Comparative Study of Particle Swarm Optimization based Unsupervised Clustering Techniques

Mr.V.K.Panchal[1] , Mr. Harish Kundra[*2] , Ms. Jagdeep Kaur[#3]

[1] Scientist 'G', Defence Research Organization, Delhi, India
[*2]HOD (Computer Sc.& Engg.) , Rayat College Of Engg. & IT, Ropar, Punjab, India
[#3] Lecturer (Computer Sc.& Engg.), Rayat College Of Engg. & IT, Ropar, Punjab, India

Summary
In order to overcome the shortcomings of traditional clustering algorithms such as local optima and sensitivity to initialization, a new Optimization technique, Particle Swarm Optimization is used in association with Unsupervised Clustering techniques in this paper. This new algorithm uses the capacity of global search in PSO algorithm and solves the problems associated with traditional clustering techniques. This merge avoids the local optima problem and increases the convergence speed. Parameters, time, distance and mean, are used to compare PSO based Fuzzy C-Means, PSO based Gustafson's-Kessel, PSO based Fuzzy K-Means with extragrades and PSO based K-Means are suitably plotted. Thus, Performance evaluation of Particle Swarm Optimization based Clustering techniques is achieved. Results of this PSO based clustering algorithm is used for remote image classification. Finally, accuracy of this image is computed along with its Kappa Coefficient.

Key words:
Particle Swarm Optimization(PSO), Fuzzy C-Means Clustering (FCM), K-Means Clustering (K-Means), Swarm Clustering, Gustaffsons-Kessel Clustering (GK), Unsupervised Clustering, Remote Sensing, Image Clustering, Image Classification.

1. Introduction

Image clustering [8] can be defined as the identification of natural groups within a multispectral data set. The algorithm that performs clustering functions to partition a set of objects (pixels) into relatively homogenous subsets based on inter-object similarities with little or no overlap. In general, clustering methods can be categorized by principle (objective function, graph theoretical, hierarchical) or by model type (deterministic, statistical, heuristic, fuzzy).

In the traditional clustering algorithm, the samples are classified in the unique cluster, which is all known as a hard division. However, there is not definite boundary in most things. The concept of fuzzy clustering applies to the essence of most things, and reflects the reality of objects better.

Clustering algorithms are usually applied to feature space, and as such they do not use any spatial information (e.g. the relative location of the patterns in the feature space).

One major limitation of many classical clustering algorithms is that they assume that the number of clusters is known. However, in practice, the number of clusters may not be known. This problem is sometimes called *unsupervised clustering*. Unsupervised prototype-based clustering aims at determining the correct number of clusters, C, without any prior knowledge about it.

2. Data Mining

Data Mining is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. In order to achieve this, data mining uses computational techniques from statistics, machine learning and pattern recognition. This paper focuses on available data mining techniques for unsupervised clustering of remote images.

3. Objectives

The objective of this paper is to enhance the quality of the satellite image by placing the pixel into it's most appropriate land cover, for this we need to-

- To develop an efficient clustering algorithm based on PSO.
- Help researchers in comparing different clustering algorithms and generate benchmarks.
- To develop an efficient clustering algorithm that can find the "optimum" number of clusters in a data set
- To show that PSO can bring out results with in reduced iterations.

Manuscript received October 5, 2009
Manuscript revised October 20, 2009

- To show that PSO is good at converging to global optima, then getting hooked up in local optima, as in case of traditional clustering techniques, thus explore a larger search area to get better and accurate results
- The main objective of using PSO is its easy understandability, with the use of simple mathematical calculations dealing with changing velocity and position.
- Provide Comparative results, in easy to understand, graphical format.

Thus, the objective of the PSO clustering algorithm is to obtain the proper centroids of clusters for minimizing the intra-cluster distance as well as maximizing the distance between clusters.

4. Clustering Types

This paper considers Non-Heirarchical/ Deterministic Partitioning Methods. It mainly concentrates on PSO based K-means, Fuzzy c-means, Gustafson's-Kessel and Fuzzy K-means with extragrades.

Data partitioning/ Non-Hierarchical algorithms divide data into several subsets. Because checking all possible subset possibilities may be computationally very consumptive, certain heuristics are used in the form of iterative optimization. Unlike hierarchical methods, in which clusters are not revisited after being constructed, relocation algorithms gradually improve clusters.

