

Survey Paper on Improving User's Search Strategy for Retrieving Desired Images

Prof. S. R. Durugkar¹, Archana Waghchawre²

Assistant Professor¹, P.G. student²

Department of Computer Engineering

Late G.N.Sapkal College of Engineering, Anjaneri, Nasik^{1,2}
University Of Pune

ABSTRACT:

As day to day we are interacting with information retrieval application. These applications are nothing but the software which is responding with top 'n' documents from its database. As an end user what we are expecting? Our expectations are such that information retrieval application should return most relevant documents within less time and with accuracy.

Though we are saying that we can retrieve the documents from other database but there are major issues in implementing such strategy.

We have proposed a novel method for automatic annotation, indexing and annotation-based retrieval of images through this paper. System or the developed application will first of all works on the user's queries and tries to find out the relevance between the index structure prepared for the collection of images. Then system will return the desired image in response to user's query with higher precision and effectiveness.

In this paper we have also surveyed various aspects such as use of information retrieval models, markov chain scheme, and application which will handle all these things.

Keywords: Information Retrieval, information retrieval models, markov chain

I.INTRODUCTION

1.1 Information retrieval

It is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on metadata.



Fig 1.1 Information retrieval

Automated information retrieval systems are used to reduce what has been called "information overload". In day to day applications we are using IR systems to provide access to books, journals and other documents. Web search engines are the one of the IR applications.

Information:

It is nothing but the meaningful data which generates some potential at the receiver to perform certain action.

1.2 Abstract Model of Information Retrieval

Abstract Model of Information Retrieval

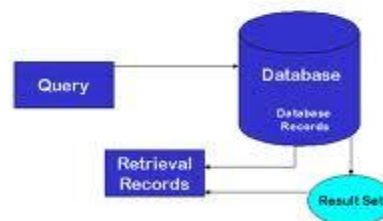


Fig 1.2 Abstract model of Information retrieval

Above diagram shows that in information retrieval end user will execute the query and wish to retrieve the relevant documents which will satisfy the user's search in less time with accuracy.

1.3 Challenges While Searching something

1. Huge Data
2. Heterogeneous Data
3. Redundant Data
4. Dangling Links
5. Unstructured Data

These many challenges we have to face as an end user.

II. IMAGE RETRIEVAL

An **image retrieval** application is a computer system for browsing, searching and retrieving images from a large database of images.

Most traditional and common methods of image retrieval utilize some method of focusing on metadata such as captioning, keywords and other aspects of that image which will describe those images so that retrieval can be performed over the annotation words.

There are types of images, **analog** and **digital**, used in variety of applications.

Analog images are the type of images that we, as humans, look at. They include such things as photographs, paintings, TV images, and all of our medical images recorded on film or displayed on various display devices, like computer monitors.

Digital images are recorded as many numbers. The image is divided into a matrix or array of small picture elements, or pixels which is represented by a numerical value. The advantage of digital images is that they can be processed, in many ways, by computer.

A digital image is a matrix of many small elements, or pixels. Each pixel is represented by a numerical value. In general, the pixel value is related to the brightness or color that we will see when the digital image is converted into an analog image for display and viewing. Generally, at the time of viewing, the actual relationship between a pixel numerical value and its displayed brightness is determined by the adjustments of the window control as discussed in other modules.

If our approach is manual then it will be time-consuming, laborious and expensive; to overcome this, there has been a large amount of research done on automatic image annotation.

Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.



Fig 2.1 retrieval Analyzer

2.1 Content Based Information Retrieval

Content-based image retrieval, a technique which uses visual contents to search images from large scale image databases

according to users' interests, has been an active and fast advancing research area

In this case of retrieval system we need to focus on the metadata of the given objects.

There fore it requires to have a prepared index like structure which will hold the metadata related with all the objects or images in case of our proposed system

In this way end user's query will be match against the index and in response to query only maximum relevant documents will be returned to user.

III.SURVEY ON MARKOV CHAINS

It is nothing but a mathematical system that goes through transitions from one state to another, among countable number of possible states. It is a arbitrary process usually characterized as memoryless:

The next state depends on the current state and not on the sequence of events that come first it. This specific kind of "memorylessness" is called the Markov property. Markov chains have many applications as statistical models of real-world processes.

