Markovian Image Retrieval

Sonal Singh

Department of Information & Communication Technology,
Manipal Institute of Technology
Karnataka, India

Abstract. This research proposes a new methodology for the retrieval of images which is based on Markov chains and its ability to store the semantics of the present situation in its states. The new algorithm called Koogle, creates a Global Markov Chain for keyword relevance and for storing the user semantics, where each state can hold more than one keyword. Since this model takes into consideration the targeted user preferences, it proves to be a better approach than most of the present methods. The research shows how multiple keywords can be included in a single state and how the new proposed ranking algorithm produces relevant results in most cases and can be used to give appropriate results for its targeted users.

1 Introduction

Content based image retrieval(CBIR) systems look for pictures based on low level features like color, texture, structure, spatial layout which can then be procured and used in the indexing the pictures. CBIR suffers with two major comebacks: the sensory gap and the semantic gap. Lack of similarity between the object in the real world and the computational information recorded by systems is termed as Sensory gap. Lack of similarity between the data extracted from the visual subject and the exegesis of the data by the users concerned is called the Semantic gap.

Another type of retrieval method exists which is the Annotation- Based Image retrieval(ABIR). It is built on the postulates of text retrieval systems. As the name suggests, ABIR is concerned with the retrieval of on-line pictures based on the tags or keywords associated with a particular image. ABIR tries to incorporate semantic data into both text based queries and the image tags. Visual data may be used to describe a picture efficiently through keywords(annotations). An image could either be manually annotated or automatically annotated.

Markov chain is a type of mathematical system whose architecture can be described as something that has a number of states and transition paths between the states, where a set of values change from one state to another. The transition from one state to another always depends on the present state(or situation) and not on the previous paths taken. This memoryless characteristic called the markov property make markov chains useful in a lot of real world scenarios.

Markov chains and specifically Markovian Semantic Indexing has been a major field of research in the past few years with advances to a huge extent. Markov chains are constructed with the keywords as states and a transition probability from one state to another is noted. This architecture accounts for the semantic information that the markov chain gathers from the user. For example, a user needs a picture of a “computer”. By “computer” the user could mean either the object computer or a picture containing the text computer. This kind of information is called the semantic information. Markov chains can help in analyzing exactly what the user wants and presenting it to him as the output. This research proposes to study Markov chains and develop a model that could support annotation based image retrieval and devise a new algorithm for the ranking of the images retrieved. The basic notion is to learn semantic information automatically from a huge set of image samples and use any specific concept model to caption fresh images. These systems attempt to reduce the semantic gap. This method is specially fit for ABIR where the per image annotation data is limited. In the past works, it has shown retrieval results more efficient in terms of precision and recall rates. The aim of the proposed research is to present and construct a new methodology for markov chain construction and ranking in the ABIR systems for image retrieval.

Organization of the rest of the paper is as follows: In the next section some of the related work has been briefly discussed. In section 3, the proposed methodology is mentioned and explained followed by the conclusion in section 4.

2 Related Work

CBIR has been explored considerably more than ABIR. In the last few years, more than 200 CBIR systems have been studied and researched. Few examples are QBIC(Query by Image Content) [1], Photobook [2], Netra [3], MARS Multimedia Analysis and Retrieval Systems [4]. Most of these quantify the image similarity and retrieve pictures that are most identical to the query image. Despite the huge amount of research, CBIR systems fail to rectify the semantic gap. This could be taken care of by annotating the image closely to the content based retrieval system. Its the most effective way to reduce the semantic gap because annotations provide the system with the semantic information that helps improve the system.

Mori et al [5] were the first to construct a method for annotation of images using grids in co-occurrences where each region within the image inherited a certain word. Duygulu et al [6] presented an innovative approach that treated image annotation as a machine translation which translates textual keywords to visual keywords. Annotation Based Image Retrieval systems are more capable because it incorporates semantic content to both the textual queries and image tags. As a result, many document retrieval and indexing techniques such as LSI [7], PLSI [8] were included into ABIR systems.

