Genetic Algorithm for Best Segmentation of Gray Level Images

Zainab Falah Hasan
University of Babylon
science collage for women / Computer science
zainab_ga@yahoo.com

Abstract

Data clustering is collecting the objects that have similar characteristic together for processing purposes.

In this work , genetic algorithm is presented for clustering of images purpose by generating efficient centers of clusters. The efficiency of generated centers is computed by using Xie-Beni index. The chromosome represents many centers needed for clustering and its length is proposed between 3 to 10. The generated centers are used for clustering of different gray levels images. Quality measures are used to evaluate images after clustering. Many experiments are used in this system. In every experiment, genetic algorithm is applied to select the best centers with different number of clusters. The results with different numbers of clusters showed the efficiency of the proposed method.

Key words : Genetic Algorithms, Clustering, Patterns Recognition.

الخلاصة

عنقدة البيانات هي تجميع الأشياء التي تملك خصائص متماثلة سوية لأغراض المعالجة.

في هذا العمل, الخوارزمية الجينية مقدمة لغرض عنقدة الصور بواسطة توليد مراكز كفؤة للعناقيد. كفاءة المراكز المتولدة تحسب باستخدام مقياس Xie-Beni. الكروموسوم يمثل العديد من المراكز المحتاجة للعنقدة وطوله مفترض بين 3 الى 10. المراكز المتولدة استخدمت لعنقدة مختلف الصور ذات المستويات الرمادية. مقاييس النوعية استخدمت لتقييم الصور بعد العنقدة. عدة تجارب استخدمت في هذا النظام. في كل تجربة, الخوارزمية الجينية تطبق لاختيار افضل مراكز مع عدد مختلف من العناقيد. النتائج بمختلف الأعداد من العناقيد أوضحت كفاءة الطريقة المفترضة.

الكلمات المفتاحية: الخوارزميات الجينية, العنقدة , تمييز النماذج.

1 Introduction

Clustering is technique of an unsupervised classification by collecting a set of patterns into groups (clusters) where the elements of one group are similar by using a defined measure. Minimizing the difference between elements of the same cluster is created by clustering of elements.

Many application domains have widely studied clustering problem, such as discovery of knowledge , mining of data , classification of data, analysis of statistical data and compression and processing of medical images. Several algorithms have been proposed for clustering. Distribution of a set of data elements into subsets have similar objects to each other and different with other subsets is the aim of all algorithms of clustering (Barker *et al.*, 2003).

Wu, *et al.*, (2003) used the K-Means algorithm for efficiently clustering large data set.

A modified genetic method for image segmentation based on feature clustering was worked by Miki and Kumiko (2004). The algorithm used many steps to create the initial population. Ya-Wei (2009) explained clustering of images by using the evolutionary algorithms and genetic algorithm.

Malay (2009) used k-means technique to Avoid empty Clusters, the modified k-means technique solved the problem of empty cluster and was semantically equivalent to the original k-means technique.

Venkateswaran and Muthukumar (2010) used method to segment medical image by algorithm of Generalized Spatial Fuzzy C- Means with genetic algorithm. In this method, objective function of clustering reached to global minima. The algorithm improved the accuracy of the segmentation of medical images although it has high computational complexity.

Mirkamal *et al.,* (2013) used Variable string length genetic algorithm to cluster medical image. Genetic algorithm was used for the determination of the optimal cluster centers.

Khan and Ravi (2013) used methods of image segmentation. They worked to segment image using the main principle of the methods.

2 Image Clustering

Classes of similar objects are created by collecting a set of objects and this process is called the clustering. A set of data objects that are similar to one another in the same cluster and in other clusters are different to the objects represents cluster. Generally, it can be classified the main clustering methods into the following types.

Methods of Hierarchical: it is dividing a given set of data objects in hierarchically format.

Methods of Partitioning: it is dividing a database of n objects into k partition , where a cluster is one partition and $k \leq n$. That is, the data classified into k groups.

Methods of Density-based: it is clustering a data in shape of spherical only and getting difficulty an arbitrary shapes at discovering clusters.

Methods based on Grid: it is quantizing the objects into grid structure that represents a finite number of cells.

Methods based on Model: it is supposing for each cluster a model and getting for the given model the data that are the best fit.

The type of available data and the main purpose of the problem determine the clustering algorithm (Jiawei and Micheline, 2006).

