Enhancing the Search Result for User Query Using Iterative User Feedback

PriyaSurana[1], PrashantJamdade[2], TejasNandedkar[3], Akshay S. Ghadge[4], NirbhayBobde[5]

Assistant Professor, Dept. of CS, PCCOE, Savitribai Phule Pune University, Nigdi, India[1]

Student's, Dept. of CS, PCCOE, Savitribai Phule Pune University, Nigdi, India[2,3,4,5]

ABSTRACT: In real scenario when users submit a search query to a search engine, each user may have different search goals. We can improve search engine relevance by analyzing user search goals. We present an approach that infers user search goals by analyzing search engine query logs. Our approach discovers different user search goals for a query by clustering the proposed feedback sessions. Users information needs can be captured with the help of feedback sessions. Feedback sessions are constructed from user click-through logs and can efficiently reflect the information needs of users on results. Secondly, we also generate pseudo-documents for better representation of the feedback sessions for clustering. Finally, we present Classified Average Precision (CAP) to evaluate the performance of inferring user search goals. For an ambiguous query, different users may have different search goals when they submit it to a search engine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. In this paper, we present a novel approach to Enhance Search Result For User Query Using Iterative User Feedback.

KEYWORDS: User search goals, feedback sessions, pseudo documents, customized and generalized results, classified average precision (CAP).

I. INTRODUCTION

We aim at discovering the number of diverse user search goals for a query and depicting each goal with some keywords automatically. We present an advance approach to infer user search goals for a query by clustering our proposed feedback sessions. Then, we implement an optimization method to map feedback sessions to pseudo-documents which can efficiently reflect user information needs. At last, we cluster these pseudo documents to infer user search goals and depict them with some goal keys. In web based search applications, user submits the query to search engine to search efficient information. 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. For example, when user submits a query star to search engine, some users are interested to know information about astronomical information and some users want to know information about any websites which has domain name as star. Therefore, it is necessary to discover different user information search goals.

II. RELATED WORK

1. A new Algorithm for Inferring User Search Goals with Feedback Sessions[1]

Different users may have different search goals when they submit it to a search en- gine. The inference and analysis of user search goals can be very useful in improving search engine relevance and user experience. A novel approach to infer user search goals by analyzing search engine query logs. First, a framework to discover different user search goals for a query by clustering the proposed feedback sessions. Feedback sessions are constructed from user click-through logs and can efficiently reflect the in- formation needs of users. Second, a novel approach to generate pseudo-documents to better represent the feedback sessions for clustering. Finally, a new criterion Classified Average Precision (CAP) to evaluate the performance of inferring user search goals.

Experimental results are presented using user click-through logs from a commercial search engine to validate the effectiveness of the methods

2. Optimizing Search Engines Using ClickthroughData[2]

An approach to automatically optimizing the retrieval quality of search engines using clickthrough data. Intuitively, a good information retrieval system should present rel- evant documents high in the ranking, with less relevant documents. The approaches to learning retrieval functions from examples exist, typically require training data gen- erated from relevance judgments by experts. This makes difficult and expensive to apply. The goal of this paper is to develop a method that utilizes clickthrough 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 clickthrough data is available in abundance and can be recorded at very low cost. Taking a Support Vector Ma- chine (SVM) approach, this paper presents a method for earning retrieval functions From a theoretical perspective, this is shown to be well-founded in a risk minimization framework. Furthermore, it is shown to be feasible even for large sets of queries and features. The theoretical results are verified in a controlled experiment. It shows that the method can effectively adapt the retrieval function of a meta-search engine to a particular group of users, outperforming Google in terms of retrieval quality after only a couple of hundred training examples.

3. Varying Approaches to Topical Web Query Classification[3]

Topical classification of web queries has drawn recent interest because of the promise it offers in improving retrieval effectiveness and efficiency. However, much of this promise depends on whether classification is performed before or after the query is used to retrieve documents. As observed two previously unaddressed issues in query classification: pre vs. post-retrieval classification effectiveness and the effect of training explicitly from classified queries vs. bridging a classifier trained using a document taxonomy. Bridging classifiers map the categories of a document taxonomy onto those of a query classification problem to provide sufficient training data. To find that training classifiers explicitly from manually classified queries outperforms the bridged classifier by 48% in F1 score. Also, a pre-retrieval classifier using only the query terms performs merely 11% worse than the bridged classifier which requires snippets from retrieved documents.

III. PROPOSED ALGORITHM

In our paper we present restructured web results in two categories such as customized view and generalized view. To get customized view user must be logged in to our search engine. For this purpose we design our system features as follows:

- **Login**

In computer security, a login or sign-up is the process by which individual access to a computer system is controlled by identifying and authenticating the user referring to credentials presented by the user. A user can log in to a system to obtain access and can then log off when the access is no longer needed. To log out is to close off one's access to a computer system after having previously logged in.

