ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Personalizing Web Search based on User Profile

[1] Sharyu Utage, [2] Vijaya Ahire

[1,2] Computer Science and Engineering, Dr BAMU
Jawaharlal Nehru Engineering College, Aurangabad.

Abstract - Web Search engine is most widely used for information retrieval from World Wide Web. These Web Search engines help user to find most useful information. When different users Searches for same information, search engine provide same result without understanding who is submitted that query. Personalized web search it is search technique for proving useful result. This paper models preference of users as hierarchical user profiles. a framework is proposed called UPS. It generalizes profile and maintaining privacy requirement specified by user at same time.

Keywords - *Profile, Privacy Protection, Personalized Web Search, UPS, Generalization, Query.*

1. Introduction

The web search engine has more popularity and importance for users information need and retrieval of useful data on the web. As the information present on web is very huge many times user might experience failure when an irrelevant result of user query is provided by search engine many of queries to search engines are short and ambiguous, and different users may have completely different information needs in same query For example, a biologist may use query “mouse” to get information about rodents, while programmers may use the same query to find information about computer peripherals.

The personalization is used for providing better search result as per user need. In personalization the User information is collected and analysis is done and according to that the search result are provided. The personalized web Search[1][2] is mainly of two categories Click-log-based method and Profile-based method. Click-log-based method is simple which it simply considers clicked pages in the users query history. Although it working well, but it work on repeated query by the same user which is big limitation of its applicability. In profile based method the improvement is done on the search experiences by user profiling technique. The overall goal of the data mining process [5] is to extract information from a data set and transform it into an understandable structure for further use. Profile based PWS is more effective for improving search result although both the techniques have the different advantage and disadvantage. The user profile is generated from the information collected from the query history, browsing data, bookmarks etc[3] but such user information can be retrieved from users private life.

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[7]. In our 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 our proposed system they are using feedback session which will reduce the noise by using the information from directly click troughs. 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.

Hence they proposed another representation method, pseudo-document. To know the keywords from the query they have to cluster the pseudo-documents. For clustering, in existing method they use K-Means clustering algorithm which is simple and effective. But it is computationally

difficult and the similarity between the pseudo terms is also important while clustering. In our proposed system they can call keywords as rifle text.

Generally, web search session contains the series of successive queries to satisfy a single need and some clicked search results. In this paper, we focus on assuming user rifle goals for a specific query. Therefore, the single session introduces only one query, which differentiates from the predictable session. The feedback session in this paper is based on a single session. And in this paper the feedback session will plays a major role.

The feedback session defines of both clicked and unclicked URLs and ends with the last URL that was clicked in a single session. It was motivated that before the last click, all the URLs have been scanned and evaluated by users. Therefore, further the clicked URLs, the unclicked ones before the last click should be a part of the user feedbacks.

For a query, users will usually have some unclear keywords representative their interests in their minds. They use these keywords to determine whether a document can satisfy their needs. We name these keywords "rifle texts". However, while rifle texts can reflect user needs of information, they are underlying and not expressed explicitly. Therefore, the feedback session can represented as pseudo-documents as substitutes to approximate rifle texts. Thus, pseudo-documents can be used to assume user rifle goals.

The rest of this paper is organized as follows. The next section briefly describes the related work. Section 3 discusses the Euclidean distance, Cosine similarity and SMTP distance measures and their semantics. Experimental results are presented in Section 4. Finally, concluding remarks are given Section 5.

2. Related Work

Previous Work on the profile-based personalized web search focused on the improving the search facility for the user. The previous work is done by two aspects, profile representation and measure of effectiveness of personalization. The recent work create profile in hierarchical Structure this Hierarchical structure constructed with hierarchy/graph, such as Wikipedia or the hierarchical profile is generated via term-frequency analysis on the user data. UPS framework can use any hierarchical representation. For the privacy protection problem in PWS application Two classes are found First class use individual identification as privacy Whereas other Consider data sensitivity as privacy. Literature works in class one solves the different level privacy problem which consist of different identity as pseudo identity, group identity, no identity, no personal information. PWS Online anonymity provide anonymity by creating a group profile of k users. By using this the relation between query and single user is made. The shortcomings of class one solution is the high cost. In class two solution, it is more self-dependent and doesn't give their complete profiles to anonymity server. Krause and Horvitz developed statistical techniques to learn a probabilistic model, and use this model to generate the near-optimal partial profile. Privacy increasing in personalized web search based on hierarchical profiles proposed a privacy protection solution for PWS. a generalized profile is obtained in effect as a rooted sub tree of the complete profile.

Context Sensitive Information Retrieval using Implicit feedback [7] [8]by Xuehua Shen & Bin Tan uses click log for improving the retrieval accuracy in an interactive information retrieval setting. Personalizing Search via Automated Analysis of Interests and Activities[3] by Jaime Teevan, Susan T. Dumais uses click log method to re-rank the web results. Adaptive Web Search Based on User Profile Generated without Any Effort from Users[6] by Kazunari Sugiyama, Kenji Hatano uses profile based method for analysis of user's browsing history in one day. For Personalized Search Identification of User Interest [8] by Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou uses click log based method to learn a user's preference based on past click history and how it can use the user preference to personalize search results.

This paper provides personalized privacy protection in PWS. Thus, this paper allows user to customize privacy requirements in hierarchical user profiles.

3. Proposed System

In this paper, they have intention at determining the number of various user rifle goals for a query and representing each goal with some keywords automatically. They first suggest a new approach to assume user rifle goals for a query by clustering the suggested feedback sessions. Then, they suggest a new optimization method to map feedback sessions to pseudo-documents. It reflects user information needs efficiently. At last, they cluster these pseudo documents to assume user rifle goals and show them with some keywords automatically.

