

Genetic Candidate Group Search Approach for Post Clustering Content based Image Retrieval

Manasee Kurkure
Department of Computer
Engineering
PCCOE Pune-44

Anuradha Thakare
Department of Computer
Engineering
PCCOE Pune-44

Santvana Gudadhe
Department of Computer
Engineering
PCCOE Pune-44

ABSTRACT

Large amounts of databases are created daily in data storage. Due to this it becomes very difficult to retrieve the images require for applications in various fields. Thus Content Based Image Retrieval Techniques play an important character in image processing. Here we will be using various masking methods to find out different features and apply different clustering algorithms. In this paper, we are proposing a hybrid model which is the combination of Genetic and Candidate Group Search algorithm. This gives us the best results in some aspects and find it suitable in point of time and accuracy. Candidate Group search Genetic Algorithm is employed to facilitate the users retrieve the images that are most relevant to the users demand.

General Terms

Data Mining Algorithm, Clustering.

Keywords

Content based Image Retrieval, clustering, Feature Extraction, Genetic Algorithm, Candidate Group search algorithm.

1. INTRODUCTION

Large amounts of databases are developed in data storage and acquisition technologies. We need to gather all the imaged precisely and effectively from these databases for diverse applications such as art collections, medical diagnosis, instruction, entertainment, fashion industry, crime prevention, etc. Hence there is a need to build up an appropriate scheme that will manage these entire databases. Two approaches are adopted for image retrieval viz., text- and substance-based. The major objective of the CBIR system is to construct meaningful descriptions of physical attributes from images to facilitate effective and efficient recovery. Generally, research activities in CBIR have progressed in three major directions: global features based, object region-level features based, and relevance feedback. Initially, developed systems are normally based on the carefully selected global image features, such as color histogram, texture or shapes, and prefixed similarity measure. CBIR is a fast advancing and developing research area in the field of image retrieval. It is a technology that consists of two stages: First is the feature extraction manages the digital image databases as per their visual contents. CBIR Typically the features that take into consideration are pattern, color, texture, shape, mean etc. On the basis of these features

thesystem then decides the similarity between different images. Image clustering and categorization is a means for high-level description of image content. The goal is to find a mapping of the images into classes (clusters) such that the set of classes provides essentially the same information about the image archive as the entire image-set collection. The generated classes provide a concise summarization and visualization of the image content that can be used for

different tasks related to image database management. Image clustering helps in the implementation of efficient retrieval algorithms and the creation of a user-friendly interface to the database.

(GAs) are inspired from phenomena found in living nature. The phenomena incorporated so far in GA models include phenomena of natural selection as there are selection and the production of variation by means of recombination and mutation, and rarely inversion, diploid and others. Most Genetic Algorithms work with one large population, i.e. in the recombination step each individual may potentially choose any other individual from the population as a mate. Then GA operators are performed to obtain the new child, offspring, the operators are:

1. Crossover: - The task of crossover is the creation of a new individual out of two individuals of the current population.
2. Mutation: - The mutation operator M (chromosome) selects a gene of that chromosome and changes the allele by an amount called the mutation variance (mv), this happens with a mutation frequency (mf).

2. RELATED WORK

In data clustering is popular technique and K-Means is one of the fastest methods for clustering, but it has some disadvantages like convergence to local optimum and sensitive to initial data. Another method is K-harmonic Means which removes first drawback of K-Means method. In this paper for solving the clustering problem proposed the new method Candidate Group Search based on K-Harmonic Mean [1]. The selection rules are used for forming the candidate group and used in the algorithm. It uses the deterministic and randomness both the properties. This yields better performance for large scale data and as compared to VNS but still it faces the local optima problem. In a data clustering similar data points are grouped into a group which are more alike to each other and it puts dissimilar data points into another group. The KHM is a popular method from clustering methods. It suffers the drawback of local optimum convergence. In this paper the Gravitational Search Algorithm (GSA), K-harmonic Means (KHM) and Genetic algorithm (GA) is employed to propose new hybrid data clustering approach called Gravitational Genetic KHM [2]. The proposed study uses the advantages of Genetic Algorithm to work off from the drawback of KHM algorithm. The result tested on the different dataset. In this paper content based image retrieval is performed using one or two low grade features such as shape, color and grain[3].They have combined the advantages and disadvantages of Bayesian and genetic algorithm techniques and content based image. The hybrid approach for the content based image retrieval is the combine effort from the genetic and Bayesian algorithm.

In paper presents a new approach called image retrieval system based on the IGA [4]. Content based image retrieval is a difficult task of capturing relevant images from a large storage space. The proposed system in this paper consists of user-oriented CBIR that uses the interactive genetic algorithm (IGA) to understand which images in the databases would be of most interest to the user. Three visual features, color, texture, and edge, of an image are utilized in the approach. IGA provides an interactive mechanism to better capture user's intention. In general, an image retrieval system usually provides a user interface for communicating with the user. It collects the required information, including the query image, from the user and displays the retrieval results to him. Here, representing and retrieving the image properties of colour, texture and edge are used using interactive genetic algorithm (IGA) for better approximation with user interaction. With the help of image compression, digital image processing, and image feature extraction techniques are more developed, CBIR maintains a steady pace of development in the research field.