As of the ABIR systems, Latent Semantic Indexing (LSI) is a technique to get relationships between terms and concepts in a document. Instead of the traditional searching techniques, LSI looks for something that is semantically close to the meaning of the searched keywords instead of simply looking for specific keywords. Singular Value Decomposition is used on term-document matrices to get the relationships between terms and concepts of the document. The process of LSI is computationally intensive and the cons become clear when dealing with large amounts of data.

Since LSI has number of shortcomings because of its unclear and insufficient theoretical base, Hofmann proposed the probabilistic LSI (PLSI) model. PLSI is an automatic document indexing technique. The basis of PLSI is similar to LSI. PLSI also handles the issue of synonymous and polysemous words just like LSI. All documents in PLSI is represented by its word frequency. Despite of being a good text analysis method, PLSI has a few limitations. Some of them are: since it provides no probabilistic model at the level of documents, it is incomplete. Over fitting issues occur if many specifications are present in the model and it is still uncertain as to how the document outside of the training data needs to be given a probability.

To help with the drawbacks of pLSI, Latent Dirichlet Allocation (LDA) an ungoverned, generative model was presented by Blei et al [9] which is quite similar to the PLSI. With a probabilistic approach, LDA is a powerful method created to model keywords of a document. Every document is a combination of a few semantic topics where every topic is then marked by distribution over keywords.

Only a few years back, Konstantinos A. Raftopoulos et al [10] found out an innovative probabilistic method called as the Markovian Semantic Indexing (MSI) which automatically annotated images, indexed them and worked on annotation based retrieval of images. The paper compared the proposed method to two of the famous methods, LSI and PLSI. MSI is preferred for images that have limited annotation data. The images are annotated in an automatic way by the users using the system which showed definite benefits of MSI over both LSI and PLSI. This research is presenting a new algorithm Koogle, a Markovian image retrieval technique for retrieving the images, taking into consideration the user preferences through markov chains.

3 Proposed Approach

The framework of the proposed approach is a image search image retrieval where the user searches for a image by submitting a query made up of keywords. The motive is to improve the results for the concerned user by presenting to the user the images that have better chances of being accepted. The system has a training phase in which relevance relations are set for the queries and the images in terms of probability weights. Through this phase, the system gathers the semantic information regarding what the user wants. In the testing phase, the system uses the gathered information to retrieve better results and increase user satisfaction.

The block diagram below shows how the system works.

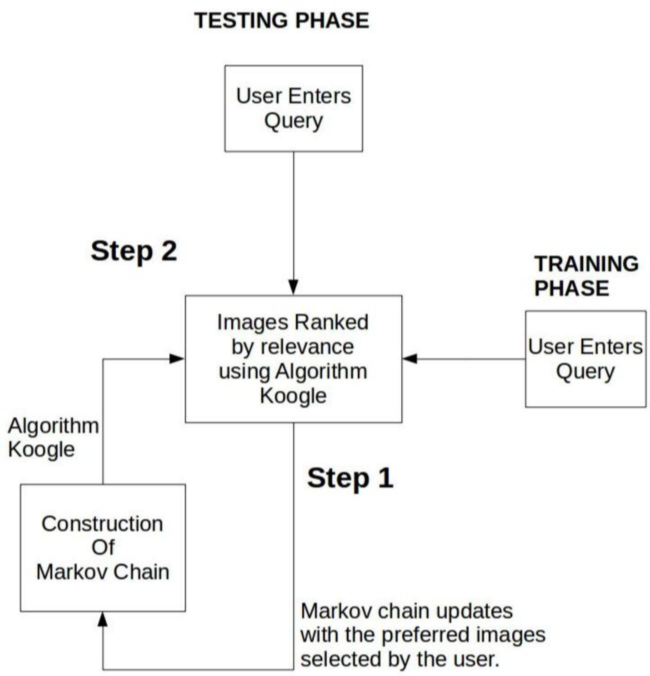

Fig 1: Working of the System

The entire procedure and its need is explained in the following steps:

