Travelling Salesman Problem Using Genetic Algorithm And Fuzzy C-Mean Clustering Algorithm

Ajendra Kumar, Preet Pal Singh, Pawan Joshi

Abstract This paper presents a new algorithm called Fuzzy C-Mean Genetic Algorithm (FCMGA) to solve TSP which is used to calculate the minimum travelling cost in TSP. FCMGA is a combination of both Fuzzy C-Mean (FCM) and Genetic Algorithm (GA). The role of FCM and GA is different in this algorithm. FCM is used to find the membership values of each chromosome, whereas GA is used to apply mutation on the variables of those chromosomes. The proposed algorithm is very helpful to find nearly optimized solutions of these types of problems, in order to give the best solution of the problem with the reduction of cost. The algorithm (FCMGA) proposed here is tested with some examples and the experiment shows that the algorithm can achieve good results as compared with GA.

Index Terms— Travelling Salesman Problem (TSP), Genetic Algorithm (GA), Fuzzy C-Mean algorithm (FCM), Fuzzy C-Mean Genetic Algorithm (FCMGA)

1. Introduction

Clustering is known to be the most vital unendorsed learning problem, so as every other issues of this kind; it deals with ruling an understanding in accumulating of unlabelled data. The intent of cluster analysis is to classify objects into subsets that have some meaning in the context of a definite problem. Clustering is a process of dividing the population or data points into a number of groups such that the points in the same group are more equivalent or we can say that more similar to other data points in the same group than those in different or other groups. In various numbers of clustering problems, clustering may be defined as distribution of n-patterns among C groups, where c is the number of clusters and the adjustment in a group are more to be expected to each other than to pattern in different group. Fuzzy C-mean (FCM) clustering algorithm is a well known and widely used clustering technique applicable in such situations. FCM algorithm is relevant to a vast area of geostatistical data analysis problems [1]. An algorithm was proposed by J.C.Bezdek in FORTRAN 4 coding by FCM clustering algorithm, in which the author described the algorithm that generates fuzzy partitions and models for any set of algebraic or numerical input [1]. FCM algorithm plays an important role in clustering analysis by classifying the unlabelled data points into similar type of clusters. The Fuzzy Probabilistic C-mean Model (FPCM), was proposed by J.M.Keller et al in 2005 in which the algorithm generates the membership and typical values of unlabeled data for noise sensitivity defect of FCM [16].

- *Ajendra Kumar is currently an assistant professor in the department of Mathematics and statistics at the Gurukula Kangri Vishwavidyalaya Haridwar India, E-mail: ajendra.gkv@gmail.com*
- *Preet Pal Singh is currently an assistant professor in the department of Mathematics at the P.T.L.M.S.(P.G) college Rishikesh, uttrakhand, India, E-mail: ppsingh003@gmail.com*
- *Pawan Joshi is currently pursuing Ph.D. in the department of Mathematics and statistics at the Gurukula Kangri Vishwavidyalaya Haridwar India, PH-+91 9997760427. E-mail: pawanjoshi425@gmail.com*

