International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Image Mining Technique for Identifying User Image Search Goal

Ms. Chaitali Jadhav[1], Mrs. Jyoti Nandimath[2]
[1]*Dept, of Computer Engineering, STES's SKNCOE,Pune,chaijadhav77@gmail.com*
[2]*Dept. of Computer Engineering, STES's SKNCOE,Pune,jyotign@gmail.com*

Abstract—As we know mining basically means searching useful knowledge from collection of data, but image mining is not same as data mining. Image mining is used to mine knowledge directly from image.Image mining is extension of data mining in the field of image processing. Mining is done by considering collection of images and its related data. It include image retrieval, image processing, data mining .The main aim of this paper is to use this mining in case of finding user image search goal when user enter any ambiguous query or which query cover broad topic to search engine.
Keywords-Support Vector Machine (SVM), Naive Bays Algorithm, Input Query.

I. INTRODUCTION

The working of web search application starts when user submit any query to search engine its nothing but what user want to search. Although it represent goal, sometime query may not represent what user actually want because query may be ambiguous or may cover broad topic and user may want information from different aspect. It's may be due to query may have different representation, different forms and different concepts. For Example, consider when user submit a query "Titan" to search engine ,in this case this query cover broad topic and have different forms, some user may want titan watch models where as some user try to find attack on titan or some may want titan robot, as shown in Fig. 1.Therefor its necessary to capture user search goal in image retrieval process. User goal is nothing but what user wants to mine from search engine. This mining will satisfy what user need. User goal is thus used for clustering purpose. Identifying user goal is effective for improving search engine relevance and user experience. It will help in following cases, first when we cluster the results according to search goal ,its possible to restructure search result, second when we cluster the result considering query keywords, and third is Re-ranking can be easy with distribution of goals. Considering this usefulness more research done on it. They can be classified into query classification, search result reorganization, session boundary detection. In case of query classification query may be Navigational or Informational consider by Lee et. al.[1].In second case Wand and Zhai [2] analyze user clicked URL directly from user click through logs. In the third class, people in the search of detecting session boundaries. Jones and Klinkner [2] predict goal and mission boundaries to hierarchically segment query logs. 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 this paper we proposed techniques for finding exact image search goal user submitted query to search engine using different algorithmic combination for clustering and classification stages. At first we are combining image visual information and click through logs by considering feedback session information in order to remove irrelevant images. Then in next stage we are using novel approach of K-

mean algorithm for clustering .At the end stage we utilize combination of SVM and Naïve Bays algorithm for classification.

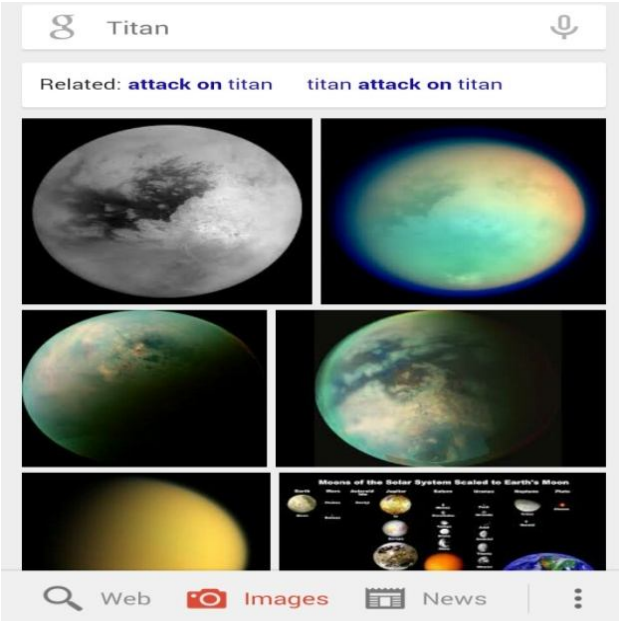

Fig.1.Different user search goal

To sum up, our work has contributions as following:

1. 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.
2. Also we propose to use SVM and Naive Bays for classification, as we know SVM is best classifier which work for unsupervised data also ,so it will give more accurate results.

