Search Result Re-ranking Based on Gap between Search Queries and Social Tags

Jun Yan[1] Ning Liu[1] Elaine Qing Chang[2] Lei Ji[1] Zheng Chen[1]
Microsoft Research Asia[1] Microsoft Corporation[2]
Sigma Center, 49 Zhichun Road Beijing, 100190, P. R. China
One Microsoft Way Redmond, WA, USA
{junyan, ningl, leiji, zhengc}@microsoft.com elchang@microsoft.com

ABSTRACT
Both search engine click-through log and social annotation have been utilized as user feedback for search result re-ranking. However, to our best knowledge, no previous study has explored the correlation between these two factors for the task of search result re-ranking. In this paper, we show that the gap between search queries and social tags of the same web page can well reflect its user preference score. Motivated by this observation, we propose a novel algorithm, called Query-Tag-Gap (QTG), to re-rank search results for better user satisfaction. Intuitively, on one hand, the search users' intentions are generally described by their queries before they read the search results. On the other hand, the web annotators semantically tag web pages after they read the content of the pages. The difference between users' recognition of the same page before and after they read it is a good reflection of user satisfaction. In this extended abstract, we formally define the query set and tag set of the same page as users' pre- and post-knowledge respectively. We empirically show the strong correlation between user satisfaction and user's knowledge gap before and after reading the page. Based on this gap, experiments have shown outstanding performance of our proposed QTG algorithm in search result re-ranking.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Information Search and Retrieval – *Retrieval models*

General Terms
Algorithms, Performance.

Keywords
Query log, Social tagging, Search result re-ranking

1. INTRODUCTION
With the growing competition in the Web search market, to provide better search user satisfaction, the search result re-ranking problem have attracted a lot of attention. Among them, one of the most commonly used algorithms is to utilize click-through log as implicit user feedback [3] for search result re-ranking. In some recent progress, using social annotation for search result ranking [2] has been validated to be effective as well. Though both the click-through information and the social tags can independently reflect users' knowledge on the corresponding search result web pages for ranking tasks, they are generated from different perspectives. For the former, it represents users' knowledge before they read the clicked Web pages while the latter represents users' knowledge after they read the corresponding pages. In this work we define the search queries implied in the click-through log as users' pre-knowledge of the Web pages, and the tags for Web pages as users' post-knowledge. The question we'd like to answer is: can the difference between users' pre- and post- knowledge reflect users' satisfaction of the search result web page?

To answer this question, the contributions of this paper are: (1) We define search queries and tags as users' pre- and post-knowledge of the web pages respectively. (2) Through studying the site search click-through log from a Microsoft commercial website, we empirically show that the gap between users' pre- and post- knowledge can well reflect user satisfaction. (3) We propose a novel and simple algorithm, called Query-Tag-Gap (QTG), for search result re-ranking based on the gap between users' pre- and post- knowledge. To our best knowledge, this is the first work which considers the correlation between search queries and tags for search result re-ranking. (4) Experiments on the search results of a commercial search engine well validated its effectiveness.

2. MOTIVATIONAL OBSERVATIONS
We collected 100 pages (which we defined as assets) from a Microsoft commercial website. These 100 assets were randomly sampled from the all assets on the site that have got more than 20 clicks in the site search click-through log within a recent month. In addition, all of these 100 assets have received at least 5 times of explicit user ratings. In other words, at least 5 users have assigned a rating score to each of these assets. The score ranges from 1 to 5 where 1 means least satisfied and 5 means most satisfied. We asked 5 volunteers to read these assets and manually assign at least 5 tags for each asset based on their own understanding. We represent each asset by its search query terms and tag terms respectively in the Bag of Words model. In the simple preliminary study, we used the simple Jaccard distance to measure the distance between the query representation and tag representation of the same asset. Figure 1 gives the scatter plot of the averaged user rating score against the distance between queries and tags.

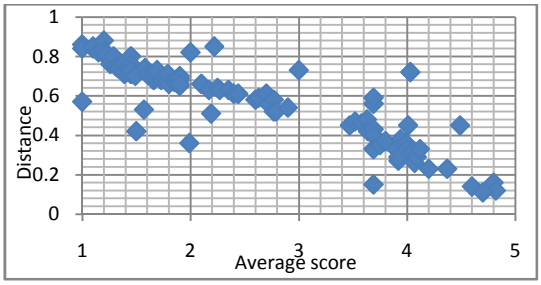

Figure 1. Scatter plot of the user satisfaction score against the distance between corresponding queries and social tags.

From this figure, we can see that for most of the assets, the distance changes almost linearly with the user rating score, which

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

reflects user satisfaction of the assets. For a given asset, the more satisfied it is by the users, the smaller distance will be between the query representation and the tag representation. This motivates us to explore the problem: Can we re-rank the search results of search engines based on user satisfaction indicated by the gap between queries and social tags?

3. THE ALGORITHM

In this Section, we propose a novel algorithm for search result re-ranking. Suppose we have a collection of web pages, which are indexed by a search engine, denoted as $P=\{p_1, p_2,\ldots\}$. Let the collection of queries that resulted clicks on a page p_i, be $Q_i=\{q_{i1}, q_{i2},\ldots\}$ and the tags that have been used to tag the page p_i, be $T_i=\{t_{i1}, t_{i2},\ldots\}$. In Section 3.1, we firstly define users' pre- and post- knowledge of the same page p_i by Q_i and T_i respectively. In Section 3.2, we introduce the detailed algorithm for search result re-ranking.

