IJDCST @Aug-Sept-2015, Issue- V-3, I-6, SW-13
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Personalized Web Search using advanced string search

Kotireddy Gurram[1], Mr. T Seshu Chakravarthy[2], T.Surekha[3]

[1] Student(Mtech),

[2]Assistant Prof, CSE department, CSE Department, Narsaraopet Engineering College, JNTU Kakinada.

[3]Associate Prof, CSE department, CSE Department, Narsaraopet Engineering College, JNTU Kakinada.

ABSTRACT: To improve the quality of various searching process in web the new searching process is called Advanced Personalized web search (PWS). Providing the privacy to the users is main aim in PWS. Existing system, PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified privacy requirements. There are two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. To improve the performance of the PWS we adopted advanced Personalized web search (APWS). Results will show the performance of the proposed system.

Keywords: Personalized web search (PWS), GreedyDP, GreedyIL.

Introduction:

One criticism of search engines is that when queries are issued, most return the same results to users. In fact, the vast majority of queries to search engines are short [12] and ambiguous [16, 7], and different users may have completely different information needs and goals under the same query [12]. 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. When such a query is submitted to a search engine, it takes a moment for a user to choose which information he/she wishes to get. On the query "free mp3 download", the users' selections can also vary though almost all of them are finding some websites to download free mp3: one may select the website "www.yourmp3.net", while another may prefer the website "www.seekasong.com". Personalized search is considered a solution to this problem since different search results based on preferences of users are provided. Various personalization strategies including [14, 9, 19] have been proposed, and personalized web search systems have been developed, but they are far from optimal. One problem of current personalized search is that most proposed methods are uniformly applied to all users and queries. In fact, we think that queries should not be handled in the same manner because we find: (1) Personalization may lack effectiveness on some queries, and there is no need for personalization on such queries. This has also been found . For example, he query "mouse" mentioned above, using personalization based on user interest profile, we could achieve greater relevance for individual users than common web search. Beyond all doubt, the personalization brings significant benefit to users in this case. Contrarily, for the query "Google", which is a typical navigational query as defined in [3, 17], almost all of the users are consistently selecting results to redirect to Google's homepage, and therefore none of the personalized strategies could provide significant benefits to users. (2) Different strategies may have variant effects on different queries. For the query "free mp3 download", using

the typical user interest profile-based personalization such as the method proposed in [6], which led to better results for the query "mouse", we may achieve poor results because the results for query "free mp3 download" are mostly classified into one topic category and the profile-based personalization is too coarse to filter out the desired results. In such a case, simply leveraging pages visited by this user in the past may achieve better performance. Furthermore, simply applying one personalization strategy on some queries without any consideration may harm user experience. For example, when a sports fan submits the query "office", he/she may not be seeking information on sports, but may be seeking help on Microsoft Office Software or any other number of office-related inquiries. In this situation, if interest-based personalization is done, many irrelevant results could erroneously be moved to the front and the user may become confused. [3] Personalization strategies may provide different effectiveness based on different search histories and under variant contexts. For example, it could be difficult to teach interests of users who have done few searches. Furthermore, as Shen noted, users often search for documents to satisfy short-term information needs, which may be inconsistent with general user interests. In such cases, long-term user profiles may be useless and short-term query context may be more useful.

Related Work:

There are several prior attempts on personalizing web search. One approach is to ask users to specify general interests. The user interests are then used to filter search results by checking content similarity between returned web pages and user interests. For example, [6] used ODP2 entries to implement personalized search based on user pro- files corresponding to topic vectors from the ODP hierarchy. Unfortunately, studies have also shown that the vast majority of users are reluctant to provide any explicit feedback on search results and their interests [4]. Many later works on personalized web search focused on how to automatically learn user preferences without any user efforts [19]. User profiles are built in the forms of user interest categories or term lists/vectors. In [19], user profiles were represented by a hierarchical category tree based on ODP and corresponding keywords associated with each category. User profiles were automatically learned from search history. In [29], user preferences were built as vectors of distinct terms and constructed by accumulating past preferences, including both long-term and short-term preferences. Tan used the methods of statistical language modeling to mine contextual information from long-term search history. In this paper, user profiles are represented as weighted topic categories, similar with those given in [6], and these profiles are also automatically learned from users' past clicked web pages. Many personalized web search strategies based on hyperlink structure of web have also been investigated. Personalized PageRank, which is a modification of the global PageRank algorithm, was first proposed for personalized web search. In [10], multiple Personalized PageRank scores, one for each main topic of ODP, were used to enable "topic sensitive" web search. Jeh and Widom [14] gave an approach that could scale well with the size of hub vectors to realize personalized search based on Topic-Sensitive PageRank. The authors of extended the well-known HITS algorithm by artificially increasing the authority and hub scores of the pages marked relevant by the user in previous searches.