The changes of state of the system or an application can be known as transitions, and then we need ti find out the probabilities associated with various state changes which are known as transition probabilities.

This method is characterized by a state space, a transition matrix describing the probabilities of particular transitions, and an initial state across the state space.

Following diagram depicts what we wish to do in our proposed system very specifically:

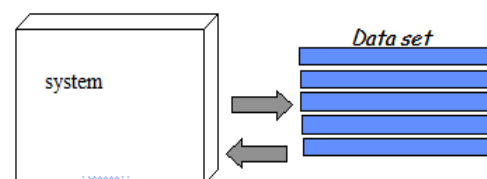


Fig. 3.1 Basic Model depicting Markov Chain

In this case we need to observe various observations such as $x_1, x_2, x_3, \dots, x_n$. we will store 'n' number of images or objects then our system will find out the most suitable or we can say the accurate 'image' in response to user's query.



Fig 3.2 Dependencies among 'n' objects

We will check whether any kind of dependencies exists in between them or not. This type of approach will helps in effective information retrieval because what we wish to do

is the in response of user’s query we will return the more relevant documents or images which will satisfy the search.

Day to day simple example for markov chain is gambler’s play; in this gambler’s pay there is chance to Gambler wins \$1 with probability p and chance of loosing \$1 with probability $1-p$

And finally Game ends when gambler goes broke, or gains a fortune of \$100

This can be depicted with the flow given below:

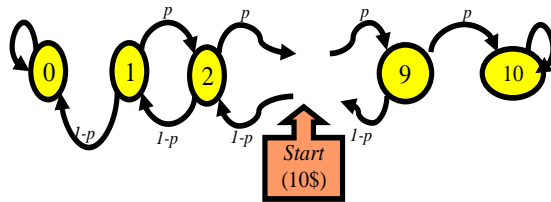


Fig 3.3 Gambling scenario

similar steps are applicable in our proposed work in which we will store ‘n’ images and try to find out the most relevant image in response to user’s query in less time.

IV. PROPOSED IDEA

Below diagram can describe the overall predicted architecture of our system.

So that one can easily find out the required modules for developing such type of systems. We introduce the Markovian Semantic Indexing scheme which is a new method for automatic annotation and annotation based image retrieval.

4.1 Indexing and retrieval in proposed system will be as follows:

It will work as follows, for all the images which are in the databases the feature vector which is one kind of representation for every image will be pre-computed and stored as index in one file.

When retrieval should be made, the image with the least distance between query image and image from database is returned to the end user this is our proposed work. Similar is depicted in the following figure.

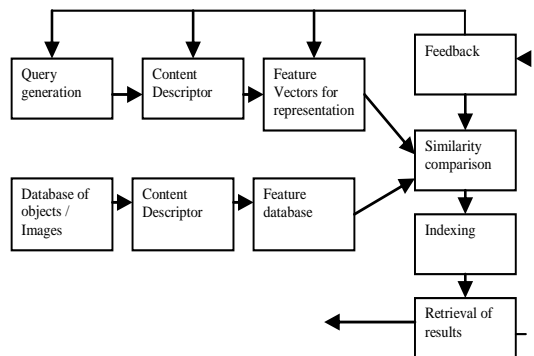


Fig 4.1 Proposed Architecture

The queries execute by the users of our system will be semantically refined, the keywords representing semantics when compared to text in documents or other vocabulary related presentations.

The objective is to progress towards the user fulfillment by returning images that have a higher probability to be accepted, meaning is that those images will be highly relevant to the user’s query.

What we are assuming is that the users will search for images by executing queries, each query being an ordered set of keywords. The system responds with a list of images. It will also possible for user to download the returned images and execution of a new query instead.

During the initial phase of the system the images will be considered with no annotation. As the users issue queries and pick images the system annotates the images in an automatic manner and at the same time establishes relevance relations between the keywords as will be explained later on in the manuscript.

4.2 Automatic image annotation

We need to consider **Automatic image annotation** which is also known as image tagging which is the method by which a computer system automatically assigns metadata – which will describe the particular image or object in more depth and also describes the additional characteristics in detail- in the form of keywords to a given image.