For image segmentation technique, fuzzy c-mean with spatial constraint is known to be very adequate and fortunate method. This method is very productive and valuable when applied on synthetic and real life applications [3, 2]. In real world application fuzzy clustering is an essential problem which has numerous uses. FCM algorithm is one of the most admired fuzzy clustering techniques because it is well-organized, straightforward and easy to execute [10]. Initial cluster centre plays a very important role in FCM algorithm [9]. By finding good initial cluster centre which is close to the genuine final cluster centre, that can reduced the dispensation time, that helps the FCM algorithm to converge swiftly [9]. In 1997 Nikhil Pal and James C. Bezdek proposed a combined mixed algorithm known as Fuzzy Probabilistic C-Mean (FPCM) model, in which the model produces both membership and possibilities along with the regular point models and cluster centres for each clusters [15]. The noise sensitivity defect of FCM algorithm is used to solve by this model and it also affected the collateral clustering problem of Probabilistic C-Mean (PCM) [15, 16]. Fuzzy clustering model is crucial and very effective tool to find the required cluster structure of given data set in pattern and image classification [20]. To escalate the act of both FCM as well as Fuzzy weighted C-Mean (FWCM) models, a new weighted fuzzy C-Mean (NW-FCM) algorithm was proposed for high dimensional multiclass pattern recognition problems [8]. In this work the author has applied the NW-FCM algorithm on synthetic and real data set which gives better result as compared to FCM and FWCM. It is also been used in Magnetic Resonance Image (MRI) of ophthalmology to differentiate the normal tissues [20, 8]. A trade-off weighted fuzzy factor and a kernel metric is used to clarify the image segmentation problem, which heavily rely on the space width of all adjacent pixels [6]. Genetic algorithm (GA) is a suitable class of transformative algorithm which uses technique stimulated by nature. GA has been used in numerous types of problem; it is an effort to optimize a specific objective function associated to a clustering problem [14]. Genetic algorithm is randomized search techniques that imitate some of the process noticed in natural evolution [7, 17]. GA is robust and probabilistic search algorithm which depends on the procedure of natural selection and survival of the fittest that is used to determine the escalation and many real life problems [5]. The main aim of GA is to produce nearly optimal solutions by letting a set of arbitrary solutions which undergoes a sequence of binary transformation guided by a section scheme influenced toward high-quality solutions [12]. GA is a process which is used to search for suitable cluster centres in the essential space such that the similar metric of the resulting clusters is optimized. In this paper we have described a common method to solve travelling salesman problem (TSP) by FCM and GA. GA has been applied on TSP by Moon.C et al in 2001, in which the ordering of vertices in a directed graph is used to minimize the travelling cost and also to develop new crossover operators for the proposed GA [13]. On applying GA in TSP, the crossover and mutation operators are evolved to contract with the TSP with unlike representation such as; binary representation, path representation, adjacency representation, ordinal representation and matrix representation [11].

Travelling salesman problem (TSP) is described as; we are given a set of places and a distance matrix which is symmetric or asymmetric, that illustrates the cost of travelling from one place to other place [4]. The goal is to find the précised circular tour, going every city exactly once, so as to reduce the total mean cost which consists of the cost of travelling from the last city back to first city [4, 11]. Generalized Traveling Salesman Problem (GTSP) is used to explain the objective of TSP, which is used to find a minimum cost tour passing through one node from each cluster [19].

This paper organized as follows; section 1, 2, 3 and 4 introduces the introduction, the basics of clustering, fuzzy c-mean clustering algorithm and genetic algorithm respectively. Section 5 presents the proposed algorithm called FCMGA which describes how the proposed algorithm works. In section 6 we have taken a problem of TSP and solved it with FCMGA .Section 7 introduces the comparison between the solutions of TSP by GA with our proposed algorithm FCMGA. Finally section 8 concludes the paper and discusses the future work.

2. CLUSTERING

Clustering is the process of dividing a given set of points or objects into a number of groups (clusters) based on some similarity or dissimilarity among them. Hence we can say that clustering is the collection of all those objects or points into different clusters based on assigning a specific rule. For example, if we divide clusters according as similarity or dissimilarity then those objects which are similar are put on same clusters and those which are dissimilar are put on different clusters. Let the set of n points $\{x_1, x_2, \ldots, x_n\}$ be the set Z and m clusters are represented by $c_1, c_2, \ldots, c_m$. Then

$c_i \neq \emptyset$ For $i = 1, 2 \ldots m$

$c_i \cap c_j = \emptyset$; $i = 1, 2 \ldots m, j = 1, 2 \ldots m$ and $i \neq j$,

$$\bigcup_{i=1}^{m} c_i = Z$$

There are two main approaches for clustering; one method is crisp clustering (or hard clustering) and the other one is fuzzy clustering. Clustering can be classified into supervised and unsupervised clustering methods [8]. For human interaction, supervised clustering method is used whereas, to detect the underlying structure in the data set for classification, pattern recognition etc…, the unsupervised clustering method is used [8].

3. FUZZY C-MEAN

FCM is a clustering algorithm, which allows data point to be designated into more than one cluster. It is an extension form of K-mean algorithm. This method was proposed by J.C.Dunn in 1973 which was later on improved by J.C.Bezdek in 1981. The algorithm works by finding membership values of each data point corresponding to each cluster centres on the basis of distance between the cluster centres and data points by specific formula given below. If the data points are near to the cluster centres then there membership values are higher corresponding to those data points which are far from that particular cluster centres.

