

A New Approach To Discover User Search Goals Using Feedback Session

Pranali Dhondiram Desai¹, Prof.Wadne Vinod Subhashrao²

Abstract: *Nowadays internet is widely used by users for getting various information with the help of search engine. Web search applications represent user information needs by submission of query to search engine. Query submitted to search engine doesn't satisfy the user information needs, because users may want to get information on various aspects when they submit the same query. Therefore finding different user search goals becomes complicated. User search goal is analyzed for efficient web searching to provide best results. The evaluation and depiction of user search goals can be very useful in improving search engine relevance and user knowledge. Feedback sessions are generated by using clickthrough logs of various search engines. We propose a new approach to find different user search goals for a query by clustering feedback sessions. Then we propose a way to generate pseudo-documents for better representation of feedback sessions for clustering. Finally we propose "Classified Average Precision (CAP)" to understand user search goals efficiently and evaluate the performance of inferring user search goals. Here user needs is highlighted and provides a user friendly search engine.*

Keywords: user search, feedback session, classified average precision

1. Introduction

In web search applications queries are given to search engines for getting user search needs. Then web search engine provides the results for the query entered by the user. But sometime queries are unable to express the exact needs of the user because different queries may represent the different aspects. Users are usually giving some keywords representing their interests in their minds. Such keywords do not match with the results produced by the search engines. For example, when we search "eclipse" in the web search engine we will find the results related to eclipse as the natural eclipse's information sites, eclipse as software from the Eclipse foundation open source community websites, the twilight saga: Eclipse as a movie's sites.

Hence it is necessary to identify the user search goals or intensions. Here search goals of the user are nothing but the information on different aspects of a query that particular groups want to search. Also user goals can be stated as clusters of the information needs for a particular query. We can improve the search engine relevance and can make search engine user friendly by analyzing the user search goals.

The user submits the query into the browser. The search

engine searches the relevant information according to the user query. The user actions are stored in the user click through logs. From the user click through logs each and every session is analyzed and generates the feedback session. The feedback session contains both the clicked and unclicked URLs and the last clicked URL in a single session. The feedback session contains the URLs and the click sequence. By analyzing the feedback sessions, the pseudo documents are generated.

The pseudo documents contains the keywords that are most clicked in a session. Likewise the pseudo documents are clustered using the clustering algorithm. The user search goals are obtained according to the feedback sessions. The restructure result is produced for the user query based on the user search goal. The CAP evaluation can be done for each user search goal and the clustering can be done to find the optimal number of users.

Some advantages of this above approach are as follows:

Reorganize web search results [4],[7],[8] according to user search goals by grouping search results with same information need. This can be useful to other users with different search goals to find easily what they want.

Query recommendation [2],[3],[5] by using user search goals depicted with some keywords. This can be helpful to other users to form their query more effective.

Re-ranking web search results according to different user search goals.

2. System Architecture

Figure 1 shows the architecture of the system. All the feedback sessions 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.

The original search results are restructured based on the user search goals inferred from the above procedure. Then, we evaluate the performance of restructuring search results by our proposed evaluation criterion CAP. And the evaluation result will be used as the feedback to select the optimal number of user search goals.

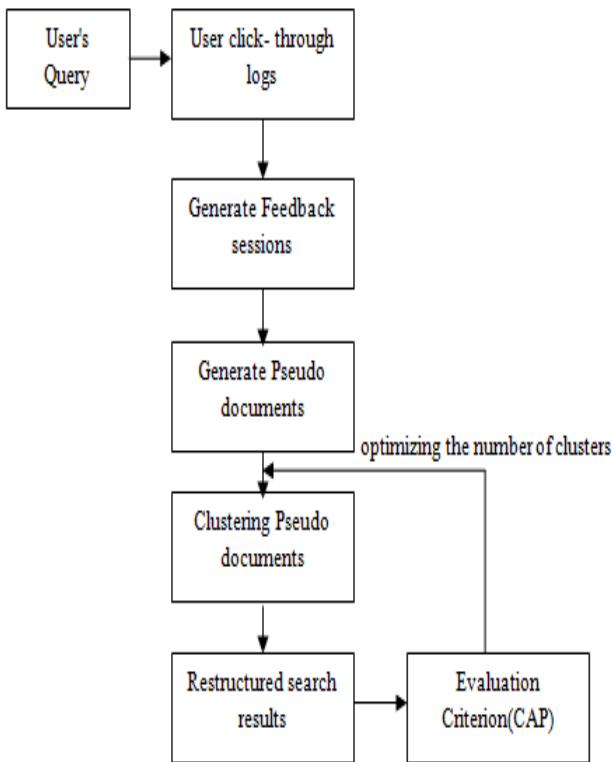


Figure 1: System Architecture

3. Representation Of Feedback Sessions

Here we can describe the proposed feedback sessions and then introduced pseudo documents to represent feedback session.

