ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

An Algorithm to Infer User Sarch Goals with Feedback Session

Sharayu Kakade, Prof.Ranjana Badre

Department of Computer Science, MIT Academy of Engineering, Pune, MH, India

ABSTRACT: Whenever a user submits his query to the search engine, he may have a particular search goal that he wants to accomplish. But the search engine may provide thousands of results for the same query. Due to this, the user has to spend a lot of time finding the information of his interest. In this paper we introduce a novel approach to find the goal behind the users query to improve the pertinence and user experience. The proposed approach is used to find various user search goals for a query by grouping the user feedback sessions. Criticism sessions are constructed from user click through logs of different search engines. The method will generate pseudo-documents to represent feedback sessions for clustering. Finally, pseudo-documents are grouped and showed with some keywords. Then these user search goals are used to restructure the web search results.

KEYWORDS: user click-through logs, feedback sessions, pseudo documents, CAP (Classified Average precision).

I. INTRODUCTION

Today internet has become the largest source of retrieving the information. But due to the sparseness of the web, it has been difficult to find the relevant information. For example, when user submits a query “java” to search engine, some users are interested to know information about programming language and some users want to know information about island of Indonesia. Therefore it is necessary to capture the different search goals for pertinent information retrieval.
In this paper, we introduce the feedback session constructed from the user click-through logs, then map the feedback sessions to pseudo documents. After this we cluster the pseudo documents which is more efficient than clustering the search results or clicked URLs directly. We use CAP (Classified Average precision) to evaluate the clustering to reconstruct the web search results.

II. RELATED WORK

Zheng Lu in March 2013 proposed framework to find various user search goals for a query by clustering the proposed feedback sessions. Sessions are constructed from click-through logs and then efficiently reflect the information requirements of users. He proposed an approach to create pseudo-documents for on behalf of the sessions for clustering. At last, he derived a new criterion (CAP) to calculate the performance of inferring user search goals[1]. Rosie Jones, Kristina Lisa Klinkner in 2008 studied real sessions labeled into hierarchical tasks, and found that timeouts, whatever their length, are of incomplete utility in identify task limitations, achieving a maximum precision up to 70%. They proposed and evaluate a method for the automated segmentation of users’ query streams into hierarchical units [2].

H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li in 2008 proposed a novel context aware query suggestion approach which contains two steps. A concept sequence suffix tree is constructed as the query suggestion model. When looking up the context in the concept sequence suffix tree, their approach suggests queries to the user in a context-aware manner[3]. U. Lee, Z. Liu, and J. Cho in 2005 studied whether and how we can automate the goal-identification process. They proposed two types of features for the goal identification task: user-click behavior and anchor-link distribution. The untried evaluation show that by combining these features we can currently identify the goals for 90% of the queries studied[6]. X. Wang and C.-X Zhai in 2007 studied the two deficiencies of clustering search results which make it not always work well: (1) the clusters discovered do not necessarily correspond to the interesting aspects of a topic from the user's perspective; and (2) the cluster labels generated are not informative enough to allow a user to identify the correct cluster[8]. In this paper, we propose to address these two deficiencies by (1)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

learning interesting aspects of a topic from Web search logs and organizing search results accordingly; and (2) generating more meaningful cluster labels using past query words entered by users.
In the traditional search engine, there was no individual feedback session as the user authentication was not provided.

No feedback

In the beginning there was no feedback session. For a particular query Q, the search engine retrieved a ranked list of all the related documents as search results R1, R2, R3… as shown in the figure. As per the cluster hypothesis, the relevant documents tend to be more similar to each than to the irrelevant ones.

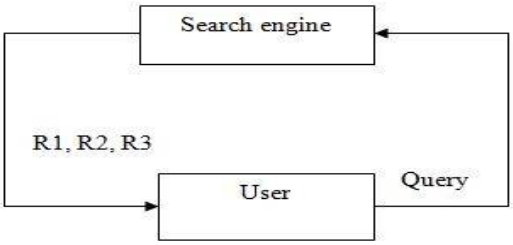

Fig. 1: No feedback

Relevance feedback

This system introduces a feedback loop. Typically this is done by extracting informative terms from the feedback documents thereby suggesting more precise terms for the user's query sessions which may better reflect the user's information needs [4][3].