4.1 K-Means Clustering

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.

2. Assign each object to the group that has the closest centroid.

3. When all objects have been assigned, recalculate the positions of the K centroids.

4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

4.2 Fuzzy C-Means Clustering

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function:

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

Where $1 <= m < \infty$

1. *Initialize $U=[\mu_{ij}]$ matrix, $U^{(0)}$*
2. *At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$*

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^{m} . x_i}{\sum_{i=1}^{N} \mu_{ij}^{m}} \qquad (2)$$

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

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{||x_i - c_j||}{||x_i - c_k||} \right]^{2/(m-1)}} \qquad (3)$$

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

4.3 Fuzzy k-means with extragrades

Fuzzy k-means with extragrades recognizes that some objects might not fit well in any of the groups that are formed and places these objects in an additional outlier group, the extragrades. DeGruijter and McBratney have assumed that there will be as many samples in the extragrade group as in a regular group. Algorithm for Fuzzy k-means with extragrades is as:-

Define 0< a <1
Initialize a
Calculate required mean extragrades membership:
u*req = 1 /(k+1)
iter_alfa = 0

Repeat {Root Finding Loop}

Iter_alfa = Iter_alfa + 1
Brent's algorithm
Update a

```
Initialize membership (U)
iter = 0
Repeat {Fuzzy k-means iteration}
   iter = iter+1
   Calculate class center (C)
   Calculate distance ||X-C||
   Update membership U'
   U=U'
Until ||U-U'|| <= tol_crit  .or. iter = Max_iter

 Calculate mean extragrades membership u*
  Calculate F(a) = |u*-u*req|

Until F(a) <= dif_tol .or. iter_alfa = Max_alfaiter
```

1) *Considerations while using distance measures:* The metric (or 'distance' measure) should be chosen with care, Euclidean, Diagonal or Mahalanobis Euclidean should not be used where different attributes have widely varying average values and standard deviations, since large numbers in one attribute will prevail over smaller numbers in another. With the diagonal and Mahalanobis metrics, the input data are transformed before use. Choosing diagonal results in transformation of the data attributes into having equal variance. Choosing Mahalanobis results in transformation of the data set to one in which all attributes have zero mean and unit variance, and correlations between variables are taken into account.

The fuzzy algorithms use measures of distance only when placing data points in groups. They are insensitive to direction. A useful (although not entirely correct) analogy is to think of the algorithms as defining circles in two-dimensional space (or spheres in three-dimensional space, or hyperspheres in multi-dimensional space) to cover the data points. Consequently, the performance of the algorithms on data sets showing a markedly stratified structure (for example, data clustered in planes, or on lines) depends critically on the choice of metric. Separations measured by diagonal or Mahalanobis distance might be better than Euclidean distance, for they compensate for marked departures from the (hyper-) spherical shape. Moreover, better separations might be given when more attributes are used, because measurements of many properties are likely to provide more distinctions and thus greater dissociation.

2) *Number Of Classes:* The analysis will commence by separating the data into the minimum number of classes (groups) and will terminate after separating the maximum number of groups. It is best to conduct an analysis by examining a few groups at a time.

3) *Numerical Parameters:* The numerical parameters chosen will affect the operation of the fuzzy algorithm. They are grouped according to whether they affect all models, or whether they are related only to those with extragrades. The following parameters apply to all models: fuzzy exponent, random start, maximum iterations and convergence limit. The parameters alpha tolerance difference and alpha iterations apply to fuzzy k-means with extragrades.

4) *Fuzzy Exponent:* The fuzzy exponent controls the degree of fuzziness of a classification. When it is made to (or close to) 1.0 a data point can belong to only one class, that is, the classification will be hard or non-fuzzy. If it exceeds 1.0, an individual may be given partial membership in more than one class, that is, the classification will be fuzzy. A fuzzy exponent of 1.3 works well with Euclidean distance but might need to be made smaller with Mahalanobis distance. If it is too small (i.e., too close to 1.0) the program will attempt divisions by zero and may crash. If too big, the individuals will be given equal memberships in all classes. When set near 1 (e.g., at 1.01) a hard classification is usually obtained. The value is not constrained at the upper end (see McBratney and Moore, 1985), and as it is increased the clustering becomes so fuzzy that no groups are distinguished. Different values should be explored.