Distance functions are used to measure the similarity or dissimilarity between two classes of patterns. The Euclidean distance, which is the most popular measure in the image analysis (Soni Madhulatha, 2012), computed as :

$$\text{Distance}(A,B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2}$$

Where A and B are patterns implemented with n features.

3 Genetic Algorithms

Genetic algorithms inspired from the evolutionary ideas of Darwin about genetic and natural selection and they are search heuristically. The Genetic algorithms main idea is designed to simulate processes required for evolution in the natural system (Pratibha and Manoj, 2010). GA has three main applications, namely, an intelligent search, optimization and robot learning.

In every generation, the Genetic algorithm consists of set of individuals (population). A potential solution to the problem acts an individual from population. The initial population is randomly generated and its size based on the nature of the problem (Pratibha and Manoj, 2010). Every individual called a chromosome that contain set of genes. The chromosome is represented by using a suitable encoding. The encoding may be a binary, integer or real form.

The efficiency of individual (solution) is computed to get the best one in the final generation using fitness measure which depends on the problem.

To get the next generation (new solutions), combination operations are applied on selected individuals from population(Rakesh and Jyotishree, 2012).

The most common selection methods are :

- Roulette wheel selection: this method uses the fitness level to select individuals (Firas and Reyadh, 2012).
- Rank selection: this method arranges the individuals in the population from best to worst based on their fitness and each individual is assigned a numerical rank based on fitness, then selection is applied based on this ranking(Firas and Reyadh, 2012).
- Tournament selection: this method randomly chooses a set of individuals and takes the best individual for reproduction(Firas and Reyadh, 2012).

After selection of individuals, combination operations are applied to generate new solutions. There are two type of operations(Muhammad and Abido, 2009):-

1) Crossover: where two individuals are generated by combining parts from two individuals. The combination positions of any two chromosomes are selected randomly (Alessandro, 2008). There are many types of crossover such as:

<u>Single Point Crossover:</u> this method chooses randomly a single point crossover and two offspring are created by exchange the data of two parents(chromosomes) after the crossover position (Alessandro, 2008).

<u>Two point crossover:</u> this method chooses randomly a two point crossover and two offspring are created by interchange the data of two parents(chromosomes) between these points of crossover. (Alessandro, 2008).

2) Mutation: where the new individuals are changed. There are many types of mutation such as:

<u>Single Point Mutation</u> which is randomly select a single gene to be mutated and change its value based on the encoding type used (Young-C. and Ying-H, 2004).

<u>Multi Point Mutation</u> which is randomly select a multi genes to be mutated and change their values based on the encoding type used (Young-C. and Ying-H., 2004).

After many generations, genetic algorithm reaches to the best solution of the problem.

4 Genetic Algorithm for Clustering

Clustering of images is summarized in the following steps:

1-Generation of Random centers: Initial Population.
2- Efficiency Computation of Random Centers(chromosome).
3- Updating of Centers: Generation of New Solutions.
4- Efficiency Measurement of New Solutions.
5-Loop the Steps 3 to 4 Until the best centers have been gotten.

4-1 Generation of Random centers (Initial Population):

In the first step, initial chromosomes(centers)are generated randomly. Length of chromosome ranges between 3 to 10 genes (centers). Every gene represents center of cluster. The value of gene is in the range of (0..255) which represents colors of the gray level image. Every chromosome works clustering to image by its genes to test its efficiency in clustering. The chromosome is shown in the following figure:

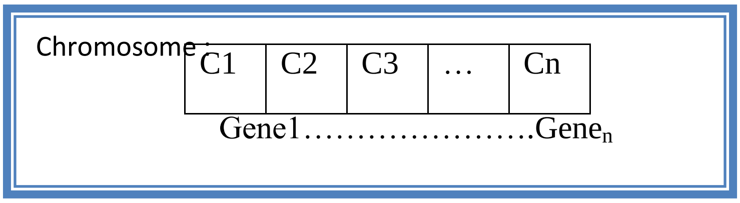


Fig. 1: Chromosome Representation.

Every gene represents color from gray level image.

4-2 Efficiency Computation of Random Centers:

The generated centers (chromosome) are evaluated their performance in clustering. This is to get the best centers for clustering. Evaluation of the centers is obtained by initially testing them in clustering of an image by determining their belonging of pixels to the nearest center using distance measure of them with this center. The nearest pixel to a center$_i$ acts its belonging to the cluster$_i$, otherwise computing its distance with another center.

