ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

A COMPARATIVE STUDY ON OPTIMIZATION TECHNIQUES FOR CLASSIFYING REMOTE SENSING IMAGES

D.Napoleon[1], M.Praneesh[2]

Assistant Professor, Department of Computer Science, Bharathiar University[1]

Assistant Professor, Department of Computer Science, Sankara College of Science and Commerce[2]

Abstract: Classification of land cover types in remotely sensed images is one of the major applications in remote sensing. This paper presents a framework for classifying the land cover information by applying Computer based optimization techniques. The proposed system was implemented and the results were obtained on different Remote sensing images. The proposed algorithm has very good efficiency and high accuracy than conventional methods.

Keywords: classification; Remote sensing images.

I. INTRODUCTION

An image may be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y, and the amplitude values of f are all finite, discrete quantities, we call the image a digital image. The field of digital image processing refers to processing digital images by means of a digital computer. Note that a digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements, pixels. Pixel is the term most widely used to denote the elements of a digital image. In remote sensing, classification is one of the processes for grouping the homogeneous pixels into meaningful categories. The classification algorithms are based on the assumption that the image to be processed contains one or more features such as spectral regions in the case of remote sensing images. In this paper, classification of satellite images is an essential process to identify different land classes. Different land classes have different properties based on which they may be identified and classified [1]. The rest of the paper is organized as follows. Section-II describes the frame work of proposed system. Section-III illustrates the experimental results obtained on remote sensing images and also analyzes the performance measures of the proposed frame work. Finally section-IV draws the conclusion of this paper.

II. METHODOLGY

The proposed architecture is designed for the classification of remote sensing images which is based on land cover information [5]. In this work we have carried out the dataset with a portion of Landsat-7 ETM + multispectral image (bands 1, 2,3,4,5 and 7) acquired over the west of Haerbin, Heilongjiang, china, on august 11, 2001. This proposed work has been represented systematically (figure-1)

Noise Removal

An adaptive noise removal filtering using the wiener filter has been applied for noise removal of images. The wiener filter can be considered as one of the most fundamental noise reduction approaches and widely based for the solution of image restoration problems. In our system we use 3x3 neighbourhoods of filtering size[2].

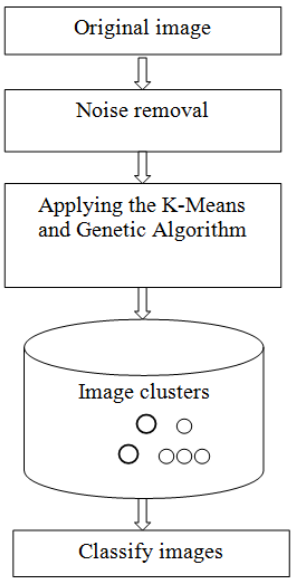

Figure 1. System Architecture

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

Noise reduction techniques are procedures that are used to remove the unwanted signals from an image. Noise reduction techniques are conceptually very similar regardless of the signal being processed, however a priori knowledge of the characteristics of an expected signal can mean that the implementations of these techniques vary greatly depending on the type of signal [3]. All denoising methods depend on a filtering parameter 'h'. This parameter measures the degree of filtering applied to the image [4]. For most methods, the parameter 'h' depends on an estimation of the noise variance σ^2.[10] The result of a denoising method D_h can be defined as a decomposition of any image 'v' as given in below Equation[5]

$w = D_h v + n (D_h, v)$

Where $D_h v$ is smoother than v. $n (D_h, v)$ is the noise guessed by the method.

B. K-Means clustering

K-means clustering is a partitioning based clustering technique of classifying/grouping items into k groups (where k is user specified number of clusters). The grouping is done by minimizing the sum of squared distances (Euclidean distances) between items and the corresponding centroid. A centroid (also called mean vector) is "the center of mass of a geometric object of uniform density". Although K-means is simple and can be used for a wide variety of data types, it is quite sensitive to initial positions of cluster centers. There are two simple approaches to cluster center initialization i.e. either to select the initial values randomly, or to choose the first k samples of the data points. As an alternative, different sets of initial values are chosen (out of the data points) and the set, which is closest to optimal, is chosen. Also, the computational complexity of original K-means algorithm is very high, especially for large data sets.