5) *Maximum iterations:* The maximum Picard iterations performed. If a solution is not reached within the number of iterations shown, increase the number.

6) *Stopping Criterion:* The program will judge that it has found a solution when successive iterations produce results that differ by less than the convergence limit.

7) *Random start:* The initial assignment of memberships to groups is performed randomly by the program.

8) *Alpha:* Parameter alpha (a) determines the relative number of extragrades to intragrades. Setting a = 1 provide the fuzzy k-means (without extragrades), while a = 0 will result in classifying all the data into extragrades. If information on the value of a is known, a can be fixed to the value, otherwise iterations is performed to search for an optimal value, which is set to 1/(no_class+1).

9) *Alpha iterations:* The extragrade algorithm requires that the average membership of the outlier (or extragrades) class be the same as the average membership of the other classes. The parameter that determines outlier group membership is called alpha (a). Since the exact relationship between alpha and the outlier group membership is not known a priori, the algorithm must arrive at the correct value for alpha using an

iterative procedure. The outer loop of the procedure serves to refine an initial estimate for alpha. With alpha fixed, the inner loop solves for the memberships in the k intragrade groups and single extragrade group. This solution is used to determine the average membership in the outlier group. If the average membership in the outlier group equals (to within a desired accuracy=difference tolerance) the average membership in all groups, the algorithm terminates. Otherwise, a new value for alpha is estimated (using earlier alpha values as a guide). The parameter alpha iterations is the maximum number of iterative alterations to alpha that the algorithm will attempt. User can see by looking at the iteration data output whether the algorithm is converging to a solution, or whether a larger number of iterations is required.

10) *Difference tolerance:* The convergence criterion for the alpha iterations.

4.4 PSO based Clustering

A single particle represents a K-cluster centroids. That is, each particle xi is constructed as xi = $(m_{i,1},\ldots m_{i,k},\ldots m_{i,K})$ where $m_{i,k}$ refers to the kth cluster centroid vector of ith particle. Therefore, a swarm represents a number of candidate data clusterings. The quality of each particle is measured using an objective function [13].

There is a matrix representing the assignment of patterns to the cluster of particle i. Each element $z_{i,k,p}$ indicates if pattern z_p belongs to cluster c_k of particle i.

The fitness function has an objective to simultaneously-

- Minimize the intra-distance between pixels and their cluster means [13].
- Maximize the inter-distance between any pair of clusters.[13].

4.5 Gustafson's-Kessel clustering

This algorithm is an extension of the Fuzzy c-means algorithm [12]. The Gustafson-Kennel[12], algorithm uses an adaptive distance norm in order to detect the clusters with the different shapes.

Fuzzy k-means with fuzzy covariance matrices (Gustafson & Kessel, 1978) calculates the Fuzzy covariance matrix for each number of classes and use it as distance norm.

Consider a set of n data points to be clustered, m_j. Number of clusters/classes are known, 2<=c<n. U_{ij} U denotes the membership of m_j in the i-th cluster. U is therefore called partition matrix. GK use Mahalanobis[12] distance –

- Initialize the partition matrix randomly such that U(0) *2 Mfc* *(4)*

Repeat for l=1,2…..

- Calculate the cluster centers.

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m x_k}{\sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m}, \quad 1<=i<=c \qquad (5)$$

- Compute the cluster covariance matrix

$$F_i^{(l)} = \frac{\sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m \left[x_k - v_i^{(l)}\right]\left[x_k - v_i^{(l)}\right]^T}{\sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m} \qquad (6)$$

Where , 1<=i<=c

- Compute the distances

$$D^2_{ikA}(x_k, v_i) = (x_k - v_i^{(l)})^T \{\rho_i det(F_i)\}^{1/n} F_i^{-1}\}(x_k - v_i^{(l)}) \qquad (7)$$

- Update the partition matrix

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c}(D_{ikA}(x_k, v_i) / Dj_k(x_k, v_i))^{2/(m-1)}} \qquad (8)$$

Where , 1<=i<=c, 1<=k<=N

until $|| U^{(l)} - U^{(l-1)}|| < \epsilon$

5. Particle Swarm Optimization

This paper concentrates on a population based optimization technique, a field of Swarm Intelligence, called Particle Swarm Optimization [9]. Particle Swarm Optimization [9] is modeled after the social behavior of flocks of birds.