Most recently, [17] developed a method to automatically estimate user hidden interests based on TopicSensitive PageRank scores of the user's past clicked pages.

Problem Definition

To protect user privacy in profile-based PWS, researchers have to consider two contradicting effects during the search process. On the one hand, they attempt to improve the search quality with the personalization utility of the user profile. They need to hide the privacy contents existing in the user profile to place the privacy risk under control. Significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) portion of the user profile, namely a generalized profile. Thus, user privacy can be protected without compromising the personalized search quality. In general, there is a tradeoff between the search quality and the level of privacy protection achieved from generalization. Unfortunately, the previous works of privacy preserving PWS are far from optimal.

Disadvantages

- The existing profile-based PWS do not support runtime profiling.
- The existing methods do not take into account the customization of privacy requirements.
- Many personalization techniques require iterative user interactions when creating personalized search results.

Existing System

To existing system UPS (User customizable Privacy-preserving Search) framework, which is a privacy-preserving personalized web search framework, which can generalize profiles for each query according to user-specified privacy requirements.

- To develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. GreedyDP tries to maximize the discriminating power (DP), GreedyIL attempts to minimize the information loss (IL).
- The framework assumes that the queries do not contain any sensitive information, and aims at protecting the authentication in individual user profiles while retaining their usefulness for PWS.
- UPS consists of a nontrusty search engine server and a number of clients. Each client (user) accessing the search service trusts no one but himself/ herself.
- The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself.
- The proxy maintains both the complete user profile, in a hierarchy of nodes with semantics, and the user-specified (customized) privacy requirements represented as a set of sensitive-nodes.
- The online phase handles queries as 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. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.

Proposed System:

APWS (Advanced Privacy-preserving Search) system,, which is a Authentication based advanced personalized web search framework, which can generalize profiles for each query according to user-specified privacy requirements.

Evaluation Measurements We use two measurements to evaluate the advanced personalized search accuracy of different strategies: rank scoring metric introduced in and average rank metric introduced in [23]. Rank Scoring Rank scoring metric proposed by Breese [15] is used to evaluate the effectiveness of the collaborative filtering systems which return an ordered list of recommended items. Sun used it to evaluate the personalized web search accuracy and we also use it in this paper. The expected utility of a ranked list of web pages is defined as

$$R_s = \sum_j \frac{\delta(s,j)}{2^{(j-1)/(\alpha-1)}}$$

Where j is the rank of a page in the list, δ(s, j) is 1 if page j is clicked in the test query s and 0 otherwise, and α is set to 5 as the authors did. The final rank scoring reflects the utilities of all test queries:

$$R = 100\frac{\sum_s R_s}{\sum_s R_s^{Max}}$$

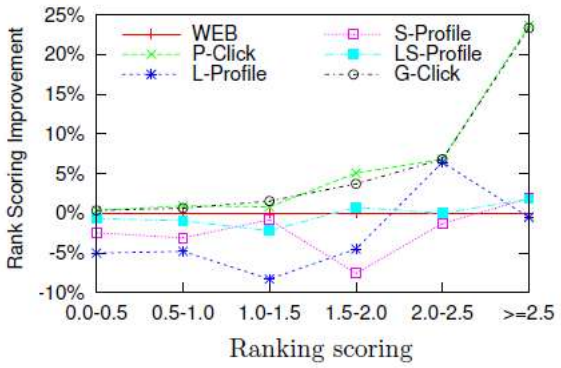
Here, R Max s is the obtained maximum possible utility when all pages which have been clicked appear at the top of the ranked list. Larger rank scoring value indicates better performance of personalized search.