The three major contributions as follows: i. they suggesting a frame work to assume different user rifle

goal line for a query by clustering the feedback sessions. ii. They suggest a new optimization method to chain the URLs in a feedback session to form a pseudo-document. iii. They suggest a new criterion CAP to evaluate user rifle goals assumption based restructuring network rifle results[12].

The system architecture shows the framework of our approach. The framework consists of two parts divided by the dashed line. In the upper part, all the feedback sessions of a query are first extracted from user click-through logs and mapped to pseudo-documents. Then, user search goals are inferred by clustering these pseudo-documents and depicted with some keywords. Since we do not know the exact number of user search goals in advance, several different values are tried and the optimal value will be determined by the feedback from the bottom part.

In the bottom part, restructured based on the user search goals inferred from the upper part taken from the original search results are. Then, evaluate the performance of restructuring search results by our proposed evaluation criterion CAP. And the evaluation result will be used as the feedback to select the optimal number of user search goals in the upper part.

The inferred user search goals are represented by the vectors and the feature representation of each URL in the search results can be computed. Then, categorize each URL into a cluster centered by the assumed rifle goals. In this paper, categorization is takes place by choosing the smallest distance between the URL vector and user-rifle-goal vectors. By this way, the search results can be restructured according to the assumed user rifle goals.

3.1 Merites

The three major contributions as follows:

1. We suggesting a frame work to assume different user rifle goal line for a query by clustering the feedback sessions.
2. We suggest a new optimization method to chain the URLs in a feedback session to form a pseudo-document.
3. We suggest a new criterion CAP to evaluate user rifle goal assumption based restructuring network rifle results.

3.2 Architecture

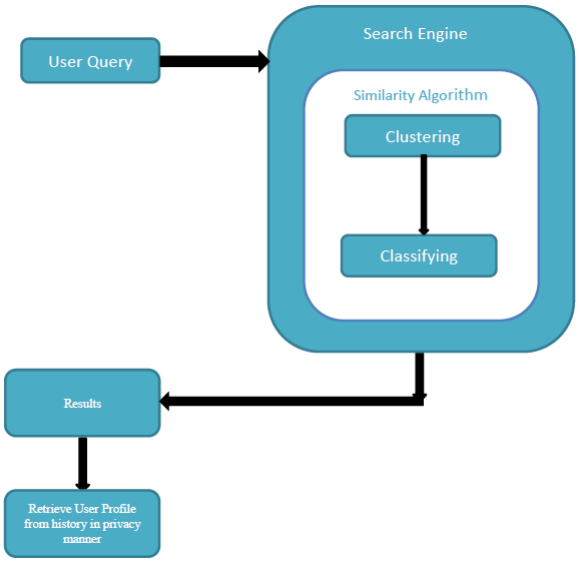

Fig.1 System Architecture

3.3 Module Descriptions

Dataset Preprocessing
In this module, 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.

User Login
In this module, 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.

Query Submission and Query Retrieval Based on User
In this module, 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.

Estimate Relevant Results
In this module, 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.

Retrieve User Profile in Privacy Manner

In this module, some adversaries to mine the history results means, only query time has been displayed. In this, other informations such as query, query results, username are not displayed by using the background knowledge.

4. System Requirements

4.1 Hardware Required

1. System: Pentium IV 2.4 GHz
2. Hard Disk: 40 GB
3. Floppy Drive: 1.44 MB
4. Monitor: 15 VGA colour
5. Mouse: Logitech.
6. RAM: 256 MB

4.2 Software Required

1. O/S : Windows XP.
2. Language: Java
3. Data Base: MySQL
4. IDE: NetBeans IDE 6.9.1

5. Evaluation

Finally, in this module we can evaluate the performance by using the parameter such as time. In the existing system, there is large amount of time has been taken to retrieve the results. In the proposed system, there is time consuming is less when compare with the existing system.

5.1 Result

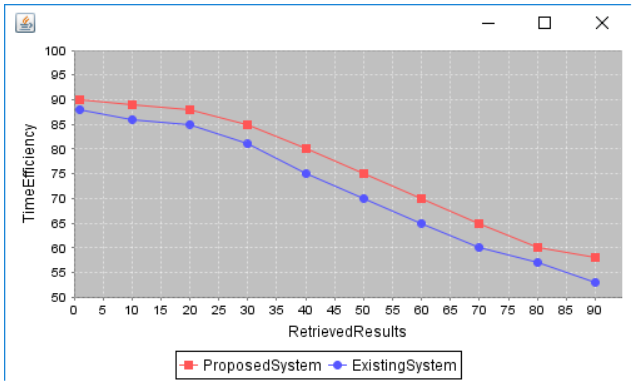

Fig. 2: Result

6. 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.

7. Future Scope

The algorithms and the data sets adopted are intended to be popular and easily accessible for anyone interested in this research area. However, it would be of greater value evaluating the performance of the measures on larger test-beds. It would be interesting to investigate the effectiveness and to try resist adversaries with border background knowledge such as richer relationship among topics such as exclusiveness, sequentially and so on or capability to capture series of queries from victim. It will be more interesting to seek more sophisticated method to build the user profile and better metrics to predict the performance of UPS.

References

[1] Lidan Shou,He Bai, Ke Chen,and Gang Chen, "Supporting privacy protection in personalized web search " ieee transactions on knowledge and data engineering vol:26 no:2 year 2014

[2] 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.

[3] 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

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

[5] 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

[6] 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

[7] 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.

[8] 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

[9] 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

[10] 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.

[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 Personalized Search and Browsing," Proc. IEEE 11th Int'l Conf. Tools with Artificial Intelligence (ICTAI '99), 1999.