This algorithm is initialized with a population of random solutions, called particles. Each particle flies through the searching space with a velocity that is dynamically adjusted. These dynamical adjustments are based on the historical behaviors of itself and other particles in the population.

$X_i = (x_{i1},x_{i2},\ldots,x_{iD})$,represents the ith particle, the best solution is $P_i = (p_{i1},p_{i2},\ldots,p_{iD})$, also called p_{besti}. The best solution of all particles is p_{gg} , also called $g_{best,}$ $V_i = (v_{i1},v_{i2},\ldots v_{iD})$ is the velocity of particle i. For every generation, the velocity changes according to the following equation:

$$v_{id} = w*v_{id} + c_1 *rand_1()*(p_{id} - x_{id}) + c_2*rand_2()*(p_{gd}-x_{id}) \qquad (9)$$

$$x_{id} = x_{id} + v_{id} \qquad (10)$$

Where d=1,2,…S, w is the inertia weight, it is a positive linear function of time changing according to generation iteration, often changing from 0.9 to 0.2. Suitable selection of inertia weight provides a balance between global and local exploration and results in fewer iterations. The acceleration constants, c_1 and c_2 represents the weighting positions. $rand_1$ and $rand_2$ are random functions which change between 0 and 1.

6. Algorithm used for Particle Swarm Optimization based Fuzzy C-Means

To evaluate each individual in the flock, the fitness function is defined

$$F(x_i) = \frac{Constant}{Discrete\ Summary\ among\ Clusters} \quad (11)$$

We use a vector $Z=(z_1,z_2,\ldots, z_i,z_c)$ to indicate a cluster center, which is a particle, z_i is the code of i_{th} clustering center. The process of PSO fuzzy c-means clustering algorithm can be described as follows:

1. Given the number of clusters c, fuzzy index m, population size N, learning factor c_1 and c_2, inertia weight w.
2. Initialize N cluster and their coding, create the first generation of particles. Each particle's p_{best} is its current location, g_{best} is the best location of all current particles.
3. Calculate the cluster centers $Z^{(k+1)}$

$$vj,\ Zj_{(k+1)} = \frac{\sum_{j=1}^{n} w_{ij}^{m} x_i}{\sum_{j=1}^{n} w_{ij}^{m}} \quad (12)$$

4. Calculate wij of each cluster center

$$Wj(k) = \frac{1}{\sum_{r=1}^{c} \left[\left(\frac{dij(k)}{dir(k)}\right)^{2/(m-1)}\right]} \quad (13)$$

$$i=1,2,\ldots\ldots n;\ j=1,2,\ldots\ldots c;$$

5. Calculate the fitness value of each particle according to equation (11). If the fitness is better than that of the current particle's best location, then update the best location of the individual particles. If the fitness of all the particles best location is better than that of the global best current location, then update the global location.
6. Update the velocity and location of each particle according to formula (9) and (10), produce the next generation of particle swarm.
7. If the current iteration number reaches the pre-set maximum, stop iteration and find the best solution in the last generation. Otherwise, return step 3.

In PSO-FCM algorithm, there is much randomicity when producing the next generation. It is not easy to fall into local minimum and has a faster rate of convergence.

7. Comparison of PSO-FCM, PSO-Fuzzy K-Means with extragrades, PSO Gustafson's Kessel and PSO based K-means

Results of all these algorithms are compared in terms of time, distance and mean values. Performance of Unsupervised Clustering Technique merged with optimization technique (Particle Swarm optimization) is evaluated. Comparative results are plotted with the help of graphs for better understandability. When time parameter is considered, PSO based Gustafson's Kessel converges within few iterations. Distance considered between two clusters is large in case of Gustafson's kessel, hence it provides well separated clusters. Fuzzy k-means with extragrades recognizes that some objects might not fit well in any of the groups that are formed and places these objects in an additional outlier group, the extragrades. Fuzzy k-means with fuzzy covariance matrices (Gustafson & Kessel, 1979) calculates the Fuzzy covariance matrix for each number of classes and use it as distance norm.

