Evaluating the Effectiveness of Web Search Metrics

J. Jayanthi and M. Ezhilmathi
Department of Computer Science and Engineering, Sona College of Technology, Salem - 636 005, Tamil Nadu, India
E-mail: srkezhil@gmail.com
(Received on 12 July 2012 and accepted on 22 September 2012)

Abstract – **Software metrics are the key performance indicators, using which the performance of a system can be assessed quantitatively. Metrics can also be applied for personalized web search which can be used to retrieve relevant results for each individual user depending on their unique profile. Although personalized search based on user profile has been under research for many years and various metrics have been proposed, it is still uncertain whether personalization is unswervingly effective on different queries for different user profiles. We present a framework for personalized search which retrieves result based on user profile. We maintain user profile in the form of Preference Network (PN). We further propose metrics for ranking the search results based on user profile.**

Keywords: **Personalization, Preference Network, Repeated Query, Self-Repeated Query, Software Metrics, User Profile**

I. Introduction

Many new disputes arise for web search with the increasing amount of information on the web. A conventional search engine returns same set of results when the same query is submitted by all users, irrespective of who submitted the query. For example, for the query "orange", some users may be interested in documents dealing with "orange" as a "fruit", some users may need document related to "orange software company", and while some other may need information about "orange mobile phones". As well, different users have utterly different information needs. Personalization is found to be a great solution to address all these problems since it can provide distinct search results depending on user profile and preference. Various personalization strategies, which include [6, 8, 9, 11, 17, 19, 20] and [21] have been proposed. But they are far from optimal [1]. Main problem of current personalized search is that most proposed algorithms are applied homogeneously to all users and queries. Our stand is that all queries should not be handled in same manner because; single personalization algorithm might not be suitable for all queries and all users.

Each algorithm has its own pros and cons. For example, for the query "orange" topical-interest-based personalization may lead to better performance but may be ineffective for the query "free games online". All relevant documents for query "free games online" are mostly classified into the same topic categories, and topical-interest-based personalization is futile in such cases. Also applying personalization techniques on certain queries may be totally ineffective. For example, on the query "orange" using personalization based on topical interests of users might give better performance for individual users than a regular web search. In contrast, for the query "Yahoo!", which is a typical navigational query as defined by Broder [7] and Lee *et al.* [21], almost all users consistently select a link to Yahoo!'s homepage. Therefore, none of the personalization strategies can provide apparent benefits to the users as demonstrated by [1].

As a solution to these problems, we develop an evaluation framework to predict the appropriate algorithm to be applied based on different criterion. We provide a strategy to:

1. Gather and model user's search history in the form of preference network (PN);
2. A rule engine deduce appropriate metrics and algorithms for each query and each user, and
3. Improve web search effectiveness by using these metrics and algorithms.

II. Related Work

The content similarity between user profile and returned web pages can be used to re-rank search results. User profiles can be obtained explicitly [8], [9] or implicitly. Majority of user are reluctant to provide explicit feedback on search results and their interests, many works in the area of personalized search focus on how to automatically learn user preferences without direct participation of users [8], [10], [11]. Dou *et al.* [1] developed an evaluation framework based on real query logs to enable large-scale evaluation of personalized search. They also evaluated five personalization algorithms and proposed new metric called click entropy [1]. WebMate [13] uses user profiles to refine user queries, but no experimental results are given. Watson [13] refines queries using a local

context, but does not learn the user profile. Inquirus 2 [4] uses users' preferences to choose data sources and refine queries, but it does not have user profiles and requires the users to provide their preferences of categories. In addition, only four non-topical categories are included in Inquirus 2. The method in [11] learns users' profiles from their surfing histories and re-ranks/filters documents returned by meta-search engine based on the profiles.

Several approaches represent user interests by using topical categories. In [5, 8, 9, 15] and [16], a user profile is usually structured as a concept/topic hierarchy. User-issued queries and user-selected snippets/documents are categorized into concept hierarchies that are accumulated to generate a user profile. When the user issues a query, each of the returned snippets/documents is also classified. The documents are re-ranked based upon how well the document categories match user interest profiles.

Some other personalized search approaches use lists of keywords to represent user interests. Sugiyama *et al.* [12] built user preferences as vectors of distinct terms and constructed them by aggregating past preferences, including both long-term and short-term preferences. Shen *et al.* [10] first used language modeling to mine contextual information from a short-term search history. Tan *et al.* then used the method to mine context from a long-term search history. Teevan *et al.* [17] and Chirita *et al.* [18] exploit rich models of user interests, built from both search-related information and other information about the user, including documents and e-mails that the user has read and created. In the work of Liu *et al.* [2], [10], keywords are associated with categories, and thus, user profiles are represented by a hierarchical category tree based on keyword categories.

