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 6, June 2016

A Novel Approach for Clustering Images Based On User Search Query Logs

Prof. Suchita Walke [1], Mohini Sarode[2]

HOD, Dept. of Information Technology, YTCEM College, Mumbai, India[1].

M. E Student, Dept. of Computer Engineering, YTGOIFE College, Mumbai, India[2]

ABSTRACT: Existing web image search engines such as Google and AltaVista return a large quantity of search results, ranked by their relevance to the given query. Web users have to go through the list and look for the desired ones. This is a time consuming task since the returned results always contain multiple topics and these topics are mixed together. Things become even worse when one topic is overwhelming but it is not what the user desires. We consider the problem of clustering Web image search results. Generally, the image search results returned by an image search engine contain multiple topics. Organizing the results into different semantic clusters facilitates users' browsing. Here we propose a novel approach to infer user search goals in image search by mining search engine query logs with semi-supervised spectral clustering. We combine information of the clicked images and then cluster the images with k-mean clustering to capture user image-search goals. Proposed method is very effective and applicable in improving the user image search goal

KEYWORDS: Click through logs, Image search goals, Sobel operator, Click session information, Density based denoising.

I. INTRODUCTION

In Web Search applications, users submit queries (i.e.some keywords) to search engines to represent their search goals. However, in many cases, queries may not exactly represent what they want since the keywords may be polysemous or cover a broad topic and users tend to formulate short queries rather than to take the trouble of constructing long and carefully stated ones.[1] Besides, even for the same query, users may have different search goals. However, sometimes queries may not exactly represent users' specific information needs since many ambiguous queries may cover a broad topic and different users may want to get information on 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 locate the homepage of a United Kingdom newspaper, while some others want to learn the natural knowledge of the sun, Therefore, it is necessary and potential to capture different user search goals in information retrieval.

We define user search goals as the information on different aspects of a query that user groups want to obtain. Information need is a user's particular desire to obtain information to satisfy his/her need. User search goals can be considered as the clusters of information needs for a query.[2] The inference and analysis of user search goals can have a lot of advantages in improving search engine relevance and user experience. In web search applications, users submit queries (i.e., some keywords) to search engines to represent their search goals.

However, in many cases, queries may not exactly represent what they want since the keywords may be polysemous or cover a broad topic and users tend to formulate short queries rather than to take the trouble of constructing long and carefully stated ones. Users have different search goals for the same query due to the following three reasons [1].

• Multi-concepts: a keyword may represent different things. For example, besides being a kind of fruit, "apple" is endowed with new concepts by Apple, Inc.

• Multi-forms: the same thing may have different forms. Take "Bumblebee" in the film Transformers as an example. It has two modes: car mode and humanoid mode. These two modes are the two formsof "Bumblebee."

• Multi-representations: in image search, the same thing can be represented from different angles of view such as the query leaf. It can be represented in a real scene or by a close-up.

Inferring user search goals is very important in improving search-engine relevance and user experience. Normally, the captured user image-search goals can be utilized in many applications. For example, we can take user image-search goals as visual query suggestions to help users reformulate their queries during image search. Besides, we can also

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 6, June 2016

categorize search results for image search according to the inferred user image-search goals to make it easier for users to browse.

II. RELATED WORK

By some specific predefined aspects, such as product intent and job intent. author focus on tagging queries with more hierarchical predefined concepts to improve feature representation of queries. These applications belong to query classification.User search goals and the number of them should be arbitrary and not predefined.The click session information is not fully utilized. Author [2] used Inferring user goals for text search, few methods were proposed in image search try tocapture user goals to give visual suggestions for a query in image search. They first selectsome tag words as textual suggestions by satisfying two properties: relatedness and informativeness. Then, they collect the images associated with a suggested keyword andcluster these images to select representative images for the keyword. In these cases, theperformance of may be decreased by using external texts as the external texts are not asreliable as tags.In [3] the research on diversity in retrieval is relevant to user goal inference. It aims to diversifythe results retrieved for an ambiguous query, with the hope that at least one of theinterpretations of the query intent will satisfy the user intent will satisfy the user.

.

III. PROPOSED ALGORITHM

The existing strategies for image search reranking suffer from the undependableness of the assumptions beneath that the initial text-based image search result's utilized within the reranking process. To improve the preciseness of the text-based image search ranking, visual re ranking has been planned to refine the search result from the text-based image program by incorporating the knowledge sent by the visual sense. Moreover, apart from the image search situation, visual re ranking may be used to improve the standard of the collected information within the method of mechanically constructing coaching information from the online for object recognition.