In paper Image Retrieval means searching, browsing, and retrieving the images from an image database [5]. Two different methods are used for image retrieval, namely text based image retrieval and content based image retrieval techniques. But now Text based search technique is old one. In Content Based Image Retrieval many visual feature like color, shape, and texture are extracted, next when we query an image its feature are compared with the stored feature and we get most similar kind of image. In this proposed method we firstly extract low level image feature like- colour histogram, colour coherence vector. Then add edge detection technique, sobel edge detection method to get better output. Finally they have use Manhattan distance to find the similar images from the database.

3. PROPOSED WORK

In proposed work shown in Fig 1 we are doing post content based image retrieval with the help of data mining algorithms. We are using the hybrid approach of Candidate Group Search and Genetic Algorithms to do clustering efficiently. The block diagram of the proposed CGSGA algorithm for post clustering of CBIR is explained in fig. 1. The steps for the proposed work are as follows: From the initial image dataset, the features are generated individually for each image. These features are then applied to different clustering algorithms like EM, K-means and CGSGA to find clusters depending upon their feature values. Hence we can be used to generate the results for comparative analysis. Thus, with the help of clustering results obtained we can fast retrieve the required data. As compared to K-means and EM, the CGSGA will do the post clustering method accurately within less computational time.

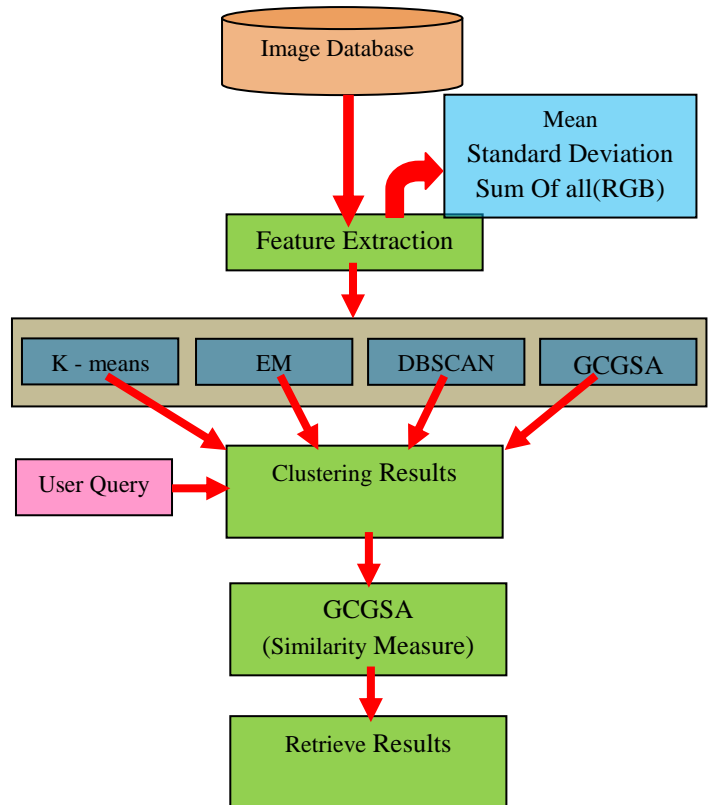


Fig 1. Proposed Model

3.1 Performance and Comparative Analysis

The performance of Candidate Group Search (CGS) and Genetic Algorithm (GA) are the clustering algorithm used for clustering data. In reaching a global optimal result of cluster we used the genetic algorithm.

3.1.1 Distance Function

In clustering data, initially, the steps of CGS algorithm are executed that is, the data set is partitioned into a number of clusters then for each data point in the distance from the center to the data point is calculated using the Euclidean distance measure.

3.1.2 Fitness Function Selection

The cluster quality is dependent on K-Harmonic Mean (KHM) as the fitness function in which lesser the KHM value than more superior the cluster quality [17]. After finding the initial centers using CGS selecting candidates to replace centers using crossover of Genetic Algorithm. Again, apply the CGS algorithm on newly generated centers. Choose the centers with highest accuracy, using the fitness function of Genetic Algorithm [1].

$$KHM = \min \sum_{i=1}^N \frac{K}{\sum_{j=1}^k \frac{1}{\|xi - cj\|^p}}$$

Each chromosome is evaluated based on the fitness function in the genetic algorithm and is given a fitness value. After that crossover operator function is used to exploit promising areas of the search space which will give improved clustering results to reach global optima.

3.1.3 The Clustering Performance Measures

F-Measure uses the concept of information retrieval. The precision and recall are calculated for clustering. The F-Measure is calculated using [8],

$$F = \sum_j \frac{mj}{m} \max_i F(i, j)$$

3.1.4 The Performance of Clustering Algorithms

The proposed algorithm CGSGA has evaluated on image datasets and compared with CGS based clustering algorithms. The experiment is done in Java and matlab. Using clustering algorithm, we obtain the results which are shown in Table 1.