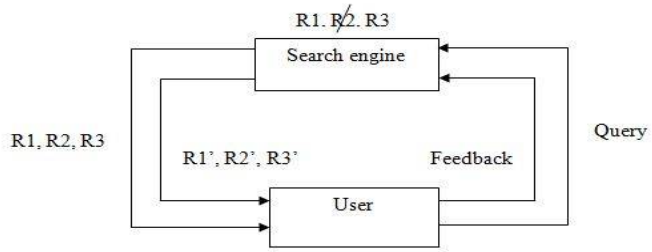

Fig. 2: Relevance feedback

Existing System

Recently most commercial search engines such as Google, Yahoo, Live Search, Ask and Baidu provide query suggestion to improve usability. That means search engine gives result as collective feedback. In collective feedback when a user issues a query, the logged database is usually looked up for feedback from the other users on the same query rather than from the same user on other queries in his/her history.

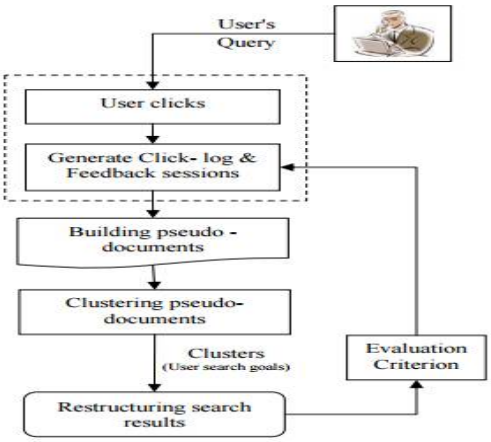

Fig. 3: Collective feedback flow

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

III. PROPOSED APPROACH

When user enters the query in the search box, the search engine provides log based term suggestion which suggest more precise terms for user's query session. It provides context-aware query session which takes into account the immediately preceding queries as context. It recommends relevant search terms, thus improving the real world search. The feedback session is defined as the series of both clicked and unclicked URL's and ends with the last URL that was clicked in a session from user clicked-through logs [1]. Every feedback session is defined within a timeout. Timeout is an elapsed time of 30 minutes (say, for threshold) between queries which signifies that the user has discontinued searching. To identify a timeout, a pair of queries is taken and decided whether the two queries come from different search goals. If the two queries belong to different search goals, the session has timeout.

After taking the feedback, the feedback session is mapped into pseudo-document which stores the enriched URL's. Enriched URL's are created by extracting titles and snippets of the URL's appearing is the feedback session[1].Then, pseudo documents are clustered , by using cosine similarity based clustering algorithm. The search results are organized by generating more meaningful cluster labels using the past query snippets and keywords stored in pseudo documents.CAP (Classified Average Precision) is also used as the evaluation criterion which evaluates according to the user implicit feedback. It is based on reconstructing web search result and used to select the best cluster to ensure that the user search goals are properly figured out.

IV. SYSTEM ARCHITECTURE

A new method is used to combine the enriched URLs in a feedback session to form a pseudo-document, which reflect the information requirements of a user and also a new criterion CAP is stated to calculate the performance of search goal inference which depends upon restructuring web search results. Thus, the number of user search goals for a query is decided. The analysis of user search goals is useful to improve search engine relevance and user experience. A framework for personalized web search is stated which take into account individual's interest into mind and enhances the traditional web search by suggesting the relevant pages of interest. An efficient model is stated to ensure good suggestions as well as promises for effective and relevant information retrieval.

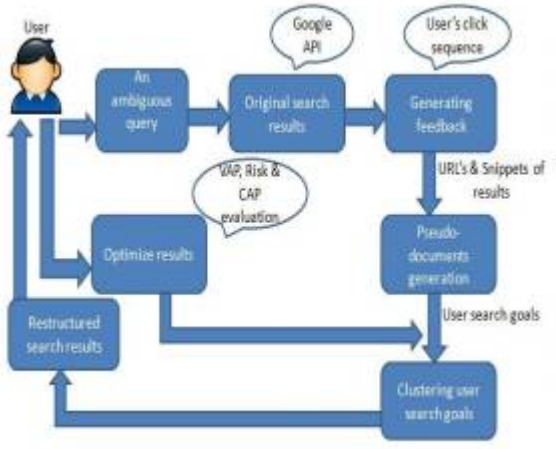

Fig.4 General Architecture of Proposed Framework

This system considers user's profile (based on user's web log navigation browsing history) and Domain Knowledge in order to perform modified web search. Using a Domain Knowledge, the system stores information about different domain categories. in order obtained from User Profile is classified into these specified categories. The studding agent learns user's choice automatically through the analysis of user routing browsing times gone by, and creates update better User Profile conditioning to the user's most recent choice. Once the user inputs query, the scheme provides good suggestions for personalized web search based on enhanced user profile. Further, the model makes good use of the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

advantages of popular hunt engines, as it can re-rank the fallout obtained by the search engine based on the enhanced user profile.
The data structure can be used to store the weights in an adjacency matrix M where each entry M_{ab} contains the value W_{ab} computed according to (3) .To limit the number of edge in such graph ,element of M_{ab} whose value is less than a threshold are too little correlated and thus discarded. This threshold is named as MinFreq in this contribution.