8. PSO Based Remote Image Classification

For remote image classification generally three main processes are there-

1. Selecting training samples for every region in the radar image.
2. Training these samples using PSO, and obtain cluster center of every region.
3. Finally, the classification of remote image with respect to cluster center is obtained.

9. Algorithm for PSO based Remote Image Classification

1. Selecting training samples for every region in the image according to the number of classes.

2. Generate the initial swarm, $X(0) = \{x_1(0), x_2(0), \ldots\ldots x_m(0)\}$, generate the initial velocity. Set $t=0$, $fbest_i = p_{besti}$, $i=1,2,\ldots.m$, $fbest = g_{best}$
3. Calculate the fitness of all the partilces in X(t).
4. For remote image classification problem,the clustering center $C=\{c_1,c_2..c_m\}$ is the mean value of all particles. The fitness of particle x(t) $=\{x_1(t),x_2(t),..x_n(t)\}$, is defined as

$$Fitness\ (x(t)) = \sum_{i=1}^{n} x_i(t) - c_i \quad (14)$$

5. Update position and velocity of all the particles and set X(t+1) to be the resulting swarm.
6. For all the particles in X(t+1), if $fitness(x_i(t))<fbest_i$, then set

$$Fbest_i = fitness(x_i(t)) \quad (15)$$
$$p_i = x_i(t+1)$$

7. If $fitness(x_i(t))<fbest$ then

$$Fbest = fitness(x_i(t)) \quad (16)$$
$$g=x_i(t+1).$$

8. If stop condition is satisfied, export X(t+1) as the output of the algorithm, Stop.
9. Otherwise, t=t+1, go to 3. Here the stop condition is to set as the maximum number of iterations T.
10. Get the optimal clustering center, which is the center of last swarm obtained.
11. Repeat Step 2 to Step 7 for every class, get the optimal clustering center for every class.
12. After training all the regions, calculate the distance between all pixels and Clustering centers, assign every pixel to one class with the least distance.

10. Classification Error matrix

Error matrices compare, on a category by category basis, the relationship between the reference field data (ground truth) and the corresponding results of a classification. These are square matrices having equal number of rows, columns and categories (whose classification accuracy is being assessed). The major diagonal of the error matrix represents the properly classified land use categories. Kappa Coefficient is calculated for the image which turns out to be 70%.

The equation for the Kappa coefficient is given as

$$K = \frac{N \sum_{i=1}^{r} Xii - \sum_{i=1}^{r} Xi{+}X{+}i}{N^2 - \sum_{i=1}^{r} Xi{+}X{+}i} \quad (17)$$

Where

r=number of rows and columns in error matrix

N =total number of observations

Xii=observation in row i and column i

Xi+=marginal total of row i and

X+i =marginal total of column i,

Another formula to find kappa coefficient is

$$K = \frac{po - pe}{1 - pe} \quad (18)$$

where

po = accuracy of the observed agreement, $\frac{\sum Xii}{N}$ (19)

pe = Estimate of chance agreement, $\frac{\sum^{N} Xi{+}X{+}i}{N^2}$ (20)

8. Overall accuracy

Both producer accuracy and users accuracy is computed for each land cover in an image.

8.1 Producers accuracy- The producer's accuracy is derived by dividing the number of correct pixels in one class divided by the total number of pixels as derived from reference data (column total in confusion matrix Table). The producer's accuracy measures how well a certain area has classified. It includes the error of omission which refers to the proportion of observed features on the ground that is not classifed in the marp. The more errors of omission exist, the lower the producer's accuracy.

Producer's accuracy (%) = 100% - error of omission(%) (21)

8.2 Users accuracy- The user's accuracy is defined by the measure of the correct classified pixels in a class divided by the total number of pixels that were classified in that class (row total in confusion matrix Table). The user's accuracy is, therefore, a measure of the reliability of the map. It informs the user how well the map represents what is really on the ground. It includes the error of commssion, which refers to the proportion of the predicted feature of the classification on the map that are not observed on the ground. The more errors of commission exist, the lower the user's accuracy.