.

Pseudocode for K-Means Algorithm

We take k number of clusters and n number of samples in an multi dimensional space. Here, k number of iterations are followed for k centroids to obtain optimal clusters, at each iteration, solution is constructed. And finally, found best optimal solution from iteration and average time is calculated at end. Below mentioned rule is followed for algorithm construction.

Step 1: Initialize all k number of clusters and n number of samples.
Step 2: Iteration(I) <=k
Step 3: Randomly select one centroid c_j, where j=1<j<k.
Step 4: Method1<=n
Step 5: Randomly choose one object o_i, where i=1<i<n.
Step 6: Object i on cluster j represent as o_{ij}.
Step 7: Result of o_{ij} is either 0 or 1
If o_{ij}=1 means, i belongs to cluster j.
If o_{ij}=0 means, i belongs to some other clusters.
Step 8: Calculate mean algorithm by
$K_j=1/n_i(o_{ij})*c_{ij}$
Step 9: Repeat Step4 to Step8 until method reaches in n sample.
Step 10: Recalculate the position of centriod by
$C_j=1/n_i(K_j*c_{ij})$.
Step 11: Repeat Step 2 to Step 10, upto centroids.
Step 12: Find the optimal clusters from the result of iteration by
Min $j= o_{ij}|K_j - C_j|^2$.

C. Genetic Algorithm

Genetic Algorithms (GA), first proposed by John Holland in the 1960s, are a category of EC that use concepts derived from evolution. Proper application of a GA finds a balance between exploration and exploitation of a given optimization problem's search space. First, a population of chromosomes is created an initialized. These chromosomes each contain a collection of genes and each gene has a value (called an allele). A single chromosome is an encoded version of a solution to the problem that the GA is attempting to optimize. The GA performs exploration or exploitation of the problems search space by evolving the population of chromosomes through a series of generations. During each generation of the GA, parent chromosomes are selected from the population. These parent chromosomes are combined to form children chromosomes and then the child chromosomes are mutated. In a generational type GA, an entirely new population for each generation is formed by creating multiple child chromosomes. For a steady state GA, the child chromosomes are used to replace members of the current population but a new population is not formed during each generation.

A very important step in the GA is the selection of parents for the next generation of chromosomes. In order to provide a guided search, which is appropriate for the given optimization problem, the selection of parents needs to be based on the quality of the solution that their chromosomes represent. A property called fitness is used to quantify the quality of a given solution and a fitness function is used to calculate the fitness value of each chromosome in a given population before parent selection is made. A variety of different selection methods are used by GA but they all use the principle that higher fit chromosomes are more likely to be chosen as parents. This fitness selection provides the GA direction for the search of an optimization problems search space.GA has been successfully implemented for various clustering problems using different chromosome encoding schemes and fitness functions. A GA performs clustering on an input set of data objects so that supervised learning can be applied to predict class labels in the second step. The input

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

for the GA is a set of data objects that have both numeric and label attributes and a desired number of clusters. The goal of the GA is to produce clusters of data objects that minimize cluster dispersion and are as pure as possible in relation to the label attributes. The GA uses a two component fitness function where the first component measures within cluster variance using a distance metric and the second component measures the similarity of the labeled attributes of the data objects. A very large input data set can be preprocessed to make a representative set that can be used by the algorithm for better time and space efficiency. In GA implement two alternate preprocessing methods for clustering algorithm such as

- The first Preprocessing method used random sampling to obtain a data set with fewer points. This reduced data set was then used in evaluating the fitness of the chromosomes.
- The second preprocessing method used summarization of the input data set and is based on the work presented in reference [1].

For this method, a grid is first constructed and then the input data set is applied to this grid. A single point location and corresponding weight is calculated for each region defined by the grid. The location of the representative point is chosen as the mean value of all the points in the region and the weight of the representative point is equal to the number of points that it replaces.

