Using A Graph-based Ontological User Profile For Personalizing Search

Mariam Daoud
IRIT, Paul Sabatier University
118, Route Narbonne
Toulouse, France
daoud@irit.fr

Lynda Tamine-Lechani
IRIT, Paul Sabatier University
118, Route Narbonne
Toulouse, France
lechani@irit.fr

Mohand Boughanem IRIT,
Paul Sabatier University
118, Route Narbonne
Toulouse, France
bougha@irit.fr

ABSTRACT

In this poster, we describe a personalized search approach, which involves a graph based user profile issued from ontology and a session boundary recognition mechanism. The user profile refers to the short term user interest and is used for re-ranking the search results of queries in the same search session. The session boundary recognition is based on tracking changes in the dominant concepts held by the query and the user profile. Experimental evaluation was carried out using the HARD 2003 TREC collection and shows that our approach is effective.

Categories and Subject Descriptors

H3.3. [**User Modeling for IR, Search Personalization**]

General Terms

Design, Experimentation, Performance

Keywords

Graph based user profile, personalization, session boundary recognition, ontology

1. INTRODUCTION

In web information retrieval, most search engines use the document content or the document structure in the information retrieval (IR) process. Such systems provide the same results for the same keyword queries even though these latter are submitted by different users with different intentions. The main reason behind this is that the search process is made out of the user context.
Personalized IR tackles this problem by integrating the user profile that refers to the user interests in the IR process. There have been several attempts at building a personalized or contextual search engine [2, 3] or session based search engines [4]. We propose a search personalization approach which consists of building and maintaining the user profile in order to re-rank search results of queries in the same search session. Our approach has the following new features:

- Representation of the user profile as a graph of interrelated concepts of the ODP[1] ontology, inferred using score propagation method through semantic links of the ontology.
- A session boundary recognition mechanism based on tracking changes in the dominant concepts held by the user profile and a new submitted query.

[1]http://www.dmoz.org

CIKM'08, October 26–30, 2008, Napa Valley, California, USA.
ACM 978-1-59593-991-3/08/10.

2. PERSONALIZING SEARCH USING A GRAPH BASED USER PROFILE

2.1 Building and maintaining the user profile

We learn a graph based user profile issued from the ODP ontology by aggregating graph based query profiles over a search session. For each query q^s submitted by the user at time s, we build the associated query profile G_q^s as a graph of semantically related concepts of the ontology. First we identify the user's documents of interest D^s using implicit feedback measures and create the query context K^s as follows:

$$K^s(t) = \frac{1}{|D^s|} \sum_{d \in D^s} w_{td} \quad (1)$$

where w_{td} is the weight of term t in document d computed using the tf*idf weighing scheme. Then, we map the query context K^s on the ontology using the cosine similarity measure as follows:

$$score(c_j) = cos(\vec{c_j}, \vec{K^s}) \quad (2)$$

where each concept c_j of the ontology is represented as a single term vector $\vec{c_j}$ issued from the web pages classified under this concept as described in previous work [1]. Based on the obtained weighted concept set, we activate for each concept its semantically related concepts using score propagation. The user profile is initialized by the profile of the first query submitted in the search session. It is then maintained by merging it with the profile of the new submitted query by accumulating the weights of possible common concepts and adding new nodes and edges issued from the query profile.

2.2 Session boundary recognition

We propose a session boundary recognition method using the *Kendall* rank correlation measure that quantifies the conceptual correlation ΔI between the user profile G_u^s and the query q^{s+1}. Here, the query is mapped onto the ontology and represented by the concept vector $q_c^{\vec{s+1}}$. We choose a threshold and believe the queries are from the same session if the similarity is above the threshold. The similarity ΔI is

computed as follows.

$$\Delta I = (\vec{q_c^{s+1}} o \vec{G_u^s}) = \frac{\sum_{c_i}\sum_{c_j} S_{c_i c_j}(\vec{q_c^{s+1}}) \times S_{c_i c_j}(\vec{G_u^s})}{\sqrt{\sum_{c_i}\sum_{c_j} S^2_{c_i c_j}(\vec{q_c^{s+1}}) \times \sum_{c_i}\sum_{c_j} S^2_{c_i c_j}(\vec{G_u^s})}} \quad (3)$$

$$S_{c_i c_j}(\vec{v}) = sign(\vec{v}(c_i) - \vec{v}(c_j)) = \frac{\vec{v}(c_i) - \vec{v}(c_j)}{|\vec{v}(c_i) - \vec{v}(c_j)|}$$

Where, c_i and c_j are two concepts issued from both the query and the user profile, $\vec{q_c^{s+1}}(c_i)$ (resp. $\vec{G_u^s}(c_i)$) is the weight of the concept c_i in $\vec{q_c^{s+1}}$ (resp. $\vec{G_u^s}$).