In our approach we utilize user profile to retrieve relevant results. User profile is maintained in the form of preference network (PN). Also, we develop a rule engine that can automatically identify the type of metric and algorithm to be applied for a query and user.

III. Proposed System

We propose an evaluation framework which can automatically identify the type of metric and algorithm to be applied based on various criterions such as user profile and user search history. The architecture of proposed system is illustrated in Figure 1.

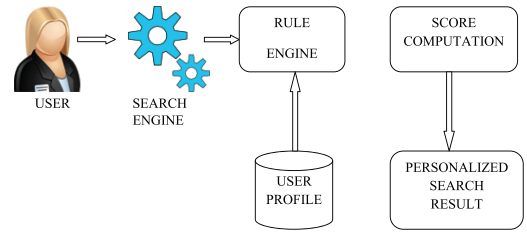

Fig.1 System Architecture

A. User Profile

User profile is maintained in the form of preference network (PN) [3]. Preference Network is constructed based on TF-IDF measure. TF-IDF measure is computed for each term in the top k documents retrieved by the web server for a query. The identical high scored terms are selected and the weights of each identical set of terms are summed up. From that list, again high weighted terms are selected to build the preference network.

The formula for Term Frequency is:

$$tf_i = \frac{n_i}{\sum_k n_k}$$

n_i = Number of occurrences of a term i

n_k = Total number of terms in a document

The formula for Inverse Document Frequency is:

$$idf_i = log\frac{N}{df_i}$$

N = Total number of relevant terms in the document

df_i = Number of documents that contain the term *i* at least once

Thus the TF-IDF weight is calculated using the formula:

$$TF - IDF\ weight = TF_i * IDF_i$$

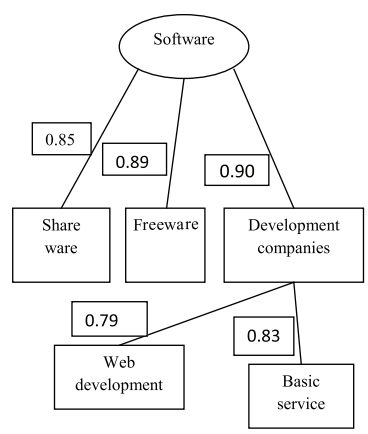

Fig.2 Preference Network for a query

B. Rule Engine

We propose a rule engine which identifies the convergence level of a user profile based on the structure and content of their preference network.

1. Profile Classification

It classifies user profile into three categories:

i. Converged profile
ii. Semi-converged profile
iii. Non-converged profile

Converged Profile (CP)

A profile is said to be converged if same set of queries are repeated over a period of time. Say, a set of 5 queries are repeatedly given in 30 sessions observed over a period of 30 days. In such cases, user profile will contain very few preference networks (here, 5).

Semi-Converged Profile (SCP)

A profile is said to be semi-converged if it has equal number of repeated queries as well as new queries. Say, in 30 sessions, 10 queries are entirely new and a set of 4 queries are given repeatedly by the user in 5 sessions. The number of preference networks will be half the count of number of sessions considered (here, 14).

Non-Converged Profile (NCP)

A profile is not converged if the user gives entirely different query in each session. Say, 30 different queries in 30 sessions. So the number of preference networks for such users will be greater than or equal to number of sessions considered.

2. Query Classification

After the successful classification of user profile, the rule engine then classifies the given query into three types:

i. Type-1: Self-Repeated Query (SRQ)
ii. Type-2: Repeated Query (RQ)
iii. Type-3: SRQ-RQ

Self-Repeated Query

When a user issues a query which is previously issued only by that user and which is not issued by any other user then it is a Self-Repeated query.

Repeated Query

If a query issued by a user is not that user's search history i.e. PN but in the PN of other users, then it is a repeated query.

SRQ-RQ

If a query issued by a user in the PN of both the current user and other users, then it belongs to this type.

C. Score Computation

For Type-1 queries, the documents are ranked in descending order of P-Click [1] scores of documents which were previously clicked by that user for the same query. The formula for calculating P-Click score is:

$$P - Click_{Doc_n}(q,p,u) = \frac{|Clicks(q,p,u)|}{|Clicks(q,\blacksquare,u)| + \beta} \quad (1)$$

|Clicks(q,p,u)| - number of clicks on web page p for the query q by the user u

|Clicks (q,p,u)| - total number of clicks for query q by u - β smoothing factor.