Pseudocode for Genetic Algorithm

Here, we take n number of chromosomes for n number of centroids,m number of Samples S_j and S_j^i is j^{th} and input attribute as I_a. First chromosome is consider as a parent chromosome and in each iteration it build child chromosomes. Below are GA rule to find cluster construction.

Step 1: Initialize n number of chromosomes to n number of centroids.
Step 2: Iteration(I)<=MaxIter
Step 3: Parent Chromosome P_i is selected randomly.
Step4: Method1<=n
Step 5: Randomly select one sample in multi dimensional space.
Step 6: Sample j is assigned to chromosome i by

$$S_j^i = 1/n_i\{\log_a(L_a)*\log_a(I_a)\}.$$

Step 7: S_j^i value is either 0 or 1
IfS_j^i =1, Sample is matched to i^{th} chromosomes.
If S_j^i = 0, Sample j belongs to some other chromosomes.
Step 8: If Chromosomes matched then fitness value of sample j to input attribute is calculated by

$$F_j(I_a) = 1/S_j^i(I_a/n_i).$$

Step 9: Fitness value of sample j for Label attribute is

$$F_j(L_a) = 1/S_j^i(L_a/n_i).$$

Step 10: Step 4 to Step 9 is repeated until Method1 reaches n samples in search space.
Step 11: Fitness value for chromosome i is calculated by

$$F_j^i = 1/n_i\{F_j(I_a)*F_j(L_a)\}.$$

Step 12: Repeat step 2 to step 10, until reach MaxIter.
Step 13: Finally, find minimum fitness value from the solution of each iteration by using frequency as

$$F=\text{Min}\{1/n(f|F_j i-S_j^i|^2,0\}.$$

Initial steps are similar K-means algorithm but Genetic Algorithm is more efficient because final solution is evaluated by frequency of fitness value and samples matching value.

III. EXPERIMENTAL RESULTS

In order to illustrate the effectiveness of the proposed approach, multi spectral images are considered to test the performance. In this work we have compared K-Means algorithm with Genetic Algorithm

A. Landsat-7 ETM+ Dataset

The First dataset is a portion of a Landsat-7 ETM + multispectral image (bands 1, 2,3,4,5 and 7) acquired over the west of Haerbin, Heilongjiang, china, on august 11, 2001. This site mainly contains two land-cover types, which are vegetation and Expose land. For this dataset the three band image classified based on spectral characteristics. Green areas represent vegetation. Dark green areas represent dry land. Slighter darker green areas on the image usually represent forest land, brighter areas represent grass land, and deep darker green areas represent paddy field. The classification results shows the following diagrams based on the different cluster index values[8]. The clustering results of K-Means and Genetic Algorithm are evaluated using overall accuracy, Kappa-value, average of producer's accuracy, average user accuracy Among them, average of the producer's accuracy, average of user accuracy, overall accuracy and kappa value are widely used in the validation of the land use/ land cover Classification. The following chart represents execution time of existing and proposed algorithm The error matrices obtained from all the considered methods are shown in table-1. for more detailed verification of the results, we assess the accuracy of the each method we calculate the producer and user accuracy.

B. Kappa Analysis

Kappa coefficients are widely used as Classification accuracy assessment for remote sensed data [5]. The result of performing kappa analysis is a KHAT statistic (an estimate of Kappa), which is computed as[10]

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

Table I. Error Matrices of the Classification for Genetic Algorithm (Landsat-7 ETM+ DATA)

CLASS	PADDY FIELD	FOREST	GRASS LAND	DRY SALT FLATS	DRY LAND	ROW TOTAL
PADDY FIELD	**116**	4	7	16	20	163
FOREST	8	**41**	9	34	3	95
GRASS LAND	12	19	**99**	13	31	174
DRY SALT FLATS	0	2	5	**73**	0	80
DRY LAND	5	7	25	0	**131**	168
COLUMN TOTAL	131	73	155	126	185	**460**
PRODUCER'S ACCURACY (%)	80.92	56.16	63.87	57.94	70.81	
USER'S ACCURACY (%)	67.09	43.16	56.90	91.25	77.98	

$$\hat{K} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+}.x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+}.x_{+i})}$$