Cosine similarity based clustering algorithm

With the proposed pseudo-documents, user search goals can be inferred. In this part, it is described how to infer user search goals and depict them with some meaningful keywords. As in (3) and (4), each feedback session is represented by a pseudo-document and the feature representation of the pseudo-document is $F_{fs.}$ The similarity between two pseudo-documents is computed as the cosine score of F_{fi} and F_{sj} , as follows:

$$Sim_{i,j}=\cos(F_{fi}, F_{sj})$$
$$= F_{fi} . F_{sj} /| F_{fi}|. |F_{sj}|$$

And the space among two feedback sessions is

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

Pseudo-documents is clustered by cosine similarity based clustering. After clustering all the pseudo-documents, every bunch can be measured as one user search goal. The center point of a bunch is computed as the average of the vectors of all the pseudo-documents in the bunch, as showing in

$$F_{center\ i}=\sum^{Ci}_{k=1} F_{fsk} /C_{i\ ,} (F_{fsk} \subset \text{Cluster i})$$

where $F_{center\ i}$ is the i^{th} cluster's center and C_i is the number of the pseudo-documents in the i^{th} cluster. $F_{center\ i}$ is utilized to conclude the search goal of the i^{th} bunch. Finally, the provisos with the highest values in the center points are used as the keywords to depict user search goals. An additional advantage of using this keyword- based description is that the extracted keywords can also be utilized to form a more meaningful query in query recommendation and thus can represent user information needs more effectively. Moreover, since the number of the feedback sessions can be obtained in each bunch, the useful distributions of user search goals can be obtained at the same time. The relation of the numeral of the reaction sessions in one cluster and the total number of all the feedback sessions is the distribution of the corresponding user search goal.

Advantages

- Feedback sessions can be considered as a process of resembling.
- Feedback session is also a meaningful combination of several URLs.
- When users submit one of the query, the look for engine can go back the fallout that are categorized into different groups according to user search goals online. Thus, users can discover what they want expediently.
- A novel optimization method is proposed to combine the enriched URLs in a feedback session to form a pseudo-document, which can efficiently reproduce the information need of a user. Thus, it is to tell what the user search goals are in detail.
- A new criterion CAP is proposed to evaluate the performance of user search goal inference based on restructuring web search results. Thus, it is easy to determine the number of user search goals for a query.

V. CONCLUSION

In this paper, it is aimed to infer the user search goals to enhance the quality of the information retrieval from the search engine. The proposed approach gives optimized web search results. It provides query enhancement for optimization. The system is less time consuming and less efforts required for searching targeted data.

REFERENCES

[1] Zheng Lu, Student Member, IEEE, HongyuanZha, Xiaokang Yang, leading Member, IEEE, Weiyao Lin, affiliate, IEEE, and Zhaohui Zheng "A New Algorithm for Inferring User Search Goals with Feedback session" VOL. 25, NO. 3, MARCH 2013

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

[2] R. Jones and K.L. Klinkner, "clear of the Session rest: Automatic Hierarchical Segmentation of Search Topics in Query Logs," Proc. 17th ACM Conf. in turn and familiarity Management (CIKM '08), pp. 699-708, 2008.
[3] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-Aware inquiry Suggestion by Mining Click-Through," Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '08), pp. 875-883, 2008.
[4] C.-K Huang, L.-F Chien, and Y.-J Oyang, "pertinent Term proposition in Interactive Web Search Based on Contextual Information in Query Session Logs," J. Am. Soc. for in sequence Science and expertise, vol. 54, no. 7, pp. 638-649, 2003.
[5] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.
[6] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," Proc. 14th Int'l Conf. World Wide Web (WWW '05), pp. 391-400, 2005.
[7] D. Shen, J. Sun, Q. Yang, and Z. Chen, "Building Bridges for Web question categorization," Proc. 29th Ann. Int'l ACM SIGIR Conf. explore and expansion in Information Retrieval (SIGIR '06), pp. 131-138, 2006.
[8] X. Wang and C.-X Zhai, "Learn from Web Search Logs to Organize Search Results," Proc. 30th Ann. Int'l ACM SIGIR Conf. investigate and Development in Information Retrieval (SIGIR '07), pp. 87-94, 2007.
[9] J.-R Wen, J.-Y Nie, and H.-J Zhang, "Clustering User query of a look for Engine," Proc. Tenth Int'l Conf. World Wide Web (WWW '01),pp.