International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 9 2754 – 2757

Reconstructing of search Results in web Search Engine Using Feedback Activities

M.V.S.Prabha
Computer Science Engineering
UCEK, Kakinada
Andhra Pradesh, India
e-mail: prabhamvs@gmail.com

A.Krishna Mohan,Associate professor
Computer Science Engineering
UCEK, Kakinada
Andhra Pradesh, India
e-mail:Krishna.anakala @gmail.com

Abstract— Now a day's information retrieval (IR) is playing vital role in web. Users forward the queries to search engine, it provides information links to users, but some uncertain queries provides different types of information on same query. To overcome this ambiguity problem, a framework has to be designed to capturing user search goals for query by using feedback activities. We propose a novel technique to align feedback activities to pseudo-documents which reflect user information needs. Ultimately, incorporating user behavior into web search ranking hysterically improves relevance providing rich user interaction features to ranker is the most effective strategy.

Keywords- *User search goals, feedback activities, pseudo-documents, restructuring search results, classified average precision, page ranking*

*****

I. INTRODUCTION

In web based search applications, user submits the query to search engine to represent the information needs of user. The information needs of different user may differ in various aspects of query information. This becomes difficult to achieve user information needs. Sometimes ambiguous queries may not exactly represented by users so it results in less understandable to search engine. To achieve the user specific information needs many ambiguous/uncertain queries may cover a broad topic and dissimilar users may want to get information on different aspects when they submit the same query.

It is advantageous to improve search engine relevance by inferring and analyzing the user search goals, as a result it also improves the user experience.

First, the web search goals can be restructured [6], [18], [20] basing on the user search goals by clustering the search results which has same search goals; therefore users can find easily in accordance.

Second, in query recommendation, keywords can be used which actually show the user search goal [2], [5], [7]; therefore the suggested queries formed help users more accurately.

Third, in re-ranking web search results, the distributions of user search goals can be useful, that contain different user search goals.

Implicit relevance feedback for ranking and personalization has become an active area of research. Recent work by Joachims and others exploring implicit feedback in controlled environments have shown the value of incorporating implicit feedback into the ranking process.

In this method we have limitations since the number of different clicks URLs of a query may be small. Another work analyzes the search results return by the search engine when a query is submitted.

- Analysis of alternatives for incorporating user behavior into web search ranking.
- An application of a robust implicit feedback model derived from mining millions of user interactions with a major web search engine.
- A large scale evaluation over real user queries and search results, showing significant improvements derived from incorporating user feedback.

II. BACK GROUND AND RELATED WORK

The figure 1 shows the framework of our approach and it consists of two parts. Those are

In the upper part all the feedback activities 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 analysis part, the original search results are restructured based on the user search goals inferred from the upper part. The search results that are restructured are then re-ranked and the evaluation result will be used as the feedback to select the optimal number of user search goals in the upper part.

