An Extension to New Algorithm for Feedback Session

Pooja Agarwal
NMIET

Swarupa Sonawane
NMIET

Pratiksha Tanpure
NMIET

ABSTRACT
Internet is widely used by users to satisfy various information needs. This paper proposed a novel approach to infer user goals. Pseudo documents are generated through feedback sessions. We introduce new criterion"classified average precision(CAP)", for evaluate performance of inferring user search goals. Results are represented on search engine to validate the effectiveness of our work. Extraction of interesting information from web data has become more popular and result of that web mining attracted lot of attention in recent times.

Keywords
User search goals,feedback session, pseudo documents,restructuring search result,classified average precision.

1. INTRODUCTION
In web based search applications, user submits the query to search engine to search efficient information. Different users have different information needs which is difficult to achieve user information needs. So to access the appropriate information needs we use web usage mining data. To satisfy the user information needs by considering the search goals with user given query, cluster the user information needs with different search goals. User search goals analysis is important to optimize search engines and effective query results organization. So we use clustering concept for search results since it is an efficient method to systematic search results. Our main goal is to discover different user search goals for a query and depict each search goal with some keywords automatically. Keywords are stored in web mining data. When the user submits the query, through the user clicks we can generate click log and also generate the feedback sessions. Once the feedback sessions are generated, we build pseudo documents and then we cluster the pseudo documents and reconstruct or restructure search results. Then the results are evaluated and it is given to feedback session generation.

2. RELATED WORK
2.1 Query Recommendation using Query Logs in Search Engines
Recently, search engines become more critical for finding information over theWorldWideWeb where web content growing fast, the users satisfaction of search engine a list of queries that are related to the user input query. The related queries are based on previously issued queries by the users. The proposed method is based on clustering process in which groups of semantically similar queries are detected. This facility provides some queries which are related to the queries submitted by users in order direct them toward their required information. This method not only discovered the related queries but also rank them according to a similarity measure. Finally the method has been evaluated using real data sets from the search engine query log.[1]

2.2 Context-Aware Query Suggestion by Mining Click-Through
Query suggestion plays an important role in improving the usability of search engines. Although some recently proposed methods can make meaningful query suggestions by mining query patterns from search logs, none of them are context aware they do not take into account the immediately preceding queries as context in query suggestion. In this , Author propose a novel context-aware query suggestion approach which is in two steps. In the one model learning step, to address data sparseness, queries are summarized into concepts by clustering a click-through bipartite. Then, from session data a concept sequence sux tree is constructed as the query suggestion model. In the online query suggestion step, a users search context is captured by mapping the query sequence submitted by the user to a sequence of concepts. By looking up the context in the concept sequence sux tree, the approach suggests queries to the user in a context-aware manner. The experimental results clearly show that they approach outperforms two baseline methods in both coverage and quality of suggestions[4].

2.3 Bringing Order to the Web Automatically Categorizing Search Results
Author developed a user interface that organizes Web search results into hierarchical categories. Text classification algorithms were used to automatically classify arbitrary search results into an existing category structure onthe-fly. A user study compared new category interface with the typical ranked list interface of search results. The study showed that the category interface is superior both in objective and subjective measures. Subjects liked thecategory interface much better than the list interface, and they were 50categories. Organizing search results allows users to focus on items in categories of interest rather than having to browse through all the results sequentially.[5]

2.4 Optimizing Search Engines using Click through Data
Author proposes a new method for evaluating the quality of retrieval functions. Unlike traditional methods that require relevance judgments by experts or explicit user feedback, it is based entirely on click- through data. This is a key advantage, since click through data can be collected at very low cost and without overhead for the user. Taking an approach from experiment design, the paper proposes an experiment setup that generates unbiased feedback about the relative quality of two search results without explicit user feedback. A theoretical analysis shows that the method gives the same results as evaluation with traditional relevance judgments under mild assumptions. An empirical analysis varies that the assumptions are indeed just and that the new method leads to conclusive results in a WWW retrieval study.

2.5 Accurately Interpreting Click through Data as Implicit Feedback

Author presents an approach to automatically optimizing the retrieval quality of search engines using click through data. Intuitively, a good information retrieval system should present relevant documents high in the ranking, with less relevant documents following below. While previous approaches to learning retrieval functions from examples exist, they typically require training data generated from relevance judgments by experts. This makes them difficult and expensive to apply. The goal of 4 this work is to develop a method that utilizes click through data for training, namely the query-log of the search engine in connection with the log of links the users clicked on in the presented ranking. Such click through data is available in abundance and can be recorded at very low cost. Taking a Support Vector Ma- chine (SVM) approach, this work presents a method for learning retrieval functions. From a theoretical perspective, this method is shown to be well founded in a risk minimization framework. Further more ,it is shown to be feasible even for large sets of queries and features. The theoretical results are varied in a controlled experiment. It shows that the method can effectively adaptthe retrieval function of a meta-search engine to a particular group of users, out- performing Google in terms of retrieval quality after only a couple of hundred training examples.

