International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Image Clustering and Classification Using Modified ABC Algorithm

S. Praveena

E.C.E, MGIT, Hyderabad, India

Abstract: ***This paper presents a hybrid clustering algorithm and feed-forward neural network classifier for regions of trees, shades, buildings and roads in a given image. It starts with the single step preprocessing procedure to make the image suitable for clustering. The pre-processed image is clustered using the hybrid genetic-Artificial Bee Colony (ABC) algorithm that is developed by hybridizing the ABC with FCM to obtain the effective clustering satellite image and classified using neural network. After applying Modified ABC method it is found that classification accuracy is improved for all regions.***

Keywords: Clustering Algorithm;Neural Network; Feature Extraction; Satellite Image Classification

1. Introduction

Image segmentation is a critical step of image analysis. The task of image segmentation can be stated as the clustering of a digital image into multiple meaningful non-overlapping regions with homogenous characteristics according to some discontinuity or similarity features like intensity, color or texture [1,2].

Generally, fuzzy based image segmentation has increased in popularity because of rapid extension of fuzzy set theory, and the development of various fuzzy set based mathematical modeling. In the earlier approach many authors proposed to segment a satellite image based on fuzzy clustering and its extensions. Fuzzy can be applied to segment an image with edge detection and also with clustering. This process is inappropriate for real world dataset in which there are no clear boundaries between the clusters. Since the inception of the fuzzy set theory thanks to Zadeh' work [3], researchers incorporate the concept of fuzzy within clustering techniques o handle the data uncertainty problem. The goal of unsupervised fuzzy clustering is to assign each data point to all different clusters with some degrees of membership.

The iterative unsupervised Fuzzy C-Means (FCM) algorithm is the most widely used clustering algorithm for image segmentation [4]. Its success is mainly attributed to the introduction of fuzziness about the pixels'membership to clusters in a way that postpones decision making about hard pixels' membership to latter.

Satellite image categorization field is quiet a challenging job. Recently, researchers have used different types of classification methods for enhancement of efficiency. Satellite image classification, an upgraded biologically motivated theory was applied by Lavika Goel [5].

A hybrid Biogeography Based Optimization (BBO) based algorithm, which is an excellent land cover classifier for satellite image has been introduced by Navdeep Kaur Johal *et al.* [6]. Parminder Singh *et al.* have presented a FPAB/BFO based algorithm for the categorization of satellite image [7]. They intend to utilize the technique of Bacterial Foraging Optimization in order to categorize the satellite image.

M. Ganesh and V. Palanisamy have approached a method known as multiple-kernel fuzzy clustering (MKFCM) for satellite image segmentation [9].

This paper proposes a Modified ABC clustering algorithm and neural network classifier for satellite image classification.
In this work:

ABC algorithm and FCM are combined to improve the classification of images.

This paper is structured as follows: Second section delineates proposed technique, third section discusses analysis and the fourth section is conclusion.

2. Proposed Modified ABC and Neural Network

This section explains the proposed satellite image classification based on Modified ABC algorithm and Feed-Forward Neural Network. The classification would be done based on building, road, shade and tree. The proposed technique is discussed in two phases which are training and testing phases.

A. Training Phase

In training phase different colors of buildings , shades, trees and roads are taken. For each image H (Hue), S(Saturation), L(Lightness), T(Tint), A, B(Color opponent dimensions), U and V(U, V from LUV color space) layers are taken. Features like mean values and maximum values of each layer are calculated. These features are given to neural networks to train.

B. Testing Phase

In this phase, input image is given for clustering. This image is made up of thousands of pixels and to classify this image based on each of this individual pixel is a hectic task and is time consuming. Processing this huge amount of data also results in increase of error rate and the degraded performance of the classifier system. Hence, we cluster the pre-processed image into clusters and then select the centroid of each of these clusters formed for the classification process. This is due to the fact, that each member in a cluster will have almost similar pixel values and differ from the centroid value of the cluster by only a small amount. Hence, this centroid value will represent all the pixels in the clusters. As a result, the classification of a centroid of a cluster will act virtually as classification of all the pixels in the cluster. This result in reducing the number of inputs to the classifier system which further reduces the classifier complexity and also the time incurred. It also results in making the system more efficient and accurate. In this research paper FCM and Modified ABC clustering techniques are used.

Initially the H, T and L layers are extracted from the pre-processed image and the layers are given separately to the ABC-FCM algorithm to cluster it. Thereafter, the clustered layers are merged one another and the feature extraction process is used on each merged clusters and then the extracted feature values of each merged clusters are applied to the trained neural networks to classify the building, road, shade and tree regions of the given input image.

C. Clustering

The clustering process is done based on Modified ABC algorithm. Consider the Modified ABC algorithm is applied on H layer. The process is explained as follows: initially fixed numbers of initial solutions (food sources) are generated randomly by giving lower bound and upper bound. Each solution represents centroids based on the required number of clusters.

After initial solutions are generated, the fitness is calculated for each solution. The calculation of fitness is as follows: initially the centroids in each solution are taken for clustering process and the clustering is done based on the minimum distance. The fitness is then calculated based on the equation given below:

$$fit_i = \sum_{j=1}^{J}\sum_{a=1}^{A}\left\|\left(x_a - C_j\right)\right\|^2$$

In the above equation fit_i denotes the fitness of i^{th} solution, where i=1,2 ---J; and x_a denotes a^{th} pixel x in j^{th} cluster; and j=1,2 ---J ; A is the total number of pixels in j^{th} cluster, where; and Cj denotes the centroid C of j^{th} cluster.

