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. 3, Issue 12, December 2015

Survey on Web Image Re-ranking Using Query Semantic Signature and Duplicate Detection

Rameshwar G. Wavhal, Prof. Basavraj Chunchure

Master of Engineering Student, Dept. of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Maharastra, India

Assistant Professor, Dept. of C.E., Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Maharastra, India

ABSTRACT: A query-specific visual similarity metric is learned from the selected examples and used to rank images. The requirement of more users' effort makes it unsuitable for web-scale commercial systems like Bing image search and Google image search, in which users' feedback has to be minimized. System addingl information to image search is important. However, the interaction has to be as simple as possible. The absolute minimum is One- Click. In this paper, we propose a novel Internet image search approach. It requires the user to give only one click on a query image and images from a pool retrieved by text-based search are re-ranked based on their visual and textual similarities to the query image. System users tolerate one-click interaction which has been used by many popular text-based search engines. For example, Google requires a user to select a suggested textual query expansion by one click to get additional results. The key problem to be solved in this paper is how to capture user intention from this one-click query image.

KEYWORDS: Keyword image,click on query image obtain,Re-ranking of query image.

I.INTRODUCTION

Many commercial Internet scale image search engines use only keywords as queries. Users type query keywords in the hope of finding a certain type of images. The search engine returns thousands of images ranked by the keywords extracted from the surrounding text. It is well known that text-based image search suffers from the ambiguity of query keywords. The keywords provided by users tend to be short. For example, the average query length of the top 1, 000 queries of Picture search is 1.368 words, and 97% of them contain only one or two words. They cannot describe the content of images accurately. The search results are noisy and consist of images with quite different semantic meanings. The top ranked images from Bing image search using "apple" as query. They belong to different categories, such as "green apple", "red apple", "apple logo", and "iphone", because of the ambiguity of the word "apple". The ambiguity issue occurs for several reasons. First, The query keywords' meanings may be richer than users' expectations. For example, the meanings of the word "apple" include apple fruit, apple computer, and apple ipod. Second, the user may not have enough knowledge on the textual description of target images. For example, if users do not know "gloomy bear" as the name of a cartoon character (shown in Figure 2(a)) and they have to input "bear" as query to search images of "gloom bear". Lastly and most importantly, in many cases it is hard for users to describe the visual content of target images using keywords accurately. In order to solve the ambiguity, additional information has to be used to capture users' search intention One way is text-based keyword expansion, making the textual description of the query more detailed. Existing linguistically-related methods find either synonyms or other linguistic-related words from thesaurus, or find words frequently co occurring with the query keywords. For example, Google image search provides the "Related Searches" feature to suggest likely keyword expansions. However, even with the same query keywords, the intention of users can be highly diverse and cannot be accurately captured by these expansions. As shown in Figure 2(b), "gloomy bear" is not among the keyword expansions suggested by Google "Related Searches". Another way is content-based image retrieval with relevance feedback. Users label multiple positive and negative image examples.

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. 3, Issue 12, December 2015

II.RELATED WORK

Generally a large-scale image search system consists of two key components—an effective image feature representation and an efficient search mechanism. It is well known that the quality of search results relies heavily on the representation power of image features. The latter, an efficient search mechanism, is critical since existing image featctivesures are mostly of high dimensions and current image databases are huge, on top of which exhaustively comparing a query with every database sample is computationally prohibitive. we represent images using the popular bag-of-visual- words (BoW) framework , where local invariant image descriptors (e.g., SIFT [3]) are extracted and quantized based on a set of visual words. The BoW features are then embedded into compact hash codes for efficient search. For this, we consider state-of-the-art techniques including semi-supervised hashing and semantic hashing with deep belief networks . Hashing is preferable over tree-based indexing structures (e.g.,kd-tree s it generally requires greatly reduced memory and also works better for high-dimensional samples. with the hash codes, image similarity can be efficiently measured. In hamming space by hamming distance, an integer value obtained by counting the number of bits at which the binary values are differentThe sheer amount of Web pages and the exponential growth of the Web suggest that users are becoming more and more dependent on the search engines' ranking methods to discover information relevant to their needs.

Typically, users expect to find such information in the top-ranked results, and more often than not they only look at the document snippets in the first few result pages and then they give up or reformulate the query. This can introduce a significant bias to their information finding process and calls for ranking methods that take into account not only the overall page quality and relevance to the query, but also the match with the users' real search intent when they formulate the query. It generally requires greatly reduced memory and also works better for high-dimensional samples. With the hash codes, image similarity can be efficiently measured (using logical XOR operations) in Hamming space by Hamming distance, an integer value obtained by counting the number of bits at which the binary values are different.

III.PROPOSED ALGORITHM

This paper uses a clustering method called K-means to classify dataset into k-clusters. Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. one of the clustering methods called K-means looking for keywords in the user profile to help in specifying the intending meaning. Because the target meaning is "computer program language", we look for slave words in the user profile that best fit this specific meaning words such as "computer", "program","awt", "application", and "swing". Content-Based Image Retrieval (CBIR) refers to image retrieval system that is based on visual properties of image objects rather than textual annotation. Contents of an image can be of various forms like, texture, color and shape etc. In this work, shape is selected as a primary feature in indexing the image database. CBIR is more robust and makes it easier for image retrieval.