User's accuracy(%) = 100% - error of commission(%) *(22)*

9. Training Sets

A training set is a portion of a data set used to fit (train) a model for prediction or classification of values that are known in the training set, but unknown in other (future) data. The training set is used in conjunction with validation and/or test sets that are used to evaluate different models.

Training sets are used in supervised learning procedures in data mining (i.e. classification of records, or prediction of target values that are continuous.)

The expert has the attribute set *P* from the Indian Remote Sensing (IRS-P6) satellite optical band image set of 1...4 *i.e.* Red (R), green (G), near-infrared (N) and middle-infrared (M) bands. The ground resolution of these images from LISS-III, sensor is 23.5m.

The land use/land cover classification (y) with independent attributes set x consists of two sets of radar sat microwave images radarsat-1(r1) and radarsat-2(r2) and digital elevation model (d) data.

The training data set consisting of 1420 locations on the image is collected taking all the 7-band images together in order to avoid any ambiguity for confirmed training sites. The knowledge resident with the expert is assumed to be the one obtained from the training set, duly verified from ground checks and is confined to the r, g, n, m bands. This data set can be represented in a tabular form similar to that of a relational database tables. Rows of the table represent the training pixels and the digital values in the columns related to the 7-bands viz., r, g, n and m. The table has, therefore, 4 attributes (r,g,n and m), termed as attributes set p.

IMG 1	RED BAND	LISS-III/IRS-P6
IMG 2	GREEN BAND	LISS-III/IRS-P6
IMG 3	NEAR INFRARED	LISS-III/IRS-P6
IMG 4	MIDDLE INFRARED	LISS-III/IRS-P6
IMG 5	LOWINCI,S1,20^O - 27^O	RADARSET-1
IMG 6	HIGHINC,S7,45^O-49^O	RADARSET-2
IMG 7	DIGITAL ELEVATION MODE	(DEM) RES:25

10. Results

From Comparative results, it is clear that PSO based Gustafson's-Kessel, performs better than PSO based fuzzy C-means and finally PSO based Fuzzy K-means with extragrades. Gustafson's-Kessel with Fuzzy K-means is an improvement over Fuzzy K-means and hence it converges very fast with well seperable clusters. Entire Code is written in Matlab 7.5.

Results of all are compared in terms of time, distance, mean values. Performance of Unsupervised Clustering Technique merged with optimization technique (Particle Swarm optimization) is evaluated. Comparative results are plotted with the help of graphs for better understandability.

Time Calculated for each to converge

PSO based Fuzzy C-Means = 2.988

PSO based Gustafson's-Kessel = 2.9325

PSO based K-Means = 23.1038

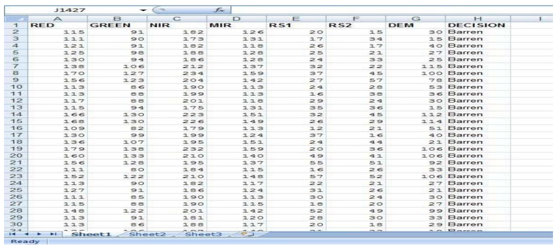
PSO Fuzzy K-Means with extragrades= 179.2178

This PSO based Clustering, has been extended to Remote Image Classifcation. Accuracy % of the image is computed.

As per my results, I found that my algorithm was very well efficient in classifying rocky region of the remote image besides other land covers.

11. Figures and Tables

Table 1: Remote Training Set



RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
115	91	182	126	20	15	30	Barren
111	90	173	131	17	34	15	Barren
121	91	182	118	26	17	40	Barren
125	98	188	128	25	21	27	Barren
130	94	186	128	24	33	25	Barren
138	106	212	137	32	22	115	Barren
170	127	234	159	37	45	100	Barren
156	123	204	142	27	57	78	Barren
113	86	190	113	24	28	53	Barren
113	88	199	113	16	38	36	Barren
117	88	201	118	29	24	30	Barren
115	94	175	131	35	36	15	Barren
166	130	223	151	32	45	112	Barren
168	130	226	149	26	29	114	Barren
109	82	179	113	12	21	51	Barren
130	99	199	124	37	16	40	Barren
136	107	195	151	24	44	21	Barren
179	138	232	159	20	36	106	Barren
160	133	210	140	49	41	106	Barren
156	128	195	137	55	51	92	Barren
111	80	184	115	16	26	33	Barren
152	122	210	148	57	52	106	Barren
113	90	182	117	22	21	27	Barren
127	91	186	124	31	26	21	Barren
111	85	190	113	30	24	30	Barren
115	88	190	115	18	20	27	Barren
148	122	201	142	52	49	99	Barren
113	91	181	120	28	30	33	Barren
113	86	188	117	20	18	29	Barren

