IMAGE RETRIEVAL SYSTEM: BASED ON USER REQUIREMENT AND INFERRING ANALYSIS TROUGH FEEDBACK

[1]Mount Steffi Varish.C, [2]Guru Rama SenthilVel

Abstract - Image Mining is a recent trended approach enveloped in extracting knowledge through preprocessing. For a similar image given as query, different users may require different search goals when they submit to a search engine. The analysis of user search goals poses advantage in improving search engine relevance and user experience. In this paper, an efficient approach is proposed to infer the search intentions by analyzing the search through query logs. The search goals are restructured to make the search more efficient. Therefore feedback session is summarized. The feedback session is obtained from user click-through logs and can efficiently reflect the information needs of the user. For evaluating the performance of the search intentions new criteria called Feature weighted Average Detection Algorithm is introduced. This algorithm is a technique to remove the irrelevant datasets for a particular search through an image. The fundamental difference between the classifications and ranking through feedback is summarized in this paper. Feature weighting alone cannot produce the clear difference to any search. Therefore, Average Detection using Correlations of the image is calculated. The performance is represented through experimental results to experience the efficiency of the proposed features.

Key Words - User Search Goals, Feedback Session, Query logs, Click-through Logs, Feature Weighted Average Detection.

I. INTRODUCTION

For several search applications that are oriented with web poses queries to submit for the search engine to get the needed information for the user.

- ***Mount Steffi Varish. C*** *is currently pursuing masters degree program in computer science engineering.*
  Ph -9944396169.
- ***Guru Rama Senthilvel****, M.E., A.P/ HOD of IT Dept.,*
  Ph-98410 65075

In many cases user doesn't get the specific information required. As the single query subject to cover broad topic, different users may require different information when the same query is submitted. For example when the query the sun is submitted to a search engine, some want to locate the homepage of a United Kingdom Newspaper, while some wants to learn the natural knowledge of the sun. The search intentions here represent the cluster of information needed for the query. The analysis of search goals has lot of advantages over the clustering of the search goals are restructuring of the search results. The restructuring of the search results is done according to the same query subjected to the search by different users for various purposes. This restructuring of the search results are used in the re-ranking of the web search applications.

This paper focuses on discovering the number of diverse user search goals for the query and depicting each goal with some keyword. To obtain the above mentioned property for a query, feedback sessions are introduced. The feedback session is defined as the series of both clicked and unclicked URL, and finally ends up with the last URL that was suggested for the keyword. Each of the feedback session can be clustered together based on the keyword. So evaluation of clustering the feedback session leads to important problem. Therefore clustering of the feedback session is done through Feature Weighted Average Detection Algorithm. This is to evaluate the performance of restructuring the search results. Since this parameter helps to detect the features of the image and cluster the similar objects, it helps to optimize the results of the user search intentions.

The paper is organized with related work in Section 2. Feedback session is explained in section 3. Design Architecture is proposed in Section 4. Section 5 describes the algorithm.

 All Rights Reserved © 2014 IJARCET

Restructuring of the search results in described in Section 6. Section 7 reveals the conclusion of the paper.

II. RELATED WORK

2. LITERATURE SURVEY

A. Feedback Session

A New Algorithm for Inferring User Search Goals with the Feedback Sessions proposed by Zheng Lu, Hongyuan Zha is based on the feedback session constructed from user click-through logs and can reflect the information needs of the users. This uses the Classified Average Precision Algorithm to structure the user search goals based on the click-through logs. The click – through logs is the number of visit the user had made with the particular URL. User Search goals can be analyzed with three classes such as Query Classification, Search Result Recognition and Session Boundary Detection. The motivation here is that before the last click all the URLs has been scanned and evaluated by the users. Each feedback can say what the user requires and what he doesn't want. Hence the analysis of the feedback sessions is much more needed. Therefore, feedback session can reflect user information.