For Type-2 and Type-3 queries, the documents are ranked in descending order of G-Score calculated using P-Click score of related documents from the profile of all the users who issued that query previously. The formula for calculating G-Score is:

$$G - Score_{Doc_n} = \frac{\sum_{i=1}^{N}(P-Click_{Doc_n})_i}{N} \quad (2)$$

$(P - Click_{Doc_n})_i$ – P-Click score of Doc_n intsc of user i

N – Total number of user profiles which contains Doc_n.

IV. Conclusion

In this paper, we proposed an evaluation framework for automatic identification of metrics and algorithms to be applied for retrieving relevant web search results for individual users. We maintain user profile in the form of Preference Networks (PN). We further proposed techniques and strategies for classifying user profiles and queries. This approach would be useful to improve search accuracy and for retrieving relevant results for each individual user depending on their preference. Future work can be extended in proposing metrics for entirely new queries which is not issued by any of the users in the data set.

REFERENCES

[1] Z. Dou, R. Song, J.-R. Wen and X. Yuan, "Evaluating the Effectiveness of Personalized Web Search", *IEEE Trans. Knowledge and Data Eng.,* Vol. 21, No. 8, pp. 1178-1190, August 2009.

[2] F. Liu, C. Yu, and W. Meng, "Personalized Web Search for Improving Retrieval Effectiveness," *IEEE Trans. Knowledge and Data Eng.,* Vol. 16, No. 1, pp. 28-40, Jan. 2004.

[3] J. Jayanthi, B. Keerthana and K. S. Jayakumar, "Building User Profile Based on Preference Networks", *International Journal of Web Engineering,* pp. 11-14, 2012.

[4] E.J. Glover, G.W. Flake, S. Lawrence, W.P. Birmingham, A.Kruger, C.L. Giles, and D.M. Pennock, "Improving Category Specific Web Search by Learning Query Modifications," *SAINT,* pp. 23-34, 2001.

[5] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing," *Web Intelligence and Agent Systems,* Vol. 1, No. 3/4, pp. 219-234, 2003.

[6] J. Pitkow, H. Schutze, 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.

[7] A. Broder, "A Taxonomy of Web Search," *ACM SIGIR Forum,* Vol. 36, No. 2, pp. 3-10, 2002.

[8] A. Pretschner and S. Gauch, "Ontology Based Personalized Search," *Proc. 11th IEEE Int'l Conf. Tools with Artificial Intelligence* (ICTAI '99), pp. 391-398, 1999.

[9] P.-A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschutter, "Using ODP Metadata to Personalize Search," *Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05),* pp. 178-185, 2005.

[10] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search," *Proc. ACM Int'l Conf. Information and Knowledge Management (*CIKM '05), pp. 824-831, 2005.

[11] F. Liu, C. Yu, and W. Meng, "Personalized Web Search by Mapping User Queries to Categories," *Proc. ACM Int'l Conf. Information and Knowledge Management* (CIKM '02), pp. 558-565, 2002.

[12] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without Any Effort from Users," *Proceedings of 13th Int'l World Wide Web Conference* (WWW '04), pp. 675-684, 2004.

[13] L. Chen and K. Sycara, "WebMate: A Personal Agent for Browsing and Searching," *Proc. Second Int'l Conf. Autonomous Agents and Multi Agent Systems,* pp. 132-139, 1998.

[14] J. Budzik and K. Hammond, "Watson: Anticipating and Contextualizing Information Needs," *Proc. 62nd Ann. Meeting Am. Soc. Information Science,* 1999.

[15] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy," *Proc. 12fth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '06),* pp. 718-723, 2006.

[16] 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 '05), pp. 449-456, 2005.

[17] P.A. Chirita, C. Firan, and W. Nejdl, "Summarizing Local Context to Personalize Global Web Search," *Proc. ACM Int'l Conf. Information and Knowledge Management* (CIKM), 2006.

[18] G. Jeh and J. Widom, "Scaling Personalized Web Search," *Proc. 12th Int'l World Wide Web Conf.* (WWW '03), pp. 271-279, 2003.

[19] P. Ferragina and A. Gulli, "A Personalized Search Engine Based on Web-Snippet Hierarchical Clustering," *Special Interest Tracks and Posters of the 14th Int'l Conf. World Wide Web* (WWW '05), pp. 801-810, 2005.

[20] J.-T. Sun, H-J. Zeng, H. Liu, Y. Lu, and Z. Chen, "CubeSVD: A Novel Approach to Personalized Web Search," *Proc. 14th Int'l World Wide Web Conf.* (WWW '05), pp. 382-390, 2005.

[21] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," *Proc. 14th Int'l World Wide Web Conf.* (WWW '05), pp. 391-400, 2005.