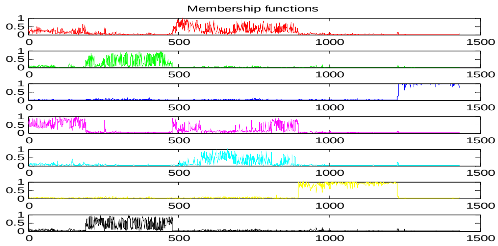


Figure.1 Plot Of membership Function Of PSO-Fuzzy C- Means

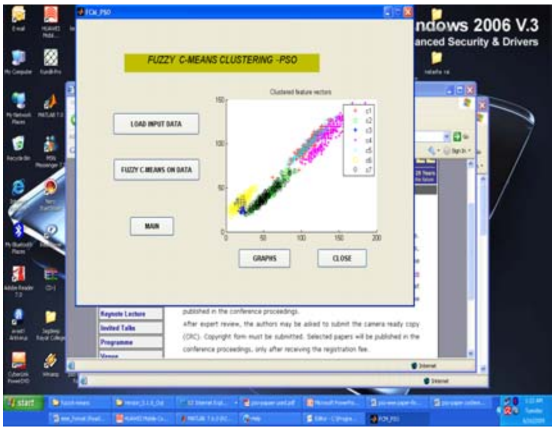


Figure.2 PSO-Fuzzy C-Means

Figure.3 Main Menu for PSO Based Clustering

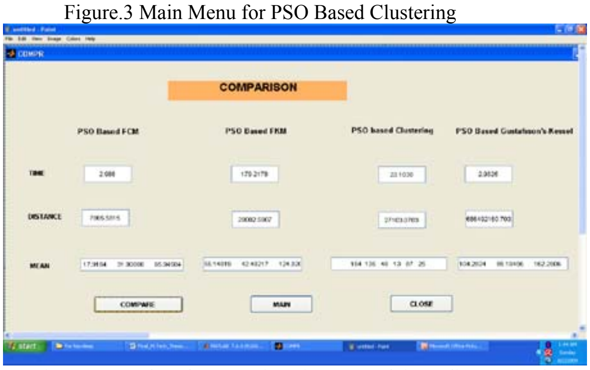


Figure.4 Results of Comparison in terms of time, distance and mean

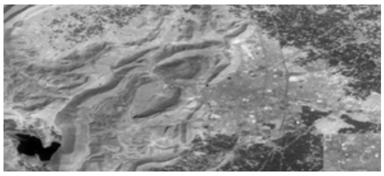


Figure.5 Original Image in NIR Region

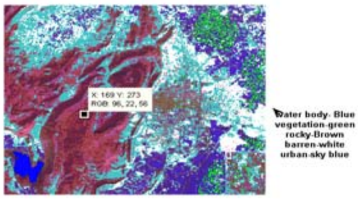


Figure. 6 Image after Classification

Table 2: Error matrix

	Vegetation	Urban	Rocky	Water	Barren	Total
Vegetation	79	0	0	47	0	126
Urban	4	22	0	0	7	33
Rocky	18	3	286	0	0	307
Water	28	0	0	59	0	87
Barren	1	7	0	0	24	32
Total	130	32	286	106	31	585

Kappa Coefficient :0.70334

Table 3 : Producer's Accuracy and User Accuracy of Various landcovers

a)	Producer's Accuracy for vegetation = 60% User Accuracy For vegetation =62.6%
b)	Producer's Accuracy for Urban = 68% User Accuracy For Urban = 66%
c)	Producer's Accuracy for rocky = 100% User Accuracy For rocky =93%
d)	Producer's Accuracy for Water = 55% User Accuracy For Water = 67.8%
e)	Producer's Accuracy for Barren = 77% User Accuracy For Barren = 75%

12. Conclusions

In this paper, I have investigated parameters to compare the results obtained from various particle swarm optimization based clustering techniques. All these techniques involve unsupervised classification. Finally, results are extended to remote image classification and accuracy of various land covers is computed.