- **User Search Goals**

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 and along with this user queries are also get saved.

- **Feedback sessions**

The feedback sessions is defined as the series of both clicked and un-clicked URLs and ends with the last URL that was clicked in a session from user click-through logs. Then we map the feedback sessions to pseudo-documents which can effectively reflect user information needs. We combine the enriched URLs in a feedback sessions to form a pseudo

document. The feedback session is based on a single session .and also it can be extended to the whole session. So besides the clicked URLs, the un-clicked ones before the last click should be a part of the user feedbacks. For inferring user search goals it is more efficient to analyze the feedback sessions than to analyze the search results or clicked URLs directly.

- **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 method id needed to represent the feedbacks in a more efficient way. The search log will be represented as 0 in the click sequence. The binary vector is used to represent the feedback sessions "1" as clicked and "0" as un-clicked. Steps to build pseudo documents:

- Represent the URL in the feedback session: It extracts the titles and snippets of the returned URLs from the feedback sessions. Each URL is represented as a small text paragraph then some textual process is implemented as text paragraphs such as transforming all the letters to lower case, stemming and removing stop words.
- Forming pseudo documents based on URL representations: Process of combining both clicked and un-clicked URLs in the feedback sessions.

- **Clustering of 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. So we do rank the suggested queries based on two criteria:

- The similarity of the queries to input query (the query submitted to the search engine).
- The support which measures how much the answers of the query have attracted the user attention.

- **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.

- Customized View
- Generalized View

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.

IV.   PSEUDO CODE

Step 1: Insert any query for searching.

Step 2: Collect the log details of user search history.

Step 3: Depend upon the users click and unclick URL's create a feedback session document.

Step 4: From the number of feedback sessions collected from multiple userd create psudo documents

Step 5: Divide this psudo Documents into temporary groups.

Step 6: Then by using K-means Algorithm create final clusters of collected URL's according to subject

Step 7: Forward this clusters for Classified Average Precision (CAP.)

Step 8: End.

V. SIMULATION RESULTS

- **Analysis Model**

In this paper we show a block diagram which depicts flow of data and information of the proposed search engine using feedback sessions.
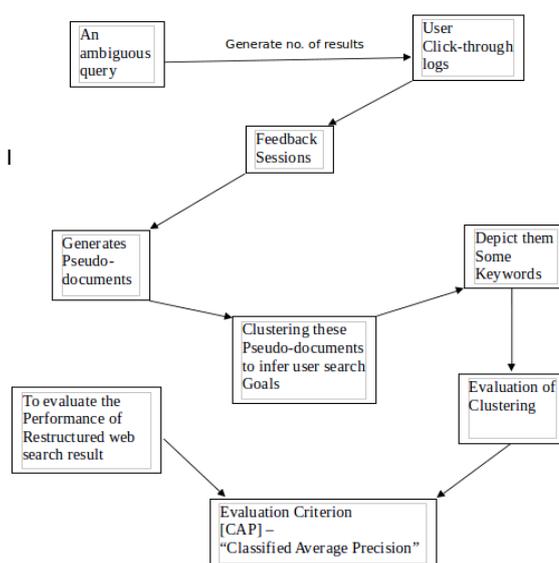

Fig. 1. Block Diagram of searching using Feedback

- **Mathematical Model**

Feedback Sessions Analysis:

We develop a model to map feedback session to pseudo document. Since feedback session is vary a lot for different click through and queries. It is unsuitable to directly use feedback session for user search goal.Therefore we use binary vector method to represent feedback session.

Search Results	Click Sequence	Binary Vector Representations
www.starsports.com	3	0
www.starradio.co.in	2	1
www.srartv.in/india	1	1
www.star.com	0	0
http://www.star.bnl.gov/	0	0

Table 1: A feedback session representation using binary vector

The binary vector [01100] can be used to represent feedback session where 1 represent click and 0 represent unclick URLs. Binary vector representation is not informative enough to tell the content of user search goal, Therefore new methods are needed to represent feedback session. For query, user will usually have some keywords representing there interest in their minds. We named that keywords as Goal Keys as user information needs they are not expressed explicitly, that's why we use pseudo document that can use to infer user search goal.