This process can be considered as a type of ‘n’-class image classification or distribution with a very large number of classes. Usually, analyzing the image in the form of extracted feature vectors which are describing the particular image and then assigning the annotation words which will be used by machine learning techniques to attempt to automatically apply annotations to new image as soon as new image will get added. One of the method which correlates between image features and training annotations, then methods used in our proposed system will try to find out the relation between documents and vocabulary.

4.3 Assumptions:

Effective image retrieval application requires locating relevant images as more and more images will be stored in the database. We describe an approach integrating text based and content based techniques, to take advantage of their complementing strengths.

The key idea of designing the image retrieval application is how to use the structure of documents to achieve effective text-based

image retrieval, how to implement the color based image retrieval technique, and also how to combine the retrieval results of these two techniques to obtain meaningful final results which will satisfy the end user.

In the color based image retrieval technique, each image in the database will be normally represented using basic aspects of the image such as color space chosen. Each color channel is quantized into some intervals. So the total number of discrete color combinations will be equal to m^3 . Basically, the system will attempt to model the user expectations by applying genetic techniques to the user's choices. The system will be based on following few points:

- ❖ An algorithm to extract the features from an image by reducing the dimensionality of the search problem.
- ❖ A user interface or GUI, allowing the user to view some images from the data base and rank them;

The interactive image retrieval application will work in the following style:

- ❖ In the preprocessing step, image coding will be performed for every image in the database,
- ❖ Representations or we can say a kind of index will be stored in a search table.
- ❖ By using clustering technique we can form a set of images.
- ❖ The user ranks the images according to their similarity to the image he is attempting to retrieve;

CONCLUSION

In this paper a retrieval study for image is proposed. The content of this paper are studied from variety of sources and also we have introduced our contribution in this system.

The main intention of this application is to develop an image retrieval application which can perform identity check of an image.

The objective is to progress towards the user fulfillment by returning images that have a higher probability to be accepted, meaning is that those images will be highly relevant to the user's query. Each feature defines a multidimensional space where images are points, and the similarity between images is computed as the distance between points. Addition to this what we are thinking is to store the sessions of end user meaning is that if user already searched for "flowers" then at the next time if he or any other user will search for the same query our system will prompt a message "Query already executed" then will display the retrieved the results within

very less time therefore significant time can be reduced.

Another thing we are thinking is without any barrier whether any barrier i.e. type of images features of those image should be retrieved and user's search should be satisfied.

We wish to develop a system that will be strong enough which will give accuracy and effectiveness though in database there will be maximum no. of images. Means without in spite of database size and any other barrier our system should give better results.

REFERENCES

1. Konstantinos A. Raftopoulos, Member, IEEE, Klimis S. Ntalianis, Dionyssios D. Sourlas, and Stefanos D. Kollias, "Mining User Queries with Markov Chains: Application to Online Image Retrieval" IEEE TRANSACTIONS & DATA ENGG, VOL. 25, NO. 2, FEBRUARY 2013
2. S. Santini and R. Jain, "Similarity Measures," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 9, pp. 871-883, Sept. 1999.
3. R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, 2008.
4. T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," Machine Learning, vol. 42, no. 1/2, pp. 177-196, 2001.
5. B.J. Jansen, A. Spink, and T. Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," Information Processing and Management, vol. 36, no. 2, pp. 207-227, 2000.
6. Reddy P.V.N.* and Satya Prasad K., Image retrieval system for natural images based on local features, Advances in Information Mining, ISSN: 0975-3265, Volume 2, Issue 2, 2010, pp-01-04
7. Victor Makarenkov, Michael Marcovich Noam Shemesh, "Content-Based Image Retrieval (CBIR)
8. F. Boschetti and S. B. Cho," First Steps Towards an Interactive Image Retrieval System for Geoscientific Images
9. Ydo Wexler & Dan Geiger, ' a tutorial on Markov Chains
- 10 Web Information Retrieval, Nick Craswell and David Hawkin
- 11 Jianfeng Gao, Jian-Yun Nie, Endong Xun*, Jian Zhang, Ming Zhou, Changning Huang,"Improving Query Translation for CLIR using Statistical Models"