Hence, in FCM algorithm it separates the finite collection of elements X= $\{x_1, x_2, \ldots, x_n\}$ on the basis of some known conditions into a number of clusters.

The main aim of this method is to minimize the value of J_m by initializing the membership matrix randomly and computing the cluster centres.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} \mu_{ij}^{m} \left\| x_i - c_j \right\|^2$$

Where, m is any fuzzy component; $m > 1$

N= Total number of data,

C= Total number of clusters,

μ_{ij} = Membership value of x_i in cluster j,

c_j = The d-dimensional centre of the cluster,

It consists of following steps:

Step1: select random cluster centre at least 2,

Step2: compute a membership matrix by following formula,

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}} ;\ 1 \leq i \leq c,\ 1 \leq j \leq n$$

where, $\left\| x_i - c_j \right\|$ is the distance from point i to current cluster centre j

Step3: calculate the new c-cluster centre by following formula,

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}{}^{m} x_i}{\sum_{i=1}^{N} \mu_{ij}{}^{m}} \quad ; 1 \leq j \leq c$$

Hence, repeat the process until we do not get same cluster centres.

4. Genetic algorithm (GA)

In computer science and operation research GA is an independent technique that encouraged by the process of natural selection that belongs to the larger class of evolutionary algorithm (EA). It is a search technique and a computational intelligence method, which is used in computer science to find approximate solution to the combinatorial optimization problems [4]. Genetic algorithms are commonly used to generate high quality solutions to escalate and analysed the problems by depending on bio inspired operators such as, mutation, crossover, and selection. The role of mutation and crossover operator are different in GA [11]. To increase the average quality of the population crossover operator is applied whereas, mutation operator is applied to avoid local optima in the algorithm and to explore new states [11].

In GA, there exist a mapping between phenotype and genotype which totally depends on the characteristics of the optimization problems. The GA method simulates the survival of the fittest among those specific solutions over generations to solve that problem [5].

The individual member of a general GA with binary representation for fix length 4 looks like in following figure, the bit values are taken randomly.

1	1	0	1

4.1. Basic operators of GA

(A). Reproduction (Selection)

Reproduction or selection is a process, which is applied on the population. It is also known as first operator of GA. Chromosomes is selected from the population of parents to apply crossover and to produce offspring. It is based on Darwin's evolution theory of "The survival of the fittest". According to the Darwin's evolution theory, the new offspring's are created by the best one who survives [5]. The selection of the strings is done by the reproduction operator for possible inclusion in the next generation which is based on the fitness value of a string, and it is calculated from the fitness function [5].

(B). Crossover-

Crossover is the second operator of GA, which makes copy of good strings. Crossover is a process in which two entities are combined to create new entity, which are copied into the new generation. It is applied to the matting pool with an expectation that it would generate better strings. To apply crossover between any two chromosomes and to produce new offspring, chromosomes are selected according to their fitness value [5]. It is also known as a child production operator [11]. It exchanges information between two parent chromosomes for generating child chromosomes [12].

(C). Mutation-

After crossover, the strings are deal with mutation of a bit. In mutation, a gene is mutated by flipping the binary values of a chromosome [12]. Mutation is also known as third operator of GA, which is used to maintain genetic multiplicity from first generation of n-population of chromosomes to the next generation. The generation of any string could be done by any given strings in the mutation operator [14]. It may produce better offspring by changing few genes from parent's chromosomes [19].

5. Fuzzy C-Mean Genetic Algorithm

Fuzzy C- Mean Genetic Algorithm (FCMGA) is an algorithm which uses both GA and FCM commonly. FCMGA clustering algorithm appropriately determines fuzzy c- clusters in R^N .

In this algorithm, each chromosome can be defined by sequence of $N \times C$ floating point numbers where, the first n-position represents n- dimensions of 1^{st} cluster and the second n- position represents n- dimensions of 2^{nd} cluster and so on. The main aim of the algorithm is to increase the fitness function of each of the chromosomes by searching for the appropriate cluster centres, to minimize the travelling cost.

5.1. Outline for FCMGA

Step1: population initialization:

C- Randomly selected distinct points from the data set, which are used to initialize the C- clusters centres encoded in each chromosome.

