ISSN:2348-2079

Volume-5 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Improve secure personal web search temporal behavior search over re-ranking with catch services model

[1]R.Navinkumar.,[2] D.Uvarani

[1] Assistant professor, Department of MCA.,

[2] PG Scholar Nandha Engineering College, Erode

Email: uvaranidamu95@gmail.com

ABSTRACT

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. Personalized web search (PWS) 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.Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. This research studies privacy protection in PWS applications that model user preferences as hierarchical user profiles. This project proposes a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user-specified privacy requirements. The proposed runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. The study presents two greedy algorithms, namely Greedy DP and Greedy IL, for runtime generalization. It also provides an online prediction mechanism for deciding whether personalizing a query is beneficial. In addition, research study to improve personalized web search model using re-ranking with catch services algorithm in web database.

INTRODUCTION

1.1 INTRODUCTION ABOUT DATA MINING

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

1.2 The Foundations of Data Mining

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery.

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996.1 In some industries, such as retail, these numbers can be much larger.

OBJECTIVE OF THESIS WORK

The specific objectives of the research work are

- Query search Keyword expansion
- Security query expansion
- Word pool expansion.
- The re-ranking result is not efficient in terms of more than two feature spaces.
- Requires training a specific search similarity for every example in the word pool, which is assumed to be fixed.
- The users' intention could not be accurately captured when the semantic meanings of the query keywords had large diversity.
- During keyword expansion, the synonym words are not added in the suggestion.

2. LITERATURE SURVEY

Zhicheng Dou, Ruihua Song and Ji-Rong Wen [1] studies this problem and provides some preliminary conclusions. They present a large-scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies (including two click-based and three profile-based ones) using MSN query logs. By analyzing the results, they reveal that personalized search has significant improvement over common web search on some queries but it has little effect on other queries (e.g., queries with small click entropy). It even harms search accuracy under some situations. They also reveal that both long-term and short-term contexts are very important in improving search performance for profile-based personalized search strategies.

- Personalization may lack eff ectiveness on some queries, and there is no need for personalization on such queries.
- Different strategies may have variant effects on different queries. In such a case, simply leveraging pages visited by this user in the past may achieve better performance

Mirco Speretta [6] describe to creating user profiles collect user information through proxy servers (to capture browsing histories) or desktop bots (to capture activities on a personal computer). Both these techniques require participation of the user to install the proxy server or the bot. In this study, they explore the use of a less-invasive means of gathering user information for personalized search. In particular, they build user profiles based on activity at the search site itself and study the use of these profiles to provide personalized search results. User profiles were created by classifying the collected information (queries or snippets) into concepts in a reference concept hierarchy. In general, personalization can be applied to search in two different ways:

To providing tools that help users organizing their own past searches, preferences, and visited URLs;

To creating and maintaining sets of user's interests, stored in profiles, that can be used by retrieval process of a search engine to provide better results.

The first approach is applied by many new toolbars and browser add-ons. The Seruku Toolbar [8] and the SurfSaver are examples of tools that try to help users to organize their search histories in a repository of URLs and web pages visited. Furl [9] is another personalization tool that storesweb pages including topics which users are interested in, however it was developed as a server-side technology rather than a desktop toolbar.

Bin Tan, Xuehua Shen, ChengXiang Zhai [10] describes the Long-term search history contains rich information about a user’s search preferences, which can be used as search context to improve retrieval performance. They propose mixture models to represent a user’s information need and apply

statistical language modeling techniques to discover relevant context from the search history, and exploit it to obtain improved estimates of the query model. They then evaluate the methods on a test set of Web search histories collected from some real users.

Kazunari Sugiyama, Kenji Hatano and Masatoshi Yoshikawa [11] Web search engines help users find useful information on the World Wide Web WWW). However, when the same query is submitted by different users, typical search engines return the same result regardless of who submitted the query.

- Relevance feedback and implicit approaches,
- User profiles based on pure browsing history, and
- User profiles based on the modified collaborative filtering.

This approach allows us to construct a more appropriate user profile and perform a fine grained search that is better adapted to each user's preferences. In the future, if broadband networks spread widely, information is expected to be provided in a variety of forms such as music, movies and so on.

Xuehua Shen, Bin Tan and ChengXiang Zhai describe a major limitation of most existing retrieval models and systems is that the retrieval decision is made based solely on the query and document collection; information about the actual user and search context is largely ignored.