Average Rank: Average rank metric is used to measure the quality of personalized search in [23, 28]. The average rank of a query s is defined as below.

$$AvgRank_s = \frac{1}{|\mathcal{P}_s|}\sum_{p\in\mathcal{P}_s} R(p)$$

Here Ps denotes the set of clicked web pages on test query s, R(p) denotes the rank of page p. The final average rank on test query set S is computed as:

$$AvgRank = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} AvgRank_s$$



Smaller average rank value indicates better placements of relevant result, or better result quality. In fact, rank scoring metric and average rank metric has similar effectiveness on evaluating personalization performance, and our experimental results show that they are consistent.

Conclusion:

In this paper, we try to investigate whether personalization is consistently effective under different situations. We develop a evaluation framework based on query logs to enable large-scale evaluation of personalized search. We use 12 days of MSN query logs to evaluate five personalized search

strategies. We find all proposed methods have significant improvements over common web search on queries with large click entropy. On the queries with small click entropy, they have similar or even worse performance than common web search. These results tell us that personalized search has different effectiveness on different queries and thus not all queries should be handled in the same manner. Click entropy can be used as a simple measurement on whether the query should be personalized and we strongly encourage the investigation of more reliable ones. Experimental results also show that click-based personalization strategies work well.

References:

[1] S. M. Beitzel, E. C. Jensen, A. Chowdhury, D. Grossman, and O. Frieder. Hourly analysis of a very large topically categorized web query log. In Proceedings of SIGIR ’04, pages 321–328, 2004.

[2] J. Boyan, D. Freitag, and T. Joachims. Evaluating retrieval performance using clickthrough data. In Proceedings of AAAI Workshop on Internet Based Information Systems, 1996.

[3] A. Broder. A taxonomy of web search. SIGIR Forum, 36(2):3–10, 2002.

[4] J. M. Carroll and M. B. Rosson. Paradox of the active user. Interfacing thought: cognitive aspects of human-computer interaction, pages 80–111, 1987.

[5] P. A. Chirita, C. Firan, and W. Nejdl. Summarizing local context to personalize global web search. In Proceedings of CIKM ’06, 2006.

[6] P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlsch¨utter. Using odp metadata to personalize search. In Proceedings of SIGIR ’05, pages 178–185, 2005.

[7] S. Cronen-Townsend and W. B. Croft. Quantifying query ambiguity. In Proceedings of HLT ’02, pages 94–98, 2002.

[8] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In Proceedings of WWW ’01, pages 613–622, 2001.

[9] P. Ferragina and A. Gulli. A personalized search engine based on web-snippet hierarchical clustering. In WWW ’05: Special interest tracks and posters of the 14th international conference on World Wide Web, pages 801–810, 2005.

[10] T. H. Haveliwala. Topic-sensitive pagerank. In Proceedings of WWW ’02, 2002.

[11] B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval: a study of user queries on the web. SIGIR Forum, 32(1):5–17, 1998.

[12] B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users, and real needs: a study and analysis of user queries on the web. Information Processing and Management, 36(2):207–227, 2000.

[13] J.C.Borda. M´emoire sur les ´elections au scrution. Histoire de l’Acad´emie Royal des Sciences, 1781.

[14] G. Jeh and J. Widom. Scaling personalized web search. In Proceedings of WWW ’03, pages 271–279, 2003.

[15] D. H. John S. Breese and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of UAI ’98, pages 43–52, 1998.

[16] R. Krovetz and W. B. Croft. Lexical ambiguity and information retrieval. Information Systems, 10(2):115–141, 1992.

[17] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In Proceedings of WWW ’05, pages 391–400, 2005.

[18] Y. Li, Z. Zheng, and H. K. Dai. Kdd cup-2005 report: facing a great challenge. SIGKDD Explor. Newsl., 7(2):91–99, 2005.

[19] F. Liu, C. Yu, and W. Meng. Personalized web search by mapping user queries to categories. In Proceedings of CIKM ’02, pages 558–565, 2002.