Step 1: The user is indirectly establishing relevance links between the query and the annotations of the selected images in terms of a probabilistic weight (training phase). Let ‘q’ be the query composed of keywords $k_1, k_2, ..., k_n$ that was searched for a total of M number of times. Let *i* be an image selected by the user and *a* its annotations composed of keywords $a_1, a_2, ..., a_n$. Suppose the user searches for the query and selects the image *i*, *m* more number of times. Then the new probability relation between the query and the image annotations assuming that the old probability was $p_{old}(q, a)$ is based on the recurrent formula:

$$p_{new}(q, a) = \frac{M \times p_{old}(q,a) + m}{M + m}$$

Through this a link is established similar to $(k_1, k_2, ..., k_n) \longrightarrow (a_1, a_2, ..., a_n)$. This process creates a Markov chain where the multiple keywords query and the image annotations form the various states of the chain. Each time the query-image pair is selected, the query state counter is incremented and the interstate link between the pair is also advanced. This way, not only the occurrences but also the sequences of the occurrence from the query to the images are recorded. The modeling approach is justified because it gathers the user’s perception of the images in relation to the query. By taking multiple keywords in one state a logical connection between the keywords is established since we know which keywords are likely to used together by the user. In the later steps, it will be explained how this property has been exploited to retrieve relevant results for the targeted user.

Step 2: From the above step, a markov chain is created with some transition probability between the states. In this step, the testing phase starts. The fact that the user grouped certain keywords and submitted, implicitly renders the keywords relative to each other. The query that the user submits is divided into multiple subsets and each subset is treated as a state in the markov chain. For each of these states(from_state), a set of annotations state(to_state) - the ones with the highest probabilities - are chosen. It should be noted that the division into subsets is done in a decreasing order and then a similar operation of finding the images for each subset is carried on from the one with the maximum number of elements to the one with the minimum number of elements, separately each time. The annotation set is called the markovian keywords and it is this set that helps retrieve the desired result images. The images are checked to contain the markovian keywords and displayed to the user appropriately. The training and the testing phase goes hand in hand so that all the preferences and changes that the user wants are recorded simultaneously by the system.

Step 3: Optimization Step. The images as a result of the markovian keywords is the outcome of the algorithm but in reality it will exist in a very large number which can make the system inefficient. Also, the user might find so many images useless. So while retrieving the images, the number of images to be retrieved is limited to a desired number n. This way, the most important images are on the top and the rest are arranged in a decreasing order of relevance and at the same time an excessive load on the system is prevented.

The experiment was performed on the dataset available at [11] which contains more than 15000 images. To understand the system well, a class of images were selected and worked upon. Recording the user preference was done by taking a log of the queries searched, user clicks and the images selected each for each of these queries. A block diagram for a better understanding of Koogle is given in figure Fig. 2. The other modules work as shown in the algorithm below.

Let the query searched by the user be composed of keywords k_1, k_2,..., k_n in order and 'markov_chain' be the chain constructed during the training phase. A transition in the 'markov_chain' happens from a 'from_state' to the 'to_state' with a transition probability of 'p'.

Algorithm Koogle:

Find the subsets of the set S = [k_1, k_2,..., k_n] in decreasing order.

For each subset 's' in S,

Retrieve the images that contain the subset 's'

markovian_keywords = get_markov_keywords(s)

Retrieve top_n_images(markovian_keywords) // n is the desired number of images

Function get_markov_keywords(subset s)

From the markov_chain,

find the transition that starts at from_state('s') and in the decreasing order of transition probabilities, get the respective to_states

Return to_state

Function top_n_images(markovian_keywords)

loaded_images[n] //data structure of size n to store images

From the database,

find the images with the annotations 'markovian_keywords' and store it 'loaded_images'