Step2: perform the chromosomes by variable $x_1, x_2, \ldots, x_n$

Step3: apply FCM algorithm to find the membership matrix (note that centre of the cluster should be at least 2).

Step4: compute fitness function for each chromosome by;

$$\text{Fitness function} = \frac{1}{\mu_i}$$

Where $\mu_i = \min(\mu_{i1}, \mu_{i2}, \ldots, \mu_{im})$; m = number of cluster centres

Step5:- select two chromosomes, which have higher fitness function and apply mutation on those two chromosomes.

Step6:- now after applying mutation, calculate the travelling cost of the chromosomes.

In the next sections, the solutions of travelling salesman problem (TSP) have been generated by FCMGA and compare the result which has been solved by GA.

6. IMPLEMENTATION OF ALGORITHM

Example 1:- Consider a distance/cost matrix by Table1.1

Distance/cost	A[1]	B[2]	C[3]	D[4]	E[5]	F[6]
A[1]	0	90	100	35	300	200
B[2]	90	0	60	120	400	290
C[3]	100	60	0	70	480	225
D[4]	35	120	70	0	320	150
E[5]	300	400	480	320	0	290
F[6]	200	290	225	150	290	0

Table 1.1

Where A, B, C, D, E, F represents different cities and the element in the matrix represents the cost of travelling one city to another.

Our aim is to find the minimum cost on travelling to these cities by FCMGA.

Let us take four different paths: P_1, P_2, P_3, P_4

$P_1 = [2,1,3,4,5,6]$

$P_2 = [1,2,3,5,4,6]$

$P_3 = [1,4,3,2,6,5]$

$P_4 = [5,3,2,1,4,6]$

Where the path cost is defined as $C(P_i) = \mu(a,b) + \mu(b,c) + \mu(c,d) + \mu(d,e) + \mu(e,f) + \mu(f,a)$

where, $P_i = [a,b,c,d,e,f]$, and $\mu(a,b)$ = distance from a to b;

$C(P_1) = 1160, C(P_2) = 1300, C(P_3) = 1045, C(P_4) = 1105.$

therefore, P_3 has least cost; $C(P_3) = 1045.$

Our goal is to find the better path, which has minimum cost than P_3 .

On applying FCMGA algorithm:

The membership matrix is shown by table 1.2;

Variables	c_1	c_2
1	0.0637	0.9363
2	0.0555	0.9445
3	0.3678	0.6322
4	0.9848	0.0152

Table 1.2

Now to compute membership value of each chromosome by

$\mu(i) = \min(\mu_{i1}, \mu_{i2});\ i = 1,2,3,4.$

Therefore,
$\mu(1) = 0.0637, \mu(2) = 0.0555, \mu(3) = 0.3678, \mu(4) = 0.0152$

Now to compute fitness function defined by:

$$f[\mu(i)] = \frac{1}{\mu(i)}$$

Hence fitness functions of variables are,

$f[\mu(1)] = 15.6986,$

$f[\mu(2)] = 18.018,$

$f[\mu(3)] = 2.7189,$

$f[\mu(4)] = 65.7895$

Now apply mutation on P_i ; i= 1, 2, 3, 4

Before mutation		After mutation	
P_i	$C(P_i)$	P_i	$C(P_i)$
[2,1,3,4,5,6]	1160	[4,1,3,2,5,6]	1035
[1,2,3,5,4,6]	1300	[1,4,3,5,2,6]	1475
[1,4,3,2,6,5]	**1045**	[1,2,3,4,6,5]	**960**
[5,3,2,1,4,6]	1105	[5,3,4,1,2,6]	1255

Table 1.3

The table 1.3 shows that after applying FCMGA the minimum travelling cost reduces from 1045 to 960. The graph representation is shown below by Fig (a).