4. EXPERIMENTAL RESULTS

The proposed CBIR with CGSGA clustering is implemented with a system of having i5 processor and main memory of 2 GB RAM using MATLAB (R2012b).

Test Bed-The Image dataset collected from CORAL database [9]. The test results are generated on 100 images. In test bed images are considered. Sample test images are shown in fig.2. The test database consists of variable images of flower, butterfly and Nature.



4.1 Result

In this paper CBIR technique is used in order to extract the feature of the image. We use the hybrid CGSGA algorithm for cluster optimization. The combination of CBIR and hybrid CGSGA gives the best result for the image Retrieval from large databases. Initially from the histogram, we extract the feature of images through image color histogram method. In this, first calculate the image histogram and count of Red, Blue, Green color pixels, from this we calculate the sum of all pixels, Mean and Standard Deviation. This is nothing but the Feature of the images and Generated new Feature vector database. On this database, we apply clustering on GSGA.

Table 1. Analysis of F-Measure Value

Sr. No.	Clustering Algorithm	Number of Iterations	Execution Time (sec)	Accuracy of Algorithm
1	K-means	10	0.01	80.02
2	CGS	10	0.13	76.15
3	GA	10	0.15	78.05
4	GCGSA	10	0.004	82.42

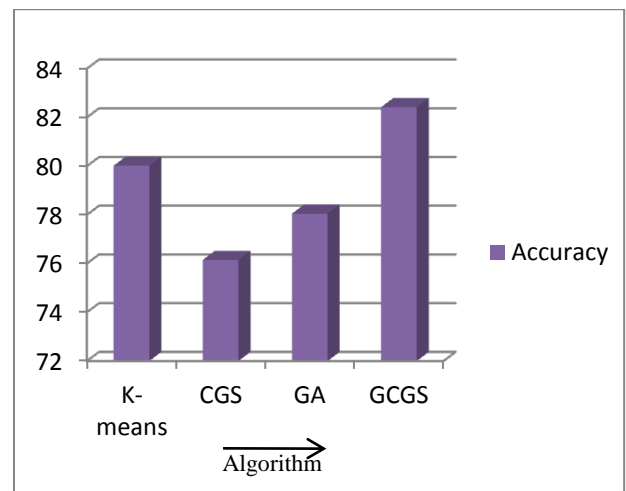


Fig 3 : Analysis of clustering algorithm in terms of Accuracy

The graphical representation of accuracy obtained by various algorithms is as shown in Fig 3. As shown in above figure we can see that the novel GCGSA gives better performance amongst all in terms of accuracy with 82.42.

5. CONCLUSION

In this work, different features are extracted and combined to form feature vectors of the image. The clustering algorithms do not always guarantee good results as the accuracy of the final clusters depends on the selection of initial centroid. The proposed method, a hybrid Candidate Group Search Genetic Algorithm called CGSGA for post clustering is implemented which results in optimized clusters. It overcomes the problem of local optimization thereby producing accurate clusters automatically. Hence GCGSA help in extracting the similarity on feature vectors which helps in fast image retrieval process from large databases.

6. REFERENCES

- [1] Cheng Huang Hung, Hua-Min Chiou, Wei-Ning Yang, "Candidate groups search for K-harmonic means data clustering," Applied Mathematical Modelling, Elsevier, 2013.
- [2] Anuradha Thakare, Rohini Hanchate, <http://research.ijcaonline.org/volume88/number17/pxc3894002.pdf>, "Introducing Hybrid model for Data Clustering using K-Harmonic Means and Gravitational Search Algorithms," 2014.

- [3] Devendra Kurwe, Prof. Anjna Jayant Deen ,Dr. Rajeev Pandey “A Hybrid Approach for content based image retrieval from large Dataset”, IJCTT) – volume 23 November 1 – May 2015.
- [4] M. Venkat Dass, Mohammed Rahmath Ali, Mohammed Mahmood Ali, “Image Retrieval Using Interactive Genetic Algorithm”, International Conference on Computational Science and Computational Intelligence-2014.
- [5] Kalyan Roy, Joydeep Mukherjee “Image Similarity Measure using Color Histogram, Colour Coherence Vector, and Sobel Method”, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064 Volume 2 Issue 1, January 2013 .
- [6] G. Sathiya, P. Kavitha, “An Efficient Enhanced K-Means Approach with Improved Initial Cluster Centers,” Middle-East Journal of Scientific Research, ISSN 1990-9233, 20 (4): 485-491, 2014.
- [7] Zulal Gungor, Alper Unler, "K-Harmonic means data clustering with tabu-search method," Applied Mathematical Modelling, Elsevier, 2007.
- [8] Yugal Kumar and G. Sahoo, “A Hybrid Data Clustering Approach Based on Cat Swarm Optimization and K-Harmonic Mean Algorithm,” Journal of Information and Computing Science Vol. 9, No. 3, 2014, pp.196-209.
- [9] <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval> Last referred on 15 september 2015.