ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Improving Image Search Using Click-Through Logs

B.Madhuri
PG Scholar,
Dept of CSE,
MVSR Engineering College,
Nadergul, Hyderabad, R.R Dist.

P.Usha Gayathri
Assistant Professor,
Dept of CSE,
MVSR Engineering College,
Nadergul,Hyd,R.R Dist.

Abstract:

Image retrieval based on visual content and re-ranking as per user image query has become the popular and effective of image retrieval techniques. Similar user query and click-through log is important for the success of an image search engine. Using this as a base and leveraging click-logs; in this paper, we focus on clustering the images produced from text-based image search based on visual features. To improve the preciseness of text-based image search, visual re-ranking has been planned to refine the search results based on similarity measurement.

Key Words:

Queries, K-means Clustering, Click-logs, Re-ranking.

1.Introduction:

The image retrieval is the process of retrieving images with respect to user intention from the large amount of databases. The user first enters query, based on the keywords in the query the search is performed and from the pool of images resulting images are displayed to the user. Initial image search is performed using surrounding text information of the image. The surrounding text information includes filename, caption or description of the image. The performance of the search decreases because of ambiguous surrounding information [8].To improve the performance of search engine new methods are developed which includes Keyword Expansion, Query Difficulty Prediction, Co-click history, Active-Reranking, Prototype Based Image Search Re-ranking. The objective of this work is to retrieve an oversized variety of pictures for the 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. The images are obtained by a text-based web search querying on the object identifier (e.g., the word apple). We first search for an image by submitting a query. The search results are returned by the search engine using text-based approach i.e. it retrieves images based on text surrounding the image, captions, keywords etc. First, the URL's of the images are downloaded in the Database, based on these URL's; the images are downloaded in to the Dataset. The images present in the dataset are clustered using k-means clustering technique and clustered images get stored into user click-through log. The task is then to remove irrelevant images and re-rank the remaining images based on the similarity measure.

2. Related Work:

The ever-growing number of digital images on the Internet (such as in the online photo sharing Website, the online photo forum and so on), retrieving relevant images from a large collection of database images has become an important research topic. Over the past decades, many image retrieval systems have been developed, such as text-based image retrieval (TBIR), content-based image retrieval (CBIR). As like a traditional method the image search is performed using text keywords in the query. There is some work related with textual information where it uses text including filename of the image, URL of the image and description or the caption of the image as a surrounding text parameter. If the text is found in these surrounding parameters of the image then that image is displayed as a result image. The textual search is used as an input for visual similarity based search by grouping the images which is having same textual tag. But some difficulty with textual search is that if the external text is ambiguous or not related with the image then this becomes a limitation for textual search [1].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Most search engine works on Text Based Approaches but there exist alternative approach, content based image retrieval that require a user to submit a query image, and return images that are similar in content . The extracted visual information is natural and objective, but completely ignores the role of human knowledge in the interpretation process. As the result, a red flower may be regarded as the same as a rising sun, and a fish the same as an airplane etc. The methods for image search re-ranking can be classified into supervised and unsupervised ones, according to whether human labeled data has been used to derive the re-ranking model or not. The unsupervised re-ranking methods do not rely on human labeling of relevant images but require prior assumptions on how to employ the information contained in the underlying text- based result for re-ranking [2].

Extraction of visual content from images is split into two parts, namely, image processing and feature extraction. The features extracted are color, texture, and shape. When users search for pictures, their intent or clarity about what they desire may vary. The clarity of intent plays a key role in a user's expectation from a search system and the nature of her interaction. It can also act as a guideline for system design [3]. A user once find candidate image of target image the re-ranking function is used by choosing that candidate image as a query image. One of the approaches described by xiao gang and xiaoou Tang which has offline and online parts [4]. To refine an image search "logs" of search engine are used. These logs are used as a relevance feedback signals to refine image distance function. This approach is similar to soft margin SVM trained with relative comparison of the images [5].

When user enters query for image search the resulting images are displayed by extracting surrounding information of the image. But if user query is not sufficient to express the intention of the user because of lack of knowledge about giving specific word for image search then the search performance also decreases [6]. To make prediction correctly about the user's query image intention the sample selection strategy is used to decrease the clients labeling efforts. To display relevant images to the user "active re-ranking" is used. In this method the user intention is captured and used for re-ranking the images. To improve the performance of search, labeling information is collected from user and new method is proposed to actively select more informative query images through structural information [7].

3. System Architecture:

The below figure explains the step by step procedure of the work done:

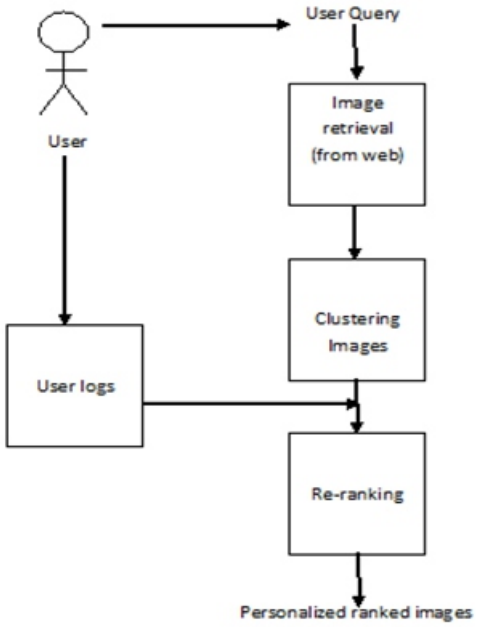

Figure: System Architecture of the project

The work done mainly in four steps those are 1. Image Retrieval using text 2. Clustering 3. User Click-through logs 4. Re-ranking.