The current work can be extended in several ways: First, they have only explored some very simple language models for incorporating implicit feedback information. It would be interesting to develop more sophisticated models to better exploit query history and click through history. For example, they may treat a clicked summary differently depending on whether the current query is a generalization or refinement of the previous query.

SYSTEM DESIGN AND IMPLEMENTATION

The existing system works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. The online phase handles queries as follows:

- When a user issues a query qi 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 Gi satisfying the privacy requirements. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search.
- No capability to capture a series of queries.
- User profile is categorized into single node in the tree structure only.
- Past query based suggestion is not given to user.

In addition with all the existing system activities, the proposed system also includes the new following processes. Users may be grouped in more than one profile so that the result includes both types of search outputs. In addition, previous query based suggestions are also provided.

- Capability to capture a series of queries.
- User profile is categorized into multiple nodes in the tree structure.
- Past query based suggestion is given to user.

SPECIFIC OBJECTIVE

- The average numbers of the relevant shots at different depths of the result list are used as the evaluation criteria.
- The fusion is carried out to the intermediate clusters only and so the calculation time and overhead is reduced much.
- Unwanted Web page links are eliminated to better extent.
- Time Reduction in searching and result Efficiency is improved
- Accurate Re-ranking of result and Multiple feature space based search is possible.

METHODOLOGY

3.3.1. PROFILE-BASED PERSONALIZATION

Previous works on profile-based PWS mainly focus on improving the search utility. The basic idea of these works is to tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal. The hierarchical profile automatically via term-frequency analysis on the user data and proposed UPS framework, do not focus on the implementation of the user profiles.

3.3.2 PRIVACY PROTECTION IN PWS SYSTEM

The third and fourth levels are impractical due to high cost in communication and cryptography. As a result any entity cannot profile a certain individual. Social networks instead of the third party to provide a distorted user profile to the web search engine. One main limitation in this work is that it builds the user profile as a finite set of attributes, and the probabilistic model is trained through predefined frequent queries.

3.3.3 USER PROFILE

A diagram of a sample user profile is illustrated in Fig. 2a, which is constructed based on the sample taxonomy repository in Fig. 2b. We can observe that the owner of this profile is mainly interested in Computer Science and Music, because the major portion of this profile is made up of fragments from taxonomies of these two topics in the sample repository.

Although a user profile H inherits from R a subset of topic nodes and their links, it does not duplicate the repository supports. Instead, each topic t 2H is labeled with a user support, denoted by sup(H, t), which describes the user’s preference on the respective topic t.

3.3.4 ATTACK MODEL

The proposed work aims at providing protection against a typical model of privacy attack, namely eavesdropping. As shown in Fig. 3, to corrupt Alice's privacy, the eavesdropper Even successfully intercepts the communication between Alice and the PWS-server via some measures, such as man-in-the-middle attack, invading the server, and so on.

3.3.5 PROFILE CONSTRUCTION

In this module, a **hierarchical user profile is constructed** and **customized with the user-specified privacy requirements**.

The first step of the offline processing is to build the original user profile in a topic hierarchy H that reveals user interests. We assume that the user's preferences are represented in a set of plain text documents, denoted by D. To construct the profile, we take the following steps

1. Parse the query.
2. Check the category names which contains query words.
3. Construct the profile tree which contains the category ids mentioned in user/category records along with children

This module handles queries as follows:

1. When a user issues a query qi 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 Gi satisfying the privacy requirements.

2. Subsequently, the **query** and the **generalized user profile** are **sent together to the PWS server for personalized search**.

CONCLUSION

The proposed system is 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. The framework allowed users to specify 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. It proposed a greedy algorithm, namely Greedy IL, for the online generalization. The experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution. The main benefits are capability to capture a series of queries, User profile is categorized into multiple nodes in the tree structure and past query based suggestion is given to user.

FUTURE ENHANCEMENT

At present, the project presented a client-side privacy protection framework called UPS for personalized web search. For future work, the thesis will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentially, and so on), or capability to capture a series of queries (relaxing the

second constraint of the adversary) from the victim. It will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

REFERENCES

1. R. KROVETZ AND W. B. CROFT. LEXICAL AMBIGUITY AND INFORMATION RETRIEVAL. INFORMATION SYSTEMS, 10(2):115–141, 1992.

2. J. TEEVAN, S. T. DUMAIS, AND E. HORVITZ. BEYOND THE COMMONS: INVESTIGATING THE VALUE OF PERSONALIZING WEB SEARCH. IN PROCEEDINGS OF PIA '05, 2005.