II. MOTIVATION

Image retrieval is a key issue of user concern. Normal way of image retrieval is the text based image retrieval technique (TBIR). TBIR needs rich semantic textual description of web images .This technique is popular but needs very specific description of the query which is tedious and not always possible. Therefore generally the process of image search includes searching of image based on keyword typed. The process that occurs in the background is not so simple though. When query is entered in the search box for searching the image, it is forwarded to the server that is connected to the internet. The server gets the URLs of the images based on the tagging of the textual word from the internet and sends them back to the client. The search engine thus navigates through the pages and collects the images. It gives
the client the top ranked image which is the one with maximum number of hits from the user and a set of Images

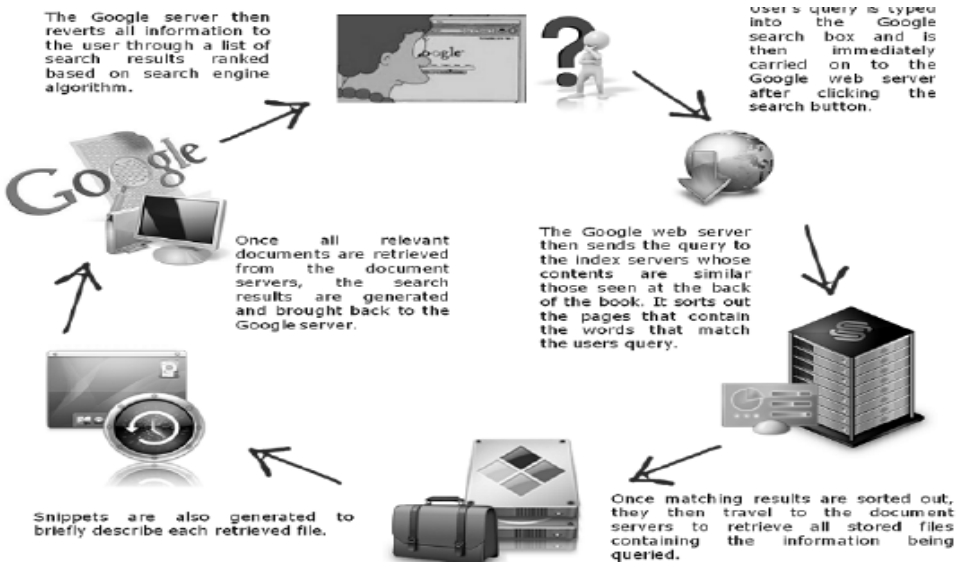

Fig.2.Google search engine [7]

. This is the technique of text based image retrieval system. But it has certain drawbacks like images obtained are many a time duplicated, of low precision, and irrelevant. This scenario may occur due to sparse and noisy textualquery. Due to this aspect user cannot be always sure of perfect images being obtained in available time. Many a times user has to surf many pages of images available to land at the perfect one. This possesses a great threat to the fast technology. Such problems surface when user needs large data to these factors of complexity, "image harvesting and retrieval" is a topic which is gaining popularity in research sector. What can be done in this respect is as follows:

1. Re-rank the images obtained on client side and provide with top rank image.
2. Use highly efficient clustering algorithm to facilitate grouping of similar images and select perfect among them.
3. Use contents of image rather than URL tagging to retrieve images from internet database.
4. Various concepts in combination to get an excellent image retrieval system.

III. STATE OF ART

Ever such researched carried on image mining. This section represents various image mining techniques. In recent years, the research on inferring user goals or intents for text search has received much attention. Many early researches define user intents as “Navigational” and “Informational” [3]. Or by some specific predefined aspects such as “Product intent” and “Job intent”. Some works focus on tagging queries with more hierarchical predefined concepts to improve feature representation of queries. However, in fact, these applications belong to query classification. User search goals and the number of them should be arbitrary and not predefined. Some works analyze the clicked documents (i.e, click content information) for a query in user click through logs to explore user goals [13]. There are number of approaches given in literature for image classification, re-ranking and image retrievals.

We concerning references and listed out some of the techniques given by different authors and inverters.