The analysis of user search goals for a question will be terribly helpful in up computer programme connection and user experience. Although the analysis on inferring user goals or intents for text search has received abundant attention, little has been projected for image search. We propose to leverage click session info, that indicates high correlations among the clicked pictures in a very session in user click-through logs, and mix it with the clicked images' visual info for inferring user image-search goals. The click session info will function past users' implicit guidance for cluster the photo graphs; a lot of precise user search goals may be obtained.

The objective of this work is to retrieve an oversized variety of pictures for a such that object category from the browser. A multimodal approach using text, metadata, and visual options is employed to assemble several high-quality pictures from the online. Candidate images are obtained by a text-based Web search querying on the object identifier (e.g., the word apple). The task is then to remove irrelevant images and cluster the remainder. [9]The Sobel operator, sometimes called the Sobel–Feldman operator or Sobelfilter, is used in image processing and computer vision, particularly within edge detection algorithms where it creates an image emphasising edges

IV. PSEUDO CODE

The aim of K-means[7] (or clustering) is if we want to group the items into k clusters such that all items in same cluster are as similar to each other as possible. And items not in same cluster are as different as possible. We use the distance measures to calculate similarity and dissimilarity. One of the important concept in K-means is that of centroid. Each cluster has a centroid. You can consider it as the point that is most representative of the cluster. Equivalently, centroid is point that is the "center" of a cluster.

Algorithm:

Step 1: Randomly choose k items and make them as initial centroids.
Step 2: For each point, the nearest centroid and assign the point to the cluster associated with the nearest centroiod.
Step 3: Update the centroid of each cluster based on the items in that cluster. Typically, the new centroid will be the average of all points in the cluster
Step 4: Repeats steps 2 and 3, till no point switches clusters.

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 6, June 2016

Pseudo-codes for Sobel edge detection method

Input: A Sample Image
Output: Detected Edges
Step 1: Accept the input image
Step 2: Apply maskGx,Gy to the input image
Step 3: Apply Sobel edge detection algorithm and the gradient
Step 4: Masks manipulation of Gx Gy separately on the input image
Step 5: Results combined to find the absolute magnitude of the gradient

$$|G| = \sqrt{Gx^2 + Gy^2}$$

Step 6: the absolute magnitude is the output edges

V. SIMULATION RESULTS

User register over the system and then only user can use the system for image search System image search will work only user login to system. As user get login to system user session get start and user can retrieve images according to his requirement. For first if user is using system for first time he will get related search result as well as irrelevent search result with ituser search query logs implemented according to user search goals is display as shown in fig.1. Here we get images related to search goals we search for apple as a mobile category it show result with apple mobile. K means algorithm is used for clustering.