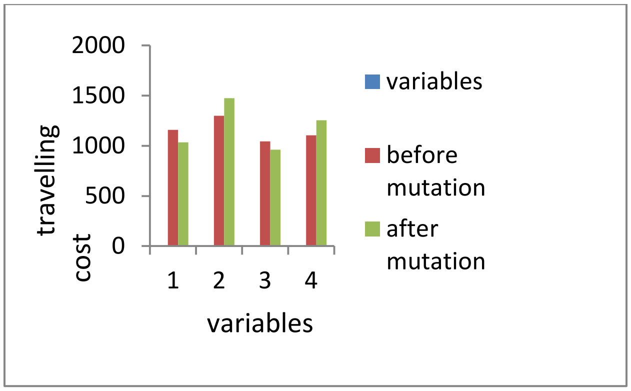

Fig (a)

Hence, FCMGA is very helpful and easy to implement for solving the travelling salesman problem. Similarly we can solve any square matrices of any order weather it is symmetric or asymmetric.

In the next section we have compared the solutions obtained by GA and FCMGA.

7. COMPARISON BETWEEN GA AND FCMGA

This section presents two comparison tables of two different order matrices (7×7) and (6×6), which has been presented by Dwivedi.V et al [4] and Moon.C et al [13] respectively.

7.1- Comparison of travelling cost between GA [4] and FCMGA given by table 1.4.

Order of matrices	By Dwivedi.V et al [4]		By FCMGA				Best path	Path cost
	Selected chromosomes	Path cost	Selected chromosomes					
			Before mutation	Minimum cost	After mutation	Minimum cost		
[7×7]	[1,3,6,4,5,7,2] [1,5,4,2,6,3,7] [1,5,6,4,3,7,2]	378 349 263	[1,3,6,4,5,7,2]	378	[1,3,4,6,5,7,2]	343	[7,6,5,3,4,2,1]	234
			[1,5,4,2,6,3,7]	329	[1,5,6,2,4,3,7]	327		
			[1,2,3,7,6,4,5]	425	[1,2,3,7,4,6,5]	386		
			[7,6,3,4,2,1,5]	368	[7,4,3,6,2,1,5]	324		
			[7,4,5,3,6,2,1]	306	[7,6,5,3,4,2,1]	234		
			[5,7,4,6,2,1,3]	423	[5,7,6,4,2,1,3]	384		
			[4,6,7,1,2,5,3]	375	[6,4,7,1,2,5,3]	369		

Table 1.4

Hence from table 1.4, it is clear that the minimum travelling cost given by Dwivedi.V et al [4] is 263, whereas the minimum travelling cost comes by FCMGA is 234. Therefore, we can say that the minimum travelling cost comes out by our proposed algorithm FCMGA.

The comparison graph of the path cost given by Dwivedi.V et al [4] and by our proposed algorithm FCMGA is given by Fig (b).

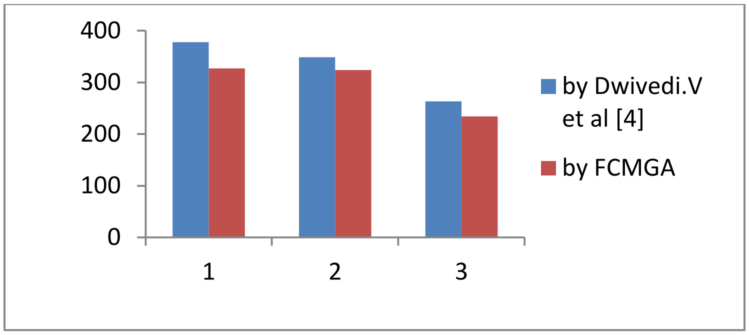

Fig (b)

7.2 Comparison of travelling cost between GA [13] and FCMGA given by table 1.5.

Order of matrix	By Moon.C et al [13]		By FCMGA				Best path	Minimum cost
	Selected chromosomes	Minimum cost	Selected chromosomes					
			Before mutation	Minimum cost	After mutation	Minimum cost		
[6×6]	[1,3,6,2,4,5]	49	[1,3,2,6,4,5] [5,3,6,2,1,4] [1,3,6,4,5,2]	53 57 48	[1,3,5,6,4,2] [2,3,6,5,1,4] [1,3,6,4,2,5]	52 58 47	[1,3,6,4,2,5]	47

Table 1.5

Hence, table 1.5 shows that the minimum travelling cost given by Moon.C et al [13] is 49; whereas the minimum travelling cost comes out by FCMGA is 47. Therefore, we can say that, FCMGA can achieve better result than GA.