Acknowledgments

I thanks, Mr. V.K.Panchal and all his team members for their support and guidance. I equally owe, my thanks to Mr. Harish Kundra and all his team members who have been a consistent source of inspiration and encouragement. I fail with my words, if I fail to thanks my parents, my family, god and my colleagues who made it possible.

References

[1] Y Shi, RC Eberhart, "A Modified Particle Swarm Optimizer", *Proceedings of IEEE International Conference on Evolutionary Computation*, World Congress on Computational Intelligence, Anchorage, Alaska, 1998.
[2] Y Shi, RC Eberhart, "Parameter Selection in Particle Swarm Optimization", *Evolutionary Programming,* Vol VII, Proceedings of Evolutionary Programming, pp 591-600, 1998.
[3] PN Suganthan, "Particle Swarm Optimizer with Neighborhood Optimizer", *Proceedings of the Congress on Evolutionary Computation,* pp 1958-1962, 1999.
[4] F Vanden Bergh, "An Analysis of Particle Swarm Optimizers", PhD Thesis, *University of Pretoria*, South Africa, 2002.
[5] Bradley, P.S., Fayyad, U.M., 1998, "Refining initial points for K-means clustering", *Proc.15th International.Conf.on Machine Learning* (ICML'98).
[6] Cheung, YM., 2003. "k*-Means: A new generalized k means clustering algorithm", *Pattern Recognition Lett.*
[7] Khan, S.S., Ahmad, A., 2004. Cluster center initialization algorithm for K-means Clustering. Pattern Recognition Lett.
[8] Jain, A.K, Murty, M.N, Flynn, P.J.: Data Clustering: A Review.
[9] By Xiangyang Wang, Jie Yang, Xiaolong teng, Weijun Xia, Richard Jensen , Paper on "Feature Selection based on Rough Sets and Particle Swarm Optimization" .
[10] Lillesand, Thomas M. and Ralph W.Kiefer, "Remote Sensing and Image Interpretation", John Wiley and Sons, Inc, 1979.
[11] Ajith Abraham,Crina Grosan Vitorino Ramos (Eds.) E- book "Swarm intelligence in Data mining" , (April,2006).
[12] Nurulhuda Zainuddin and Ali Selamat, "Image Information Retrieval Using Gustafson-Kessel Relevance Feedback", the 2nd Regional conference on Artificial Life and Robotics 14-15 July, 2006, Songkhla, Thailand.
[13] M Omran, A Salman, AP Engelbrecth , "Image Classification using Particle Swarm Optimization".
[14] A Clustering, "A tutorial on K-Means Clustering", 2008.
[15] A Clustering, "A tutorial on Fuzzy c-Means Clustering", 2008.

Presently working as Scientist 'G' in DRDO. He completed his M.S. From BITS,Pilani.Ph.D. in 'soft computing applications in Remote Sensing' from Jawaharlal Nehru University, Delhi.Several publications in National & International journals. Chaired sessions in different National/International Conferences.

Er. Harish Kundra is an Assistant Professor and Head of Department in the Computer Sc. & IT Department of Rayat Institute of Engineering & IT, Railmajra, Punjab, India. I graduated from Baba Banda Singh Bahadur Engineering College, Fatehgarhsahib in Computer Science & Engineering. in 2000 and received Master's Degree in Computer Science and Engineering from Vinayka Mission Research Foundation (Deemed University) , Tamil Nadu in 2007 . I have presented 15 papers in International /National Journals and 50 papers in International/National Conferences. I am a lifetime member of Indian Society of Technical Education. I guided more than 100 projects at graduate and master's level and also 4 dissertations at Master'slevel.

Er. Jagdeep Kaur is Lecturer in the Computer Sc. & IT Department of Rayat Institute of Engineering & IT, Railmajra, Punjab, India. I graduated from Gulbarga University, Karnatka in Computer Science & Engineering in 1997 and received Master's Degree in Computer Science and Engineering from Rayat Institute of Engineering & IT, Railmajra, Punjab in 2009. I am a lifetime member of Indian Society of Technical Education.