3. SYSTEM DESCRIPTION

The proposed system consists of both clicked and unclicked URL's and ends with last URL that was clicked in single session. Each feedback session gives information what user requires and what user does not require. There are many different feedback session in user click through logs. The aim of paper is to infer different. user search goals for query by clustering feedback sessions.
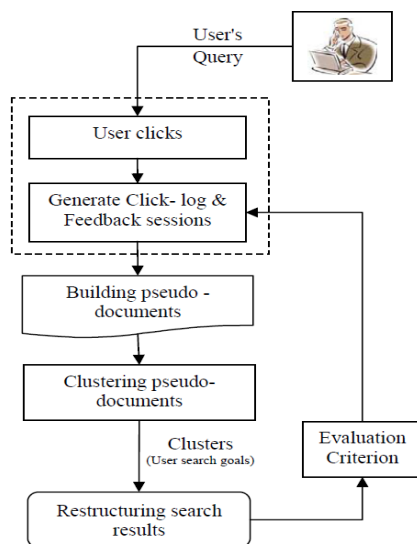

Fig1: Architecture

3.1 Click through data

In web search environment there are many ambiguous queries and user clicks. In this the URL clicked by user are recorded in user click through data. When the query is issued ,user scans links to document in result from top to bottom and then user click on the URL which is look relevant of informed choice and skips other.

3.2 Feedback session

Feedback session is both clicked and unclicked URL's and ends with last URL which was clicked in session from user. clustering feedback session more efficient then clustering search results

3.3 Pseudo documents

The pseudo document contain all keyword and which are based on feedback session. The keywords are related to ambiguous query. We perform clustering based on different values for knowing exact user search goal. The optimal value determined through evaluation criteria.

3.4 Clustering the pseudo documents

The Pseudo documents are clustered into K means clustering It performs clustering based on the five values. 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. Similar queries may not share query-terms but they do share terms in the documents selected by the users. Thus we avoids the problems of comparing and clustering sparse collection of vectors in which similar queries are difficult to find a problem that appears in previous works on clustering.
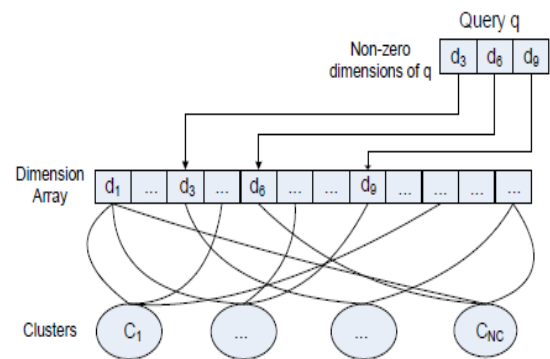

Figure 2: clustering

3.5 Final restructured 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. 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.

4. CONCLUSION

As the web and its usage continues to grow , so grows the opportunity to analyze web data and extract all manner of useful knowledge. In this approach we proposed to infer user search goals. For a user query by clustering its feedback session, it also proposed a novel optimization method to map feedback session.

5. REFERENCES

[1] Hamada M.Zahera, Gamal F. El Hady, Waiel.F Abd El-Wahed , Query Recommendation Using Query Logs in Search Engines, Proc. Intl Conf. Current Trends in

Database Technology (EDBT 04), pp. 588- 596, 2004.

[2] 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 Intl Conf. Knowledge Discovery and Data Mining (SIGKDD08), pp. 875-883, 2008.

[3] H. Chen and S. Dumais, Bringing Order to the Web: Auto- matically Categorizing Search Results, Proc.SIGCHI Conf. Human Factors in Computing Systems (SIGCHI 00), pp. 145-152, 2000

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

Performance Using Click- through Data, Text Mining,

[5] T. Joachims, Optimizing Search Engines Using Clickthrough Data, Proc. Eighth ACM SIGKDD Intl Conf.Knowledge Discovery and Data Mining (SIGKDD02), pp. 133-142, 2002.

[6] T. Joachims, L. Granka, B. Pang, H. Hembrooke, andG. Gay, Accurately Interpreting Clickthrough Data asImplicit Feed- back, Proc. 28th Ann. Intl ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR 05), pp. 154-161, 2005.

[7] R. Jones, B. Rey, O. Madani, and W. Greiner, Generating Query Substitutions, Proc. 15th Intl Conf. World Wide Web (WWW 06), pp. 387-396, 2006.