8. CONCLUSION

This paper presents the utility of a Fuzzy C-mean Genetic Algorithm (FCMGA) which is used to reduce the travelling cost in Travelling Salesman Problem (TSP). In this algorithm cluster centres and membership matrix are used to find the fitness function of the variables by FCM algorithm. We can also say that, FCM is used for reproduction whereas; GA is used for mutation and crossover process. Hence experimental result shows that our proposed algorithm (FCMGA) is better in terms of quality of solutions and cost as compared to GA and FCM clustering algorithm. Our future plan is to use the algorithm in the medical or pharmacy field to reduce the expenditure of production of medicines.

REFERENCES

[1]. Bezdek, J.C., Ehrlich, R., Full, W., 1984. The Fuzzy C-Mean clustering algorithm. Computer and Geosciences 10, 191-203.

[2]. Cai, W., Chen, S., Zhang, D., 2007. Fast and Robust Fuzzy C-Means Clustering Algorithms Incorporating Local Information for Image Segmentation. Pattern recognition 40, 825-838.

[3]. Chuang, K.S., Tzeng, H.L., Chen, S., Wu, J., Chen, T.J., 2005. Fuzzy C-means clusters with spatial information for image segmentation. Computerized medical imaging and graphics 30, 9-15.

[4]. Dwivedi, V., Chauhan, T., Saxena, S., Agrawal, P., 2012. Traveling Salesman Problem using Genetic Algorithm. International journal of computer applications (IJCA).

[5]. Gambhava, N., Sanghani, G., 2003. Traveling Salesman Problem using Genetic Algorithm.

[6]. Gong, M., Liang, Y., Shi, J., Ma, W., Ma, J., 2013. Fuzzy C-Means Clustering with Local Information and Kernel Metric for Image Segmentation. IEEE Transactions on image processing 22(2), 573-584.

[7]. Grefenstette, J., Gopal, R., Rosmaita, B., Gucht, D.V., 1985. Genetic algorithm for travelling salesman problem. (IEEE conference paper), 160-168.

[8]. Hung, C.C., Kulkarni, S., Kuo, B.C., 2011. A New Weighted Fuzzy C-Means Clustering Algorithm for Remotely Sensed Image Classification. IEEE Journal of selected topics in signal processing 5(3), 543-553.

[9]. Hung, M.C., Yang, D.L., 2001. An Efficient Fuzzy C-Means Clustering Algorithm. IEEE, 225-232.

[10]. Izakian, H., Abraham, A., 2011. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. Expert systems with applications 38, 1835-1838.

[11]. Larranaga, p., Kuijpers, C.M.H., Murga, R.H., Inza,I., Dizdarevic, S., 1999. Genetic Algorithm for Travelling

Salesman Problem: A Review of Representations and Operators. Artificial intelligence 13, 129-170.
[12]. Maulik, u., Bandyopadhyay, S., 2000. Genetic Algorithm-Based Clustering Algorithm. The journal of the pattern recognition 33, 1455-1465.
[13]. Moon, C., Kim, J., Choi, J., Seo, Y., 2002. An Efficient Genetic Algorithm for Travelling Salesman Problem with Precedence Constraints. European journal of operational research 140, 606-617.
[14]. Murthy, C.A., Chowdhury, N., 1996. In search for optimal clusters using genetic algorithm. Pattern recognition letters 17, 825-832.
[15]. Pal, N., Bezdek, J.C., 1997. A mixed c-mean clustering model. IEEE, 11-21.
[16]. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C., 2005. A Probabilistic Fuzzy C-Mean clustering algorithm. IEEE Transactions on fuzzy systems 13(4), 517-530.
[17]. Potvin, J.Y., 1996. Genetic algorithm for the travelling salesman problem. Annals of operational research 63, 339-370.
[18]. Rezaee, M.R., Lelieveldt, B.P.F., Reiber, J.H.C., 1998. A new cluster validity index for the fuzzy *c*-mean. Pattern recognition letters 19, 237-246.
[19]. Snyder, L.V., Daskin, M.S., 2006. A Random Key Genetic Algorithm for the Generalized Travelling Salesman Problem. European journal of operational research 174, 38-53.
[20]. Wu, K.L., Yang, M.S., 2002. Alternative C-Means Clustering Algorithm. Pattern recognition 35, 2267-2278.