B. Query Recommendation

Query Recommendation Using Query Logs in Search Engines proposed by R.Baeza-Yates, C. Hurtado describes that for a given query which is submitted to the search engine will list group of searches related to that query. The related queries are based on the previously issued queries and can be issued by the user to the search engine to tune or redirect the search process. This method is based on query clustering process in which the groups of semantically similar queries are identified. The clustering process uses the content of the historical preferences of users registered in the query logs of the search engine. The method not only discovers the related queries but also ranks them according to the relevance.

C. Optimization of Search Engine

Optimizing Search Engines Using Click Through Data by T.Joachims proposed an approach that automatically optimizes the retrieval quality of the search engine using click-through data. A good information retrieval system should present the relevant document high in ranking with the less relevance documents. The aim of the query search is to develop a method that utilizes click-through data for training the query log of search engine in connection with the log of the links that the users clicked on the presented ranking. SVM approach is to present the method of learning retrieval functions. This approach adopts the retrieval function of the search engine to particular group of the users. Optimization is of evaluating the number of service method to categorize the obtained results based on the values obtained during the performance calculated.

D. Implicit Feedback

Accurately Interpreting Click through Data as implicit Feedback examines the reliability of implicit feedback generated from click through data in www search. Analyzing the user search process using tracking and comparing implicit feedback against the manual relevance judgments, the clicks are informative but biased. While this makes the interpretation of the clicks as more absolute relevance judgments difficult, the relevance preferences derived from the clicks are reasonably more accurate. Implicit feedback is potentially more noisy and difficult. There are two ways of clicking decision biased which is trust and quality bias. The trust bias leads to more click on the links ranked highly by Google. Quality bias is that the user clicking decision is not only influenced by the relevance of the clicked links but also on the overall quality and ranking of the abstracts.

E. Content Based Image Retrieval

A Survey on: Content based Image Retrieval Systems introduces the CBIR technique that helps to organize the digital image archives by their visual content. From this statement, anything ranging from image similarity function to a robust image annotation engine falls under the preview of CBIR. CBIR uses visual content called the features to search images from the large scale image databases according to the user request in the form of the query image. An algorithm called edge histogram is used to extract the features of the images. The content based retrieval can be of color, shape, texture. The fuzzy inference integrates the perfect feature of the image in the content based retrieval. CBIR is the most used technical aspect to be used in retrieval of image from the web services.

F. Image Retrieval Systems

Image Retrieval: Current Issues, Techniques, Promising Issues and other Trends

 All Rights Reserved © 2014 IJARCET

focuses mainly on image retrieval for both text and images. Most image retrieval systems support random browsing , search and navigation of the images with the custom searches. QBIC standing for the query by the image content, is the first commercial content based image retrieval systems. Its system framework and techniques have profound effects on later image retrieval systems. Virage is a content-based image search engine developed at virage Inc. Virage supports visual queries based on the color, composition, texture, structure. It also supports arbitrary combinations of the above mentioned atomic queries.

G. Retrieval Evaluation

Evaluating retrieval through click-through data describes unlike the traditional methods that require relevance judgments by experts or explicit user feedback. It is based on the data that can b collected at very low cost and without overhead for the user. Taking the approach from the experimental design, the paper proposes an experimental setup that generates unbiased feedback about the relative quality of two search results without explicit user feedback. A theoretical relevance shows that the method gives the same results as the evaluation with relevance judgments under mild assumptions. An empirical analysis verifies that the assumptions are indeed justified and the new method leaves to the conclusive results.

H. Fuzzy Keyword Search

The Efficient Interactive Fuzzy keyword search uses traditional information systems that return answers after a user submits a complete query. Users often feel left in dark when they have limited knowledge about the underlying data, and have to use a try and see approach for finding information. A recent trend in supporting auto complete in these systems is the first step to solve the problems. The auto complete interfaces are extended by allowing the keywords to appear in multi attributes and finding relevance to matching the keywords.

III. FEEDBACK SESSION

A session for web search generally involves the series of the successive queries to satisfy a single information need and some clicked search results. This paper focuses on the user search goals for a single query. The feedback session in this paper is based on the single session and is extended to the whole session.