2.3 Search personalization

We personalized search results of query q^{s+1} in the same search session using the user profile G_u^s by combining for each retrieved result d_k, the initial score S_i and a contextual score S_c as follows:

$$S_f(d_k) = \gamma * S_i(q, d_k) + (1 - \gamma) * S_c(d_k, G^s) \quad (4)$$

$$0 < \gamma < 1$$

The contextual score S_c of result d_k is computed using the cosine similarity measure between the result d_k and the concepts of the user profile G_u^s as follows:

$$S_c(d_k, G_u^s) = \sum_{j=1..3} score(c_j) * cos(\vec{d_k}, \vec{c_j}) \quad (5)$$

Where c_j is a concept in the user profile, $score(c_j)$ is the weight of concept c_j in the user profile G_u^s.

3. EXPERIMENTAL EVALUATION

We have conducted two set of experiments to evaluate the session boundary accuracy and the effectiveness of search personalization. We used the data collection provided by TREC[2] 2003 HARD Track. As our approach requires a session based evaluation setup, we simulate a search session by a single topic and related queries by creating three subtopics of the main topic. For each topic we selected randomly a relevant document set called the profile set provided by TREC data. This set is divided into three subsets where the top three representative terms of each subset constitute the subtopic.

3.1 Session boundary accuracy

We aligned a *training subtopic sequence* in order to identify the best threshold value for the session boundary recognition. First, we computed the Kendall correlations between each subtopic and the user profile built across subtopics related the same topic and obtained values in $[-0.6 \ +0.01]$. Then we evaluated the session boundary accuracy in terms of the fraction of correctly identified related queries ($P_{intra}(\sigma)$) and the fraction of correctly identified session boundaries $P_{inter}(\sigma)$ for each σ value in ($[-0.6 \ +0.01]$). Results show that the threshold value -0.34 achieves the optimal session boundary accuracy maximizing both P_{intra} at 53.33% and P_{inter} at 85.71%.

[2]Text REtrieval Conference:http://trec.nist.gov

3.2 Search personalization effectiveness

We evaluated the personalized search effectiveness along a *test subtopic sequence* using the optimal threshold value (-0.34) identified above. We compared the personalized search to the baseline search based on the OKAPI BM25 retrieval model. Evaluation measures are the Top-n recall and the Top-n precision done as follows:

$$Top{-}nRecall = \frac{RelDoc_n}{RelDoc_{total}}, Top{-}nPrecision = \frac{RelDoc_n}{n} \quad (6)$$

where $RelDoc_n$ is the number of relevant documents in the top n retrieved search results, $RelDoc_{total}$ is the total number of relevant documents of the *subtopic*, by excluding the profile set associated to its topic. Results show clearly a significant improvement of personalized search comparatively to the typical search.

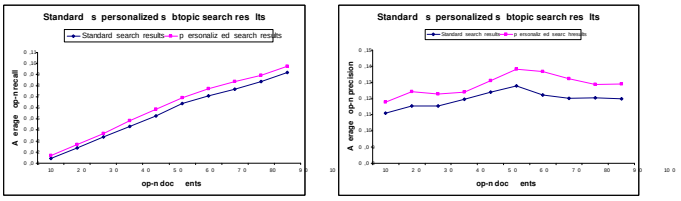

Figure 1: Average Top-n recall and Top-n precision comparison between the personalized and the standard search

4. CONCLUSION

In this poster, we described a personalized search approach that builds a graph based user profile in a specific search session. We proposed a session boundary recognition mechanism in order to personalize search results of queries in the same search session. We have also conducted an evaluation of our approach that shows it effectiveness. It is important to pursue the evaluation to study the user profile accuracy and the evaluation of session boundary using data provided by search engine log file.

5. REFERENCES

[1] M. Daoud, L. Tamine-Lechani, and M. Boughanem. Learning user interests for a session-based personalized search. In *Proceedings of the the Second International IIiX Symposium (IIIX'08). To appear*, London, U.K., 2008.

[2] S. Gauch, J. Chaffee, and A. Pretschner. Ontology-based personalized search and browsing. *Web Intelli. and Agent Sys.*, 1(3-4):219–234, 2003.

[3] F. Liu, C. Yu, and W. Meng. Personalized web search for improving retrieval effectiveness. *IEEE Transactions on Knowledge and Data Engineering*, 16(1):28–40, 2004.

[4] X. Shen, B. Tan, and C. Zhai. Context-sensitive information retrieval using implicit feedback. In *the international ACM SIGIR conference on Research and development in information retrieval*, pages 43–50, New York, NY, USA, 2005. ACM.