3. X. SHEN, B. TAN, AND C. ZHAI. IMPLICIT USER MODELING FOR PERSONALIZED SEARCH. IN PROCEEDINGS OF CIKM '05, PAGES 824–831, 2005.

4. X. SHEN, B. TAN, AND C. ZHAI. CONTEXT-SENSITIVE INFORMATION RETRIEVAL USING IMPLICIT FEEDBACK. IN PROCEEDINGS OF SIGIR '05, PAGES 43–50, 2005.

5. M. SPERTTA AND S. GACH, "PERSONALIZING SEARCH BASED ON USER SEARCH HISTORIES," PROC. IEEE/WIC/ACM INT'L CONF. WEB INTELLIGENCE (WI), 2005.

6. ONSTAT. MOST PEOPLE USE 2 WORD PHRASES IN SEARCH ENGINES ACCORDING TO ONESTAT.COM, FEBRUARY 2004. HTTP://WWW.ONSTAT.COM.

7. SERUKU. SERUKU TOOLBAR, 2004. HTTP://WWW.SERUKU.COM/INDEX.HTML.

8. MIKE GILES. FURL, 2003. HTTP://WWW.FURL.NET.

9. 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.

10. J. ALLAN ET AL. CHALLENGES IN INFORMATION RETRIEVAL. IN SIGIR FORUM, VOLUME 37, 2003.

11. T. JOACHIMS. OPTIMIZING SEARCH ENGINES USING CLICKTHROUGH DATA. IN PROCEEDINGS OF SIGKDD 2002, PAGES 133–142, 2002.

12. X. SHEN, B. TAN, AND C. ZHAI. UCAIR TOOLBAR: A PERSONALIZED SEARCH TOOLBAR (DEMO). IN PROCEEDINGS OF SIGIR 2005, PAGE 681, 2005.

13. X. SHEN, B. TAN, AND C. ZHAI. CONTEXT-SENSITIVE INFORMATION RETRIEVAL USING IMPLICIT FEEDBACK. IN PROCEEDINGS OF SIGIR 2005, PAGES 43–50, 2005.

14. 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.

15. R. COOLEY, B. MOBASHER, AND J. SRIVASTAVA. DATA PREPARATION FOR MINING WORLD WIDE WEB BROWSING PATTERNS. KNOWLEDGE AND INFORMATION SYSTEMS, 1(1):5–32, 1999.

16. C. SHAHABI AND Y.-S. CHEN. AN ADAPTIVE RECOMMENDATION SYSTEM WITHOUT EXPLICIT ACQUISITION OF USER RELEVANCE FEEDBACK. DISTRIBUTED AND PARALLEL DATABASES, 14(3):173–192, 2003.

17. K. BHARAT, T. KAMBA, AND M. ALBERS. PERSONALIZED INTERACTIVE NEWS ON THE WEB. MULTIMEDIA SYSTEMS, 6(5):349–358, 1998.

18. U. MANBER, A. PATEL, AND J. ROBISON. EXPERIENCE WITH PERSONALIZATION ON YAHOO! COMMUNICATIONS OF THE ACM, 43(8):35–39, 2000.

    FIELD, P. HARTEL, AND W. MOOIJ. PERSONAL DJ, AN ARCHITECTURE FOR PERSONALISED CONTENT DELIVERY. IN PROC. OF THE 10TH INTERNATIONAL WORLD WIDE WEB CONFERENCE (WWW10), PAGES 1–8, 2001.

19. 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.

20. G. NUNBERG. AS GOOGLE GOES, SO GOES THE NATION. NEW YORK TIMES, MAY 2003.

21. 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.

22. J. Allan and et al. Challenges in information retrieval and language modeling. Workshop at University of Amherst, 2002.

23. J. Rocchio. Relevance feedback information retrieval. In The Smart Retrieval System-Experiments in Automatic Document Processing, pages 313–323, Kansas City, MO, 1971. Prentice-Hall.

24. D. Kelly and N. J. Belkin. Display time as implicit feedback: Understanding task effects. In Proceedings of SIGIR 2004, 2004.

25. C. Huang, L. Chien, and Y. Oyang. Query session based term suggestion for interactive web search. In Proceedings of WWW 2001, 2001.

26. R. W. White, J. M. Jose, C. J. van Rijsbergen, and I. Ruthven. A simulated study of implicit feedback models. In Proceedings of ECIR 2004, pages 311–326, 2004.