Where r is the number of rows in the error matrix (also called as confusion matrix), x_{ii} is the number of observations in row i and column$i$, x_{i+} and x_{+i} are the marginal totals of row i and column$i$, respectively, and N is the total number of observations. The overall accuracy and kappa analysis results are tabled (Table-2)

Table II. Accuracy

parameter	K-Means	GA
execution numbers	55	5
execution time	70	54
overall accuracy	65.16	69.19
kappa value	0.584	0.631

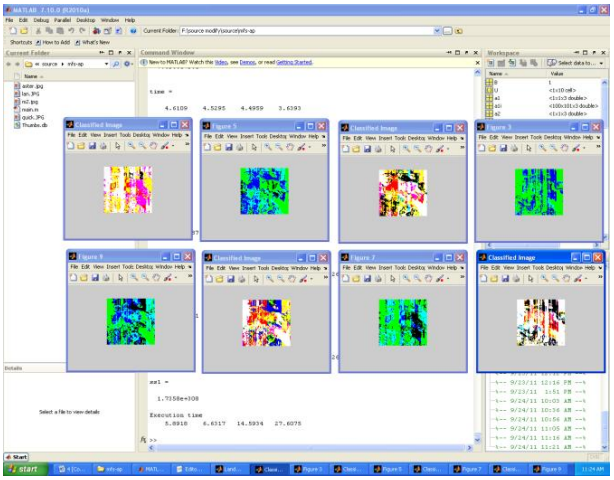

Fig-2 Results of Proposed Algorithm

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 9, September 2013

IV. CONCLUSION

In this work a fast and efficient remote sensing classification system based on land cover information was proposed using Genetic Algorithm. This work illustrates and concluded that the system performance for land cover classification in GA is better than K-Means Clustering algorithm. The classified images can also be compared with ground truth information physically.

V. REFERENCES

[1] Zhenjie Zhang, Bing Tian Dai, Anthony K.H. Tung "On the Lower Bound of Local Optimums in K-Means Algorithm" volume 3 pg-no 323-328,2006.

[2] Ratika Pradhan, M.K Ghose , A.Jeyaram,"Land cover classification of remotely sensed satellite data baysiean and hybrid classifier" International journal of computer Applications,vol-7 no-11.october 2010350.

[3] R. Nicole, "Title of paper with only first word W.Perrizo, Q.Dung, Q.Dung, and A.Roy, "On mining Satellite and other remotely Sensed Images", Proc.SIGMOD Workshop on Research Issues in Data mining and knowledge Discovery. Page 33-40, May 2001

[4] Venkatesh Katari,Suresh Chandra Satapathy, JVR Murthy,PVGD Prasad Reddy, "Hybridized Improved Genetic Algorithm with Variable Length Chromosome for Image Clustering", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.11, November 2007

[5] Ya-Wei Ho; Chih-Hung Wu; Chih-Chin Lai*," Aerial image clustering using genetic algorithm*, "IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2009.

[6] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "*An efficient k-means clustering algorithm: analysis and implementation*," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002, pp. 881-892.

[7] L. Abul, R. Alhajj, F. Polat and K. Barker "*Cluster Validity Analysis Using Sub sampling*," in proceedings

of IEEE International Conference on Systems, Man, and Cybernetics, Washington DC, Oct. 2003 Volume 2: pp. 1435-1440.

[8] E.P Nobi, R.uma maheswari, " Land use and land cover assessment along pondicherry and its suuroundings using indian remote sensing sateliite andGIS., American-Eurasian Journal of Scientific Research 4(2) : 54-58,2009.

[9] L.Yu, L.Jonathan, D.Haibiri and G.Xiangqian, "A Fuzzy relation based algorithm for segmenting colour aerial images of urban environment" proc. ISPRS Technical Commission II Midterm Symposium on intergrated system for spatial data production, custodian and decision support, Xi'an PR,China,pp 271-274,2002.