3.1 Picture Collage:
In this paper, they address a novel problem of automatically creating a picture collage from a group of images. Picture collage is a kind of visual image summary to arrange all input images on a given canvas, allowing overlay, to maximize visible visual information. We formulate the picture collage creation problem in a Bayesian framework.

3.2 Inferring Users' Image-Search Goals with Pseudo-images:
Zheng Lu1, Xiaokang Yang1, Weiyao Lin1, Xiaolin Chen1, Hongyuan Zha:.
In this paper, they propose a novel approach to capture user search goals in image search by exploring pseudo-images which are extracted by mining single sessions in user click-through logs to reflect user information needs. Moreover, they also propose a novel evaluation criterion to determine the number of user search goals for a query. Experimental results demonstrate the effectiveness of the proposed method.

3.3 Personalized Image Search through Tag-based User Profile on Social Websites :
In this paper they exploit the Social annotations and Novel Framework for considering the user query relevance and user specific-topic to learn personalized image search. The proposed framework contains two techniques:

1. Utility and Prediction model for social annotations.
2. We introduce a Hit Matrix technique for user query relevance and preference into the specific topic space. Performance evaluation shows that our proposed method outperforms the existing method and also shows that the developed model demonstrates the effectives of the Personalized Image Search

3.4 Towards a comprehensive survey of the semantic gap in visual image retrieval:
Peter Enserand **Christine Sandom** This paper adopts the premise that the 'semantic gap' is an incompletely surveyed feature in the landscape of visual image retrieval, and proposes a framework within which this deficiency might be made good. Simple classifications of types of image and of types of user are proposed. Consideration is then given in outline to how semantic content is realized by each class of user within each class of image. The argument is advanced that this realization finds expression in perceptual, generic interpretive and specific interpretive content.

3.5 The effect of specialized multimedia collections on web searching:
Bernard J. Jansen In this paper, they report the results of a research study evaluating the effect of separate multimedia Web collections on individual searching behavior. The Alta Vista search engine has an extensive multimedia collection and uses tabs to search specific collections. The motivating questions for this research are: (1) What are the characteristics of multimedia searching on Alta Vista and (2) What are the effects on Web searching of separate multimedia collections. The results of our research show that multimedia searching is complex relative to general Web searching.

3.6 Searching for multimedia: analysis of audio, video and image Web queries

Bernard J. Jansen a, Abby Goodrum b and Amanda Spink c In this case, the problem seems to be in representing audio and image information needs with textual queries, or with representing retrieved multimedia documents as short textual abstracts. In order to express a non-textual information need in only textual terms, the user takes on an additional cognitive load. In order to make relevance judgments, the user must visually inspect the full record in order to know if the retrieved document contains the requested multimedia information.

3.7 Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information :

Deng Cai1 Xiaofei He2 Zhiwei Li Wei-Ying Ma and Ji-Rong Wen

In this paper, they described a method to organize WWW image search results. Based on the web context, they proposed three representations for web image, i.e. representation based on visual feature, representation based on textual feature and representation induced from image link graph. Spectral techniques were applied to cluster the search results into different semantic categories. For each category, several images were selected as representative images according to their Image Ranks, which enables the user to quick understanding the main topics of the search results.

IV. PROPOSED WORK

Inferring User Image-Search is mainly used to produce the pure images from the search engine. The proposed system has mainly four modules:

Single word query, Download related images, Classification and Re-ranking.

And the block diagram of our proposed method was shown below at Fig 3.In first module user is submitting query to search engine which is then navigated through no. of pages and URL same as Google.