3.1 Pre- and Post- Knowledge of Users

We define the queries which have led clicks to a web page as the pre-knowledge of users to this page.

Definition-1: The *pre-knowledge* of a web page p_i is defined as the set of queries Q_i, which have resulted user clicks to it, with the corresponding term distribution. We use $D(Q_i)$ to represent the term distribution in Q_i, i.e. the probability of each term to appear in Q_i.

On the other hand, there are more and more web annotators who semantically express their understanding of a web page by social tags after they read the content of the page. We define the tags used as users' post-knowledge of the page.

Definition-2: The *post-knowledge* of a web page p_i is defined as the set of tags T_i which are used to annotate the page with the corresponding term distribution. We use $D(T_i)$ to represent the term distribution in T_i, i.e. the probability of each team to be used to tag p_i.

3.2 Search Result Re-ranking

The preliminary experiment introduced in Section 2 shows that the difference between users' understanding of the same page before and after they read it is a good reflection of user satisfaction. This motivates us to consider the gap between the users' pre- and post- knowledge for search result re-ranking, which aims to better satisfy the search users. Given search query q , let $\Pr(q|Q_i)$ and $\Pr(q|T_i)$ be the probability of search query set Q_i and social tag set T_i. We use the difference between these two probabilities, i.e.

$$S(p_i|q) = |\Pr(q|Q_i) - \Pr(q|T_i)| \quad (1)$$

to calculate the penalty score of returning page p_i as search results of the given query q. The remaining problem is how to estimate the conditional probabilities $\Pr(q|Q_i)$ and $\Pr(q|T_i)$. For simplicity, we assume all terms in the queries are independent. Suppose the query q consists of m terms, which are $t_1, t_2,\ldots, t_m$, then we have,

$$\Pr(q|Q_i) = \sum_{j=1}^{m} \Pr(t_j|Q_i) \text{ and } \Pr(q|T_i) = \sum_{j=1}^{m} \Pr(t_j|T_i) \quad (2)$$

Through this way, we can assign a penalty score to each page for a given query q. Suppose the search results of q are p_i, $i=1,2,\ldots,n$. We propose to re-rank the search results by $S(p_i|q)$ in a increasing order.

4. EXPERIMENTS

In this Section, we show the experimental results of our proposed algorithm on re-ranking the search results of a commercial search engine. We randomly selected 50 queries in the search engine query log. For each query, we collected the top 10 search results. All queries that have resulted clicks on these pages within a time range of one month were used to calculate user *pre-knowledge* of these pages. The social tags of these pages that we collected from del.icio.us were used to calculate the *post-knowledge* of these pages. We asked 5 volunteers to score the search results from score "1" to "5", where "5" means the best search result and "1" means the worst search result. The average score of each page is used as ground truth to rank the search results for these 50 sampled queries. The evaluation metrics we used in the experiments are Precision at n (P@n) and Normalized Discount Cumulative Gain (NDCG) [1]. The log based implicit user feedback for search result re-ranking [3], which is represented by "Log", is utilized as a baseline for comparison purpose. The experimental results are shown in Table 1, where "baseline" stands for the search results of our used search engine without re-ranking. From the results we can see that the proposed QTG algorithm can effectively re-rank the search results. The statistical t-test with the results 0.00073 and 0.00004 validated that the improvements of our proposed algorithm are significant in contrast to the two baseline algorithms.

Table 1. Results of search result re-ranking

	Live	Log	QTG
P@1	0.786	0.799	**0.806**
P@5	0.746	0.753	**0.765**
NDCG	0.633	0.642	**0.659**

5. CONCLUSION

In this work, we studied the site search click-through log of a commercial website, which has user rating information, to show the fact that the gap between query and social tags' distribution can well reflect the search users' satisfaction. Motivated by this observation, we define the query distribution of a web page as the users' pre-knowledge of it. In the contrary, we define the social tag distribution of this web page as users' post-knowledge of it. Based on the gap between the pre- and post- knowledge, we propose a novel and simple algorithm, called QTG, for search result re-ranking. Experimental results on real web search data well validated our proposed algorithm. In our next step work, we will explore how to integrate our algorithm together with other search result re-ranking algorithms to achieve better performance.

6. REFERENCES

[1] Järvelin, K. and Kekalainen, J. IR evaluation methods for retrieving highly relevant documents. In Proceedings of the *23rd* annual international ACM SIGIR conference (2000), ACM Press, pp 41~48.

[2] Shenghua B., Guirong X., Xiaoyuan W., Yong Y., Ben F., Zhong S., Optimizing web search using social annotations, in Proceedings of the 16th international conference on World Wide Web (2007) ACM Press, pp 501~510.

[3] Thorsten J., Laura G., Bing P., Helene H., Geri G., Accurately Interpreting Clickthrough Data as Implicit, in Proceedings of the 28th annual international ACM SIGIR conference (2005), ACM Press, pp 154~161.