The similarity used for search criteria could be Meta tags, color distribution in images, region/shape attributes, etc. Unfortunately, image retrieval systems have not kept pace with the collections they are searching. The shortcomings of these systems are due both to the image representations they use and to their methods of accessing those representations to find images. The problems of image retrieval are becoming widely recognized, and the search for solutions an increasingly.

with large scale storing of images the need to have an efficient method of image searching and retrieval has increased. It can simplify many tasks in many application areas such as biomedicine, forensics, artificial intelligence, military, education, web image searching. Most of the image retrieval systems present today are text-based, in which images are manually annotated by text-based keywords and when we query by a keyword, instead of looking into the contents of the image, this system matches the query to the keywords present in the database.This technique has its some disadvantages:Firstly, considering the huge collection of images present, it is not feasible to manually annotate them. Secondly, the rich features present in an image cannot be described by keywords completely. This system distinguishes the different regions present in an image based on their similarity in color, pattern, texture, shape, etc. and decides the similarity between two images by reckoning the closeness of these different regions The CBIR approach is much closer to how we humans distinguish images. Thus, we overcome the difficulties present in text-based image retrieval because low-

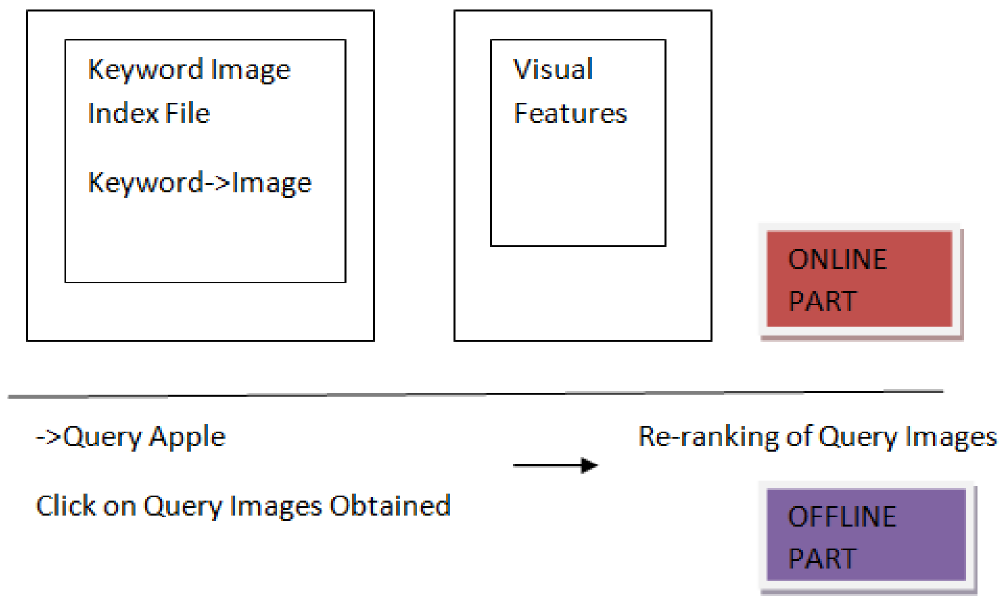
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

level image features can be automatically extracted from the images by using CBIR and to some extent they describe the image in a more detail compared to the text-based .



IV.ACKNOWLEDGEMENT

We would like to take this opportunity to express my sincere gratitude to my Project Guide Prof.Basavraj Chunchure (Assistant Professor, Computer Engineering Department,SPCOE) for his encouragement, guidance, and insight throughout the research and in the preparation of this dissertation. He truly exemplifies the merit of technical excellence and academic wisdom.

V.CONCLUSION AND FUTURE WORK

Web image Reranking is the pattern in which it gets highest images to the user in that all images can be recovered and got the maximum output We propose a novel image re-ranking framework, which learns query-specific semantic spaces to significantly improve the effectiveness and efficiency of online image reranking. The visual features of images are projected into their related visual semantic spaces automatically learned. a novel framework for query-adaptive image search with hash codes. By harnessing a large set of predefined semantic concept classes, our approach is able to predict query-adaptive bitwise weights of hash codes in real-time, with which search results can be rapidly ranked by weighted Hamming distance at finer-grained hash code level. This capability largely alleviates the effect of a coarse ranking problem that is common in hashing-based image search. A novel framework for query-adaptive image search with hash codes. By harnessing a large set of predefined semantic concept classes, our approach is able to predict query-adaptive bitwise weights of hash codes in real-time, with which search results can be rapidly ranked by weighted Hamming distance at finer-grained hash code level. This capability largely alleviates the effect of a coarse ranking problem that is common in hashing-based image search. The extracted semantic signatures can be 70 times shorter than the original visual feature on average, while achieve 20%□35% relative improvement on re-ranking precisions over state-of -the- art methods.

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. 3, Issue 12, December 2015

REFERENCES

[1]Image Retrival Using Semantics Of Query Image Siddanagowda G R1 Santosh S2 Sandeep Kumar3 Raghu M T4, Internatinal Journal of Darshan Institute On Engineering areaerch And Emerging Technologies

[2]Re-ranking of Images Using Semantic Signatures with Duplicate Images Removal & K-means Clustering Sayli Baxi1,S.V.Dabhade International Journal of Engineering And Computer Science ISSN:2319-7242 volumee 3 ISSUE 5 may 2014

[3]Query-Adaptive Image Search With Hash Codes Yu-Gang Jiang,JunWang,Xiangyang Xue,IEEE Member and Shih-Fu Chang IEEE TRANSACTIONS ON MULTIMEDIA

[4]Improving Web Image Search Re-Ranking Using Hybrid Approach Kirti Yadav, Sudhir Singh , Intrernational Journal of Advanced Research in Computer Science and Software Engineering

[5]Visual Semantic Complex Network for Web Images Shi Qiu1 Xiaongang Wang2,3 and Xiaoou Tang1,3, ICCV 2013

[6]Efficient Re-Ranking of Images from the Web Using Bag Based Method S.Keerthana,R .C. Narayanan k.k. Krishnamoorthy International Journal of Computer Applications volume 42 No 21 March 2012