3.1 User search logs

The user enters the queries to the search engine. The queries are maintained as a log and the results will be produced based on the keywords. The search goals for a query and depicting each goal with some keywords automatically. The user's queries are saved.

3.2 Feedback sessions

For web searching a session is a series of successive queries to satisfy a single information need and some clicked search results. The user submits the query into the browser. The search engine searches the relevant information according to the user query [6].

The user actions are stored in the user click through logs. From the user click through logs each and every session is analyzed and generates the feedback session. The feedback session contains both the clicked and unclicked URLs and ends with the last clicked URL in a single session. The feedback session contains the URLs and the click sequence. 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.

It is more efficient to analyze the feedback sessions than to analyze the search results or clicked URLs directly.

3.3 Generating Pseudo Documents

The feedback sessions vary a lot for different clicks through and queries, it is not suitable to directly use the feedback sessions. Some procedure is needed to represent the feedbacks in a more efficient way. We can use binary vector method to represent a feedback session. The feedback sessions 1 as clicked and 0 as unclicked is denoted by using binary vector. The binary vector representation is not useful enough to tell the contents of user needs. Here we consider new approach to generate pseudo documents from feedback sessions.

Steps to construct pseudo documents:

➤ Represent the URL in the feedback session:

It extracts the titles and snippets of the returned URL's from the feedback sessions. Each URL is represented as a small text paragraph containing title and snippet. Then some textual process is implemented as text paragraphs such as transforming all the letters to lower case, stemming and removing stop words. And then TF-IDF vector of URL's titles and snippets are formed.

➤ Forming pseudo documents based on URL representations:

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

4. Clustering Pseudo Documents

Clustering is the process of grouping the data into classes or clusters. The Pseudo documents are clustered by using K means clustering algorithm. The K-means algorithm is simple and effective. The terms with the highest values in the center points are used as the keywords to depict user search goals. The clustering is the process based on a term-weight vector representation of queries, obtained from the aggregation of the term-weight vectors of the clicked URLs for the query.

After clustering all the pseudo-documents, each cluster can be considered as one user search goal. Clustering feedback sessions is more efficient than clustering search results or clicked URLs directly.

5. Evaluation Criterion And Restructuring Search Results

The results are restructured based on the evaluation of web search goals. This approach is called CAP(Classified Average Precision). Search engines will returns millions of search results so it is necessary to organize them to make it easier for users to find what they want.

Restructuring Web Search Results it is an application of inferring user search goals. The inferred ones are represented by the feature representation of each URL in the search result. Then categorize them into a cluster centered by the inferred search goals.

Average Precision evaluates according to user implicit feedbacks [1]. It is the average of precisions computed at the point of each relevant document in the ranked sequence. The URLs in the single session are restructured into two classes. Voted AP is the AP of class including more clicks namely votes.

Finally experimental result measures the clustering results with parameters like classified average precision (CAP), Voted AP (VAP), risk to avoid classifying search results and average precision (AP).

The user search goals are represented as the vectors. So we perform categorization by choosing the smallest distance between the URL vector and user search goal vectors. By this way the results can be restructured according to the inferred user search goals.

6. Conclusion

In this Project, a novel approach has been proposed to infer user search goals for a query by clustering its feedback sessions represented by pseudo-documents. Feedback sessions is constructed by considering Both the clicked URLs and the unclicked ones before the last click. We introduce feedback sessions to be analyzed to infer user search goals rather than search results or clicked URLs. Therefore, feedback sessions can reflect user information needs more efficiently. Then we generate pseudo documents from feedback session related to user search goals. The pseudo-documents can enrich the URLs with additional textual contents including the titles and snippets. User search goals can be represented with some keywords by using pseudo documents. Lastly, a new criterion CAP is formulated to evaluate the performance of user search goal inference.

The complexity of this approach is low and it can be implemented in reality easily. The running time depends on the number of feedback sessions. The proposed approach can discover user search goals for some popular queries offline at first. When user submit one of these queries, search engine can return the restructured search result. Thus users can find what they want ,conveniently.

References

- [1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. ACM Press, 1999. 512 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013
- [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] 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 Data Mining (SIGKDD '08)*, pp. 875-883, 2008.
- [4] H. Chen and S. Dumais, "Bringing Order to the Web: Automatically Categorizing Search Results," *Proc. SIGCHI Conf. Human Factors in Computing Systems (SIGCHI '00)*, pp. 145-152, 2000.
- [5] C.-K. Huang, L.-F. Chien, and Y.-J. Oyang, "Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs," *J. Am. Soc. for Information Science and Technology*, vol. 54, no. 7, pp. 638-649, 2003.
- [6] R. Jones and K.L. Klinkner, "Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs," *Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08)*, pp. 699-708, 2008.
- [7] 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.
- [8] 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.