

Identification of User Search Targets Using Feed Backs

¹Gujjarlapudi Srilatha, ²S.K.Chaitanya Rudraraju

^{1,2}Sir C.R.Reddy College of Engineering, Eluru, Andhra Pradesh, India

Abstract

Different users may have different search objectives and goals for a huge and confusing search item. The search engine performance can be improved by identifying and analyzing the search goals. In this paper, we propose a studied the approach to identify the user search goals by analyzing search engine query logs. The search goals of different users by clustering the proposed feedback from the search sessions.. to get the best results it is necessary to capture different user search goals. These user goals are nothing but information on different aspects of a query that different users want to obtain. The judgment and analysis of user search goals can be improved by the relevant result obtained from search engine and user's feedback. Here, feedback sessions are used to discover different user search goals based on series of both clicked and un clicked URL's. The pseudo-documents are generated to better represent feedback sessions which can reflect the information need of user. With this the original search results are restructured and to evaluate the performance of restructured search results, classified average precision is used.

Keywords

Search Goals, Feedback Sessions, Pseudo-Documents

I. Introduction

Web mining is one of the applications of data mining techniques to discover knowledge from the web. In web search, users are submitted queries to the search engines to get relevant information. But many search engines results are not informative and failed to produce results according to the user search goals. Users are usually giving some vague keywords representing their interests in their minds. Such keywords do not match with the results produced by the search engines. Many works about user search goals analysis should be carried out. Some users give ambiguous queries to the search engines they get mostly the irrelevant results. User search goals are classified as Navigational and Informational, the queries that seek a single website or webpage and queries that reflect the intent of the user to perform a particular transaction respectively. Many related works have been carried out according to the web search applications and the user search goals. In previous works, clustering is done on a set of top ranked results. The user search logs information is not analyzed and the feedback sessions are not considered. Analyzing the clicked URLs only from the web search logs. They only identify whether a pair of queries belong to the same goal or mission and does not care about what the goal is in detail. Semantic based web search for a particular query and the similarity between the words are carried out. Various algorithms such as star clustering algorithm, k-means clustering algorithm are used for clustering the pseudo documents but it also does not cluster the relevant information according to the user search goals. In clustering the cluster labels discovered are also not informative. User search goal is the information on different aspects of a query that users wants to obtain. Information need is a user's desire to obtain the relevant information to satisfy his need. To cluster web search results, the URLs are analyzed by extracting the titles and snippets. But all those works produced noisy results and does not obtain the user search goals precisely. When more irrelevant and relevant results are produced by the search engines it is time

consuming to browse. When 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 un clicked 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. The feedback session contains the URLs with the click sequence.

User Click-through data log contains data about the interactions between users and Web search engines. It is one of the most extensive surveys of user experience. For researchers it helps to understand human interaction with IR results. The user click through logs contains all the user actions. It contains the session id, query term, position of the URL, click sequence and the URL. The feedback sessions is discovered from each and every session from the user click through logs. The feedback sessions consists of the URLs that users visited and unvisited. Using the click sequence, the order in which the URLs are visited by the users the feedback sessions are generated. The feedback sessions consists of URLs that contains the URLs from first URL and up to the last visited URL. The feedback session is based on the users browsing actions that are stored in user click through logs according to the particular query.

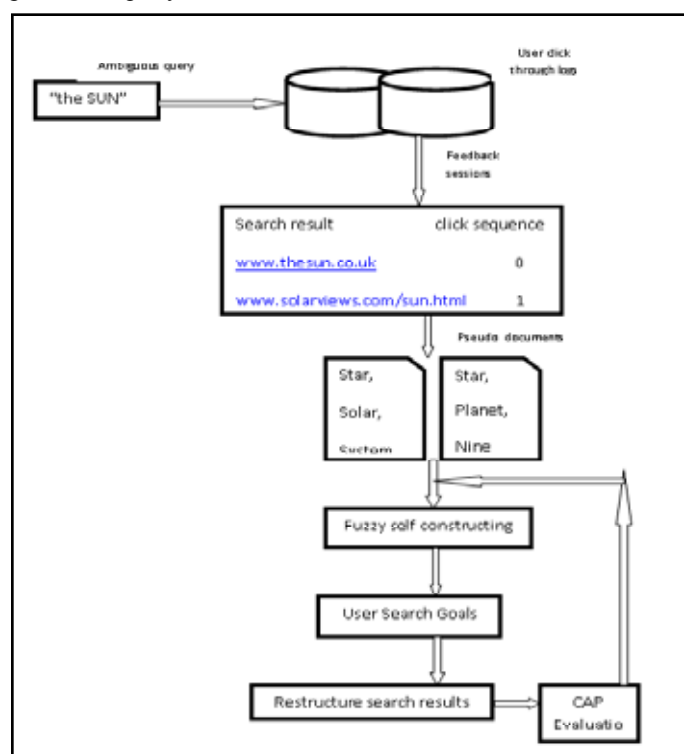


Fig. 1: Framework of the Approach

A novel optimization method to combine the enriched URLs in a feedback session to form a pseudo-document. This can effectively reflect the information need of a user. The URLs are enriched from the feedback sessions based on the click sequence. The enriched URLs with more value in click sequence are mapped to pseudo-documents. The pseudo documents are depicted with some keywords based on the URL. At last, cluster these pseudo documents to infer user search goals and depict them with some keywords. For clustering the pseudo-documents the fuzzy c-means clustering algorithm is used. The clustered pseudo documents are stored in the user search goals. From the user search goals the restructured results are produced. A new criterion CAP to evaluate the performance of user search goal inference based on restructuring web search results. Thus the restructured web search result is produced. This proposed novel criterion CAP to evaluate the restructure results.

I. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.

II. Conclusion

A novel approach for infer user search goals for an ambiguous query by clustering its feedback sessions represented by pseudo documents. The feedback sessions to be analyzed to infer user search goals rather than search results or clicked URLs. Both the clicked URLs and the un clicked ones before the last click are considered as user implicit feedbacks and taken into account to construct feedback sessions. Therefore, feedback sessions can reflect user information needs more efficiently. The pseudo documents can enrich the URLs with additional textual contents including the titles and snippets. Based on these pseudo documents, user search goals can then be discovered and depicted with some keywords. The pseudo documents are clustered using the fuzzy clustering algorithm. The cluster labels are discovered precisely. Finally, CAP is formulated to evaluate the performance of user search goal inference. The restructured web search result is produced for every user search query. The result produced is efficient and time consuming for users.

References

- [1] R. Baeza-Yates, B. Ribeiro-Neto, "Modern Information Retrieval", ACM Press, 1999.
- [2] X. Wang, 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.
- [3] H.-J Zeng, Q.-C He, Z. Chen, W.-Y Ma, 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.
- [4] H. Chen, 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] T. Joachims, "Optimizing Search Engines Using Clickthrough Data", Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02), pp. 133-142, 2002.
- [6] T. Joachims, L. Granka, B. Pang, H. Hembrooke, G. Gay, "Accurately Interpreting Clickthrough Data as Implicit Feedback", Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05), pp. 154-161, 2005.
- [7] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data", Text Mining, J. Franke, G. Nakhaeizadeh,