The feedback session consists of both clicked and unclicked URLs and ends up with the URL that was clicked in the single session. It is illustrated that the keyword submitted to the search is provided with the number of URL related to the particular keyword. Based on the click through logs the web search results are restructured. Inside the feedback session the clicked URL tells what the user require and the unclicked ones says what the user is not needed with. Hence inferring of the user search goals here is to analyze feedback session to provide efficient search results.

A. Feedback Representation

Before inferring the user search goals through feedback, some methods that were used to represent the search criteria is discussed. One among such representation is that the binary vector representation. The binary vector [1 0 0 1 1 0 0] can be used to represent the feedback session, where 1 represents the clicked and 0 represents the unclicked URL. Feedback with vector is represented in the figure 1.

However the binary vector representation does not yield clear information for restructuring of the results, an alternate method is analyzed.

B. Representing the URLs in the Feedback

In the first step, the URLs are enriched with the additional textual contents by extracting the titles, snippets of the returned URLS appearing the feedback session. In this way, each URL in a feedback session is represented by a small text paragraph that consists of the titles and snippets. Then some textual processes were implemented to those text paragraphs such as transforming all the letters to lower cases and remaining stop words. Feedback sessions can be considered as a process of resampling the data.

Search results	Click sequence	Binary vector
www.thesun.co.uk/	0	0
www.nineplanets.org/sol.html	1	1
www.solarviews.com/eng/sun.htm	2	1
en.wikipedia.org/wiki/Sun	0	0
www.thesunmagazine.org/	0	0
www.space.com/sun/	0	0
en.wikipedia.org/wiki/The_Sun_(newspaper)	3	1

Fig.1 The binary vector representation of the feedback session

Feedback session is also meaningful combination of several URLs. Therefore, it can reflect user required information more precisely and there are plenty of feedback sessions to be

 All Rights Reserved © 2014 IJARCET

analyzed. Feedback session can be analyzed in three cases. In case1 one term appears in all clicked URLs and does not appear in any unclicked ones. In this case, people usually skip because the unclicked URLs does not contain this term. In case3 both the clicked and unclicked URLs are about one single subject and the term is not distinguishable. Feedback can be managed for images also based on feature extraction through similar attributes. This is described in figure 2.

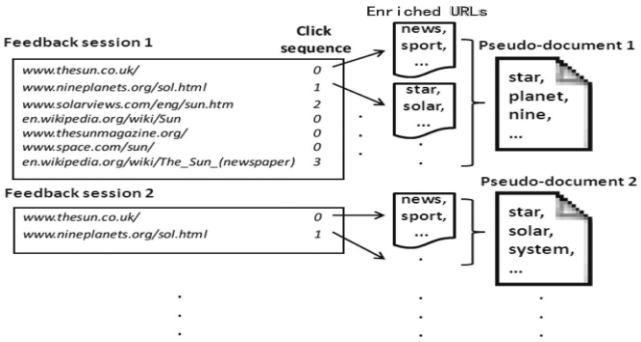

Fig 2. Structure of Feedback session

IV. DESIGN ARCHITECTURE

Figure 3 shows the framework of the approach. The proposed method consists of the feedback session of the image. This proposed method performs the weighting features based on the coefficient of the feedback session. The main difference between the proposed and the existing method is the Feature weight Average detection and the similarity measurement staging.

The Feature Weighted Average Detection Algorithm is to store the weights of each feature of the image. The similarity measurement functions to cluster the similar objects in the image. The Feature extraction here is composed of the two components i.e. Feature weighted Average Detection and the similarity measurement store factors. Feedback management is introduced with the feature weighted algorithm in order to restructure the search results within the short period of time. This feedback is to determine the accuracy in the results based on user point of view.