 ISSN: 2321-8169
2754 – 2757

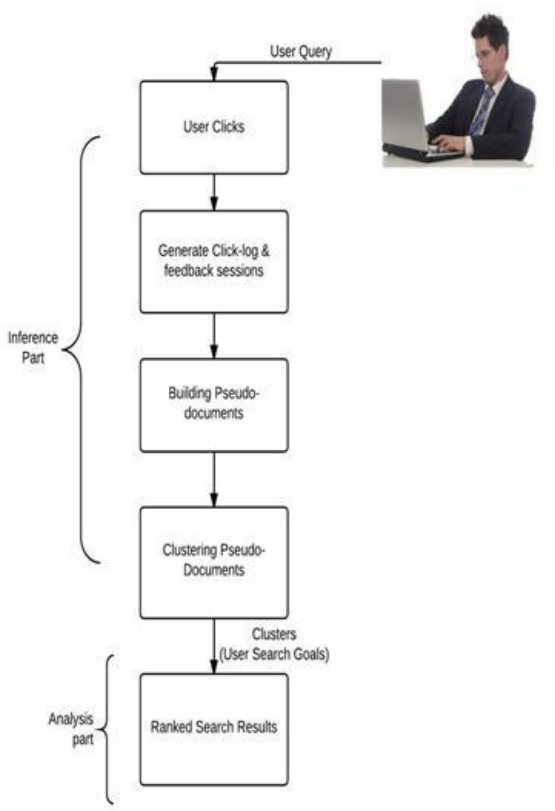

Fig 1: The framework of our approach

A. *Ranking enhanced with Feedback results*

The ranking strategy in the latest web search engines is based on large number of characteristics such as, content-based and query-independent page quality features. In most cases these techniques are developed for activating the particular ranking function modules that merge these feature values. The requirement of this model is a tough ranking methodology to missing values: greater than half of the queries to web search engine are same, with having no previously given feedback. This model requires a ranking algorithm to be robust to missing values: more than 50% of queries to web search engines are unique, with no previous implicit feedback available.

Our approach has a key advantage that used recent advanced methods in machine learning, called trainable ranking algorithms for IR (information retrieval) and web search. A key aspect of our approach is exploiting recent advances in machine learning, namely trainable ranking algorithms for web search and information retrieval and classical results.

The settings in our method are made in such a way that the explicit human relevance evaluations are available for a set of web search queries and results. In our settings explicit human relevance judgments (labels) are available for a set of web search queries and results. Therefore intelligent choice is to use supervised machine learning technique to learn and adapt a ranking function that best forecasts relevance judgments.

III. EXISTING SYSTEM

Existing methods only produces the result with higher level of the documents only and it doesn't make the results for all search based user goals.

User search goal analysis is important to optimize search engine and effective query results organization. When query is submitted to search engine, the returned web pages of search results are analyzed. Since it does not consider user feedback, many unuseful and noisy search results that are not clicked by user may be analyzed. This may degrade the search goals.

IV. PROPSED SYSTEM

We first describe the proposing feedback activities and then we introduce the proposed pseudo documents to represent feedback activities.

A. *Feedback activities*

We focus on inferring user search goals for a particular query. Therefore, the single session containing only one query is introduced, which distinguishes from the conventional session. Meanwhile the feedback activities in this paper are based on a single session, although it can be extended to the whole session. The proposed feedback activities consist of both clicked and unclicked URLs and ends with the last URL that was clicked in a single session. It is motivated that before the last click, all the URLs have been scanned and evaluated by users. Therefore, besides the clicked URLs, the unclicked ones before the last click should be a part of the user feedbacks. Below diagram shows the actual flow of the proposed scheme.

1) *Map Feedback activities to Pseudo-Documents*

Since feedback activities vary a lot for different click-trough's and queries, it is unsuitable to directly use feedback activities for inferring user search goals. Some representation method is needed to describe feedback activities in a more efficient and coherent way. There can be many kinds of feature representations of feedback activities.

However since different feedback activities have different numbers of URLs, the binary vectors of different feedback activities may have different dimensions. Moreover, binary vector representation is not informative enough to tell the contents of user search goals. We introduce pseudo-documents as surrogates to approximate goal texts. Thus, pseudo-documents can be used to infer user search goals.

2) *Forming pseudo-document based on URL representations*

In order to obtain the feature representation of feedback activities we propose an optimization method to combine both clicked and unclicked URLs in the feedback activities.

Our method can address this problem. Let us analyze the problem from three cases.

Case 1: One term appears in all the clicked URLs and does not appear in any unclicked ones.

Case 2: One term appears in both the clicked URLs and a subset of the unclicked ones.

Case 3: One term appears in both the clicked URLs and almost all the unclicked ones.