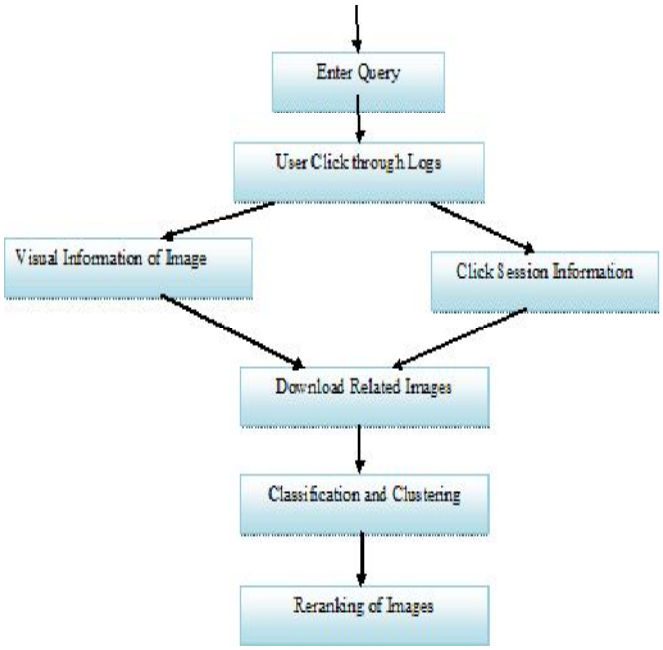

Fig.3.System Flow

From this navigation it will collect visual information of image and combine it with click session information, which consider feedback session information. By combining this two approaches we can eliminate the noisy images n download related images which containing same URL as textual part of submitted query. Next stage is classification and clustering , Image classification and image clustering are the supervised and unsupervised classification of images into groups respectively. In supervised classification, one is provided with a collection of labeled (pre-classified) images, and the problem is to label newly encountered, unlabeled images. Typically, the given labeled (training) images are used to do the machine learning of the class description which in turn are used to label a new image.
In this stage we are using learning algorithms like SVM and Naive Bays algorithm. Last stage is Reranking the target is to re-rank the retrieved pictures. So every feature is treated as binary:"True" if it contains the question word (e.g., tablet) and "False" otherwise. To re-rank pictures for one explicit category (e.g., dog), we tend to don't use the complete pictures for that category. Therefore, we train the classifier using all available comments except the class we want to re-rank. Finally we get the pure image as output.

CONCLUSION

In this paper we have highlighted how image mining techniques are used to identify user image search goal. Here we are using novel approach for classification and clustering which will give more accurate results. Complexity of this is low so it can be utilize in reality. Here unsupervised classification also possible so user can find his/her goal conveniently. Furthermore, a click-classification incoherence based approach is also proposed to automatically select the optimal search goal numbers. Experimental results demonstrate that our method can infer user image-search goals precisely. It is worth noting that the proposed method in this paper focuses on analyzing a particular query appearing in the query logs. Inferring user image-search goals for those popular queries can be very useful.

REFERENCES

[1]X. Wang and C.-X Zhai, "Learn from Web Search Logs toOrganize Search Results," Proc. 30th Ann. Int'l ACM SIGIR Conf.Research and Development in Information Retrieval (SIGIR '07),pp. 87-94, 2007.
[2][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.
[3]Winston H.Hsu,LyndonS.Kennedy,Shih-Fu Chang."Reranking Methods for Visual Search", *2007 IEEE Computer Society Conference.*
[4]B. Jansen, A. Goodrum, and A. Spink, "Searching for multimedia: analysis of audio, video and image web queries," *World Wide Web*, vol. 3, no. 4, pp. 249–254, 2000.
[5] B. Jansen, A. Spink, and J. Pedersen, "The effect of specialized multimedia collections on web searching," *Journal of Web Engineering*, vol. 3, no. 3-4, pp. 182–199, 2004.
[6] D. Rose and D. Levinson, "Understanding user goals in web search," in *Proceedings of the 13th international conference on World Wide Web*. ACM, 2004, pp. 13–19.
[7] D. Cai, X. He, Z. Li, W. Ma, and J. Wen, "Hierarchical clustering of www image search results using visual, textual and link information," in *Proceedings of the 12th annual ACM international conference on Multimedia*. ACM, 2004, pp. 952–959.
[8] J. Wang, L. Quan, J. Sun, X. Tang, and H. Shum, "Picture collage," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 347–354.
[9]Winston H.Hsu,LyndonS.Kennedy,Shih-Fu-Chang."Reranking Methods For Visual Search", *2007 IEEE Computer Society Conference.*