The building pseudo document include

1. Representing the URLs in feedback session

2. Forming pseudo document based on URL representation
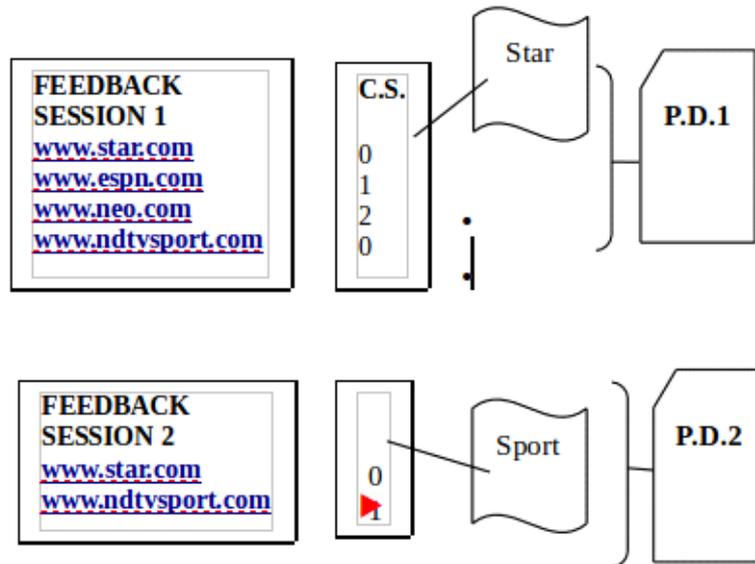With the help of this two-step we can develop a pseudo document.

Fig 2: Mapping of feedback session to pseudo document

Representing the URLs in feedback session :
URLs in Feedback session represented by a Term Frequency- Inverse Document Frequency [TF-IDF] as

$$T_{ui} = [t_{w1}, t_{w2}, \ldots\ldots\ldots, t_{wn}]^T,$$
$$S_{ui} = [s_{w1}, s_{w2}, \ldots\ldots\ldots, s_{wn}]^T$$

Where, T_{ui} and S_{ui} is TF-IDE vector of URL's title and snippet respectively. $w_j (j=1,2,\ldots n)$ is the jth term appearing in enriched URL's. Therefore, we represent enriched URL by

$$F_{ui} = w_t T_{ui} + w_s S_{ui} = [f_{w1}, f_{w2} \ldots\ldots f_{wn}]^T \quad (1)$$

Where, F_{ui} means the feature representation of the ith URL in the feedback session and w_t, w_s are the weights of the titles and the snippets, respectively.

Restructured Web Search Result:

We represent user search results in two views:
1) Generalized view and
2) Customized view

- **Generalized View**:

Like a most of the common web browsers search result this view gives all possible web search results without any concern regarded to categorisation. This view do not required any type of Login type verification for searching.

- **Customized view:**

This view represent a main objective. For customized view user must have to login verification by our search engine. Results obtained through this view are categorized into different classes. Depending upon users previous sessions the pseudo document is maintain which carry the count for each specific URLs. This count gives the priority to each URLs which affect in the incoming search results. As higher the count URL get higher priority and top position in search results.

In this view user can gives rating to each website. This rating by each user evaluated and reflected in customized search results.

VI.   CONCLUSION AND FUTURE WORK

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. First, we introduce feedback sessions to be analyzed to infer user search goals rather than search results or clicked URLs. 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 sessions. Therefore, feedback sessions can reflect user information needs more efficiently. Second, we map feedback sessions to pseudodocuments to approximate goal texts in user minds.

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. Finally, a new criterion CAP is formulated to evaluate the performance of user earch goal inference. Experimental results on user click-through logs from a commercial search engine demonstrate the effectiveness of our proposed methods. The complexity of our approach is low and our approach can be used in reality easily. For each query, the running time depends on the number of feedback sessions. However, the dimension of Ffs in (3) and (5) is not very high. Therefore, the running time is usually short. In reality, our approach can discover user search goals for some popular queries offline at first. Then, when users submit one of the queries, the search engine can return the results that are categorized into different groups according to user search goals online. Thus, users can find what they want conveniently.

REFERENCES

[1]Zheng Lu, HongyuanZha, XiaokangYang,"A New Algorithm for Inferring User Search Goals with Feedback Sessions". IEEE, March 2013.
[2] T. Joachims, Optimizing Search Engines Using Clickthrough Data, Proc. Eighth ACM SIGKDD Intl Conf. Knowledge Discovery and Data Mining (SIGKDD 02), pp. 133-142, 2002.
[3] S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, Varying Approaches to Topical Web Query Classification,Proc. 30th Ann. Intl ACM SIGIR Conf. Research and Development (SIGIR 07), pp. 783-784, 2007

BIOGRAPHY

Mrs.PriyaSuranais an Assistant Professor at computer engineering department of Pimpri-chinchawadcpllege of engineering, Nigdi, Pune, India - 411044.

Mr.PrashantJamdade, Mr.TejasNandedkar, Mr.AkshayS.Ghadge, Mr.NirbhayBobdeare the student at computer engineering department of pimpri-chinchawad college of engineering, Nigdi, Pune, India - 411044.