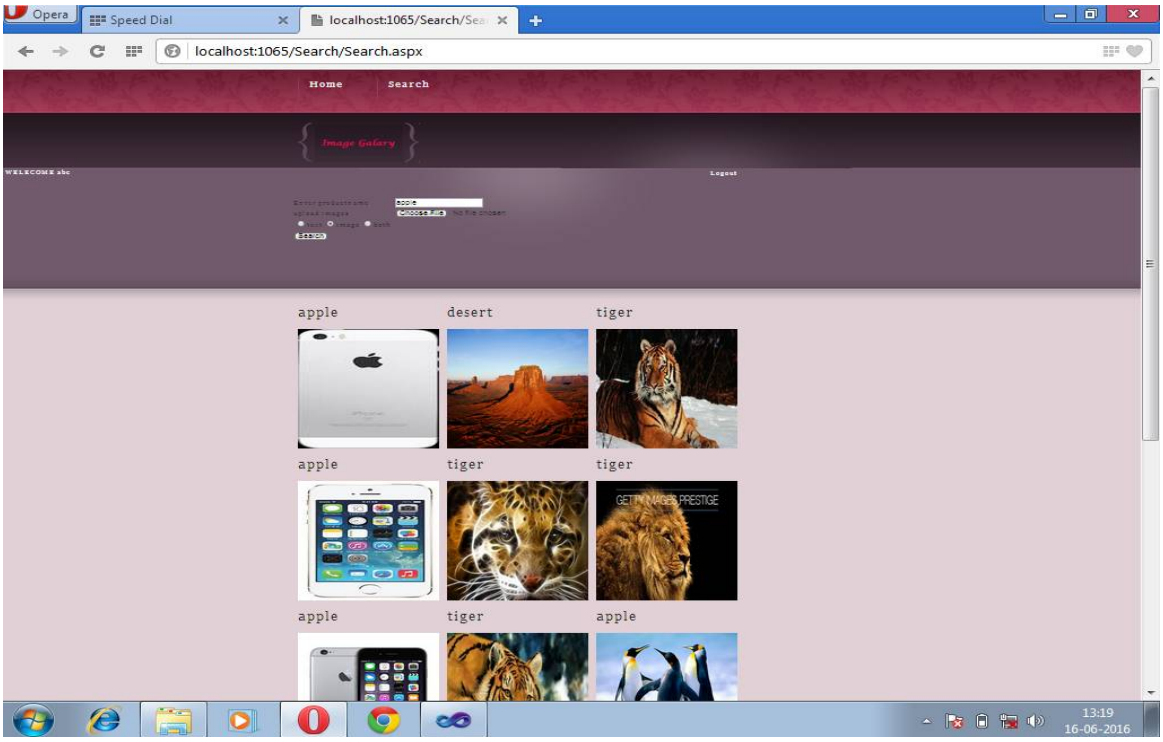

Fig.1.User search query log result

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 6, June 2016

Similarly canny edge detector is used for smoothing the images. To show edge detective image first we have to upload image after login process as shown in fig.2.

Fig.2.Original image

Sobel operator is used for edge detection. When we upload image at that time that image is saved as shown in fig.3.

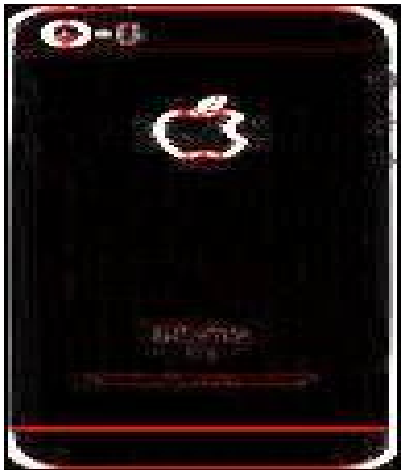

Fig.3.edge detective image

VI. CONCLUSION AND FUTURE WORK

We proposed to leverage click session information to infer user image-search goals. Click session information can serve as the implicit guidance of the past users to help clustering. Based on this framework, we implemented strategies such as click session information. Further-more, a click classification incoherence based approach was also proposed to automatically select the optimal search goal numbers. Proposed algorithm k means and sobel edge detection methods efficiency is high.In development it give us the effective results to infer user image search goal.The system should be robust and performance of the proposed method is better compared with existing methods

REFERENCES

1. Zheng Lu, Xiaokang Yang, Senior MemberIEEE, Weiyao Lin, Hongyuan Zha, and Xiaolin Chen "Inferring User Image-Search Goals Under the Implicit Guidance of Users", IEEE Transaction on Circuits and SystemforvideoTechnology, VOL. 24, NO. 3, MARCH 2014.

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 6, June 2016

2. 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," IEEE Transactions on Knowledge and Data Engineering, Vol.25, No. 3,March 2013.
3. Mrs Monika Jain, Dr. S.K.Singh "A Survey On: Content Based Image Retrieval Systems Using Clustering Techniques for large Data sets",International Journal of Managing Information Technology (IJMIT) Vol.3, No.4, November 2011
4. Deng CaiXiaofei He Zhiwei Li Wei-Ying Ma and Ji-Rong Wen "Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information", Microsoft Research ,Asia
5. Barbara Poblete Benjamin Bustos Marcelo Mendoza Juan Manuel Barrios"Visual-Semantic Graphs:Using Queries to Reduce the Semantic Gap on the Importanceof the work or suggest applications and extensions"CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.
6. Grishma Y. Bobhate,Prof. Usha A. Jogalekar,"An Efficient Algorithm to Reduce the Semantic Gap between Image Contents and Tags" International Journal of Computer Applications (0975 – 8887) Volume 72– No.11, May 2013
7. Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE, NathanS.Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and AngelaY.Wu,.Senior Member, IEEE "An Efficient K-Means Clustering Algorithm: Analysis an Implementation"IEEE Transactions onpattern analysisand machine intelligence, Vol. 24, NO. 7, JULY 2002
8. Mr. Manoj K.Vairalkar,Prof. S.U.Nimbhorkar "Edge Detection of Images Using Sobel Operator"International Journal of Emerging Technology and Advanced EngineeringISSN 2250-2459, Volume 2, Issue 1, January 2012.

BIOGRAPHY

Mohini Vilas Sarodereceived bachelor's degree from Mumbai University in 2012. Now a P.G student in YTGOIFE College,Mumbai University. Research interest in Data mining, image processing.