After that, the chromosome uses Xie-Beni index to compute its fitness and evaluate its performance in clustering test. This is the quotient between the mean quadratic error and the minimum distance of the minimal squared between the objects in the clusters. The mean quadratic error is simply the mean of the squared distances of all the pixels with respect to the cluster center they belong to.

4-3 Updating of Centers

Updating of centers is done to generate new centers using crossover and mutation operations to get the best centers. Binary tournament Selection is employed to choose chromosomes for combination operations. One-x crossover type is used to get new chromosomes where combination of two chromosomes(set of centers) generates other centers for clustering. After this, 2m mutation type is applied on the resulting chromosomes. The new chromosomes are tested in clustering of the image where they may have the best centers.

4-4 Efficiency Measurement of New Solutions

This step similar to 4-2 , the fitness is computed to the new population . After getting of the best chromosome (best centers) , updating of new chromosomes (sets of centers) is stopped and the image is clustered using the best selected centers from genetic algorithm. Otherwise , loop step 4-3 to 4-4 until no other best centers has been gotten.

5 Results

Quality measures are computed for clustered images to evaluate the performance of this system that are:

1- Image Fidelity (IF)=$1-\sum_k\sum_j [(X(j,k)-Y(j,k)^2]/\sum_k\sum_j [X(j,k)^2]$
2- Structural Content (SC)= $\sum_k\sum_j [X(j,k)^2]/[\sum_k\sum_j Y(j,k)^2]$
3- Normalized Correlation Coefficient (NK)= $\sum_k\sum_j [(X(j,k)*Y(j,k)^2]/ \sum_k\sum_j [X(j,k)^2]$

Where j=1..N, k=1..M , N and M are dimensions of the image and X is the original image and Y is the clustered image.

The results when using genetic algorithm for two different images of size 255*255 and for every number of clusters are shown in the table 1. Table 2 explains quality measures of clustered images using different number of clusters ranges between 3 to 10, the original images are shown in the fig. 2 and fig. 3:

Fig. 2 The Original Image1.

Fig. 3 The Original Image2.

Table1: Clustered Images with Different Number of clusters and their Selected Centers by Genetic Algorithm.

No. of clusters	Clustered image1	Selected Centers	Clustered image2	Selected Centers
3		19, 71, 202		27, 205, 160
4		27, 185, 219, 92		249, 71, 125,172

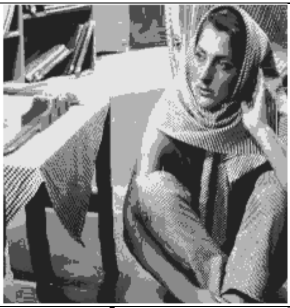
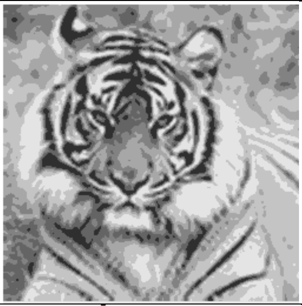
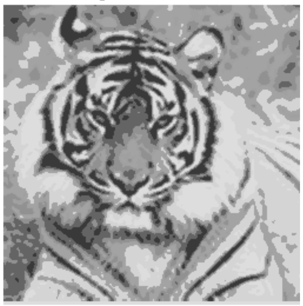
5		40, 88, 125, 224, 186		62, 190, 240, 116, 160
6		229, 65, 76, 35, 137, 179		100, 47, 130, 228, 196, 167
7		222, 235, 15, 175, 146, 114, 64		145, 221, 100, 212, 163, 160, 57
8		82, 228, 198, 50, 126, 30, 169, 115		173, 127, 180, 126, 66, 241, 147, 198
9		103, 88, 15, 221, 50, 91, 140, 233, 181		211, 159, 124, 171, 201, 49, 228, 186, 104
10		29, 91, 223, 87, 17, 133, 102, 241, 173, 201		82, 200, 207, 155, 15, 177, 191, 163, 105, 238

Table2: The Performance of Segmentation for Images by Genetic Algorithms using Quality Measures.

