maintaining user profiles and privacy are used to predict the performance of the user profiles.

V. REFERENCES

[1] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Individualized 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, K. Hatano, and M. Yoshikawa, "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 Individualized 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 Individualized 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, "Individualized Search," Comm. ACM, vol. 45, no. 9, pp. 50-55, 2002.

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

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

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

[13] 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.

[14] 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.

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

[16] 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.

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

[18] 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.

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

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

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

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

[23] 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.

[24] 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.

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

[26] J. Teevan, S.T. Dumais, and D.J. Liebling, "To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 163-170, 2008.

[27] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "Ups: Efficient Privacy Protection in Individualized web Search," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information, pp. 615624, 2011.

[28] J. Conrath, "Semantic Similarity based on Corpus Statistics and Lexical Taxonomy," Proc. Int'l Conf. Research Computational Linguistics (ROCLING X), 1997.

[29] D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, "Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies," Proc. Int'l Conf. Web Search and Data Mining (WSDM), pp. 139-148, 2008.