Return loaded_images

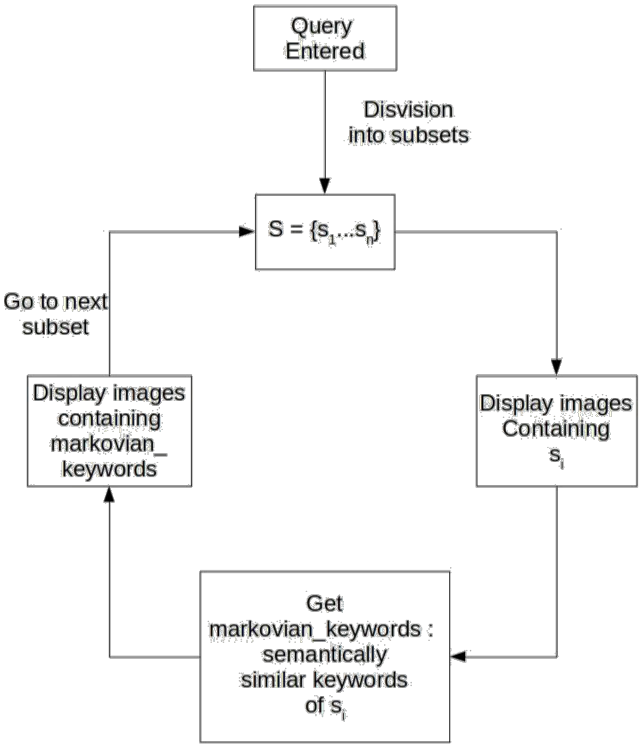

Fig 2: Block diagram explaining Koogle

The algorithm produces a series of images in the order of highest relevance to lowest relevance.

4 Conclusion and Future Work

The research proposed Markovian Image retrieval, an approach that mines user queries by incorporating keyword relevance with the help of markov chains and deducing a new ranking system based on the same. The proposed system is dynamically trained by the queries of the same user that the system will serve. As a result, the focusing will be more or less accurate than most of the existing systems. A reliable ranking algorithm is proposed which ranks the images retrieved in order of relevance to the targeted user. The algorithm works on the markov chain containing the user preferences constructed by the system itself. The system also shows how two or more keywords can be grouped in a single state of the markov chain and be used to establish keyword relevance.

As of now the proposed algorithm retrieves images that contain the actual keywords(A_k) and the images that contain the semantically similar keywords(S_k). The system can be extended into retrieving images that are semantically similar to keywords 'S_k' and so on. The system is being integrated with Automatic image annotation of the images used and being extended to the Markovian Semantic Indexing as well. The results be will studied which could reveal certain advantages of the proposed system in the future.

References

1. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D.,And Yanker, P. 1995. Query by image and video content: The QBIC system. IEEE Comput. 28, 9, 23–32.
2. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. IJCV, 18(3):233–254, 1996.
3. Ma,W. And Manjunath, B. 1997. Netra: A toolbox for navigating large image databases. In Proceedings of the IEEE International Conference on Image Processing (ICIP).
4. The MARS web site: http://www.ifp.illinois.edu/~qitian/MARS.html
5. Y.Mori, H.Takahashi and R.Oka, Image-to-word transformation based on dividing and vector quantizing images with words. First International Workshop on multimedia Intelligent Storage and Retrieval Management (1999).
6. P. Duygulu, K. Barnard, J.F.G. de Freitas and D.A. Forsyth "Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary." In proceedings of 7th European Conference on Computer Vision, 2002.
7. S. Deerwester, S. T. Tumais, T. K. Landauer, G. W. Furnas, and R.A. Harshman Indexing by latent semantic analysis, J. Soc. Inform. Sci. 41, 6 (1990), 391_407.
8. T. Hofmann, "Probabilistic Latent Semantic Indexing," Proc. 22ndInt'l Conf. Research and Development in Information Retrieval (SIGIR'99), 1999.
9. D.M. Blei and A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.
10. Konstantinos A. Raftopoulos, Member, IEEE, Klimis S. Ntalianis, Dionyssios D. Sourlas, and Stefanos D. Kollias, Member, IEEE Mining User Queries with Markov Chains: Application to Online Image Retrieval IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 2, February 2013
11. The Open Images dataset is available at: https://github.com/openimages/dataset.