Proceedings of 1st Shri Chhatrapati Shivaji Maharaj QIP Conference on Engineering Innovations
Organized by Shri. Chhatrapati Shivaji Maharaj College of Engineering, Nepti, Ahmednagar
In Association with JournalNX - A Multidisciplinary Peer Reviewed Journal, ISSN No: 2581-4230
21st - 22nd February, 2018

A SURVEY ON: INFERRING USER SEARCH GOALS ENGINE THROUGH IMAGE CLICK THROUGH DATA

Prof.Agale Deepali B.
Department of Computer Engineering ,Savitribai Phule University SCSMCOE,Nepti,Ahmednagar
deepa.agale@gmail.com

Prof.Aher Shital M.
Department of Computer Engineering, Savitribai Phule University SCSMCOE,Nepti,Ahmednagar
shitalaher5445@gmail.com

Abstract— Different users may have different search goals when they submit broad-topic and ambiguous query, to a search engine. The inference and analysis of user search goals can be very useful in improving performance of search engine. To infer user search goals by analyzing search engine query logs a novel approach is proposed. First, we propose a framework to find out different user search goals for a query by clustering the proposed feedback sessions. Second, we propose a novel approach to generate pseudo-documents by using feedback sessions for clustering. Thus the project focuses on combine approach of web usage mining and web content mining to improve the search engine results by inferring user search goals. In the proposed work a new approach is introduced to re-order the search results based on the contents and user interest rather than keyword and page ranking provided by search engines. This paper presents a survey on inferring the user meant image retrieval under user feedback information.

Keywords— User Search Goals Feedback Sessions, Pseudo- Documents, Restructuring Search Results, Classified Average Precision,Re-ranking.

1.INTRODUCTION

Many ambiguous queries may cover a broad topic and different users may want to get information about different aspects when they submit the same query. For example, when the query "the sun" is submitted to a search engine, some users want to get information about a United Kingdom newspaper, while some others want to learn about the natural knowledge of the sun. In this paper, we aim at searching the number of diverse user search goals for a query and depicting each goal with some keywords automatically. First we propose a new approach to infer user search goals for ambiguous query by clustering our proposed feedback sessions. The feedback session is defined as the series of clicked and unclicked URLs and ends with the last URL that was clicked in a session from user click- through logs. Then, we propose an optimization method to map feedback sessions to pseudo-documents. Then, we cluster these pseudo documents to infer user search goals and depict them with some keywords. Since, we also propose an evaluation criterion classified average precision (CAP) to evaluate the performance of the restructured web search results. The above technique is only for text type of data. So,in the proposed sceanario. we are also including the image format data, which compares the different kind of images.

The objectives of the proposed work are as below:

1. **Feedback Session:** The proposed feedback session consists of both clicked and unclicked URLs and ends with the URL that was clicked in a single session at last. It is motivated that before the last click, all the URLs have been scanned and evaluated by users. In this way it shows the listed clicked and unclicked URL's by user.

2. **Optimization method to map:** feedback session is mapped onto pseudo document, which consists of titles and snippets by using optimization method.

3. **Generate image visual data:** The visual features of an image is useful to reduce semantic gap between image and a user query. For this purpose a low level feature extraction is done. This visual data is extracted from images from user click-through logs. The images clicked by users with common search goal have common visual pattern whereas the images clicked by users with different search goals have different visual patterns.

2. RELATED WORK

1) In recent years, many works have been done to infer the so called user goals or intents of a query [14], [15], [18]. But in fact, their works belong to query classification. Some works analyze the search results returned by the search engine directly to exploit different query aspects [7], [21]. However, query aspects without user feedback have limitations to improve search engine relevance. Some work stake user feedback into account and analyze the different clicked URLs of a query in user click-through logs directly, nevertheless the number of different clicked URLs of a query may be not big enough to get ideal results. Wang and Zhai clustered queries and learned aspects of these similar queries [19], which solves the problem in part. However, their method does not work if we try to discover user search goals of one single query in the query cluster rather than a cluster of similar queries. For example, in [18], the query "car" is clustered with some other queries, such as "car rental," "used car," "car crash," and "car audio." Thus, the different aspects of the query "car" are able to be learned through their method. However, the query "used car" in the cluster can also have different aspects, which are difficult to be learned by their method.