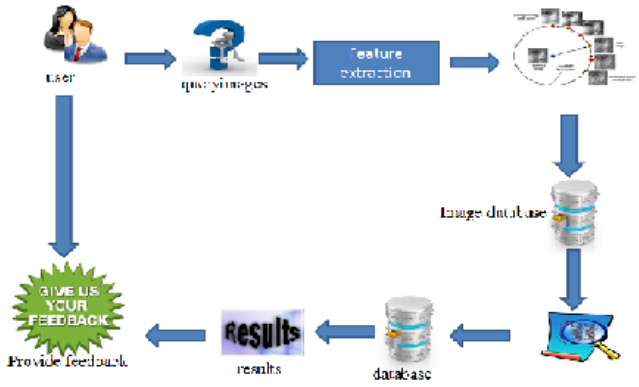

Fig 3 Design Architecture of the system

V. ALGORITHM

In this section the algorithms that are used to evaluate the performance of the system is being described.

A. Feature Weighted Average Detection Algorithm

The basic idea behind the Feature Weighted Average Detection Algorithm is taken from the image attributes. The coefficient of the attributes for each image using correlation technique is calculated. The formula to calculate the correlation of the attributes in the image is:

$$R(y,12\ldots k)=\sqrt{1-(1-ry1)^2\ldots\ldots\ldots\ldots..(1-ryk\ldots 2(k-1))}$$

Here ry1, ry2… ryk..(k-1) are the co-efficient of single correlation. From correlation, the weights of the image based on the features are calculated.

B. Similarity Measurement

The matching procedure is found by means of similarity in the query image and the images in the database. The similar attributes are clustered together and the features are analyzed based on the need of the user.

C. Evaluation of Weighted Algorithm

To evaluate proposed feature weighted method, the data in the database are clustered based on the similarity of the objects. So the improved k-means algorithm (IK-Means) is proposed. This algorithm uses the distance value to cluster the image attributes. The feature weighted technique along with the feedback provides efficient search results while search intentions are processed. The steps involved in IK-Means Algorithm are:

1. A new cluster is created by specifying threshold.

 All Rights Reserved © 2014 IJARCET

2. Create_ new_ cluster=true

3. Assign d (i) to c(j)

4. For i form 2 to n

A. For all clusters

(i) Calculate distance (d (i), c (j))

(ii) If (distance (d (i), c (j) <= threshold)

And

B. Create_new_cluster=false

C. If (Create_new_cluster=True) then create a cluster k= k+1, c (k) =d (i)

5. Run the new cluster

6. End

Here d (i) is the record in the database and c(j) is the centroid of the cluster. In order to test the performance of the weighted features based IK-Means Algorithm , experiment results of both IK-Means algorithm and weighted features based IK-Means algorithm, a dataset with 10 points was clustered in the experiment. Such datasets on clustering form different groups with similar object in the cluster. This is explained in the figure 3 and figure 4.

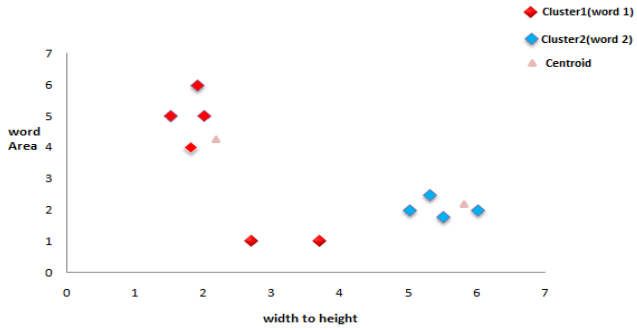

Fig 4. Clustering without Weighted Features

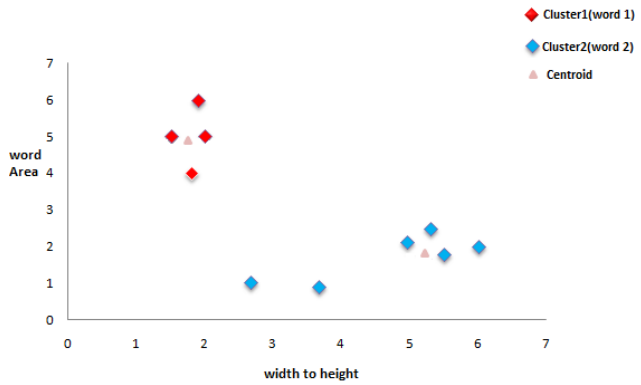

Fig 5 Clustering with Weighted Features

Figure 4 shows the clustering structure of the images based on the similarity of the objects in the images without the weighted average of the cluster. Figure 5 describes the clustering structure of the objects with weighted features of the image.