No. of clusters	Image1 clustering			Image2 clustering		
	IF	SC	NK	IF	SC	NK
3	0.998	0.995	0.979	0.996	0.997	0.992
4	0.994	0.999	0.990	0.997	0.996	0.994
5	0.995	0.998	0.996	0.995	0.998	0.997
6	0.998	0.995	0.997	0.998	0.995	0.999
7	0.996	0.997	0.997	0.994	0.999	0.996
8	0.997	0.996	0.999	0.995	0.998	0.997
9	0.994	0.999	0.997	0.994	0.999	0.998
10	0.998	0.995	0.999	0.995	0.999	0.998

6 Conclusions

The Genetic algorithm is applied in this work on different gray level images to get efficient centers for clustering. Xie Beni Index is participated in improving of performance of the system by computing the fitness of chromosomes for clustering. This method is efficiently clustered images and succeeded to find the optimal centers for different number of clusters. Genetic algorithm has many generations to get the best solution and it is continue with generation of solutions until no more best solutions has been gotten and this is appearing its efficiency than other clustering methods. Quality measures is also used to evaluate the performance of this method and they are showed efficiency of the system with different numbers of centers.

7 References

Abul, L.; Alhajj, R.; Polat F. and K. Barker , 2003, Cluster Validity Analysis Using Sub sampling, in proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Washington DC, Volume 2.

Akbar Shahrzad Khashandarag, Mirkamal Mirnia and Aidin Sakhavati, 2013, Medical Image Clustering Using Genetic Algorithm, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 1

Alessandro V. A., Luciano d. E., Andr´e L. L. A., Luciana P. A. and Carlos H. N. R. , 2008, Analysis of Selection and Crossover Methods used by Genetic Algorithm-based Heuristic to solve the LSP Allocation Problem in MPLS Networks under Capacity Constraints, International Conference on Engineering Optimization Rio de Janeiro, Brazil, 01 - 05.

Amit K. , 2000, Artificial Intelligence and Soft Computing, CRC Press LLC, Vol. 2.

Firas Alabsi and Reyadh Naoum, 2012, Comparison of Selection Methods and Crossover Operations using Steady State Genetic Based Intrusion Detection System, Journal of Emerging Trends in Computing and Information Sciences, ISSN 2079-8407 Vol. 3, No.7.

Jiawei Han and Micheline Kamber, 2006, Data Mining: concepts and Techniques, ELSEVIER.

Khan A. M. and Ravi. S, 2013, Image Segmentation Methods: A Comparative Study, International Journal of Soft Computing and Engineering (IJSCE).

Malay K. Pakhira, 2009, Modified k-means Algorithm to Avoid Empty Clusters , International Journal of Recent Trends in Engineering, Vol 1, No. 1.

Miki Aoyagi and Kumiko Tsuji, 2004**,** A modified genetic algorithm for image segmentation based on feature clustering, International Symposium on Information Theory and its Applications.

Muhammad T. and M. A. Abido, 2009, Assessment of Genetic Algorithm Selection, Crossover and Mutation Techniques in Reactive Power Optimization, IEEE.

Pratibha B. and Manoj K. , 2010, Genetic Algorithm– an Approach to Solve Global Optimization Problems, Indian Journal of Computer Science and Engineering, Vol. 1 No. 3 pp. 199-206.

Rakesh K. and Jyotishree, 2012, Effect Of Polygamy With Selection In Genetic Algorithms, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Vol. 2.

Soni T. Madhulatha, 2012, An Overview On Clustering Methods, IOSR Journal of Engineering, Vol. 2(4) pp: 719-725.

Venkateswaran R. and Muthukumar, S. 2010, Genetic Approach on Medical Image Segmentation by Generalized Spatial Fuzzy C- Means Algorithm , IEEE International Conference on Computational Intelligence and Computing Research.

Wu, F. X.; Zhang, W.J. and A.L. Kusalik, 2003, Determination of the minimum samples size in micro array experiments to cluster genes using K-means clustering, in Proceedings of 3rd IEEE Symposium on Bioinformatics and Bioengineering.

Ya-Wei Ho, Chih-Hung Wu and Chih-Chin Lai, 2009, Aerial image clustering using genetic algorithm, IEEE transactions on Pattern Analysis and Machine Intelligence.

Young C. and Ying-H.C., 2004, A New Efficient Encoding Mode of Genetic Algorithms for the Generalized Plant Allocation Problem, Journal of Information Science and Engineering, Vol. 20.