2) Some other works introduce search goals and missions to detect session boundary hierarchically [12].

However, their method only identifies whether a pair of queries belong to the same goal or mission and does not care what the goal is in detail. A prior utilization of user click through logs is to obtain user implicit feedback to enlarge training data when learning ranking functions in information retrieval. Thorsten Joachims did many works on how to use implicit feedback to improve the retrieval quality [9], [10],[11]. They consider feedback sessions as user implicit feedback and propose a novel optimization method to combine both clicked and unclicked URLs in feedback sessions to find out what users really require and what they do not care. One application of user search goals is restructuring web search results. There are also some related works focusing on organizing the search results [7], [8], [21]. They infer user search goals from user click-through logs and restructure the search results according to the inferred user search goals.

3) Image Search Re-ranking With Query-Dependent Click- Based Relevance Feedback: A method referred to as Click based relevance feedback is employed for image search re-ranking[13]. It will click primarily based connection feedback(CBRF) instead of pseudo connection feedback(PRF). In CBRF, the pseudo positive data are clicked pictures and haphazardly chosen pictures for other queries as pseudo negative information. It apply multiple kernel learning algorithmic rule to be query dependant fusion weights.

4) Inferring image results under implicit guidance of users: A work on giving image results under implicit guidance of users considers past users guidance[8] which exploits user click information and the multiple visual features. It cannot infer image goals for group of similar queries if exact query match not found. It analyze the click through log but it does not perform ranking of the goal images by their distributions. In the proposed work, the distribution of the search goals is used for ranking of images. Mining Latent Attributes From Click-Through Logs for Image Recognition: The attribute vocabulary are retrieved
from click through log by matrix factorization[17]. This matrix contains query and images for that query. It gives all latent topics from click through logs but not optimized number of image goals for a user. It considers interactions among user clicks as well as correlation between queries.

3.SURVAY OF PROPOSED WORK

A. Feedback Session: The proposed feedback session consists of both clicked and unclicked URLs and ends with the URL that was clicked in a single session at last. It is motivated that before the last click, all the URLs have been scanned and evaluated by users. In this way it shows the listed clicked and unclicked URL's by user.

B. Optimization method to map: feedback session is mapped onto pseudo document, which consists of titles and snippets by using optimization method.

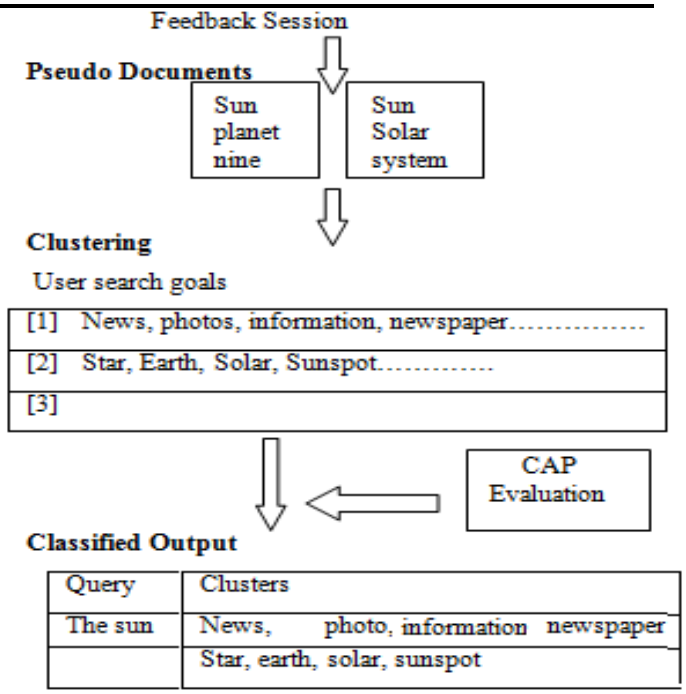

Figure 3.1: Framework of existing system

The overall system architecture is as shown in figure 3.1 below. First block of figure displays the feedback session of search results. Then by using optimization method titles and snippets are extracted from feedback session, this collection of titles and snippets is called as pseudo document. After that clustering algorithm applied on to these pseudo document. Finally clustered data is obtained and at the same time by using evaluation criteria performance of algorithm is checked.

