Optimization of Fuzzy C Means Clustering using Genetic Algorithm for an Image

Jyoti Tyagi
Inderprastha Engineering College
M.Tech
Computer Science Engineering

Neeta Verma
Inderprastha Engineering College
Associate Professor
Computer Science Department

ABSTRACT
Fuzzy C-Means Clustering algorithm (FCM) is a method that is frequently used in pattern recognition. It has the advantage of giving good modeling results in many cases, This paper presents the optimization of cluster center of Fuzzy C-Means algorithm by evolutionary methods, this in order to automatically select the best of cluster center with maximum probability. Optimization methods used to realization of this paper were genetic algorithms and for selection method roulette wheel method is used.

Keywords
cluster center, membership matrix, objective function, genetic algorithm, optimization, roulette wheel selection.

1. INTRODUCTION
Clusters of data arise from the need to find interesting patterns or groups of data with similar characteristics within a given data set. Fuzzy clustering aims at partitioning a data set into homogeneous fuzzy clusters. The most widely used algorithm to realize fuzzy clustering is the Fuzzy C-Means (FCM) algorithm proposed by Bezdek (1981) [15]. This algorithm has been the base to developing other clustering algorithms.

Image segmentation is an process of dividing the image into different types of classes and groups the homogeneous pixels into a cluster [14]. In image only the pixel attribute is considered for doing clustering. Clustering can be done with various algorithm e.g K Means clustering, Fuzzy C Means clustering [15,13]. In K Means clustering k centroids are initialized i.e one for each cluster [13] , in this each pixel of image will belong to exactly one cluster for and the clusters are crisp in nature [13]. In this paper we have used Fuzzy C Means clustering algorithm [8]. Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In some cases Fuzzy Clustering is more natural than K Means clustering [7]. In Fuzzy Clustering the number of clusters and weight exponent m=2 are predefined [14]. Fuzzy Set theory produces an idea of partial membership of belonging which is based on the membership function [15]. The fuzzy nature and the ease of implementation had made this algorithm very attractive for a lot of researchers. The algorithm FCM is a method of clustering which allow one piece of data to belong to two or more cluster [8] with the help of membership function. The genetic algorithm is used for the optimization of Fuzzy C Means algorithm. In genetic algorithm there are various selection process [6].The individual from are evaluated from the fitness function. Roulette Wheel is on the selection process that is used in this paper [1].

2. FUZZY C-MEANS CLUSTERING
FUZZY C-MEANS CLUSTERING is introduced by Bezdek in 1981,which was extended from Hard C-Mean clustering method. Clustering is a method of dividing data points into homogenous cluster and the data point which are similar, belongs to same cluster and dissimilar data point belong to different cluster. It is an approach of unsupervised learning. Clustering is used in various areas like medicine, geology, business, engineering systems and image processing, etc. This algorithm is based on distance between various data points. The clusters are formed according to the distance between data points and the cluster centers are formed for each cluster. The degree of membership of each data item to the cluster is calculated which decides the cluster to which that data item is supposed to belong. In fuzzy clustering objects are not fully forced to belong one of the class but rather they are assigned membership degree between 0 and 1indicating their partial membership , so a data point can belong to all cluster with different membership values between 0 and 1. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center.
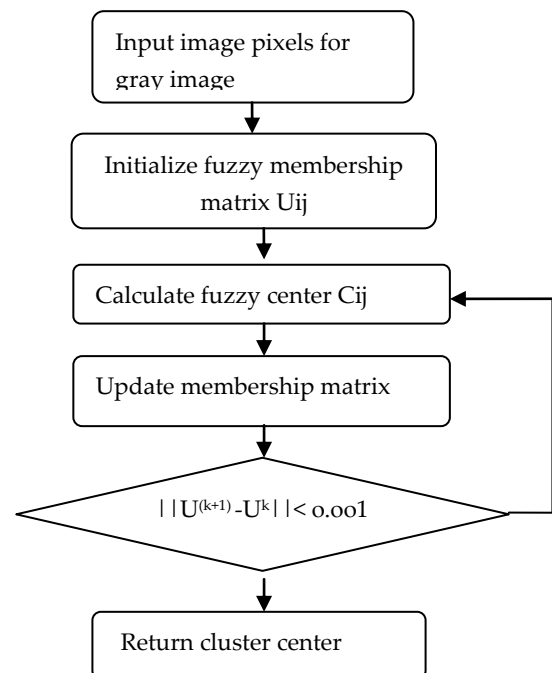
The algorithm is composed of the following steps:

1.Initialize U=[u_{ij}] matrix, $U^{(0)}$

2. At k-step: calculate the centers vectors $C^{(k)}$=[c_j] with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} . x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

Where m, is constant,

u_{ij} is the degree of membership of x_i in the cluster j,

x_i is the ith of d-dimensional measured data,

c_j is the center of the cluster

3. Update,$U^{(k)}$,$U^{(k+1)}$

$$\mu_{ij} = \frac{1}{\sum_{k-1}^{c} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right) \left(\frac{\|x_i - c_j\|}{\|x_i - c_j\|} \right)}$$

4. If $\| U^{(k+1)} - U^{(k)} \| <$ then STOP; otherwise return to step 2 .

3. GENETIC ALGORITHMS

Genetic algorithms are inspired by Darwin's theory about evolution. Solution to a problem solved by genetic algorithms is evolved. Algorithm is started with a set of solutions (represented by chromosomes) called population. New population is formed from the solution of one population. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are selected from the current population (based on their fitness), and modified to form a new population. With a hope, that the new population will be better than the old one. Solutions which are selected to form new solutions (offspring) are selected according to their fitness - the more suitable they are the more chances they have to reproduce. This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied.

The selection process of which individuals will be terminated, and which will be parents for the next generation. Evolution is a process of adaption. The fitness function that is used for evaluating the individuals tells us something about the requirements that are needed to survive in the environment. To obtain a higher fitness value, the population needs to adapt increasingly more to the environment.

4. ROULETTE WHEEL SELECTION (RWS)

Roulette method selection method works similarly like a roulette wheel, where the best individual is chosen from a group according to the fitness value of each individual. Basically selection process select the best individual from one generation to create the basis for the next generation. Fittest individual have a greater chance of survival than the weaker ones. The size of each individuals „slice" of the roulette wheel will be inversely proportional to their fitness value. Once the „slices" have been determined, a number is generated at random. The individual with the range of numbers that contains this randomly generated number will be one parent. This continues until the desired number of parents is found. Based on the value of fitness function, roulette wheel method selects the next best possible solution chromosome that will
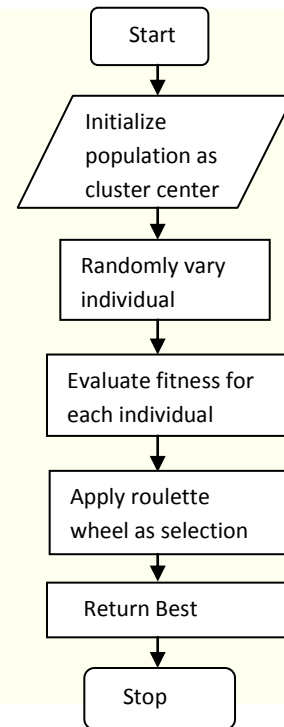
The conspicuous characteristic of this selection method is the fact that it gives to each individual i of the current population a probability p(i) of being selected, proportional to its fitness f(i)

$$p(i) = \frac{f(i)}{\sum_{j=1}^{n} f(j)}$$

Where n denotes the population size in terms of the number of individuals and f(i) is the fitness function. The RWS can be implemented according to the following pseudo-code
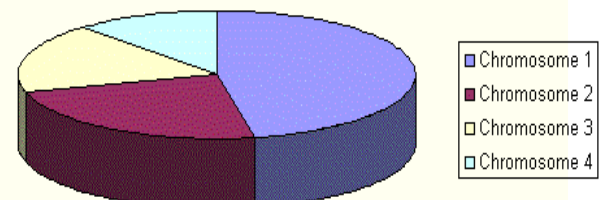

Figure:1 Roulette wheel selction

RWS_ pseudo-code

```
{
    • Calculate the sum    s = ∑ⁿᵢ₌₁ f(i)
    • For each individual 1 ≤ i ≤ n do {
        - Generate a random number α Є [0,1];
        - iSum=0;  j=0;
        - Do  {
                • iSum← iSum + f(i);
                • j=j+1;

            } while ( isum <α and j<n)

        -    Select the individual j; }
}
```

5. FITNESS FUNCTION

The fitness function plays a very important role to obtain the best solutions within a large search space. Good fitness functions will help GA to explore the search space more effectively and efficiently. Bad fitness functions, on the other hand, can easily make GA get trapped in a local optimum solution and lose the discovery power. The "fitness function" is responsible for performing this evaluation and returning a positive integer number, or "fitness value", that reflects how optimal the solution is: the higher the number, the better the solution. The fitness values are then used in a process of natural selection to choose which potential solutions will continue on to the next generation, and which will die out. However, that natural selection process does not merely choose the top *x* number of solutions; the solutions are instead chosen statistically such that it is more likely that a solution with a higher fitness value will be chosen, but it is not guaranteed. This tends to correspond to the natural world.