B. Inferring user search goals byclustering pseudo documents

Inference on user search goals can be done. We can infer user search goals. This section contains the description of how inference is applied on the user search goals and how are they depicted into some meaningful pseudo words. In this section we will describes how to infer user search goals and depict them with some meaningful keywords. Various pseudo-documents are clustered according to respective categories. After the process of clustering these pseudo documents, the group belonging to one cluster differentiates different user search goals. After clustering all the pseudo-documents, each cluster can be considered as one user search goal. Clustering is performed by computing the average of vectors of all the pseudo-documents that are belonging to the same cluster. The center point of a cluster is computes as the average of the vectors of all the pseudo-documents in the cluster.

1) Ranking Results Retrieval algorithm:

Step 1: For each new query q
Step 2: We calculate its textual similarity w_{qi} with each cluster
Step 3: We produce N different rankings R_{qi} with r_{qij} being the rank of result j, for query q after re-ranking results using the ranking function trained on cluster

$$rank(q, j) = \sum_{i=1}^{N} w_{qi} r_{qij}$$

- w_{qi} represents how similar is the content of cluster i with query q.
- r_{qij} gives the result rank when using the ranking function of cluster.
- We combine all produced rankings according to how similar they are to the query.

2) Equations

1.

$$\mathbf{T}_{u_i} = [t_{w_1}, t_{w_2}, \ldots, t_{w_n}]^T,$$
$$\mathbf{S}_{u_i} = [s_{w_1}, s_{w_2}, \ldots, s_{w_n}]^T,$$

Where T_{ui} and S_{ui} i are the TF-IDF vectors of the URL's title and snippet T_{ui} means the ith URL in the feedback activites. And W_j is the jth term appearing in the enriched URLs.

2.

$$\mathbf{F}_{u_i} = \omega_t \mathbf{T}_{u_i} + \omega_s \mathbf{S}_{u_i} = [f_{w_1}, f_{w_2}, \ldots, f_{w_n}]^T$$

F means the feature representation of the ith URL in the feedback activites, and W_t and W_s are the weights of thetitles and the snippets.

3. The similarity between two pseudo-documents is computed as the cosine score:

$$Sim_{i,j} = \cos(\mathbf{F}_{fs_i}, \mathbf{F}_{fs_j}) = \frac{\mathbf{F}_{fs_i} \cdot \mathbf{F}_{fs_j}}{|\mathbf{F}_{fs_i}||\mathbf{F}_{fs_j}|}.$$

4. Distance between two feedback activities

$$Dis_{i,j} = 1 - Sim_{i,j}.$$

5. After clustering all the pseudo-documents, each cluster can be considered as one user search goal. The center point of a cluster is computed as the average of the vectors of all the pseudo-documents in the cluster.

$$\mathbf{F}_{center_i} = \frac{\sum_{k=1}^{C_i} \mathbf{F}_{fs_k}}{C_i}, (\mathbf{F}_{fs_k} \subset Cluster\ i)$$

3) RR Algorithm

Step1: User issues ambiguous query
Step2: Evaluating feedback activities based on user query [EQ-1]
Step3: Maintaining the click sequences of feedback activities [EQ-2]
Step4: Binary vector method for feedback activities [EQ-3]
Step5: Embedding feedback activities to pseudo documents [EQ-3]
Step6: Construction of pseudo documents by using URL's in feedback activities [EQ-4]
Step7: K-Means algorithm for cluster pseudo documents [EQ-5]
Step8: Displaying re-structured ranked search results based on user query.

V. EXPERIMENTAL RESULTS

We will show experiments of our proposed algorithm. The data set that we used is based on the click through logs from a commercial search engine collected over a period of two months, including totally 2,300 different queries, 2.5 million single sessions and 2.93 million clicks. On average, each query has 1,087 single sessions and1, 274 clicks. In our approach, we have two parameters to be fixed: Kin K-means clustering and ŏ. When

clustering feedback activities of a query, we try five different K (1,2,...,5) in K-means clustering.