3. ALGORITHM FOR CLUSTERING:

For Bisecting algorithm you must cluster documents using kmeans algorithm and then on the result of k-means algorithm you can apply bisecting algorithm.

Steps for bisecting Algorithm:

The idea is iteratively splitting your cloud of points in 2 parts. In other words, you build a random binary tree where each splitting (a node with two children) corresponds to splitting the points of your cloud in You begin with a cloud of points.

- Compute its centroid w
- Select randomly a point cL among the points of the cloud
- Construct point cR as the symmetric point of cL when compared to w (the segment cL->w is the same as w->cR)
- Separate the points of your cloud in two, the ones closest to cR belong to the subcloud R, and the ones closest to cL belongs to the subcloud L
- Reiterate for the subclouds R and L

Note:You can discard the random points once you've used them already. However, keep the centroids of all the subcoulds.

Stop at point when your subclouds contain exactly one point.

C. Generate image visual data: The visual features of an image is useful to reduce semantic gap between image and a user query. For this purpose a low level feature extraction is done. This visual data is extracted from images from user click-through logs. The images clicked by users with common search goal have common visual pattern whereas the images clicked by users with different search goals have different visual patterns.

D. Extract click session data: Then extract the click session data from user click-through logs. The clicked pictures in a session have high correlation. In real situation, several user might click on some irrelevant images. For instance, not withstanding a user solely desires to search the fruit apple at the start once he submits the query, he may click some pictures regarding the logo of apple.

E. Data Combining: Clustering is employed to mix every image visual information and click session information. Spectral clustering is introduced as a result of clusters representing completely different user goals might have different shapes in visual feature house once cluster for instance, the shapes of the clusters green apples, red apples and red laptops area unit spherical. The string connecting two points implies that these two pictures seem simultaneously in a minimum of one session. Therefore new cluster green and red apples will be formed.

4. CONCLUSION

This paper gives detailed description of existing text based search engine as well as a detailed survey on inferring user image search goals. A survey on existing techniques for web log analysis is done. Also survey on click session information is carried out. Some existing techniques for image retrieval and image ranking are introduced. In the proposed work the click session information and image visual data will be combined to infer user image-search goals. Click session data will serve as the implicit information of the past queries to assist clustering supported this framework.

5. FUTURE WORK

For the new query not appearing in the query log, the new queries can be classified into a query cluster initially. Then the user search goals for the query cluster will be considered for this new query.

ACKNOWLEDGMENT

We thankful to all our collogues for directly or indirectly supporting us for selecting such topic..

REFERENCES

1) Zheng Lu, Student Member, IEEE, Hongyuan Zha,Xiaokang Yang, Senior Member, IEEE, Weiyao Lin, Member, IEEE, and Zhaohui Zheng**,"** A New Algorithm for Inferring User Search Goals with Feedback Sessions"
2) S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, "Varying Approaches to Topical Web Query Classification," Proc. 30th Ann.
3) Int'l ACM SIGIR Conf. Research and Development (SIGIR '07), pp. 783-784,2007.
4) B. Poblete and B.-Y Ricardo, "Query-Sets: Using Implicit Feedback and Query Patterns to Organize Web Documents," Proc. 17th Int'l Conf. World Wide Web(WWW '08), pp. 41-50, 2008.
5) D. Shen, J. Sun, Q. Yang, and Z. Chen, "Building Bridges for Web Query Classification," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '06), pp. 131-138, 2006.
6) 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.
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) 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.
9) 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.
10) D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '00), pp. 407-416, 2000.
11) M. Pasca and B.-V Durme, "What You Seek Is what You Get: Extraction of Class Attributes from Query Logs," Proc. 20th Int'l Joint Conf. Artificial Intelligence (IJCAI '07), pp. 2832-2837, 2007.
12) A. Bonnaccorsi, "On the Relationship between Firm Size and Export Intensity," Journal of International Business