A common metaphor for the selection process is that of a large roulette wheel. Remembering that fitness values are positive integers, imagine that each potential solution gets a number of slots on the wheel equal to its fitness value. Then the wheel is spun and the solution on which it stops is selected. Statistically speaking, solutions with a higher fitness value will have a greater chance of being selected since they occupy more slots on the wheel, but even solutions with just a single slot still have a chance.

The fitness function used here is :

$$\pi = -4sin\theta1 - 6cos\theta1 - 4sin\theta2 - 2cos\theta2$$

S. NO	Binary coding	Angles	S. No	Binary coding	Angles
1.	0000	0	9.	1000	48
2.	0001	6	10.	1001	54
3.	0010	12	11.	1010	60
4.	0011	18	12.	1011	66
5.	0100	24	13.	1100	72
6.	0101	30	14.	1101	78
7.	0110	36	15.	1110	84
8.	0111	42	16.	1111	90

Table 1: angles corresponding to binary coding.

6. EXPERIMENTAL RESULT

This paper gives an approach to find the best fittest cluster center out of various centers obtained through implementation of Fuzzy C Means clustering algorithm. Table 2 comprises of different cluster obtained from FCM algorithm, the roulette wheel fitness function, and the probability of all cluster. The experimental conduct shows that the cluster 4 with cluster center value 92 have highest probability when applied to roulette wheel selection approach.

S.No	center	F= -f-7	Probability	f
1	66	2.8962	0.0819	8.19
2	26	3.8493	0.1088	10.88
3	121	4.547	0.1286	12.86
4	92	4.6184	0.1306	13.06
5	44	4.1228	0.1166	11.66
6	237	1.9337	0.0547	5.47
7	188	3.5169	0.0994	9.94
8	171	3.9318	0.1112	11.12
9	148	3.2168	0.091	9.1
10	5	2.7321	0.0773	7.73

Table 2: Results of cluster selection approach.

7. CONCLUSION

The fuzzy C Means algorithm is an iterative approach to find the optimal number of cluster. The main aim of this algorithm is to find the different cluster centers value, it may be possible that some of the clusters are not effective, so to find the best cluster center genetic algorithm is used. There are various selection process available in genetic algorithm, one is roulette wheel selection. A fitness function F is used to for the selection process. This paper gives a approach to find the best cluster value.

Fig 2: image used for clustering source"
www.smartchoicemri.com"

8. REFERENCES

[1] Adam Lipowski , Roulette-wheel selection via stochastic acceptance.

[2] Castillo O "Optimization of Fuzzy C-Means Algorithm using Evolutionary Methods,2010

[3] D.Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley, 1989.

[4] Dong-Chul Park "Intuitive Fuzzy C-Means Algorithm" IEEE conference 2009 PP – 83-87

[5] George J. Klir and Bo Yuan "Fuzzy Sets and Fuzzy Logic" Prentice-Hall India Publications-2005

[6] Khalid Jebari, Selection Methods for Genetic Algorithms , IJES ,2013.

[7] M Alata "Using GA for Optimization of Fuzzy C- Means Clustering Algorithm. Research Journal of Applied Sciences ,(2013)

[8] Rajasekaran S Neural Network , Fuzzy Logic, and Genetic Algorithms book

[9] Rakesh Kumar "Blending Roulette Wheel Selection & Rank Selection in Genetic Algorithms" International journal of Machine Learning and Computing , 2012.

[10] R.Suganya " Fuzzy C- Means Algorithm- A Review , International Journal of Scientific and Research Publications , 2012.

[11] R. Kruse, C. Döring, M. J. Lesot; ―Fundamentals of Fuzzy Clustering‖; In: Advances in Fuzzy Clustering and its Applications; John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2007, Pages 3-30.

[12] Stelios Krinidis and Vassilios Chatzis " A Robust Fuzzy Local Information C-Means Clustering Algorithm" 2009 IEEE journal Transactions on Image Processing pp-01-11

[13] Soumi Ghosh " Comparative Analysis of K-Means and Fuzzy C-Means Algorithms"IJACSA,2013

[14] Venkateswaran R "Genetic Approach on Medical Segmentation by Generalized Spatial Fuzzy C-Means Algorithm"IEEE International Conference,(2010)

[15] Yang Y "Image Segmentation By Fuzzy C-Means Clustering Algorithm With a novel penalty term" , (2010)