VI. RESTRUCTURING SEARCH RESULTS

Restructuring the search results is an application of inferring user search goals. The technique to restructure the search results to infer the user goals are introduced. Then the evaluation based on restructuring search results is also described.

The inferred user search goals are represented by the vectors and the feature representation of each image in search results can be computed. In this paper, categorization by choosing the smallest distance between the images of the vector and the user search goal vectors. By this way, the search results can be restructured according to the inferred user search goals.

VII. CONCLUSION

In this paper, a novel approach has been proposed to infer user search goals for a query by clustering its feedback sessions. Feedback session is analyzed to infer user search goals rather than the search results. Both the clicked and unclicked URL and images are considered to construct the feedback sessions. Therefore the feedback sessions can reflect user information needs more efficiently.

Then the images are given as the query to find the matching attributes of the image to restructure the search results. A new criterion called Feature Weighted Average Detection Algorithm is used to evaluate the performance of

 All Rights Reserved © 2014 IJARCET

the user search goals to increase the efficiency in search.

The complexity of the approach is low as it provides the optimal result. This approach depends on feedback to determine the run time. Hence this approach can be used both in online and offline searches. Thus, the users can get the needed goal more efficiently.

REFERENCES

[1] Baez- Yates, C.Hurtado, and Mendoza, "Query Recommendation Using Logs in Search Engines , *PROC International Conference on Trends in Database Technology* , pp 588-596, 2004.

[2] D. Beeferman and A. Berger ,"Agglomerative Clustering of a Search Engine Query Log", *ACM International Conference Knowledge and Data Mining*, pp 407 -416, 2000.

[3] T. Joachim," Evaluating Retrieval performance Using Click through Data", Text Mining J. Franke, G.Nakhaeizadeh, and I.Renz eds pp 79-96, physical Springer Verlag, 2003.

[4] T .Joachims, L.Granka, B.Pang, H.Hembrooke, and G.Gay, "Accurately Interpreting Click through Data as Implicit Feedback ", *Proc International ACM SIGIR Conference Research and Development in Information Retrieval,* pp 154-161, 2005.

[5] Yong Rui,Thomas S.Huang "Image Retrieval: Current Techinques,Promising Directions and Open Issues", *Journal of Visual Communication and Image Representation*,1999

[6] Zheng Lu, Hongyuan Zha, Xiaokang Yang, Weiyao Lin "A new Algorithm for Inferring User Search Goals with Feedback Sessions*" IEEE Transaction on Knowledge and Data Engineering, vol 25, No march 2013.*

[7] Zhang, J R Wen, J Nie "Evaluating Retrieval Performance using Click-Through Data" *Proc. Tenth International Conference, World Wide Web* pp162-168,2001.

[8] Zhai, X.Wang "Learn From Web Search Logs to Organize Search Results", Proc. 30th *Ann. International ACM SIGIR Conference on Research and Development* pp 131-138, 2006.

[1]Mount Steffi Varish.C received degree B.Tech Software Engineering from Regional Centre of Anna University Tirunelveli , in 2012. Now pursuing M.E Computer Science and Engineering in Meenakshi College of Engineering, Anna university, Chennai. Ph- 9944396169.

[2]Guru Rama Senthilvel received B.Sc., (Physics) from Madurai Kamaraj University in 1992, MCA., from Madurai Kamaraj University Campus in1996 and M.E (CSE) from Anna University, 2007. Currently,he is the Head of the department in Information technology.Ph-9841065075.

 All Rights Reserved © 2014 IJARCET