1.Image Retrieval using text:

User gives a single word query to image search engine to search for related images using text based web search. The search results are returned by the search engine. First the URL's of the images are downloaded in the database, based on these URL's; the images are downloaded in to the dataset. By gathering all the URL's our application is retrieving all our related images.

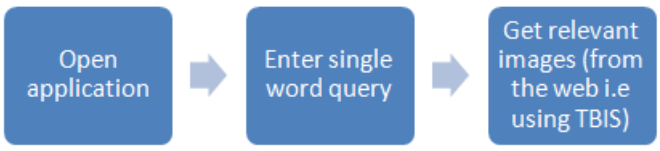

Figure 1: Single word query.

2.Clustering:

The K-means clustering algorithm is a simple method for estimating the mean (vectors) of a set of K clusters.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. Algorithmic steps for k-means clustering Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

Where, 'ci' represents the number of data points in ith cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3).

Figure 2: K-means Clustering

3.User Click-through Logs :

Click-through Log data is important for improving search performance as well as understanding user behaviors. Search Engines retrieve relevant information for a given query and return the results. Users will examine the list and click on the "perceived" relevant results, which generates click logs as implicit relevance feedback to search engines. The click-through information from the past users can provide good guidance about the semantic correlation among images.After clustering all the images present in the dataset, find the mean of all the k-clusters and display in user click-through log. Now user can select the desired images by clicking on the checkboxes.

Figure 3: User Click-through log

4.Image Re-ranking:

After user-clicks, it's time for ranking the images i.e.to find nearest neighbors based on similarity measure. The ranking is done based on Similarity measure which is the image distance i.e. Euclidean distance which compares the similarity of two images (query image and the dataset image) in various dimensions such as color, texture and shape.

ED = √∑[qi-pi]2 ,where qi and pi are feature vectors of query image and dataset image.

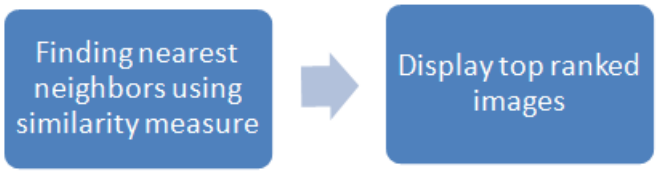

Figure 4: Re-ranking

4.Results:

This method was tested with several user search queries. The work done in this paper is uncomplicated to understand and it is implemented with the help of JSE1.6 and run on desktop PC with 3.09 GHz Intel and 512MB RAM. The output is more accurate and thus we can significantly improve the performance of inferring user search goals.

5.Conclusion:

To get the relevant images from the search engine and clustering the images for a query based on visual content to infer user image search goals. Image visual information is obtained with the help of visual image features, we calculate the image distance for finding the similarity of the images. Image Re-ranking is applied, where the top ranked images are retrieved which reaches the user image search goals. The retrieved images are relevant to the users query.

Experimental results demonstrate that the method can infer user image-search goals precisely. It is worth nothing that the work done 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.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

6.References:

[1] ZhengLu,XiaokangYang,Seniormember,IEEEWeiyao Lin, HongyuanZha and Xiaolin Chen "Inferring User Image-Search Goals under the Implicit Guidance of Users",2013.

[2] Sabitha M G1, R Hariharan2 "Hybrid Approach for Image Search re-ranking" Volume 2 Issue 5, May 2013.

[3] R.Datta, D.Joshi, J.Li, andJ.Z.Wang, ,,,,Image retrieval: Ideas, influences, and trends of the new age"""", ACM Comput. Surveys, vol. 40, no. 2, pp. 1–5, 2008.

[4] Wang, Member, IEEE , Shi Qiu, Ke Liu, and Xiaoou Tang, Fellow, IEEE, Web Image Re-Ranking, Using Query-Specific Semantic Signatures, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 36, No. 4, April 2014.

[5] Yogdong Zhang, Senior Member, IEEE, Xiaopeng Yang, and Tao Mei, Senior Member, IEEE, Image Search Re-ranking with Query-dependent Click-based Relevance Feedback, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1057-7149 (c) 2013 IEEE.

[6] Xiaoou Tang, Fellow, IEEE, Ke Liu, Jingyu Cui, Student Member, IEEE, Fang Wen, Member, IEEE, and xiaogang Wang, Member, IEEE, Intent Search: Capturing User Intention for One-Click Internet Image Search , IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 34, NO. 7, JULY 2012.

[7] Xinmei Tian, Dacheng Tao, Member, IEEE, Xian-Sheng Hua, Member, IEEE, and Xiuqing Wu, Active Re-ranking For Web Image Search, IEEE Transactions On Image Processing, Vol. 19, No. 3, March 2010.

[8] Linjun Yang, Member, IEEE, and Alan Hanjalic, Senior Member, IEEE, Prototype-Based Image Search Re-ranking , IEEE Transactions On Multimedia, Vol. 14, No. 3, June 2012 871.