1) Intuitive Results of Inferring User Search Goals

1. We incorporate inference on user search goals for a query using clustering with the help of feedback activities. We infer user search goals for a query by clustering its feedback activities. User search goals are represents by the center points of different clusters. This confirms that our approach can infer user search goals properly and depict them with some keywords meaningfully
2. Risk and VAP are used to evaluate the performance of restructuring search results together. Each point represents the average Risk and VAP of a query. If the search results of a query are restructured properly, Risk should be small and VA should be high and the point should tend to be at the top left corner.

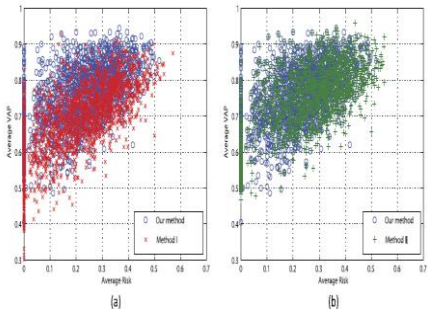

Fig. Comparison of three methods for 1,720 queries.

VI. CONCLUSION

In this paper, we proposed a novel approach that incorporates inference on user search goals for a user query by using supervised machine learning clustering user query feedback activities that correspond to pseudo-documents. In this paper, a novel approach has been proposed to infer user search goals for a query by clustering its feedback activities represented by pseudo-documents.

Firstly, we suggest feedback activities to be interpreted to infer on user search goals which is an alternative to plain search results or URL in place. First, we introduce feedback activities to be analyzed to infer user search goals rather than search results or clicked URLs. Implicit feedbacks are both clicked and unclicked URLs before the last click. These implicit feedbacks are responsible for building the feedback activities. Both the clicked URLs and the unclicked ones before the last click are considered as user implicit feedbacks and taken into account to construct feedback activities. These feedback activities or feedback activities actually describe the user information needs and goals efficiently. Therefore feedback activities can reflect user information needs more efficiently.

Secondly, we align the feedback activities to the pseudo-documents that contain the key words to exact goal texts in user minds. The pseudo-documents are to upgrade the URLs with extra textual information that covers the titles as well as snippets second; we map feedback activities to pseudo-documents to approximate goal texts in user minds. The pseudo-documents can enrich the URLs with additional textual contents including the titles and snippets. Taking the support of these pseudo-documents user search goals is then identified and this depicts the user search goals with some keywords.Based on these pseudo-documents user search goals can then be Discover and depicters with some keywords.

VII. REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval.ACM Press, 1999.

[2] R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query Recommendation Using Query Logs in Search Engines," Proc. Int'l Conf.Current Trends in Database Technology (EDBT '04), pp. 588-596,2004.

[3] D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '00), pp. 407-416,2000.

[4] S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, "VaryingApproaches to Topical Web Query Classification," Proc. 30th Ann.Int'l ACM SIGIR Conf. Research and Development (SIGIR '07),pp. 783-784, 2007.

[5] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li,"Context-Aware Query Suggestion by Mining Click-Through,"Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and DataMining (SIGKDD '08), pp. 875-883, 2008.

[6] H. Chen and S. Dumais, "Bringing Order to the Web: AutomaticallyCategorizing Search Results," Proc. SIGCHI Conf. HumanFactors in Computing Systems (SIGCHI '00), pp. 145-152, 2000.

[7] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant TermSuggestion in Interactive Web Search Based on ContextualInformation in Query Session Logs," J. Am. Soc. for InformationScience and Technology, vol. 54, no. 7, pp. 638-649, 2003.

[8] T. Joachims, "Evaluating Retrieval Performance Using Click through Data," Text Mining, J. Franke, G. Nakhaeizadeh, andI. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.

[9] X. Wang and C.-X Zhai, "Learn from Web Search Logs to Organize Search Results," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '07), pp. 87-94, 2007.

[10] H.-J Zeng, Q.-C He, Z. Chen, W.-Y Ma, and J. Ma, "Learning to Cluster Web Search Results," Proc. 27th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '04), pp. 210-217, 2004.