D. Employed Bee Operation

The employed bee then makes modification on the solution in its memory based on the local visual information and then calculates the nectar amount (fitness) of the new solution. If the nectar amount of the new solution is better than the old one, the bee would memorize the new one and forgets the old one. Otherwise it would keep the position of the old one in its memory. The employed bee operation is performed on each solution. To produce a candidate food position from old one in memory, the ABC uses the following expression:

$$S_{ij}^{new} = S_{ij} + \phi_{ij}\left(S_{ij} - S_{kj}\right)$$

In the above equation, $k \in \{1,2,....,I\}$ and $j \in \{1,2,....,J\}$ are randomly chosen index. Though k is determined randomly, it is different from i. i.e. Sij denotes the j^{th} centroid of i^{th} solution S; and S_{kj} denotes the j^{th} centroid of k^{th} solution S. The Φij in the above equation is a random number between (-1,1) and it controls the production of neighbor food sources around S_{ij} and represents the comparison of two food positions visible to a bee. Using the above equation the j^{th} centroid of i^{th} solution S would get altered. After the employed bee operation is performed on each solution, the fitness is calculated for each newly formed solution. If the nectar amount of the newly formed solution is better than the old one, the employed bee would memorize the new one and forgets the old one. The employed bees then share the nectar (fitness) information with the onlooker bees on the dance area.

E. Onlooker Bee Operation

The onlooker bee then evaluates the nectar information taken from all employed bees and chooses a food source with a probability to its nectar amount. The probability value is calculated for each solution and it calculated by the following equation:

$$\Pr_i = \left(\frac{0.25}{\max(fit)}\right) \times fit_i + 0.1$$

In the above equation Pr_i is the probability of i^{th} solution; max(fit) is the maximum fitness value among all the solutions; and fit_i is the fitness value of i^{th} solution. After calculating the probability of i^{th} solution, the onlooker bee would check whether $Pr_i > rand$, where *rand* is a randomly generated number between zero and one. If it so, the onlooker bee would produce a new solution instead of this i^{th} solution. The new solution is formed based on the operation performed by the employed bee i.e. based on S_{ij}^{new} calculation. Then it would calculate the fitness (nectar amount) for the newly generated solution and compare with the old one. If the fitness of the newly formed solution is better than the old one, it would memorize the new one and forgets the old one.

F. Scout Bee Operation

The food source of which the nectar is abandoned by the bees is replaced with a new food source by the scouts i.e. the solutions which are not altered by any one of operations (which are employed bee operation and onlooker bee operation) is replaced by a new solution using scout bees. Consider i^{th} solution is not altered using either of employed bee operation and onlooker bee operation, the scout bee operation is performed on the i^{th} solution as defined below:

$$S_i^j = S_{\min}^j + rand(0,1)\left(S_{\max}^j - S_{\min}^j\right)$$

In the above equation S_{ji} is the j^{th} centroid of i^{th} solution; $S_{\min}^j$ is the minimum j^{th} centroid value among all the solutions; rand(0,1) is the random value between 0 and 1; and $S_{\max}^j$ is the maximum j^{th} centroid value among all the solutions. The scout bee operation is performed only if there has any abandoned solution. ABC operation is repeated until the iteration number set and a solution that has best fitness in the final iteration is taken for the FCM operation.

ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

G. FCM Operation

The best solution obtained from the ABC process is taken for the FCM operation. The FCM is a clustering technique that allows a single pixel to belong to two or more clusters. The degree of membership matrix is formed based on the solution obtained from the ABC process.

Similarly the Modified ABC algorithm is applied on T and L layers separately to cluster the pixels in it.

H. Classification

The feature extraction process is done in the same way as in the training phase on each merged group i.e. the H, S, L, T, A, B, U and V layers are extracted from each merged group to calculate the feature values. The feature values from the feature extraction process of each merged group are then given to the trained neural network to classify the regions. In our training process we use three neural networks to classify the regions. The process of classification is as follows: the feature value of the first merged group is given to the first neural network to check whether it has building region or not, because the first neural network is trained to identify building region. If not, the feature value is then given to the second neural network to check whether it has road region, because the second neural network is trained for road region. If not, the feature value is given to the third neural network to check whether it has shade region, because the third neural network is trained for shade region. If not, the system would take it as tree region. Similarly, the feature value obtained for each merged group is given to the trained neural network to classify the regions.

3. Results and Discussion

This section delineates the results obtained for our proposed technique compared with the existing segmentation techniques. The performances are compared in terms of external metrics and internal metrics. The external metric is accuracy performs the evaluations based on ground truth.

A. Evaluation Metrics

The calculations of metrics are as follows:

$$accuracy = \frac{number of\ true\ positives + number of\ true\ negatives}{number of\ true\ positives + false\ negatives + true\ negatives + false\ positives}$$

B. Performance of proposed technique

The performance of proposed technique is compared with the existing clustering algorithm like FCM in terms of external metrics. The Fig.5(a) shows the satellite images taken for experimentation and the Fig.5(b) shows the classified regions for image.

Figure 5(a) Satellite Image

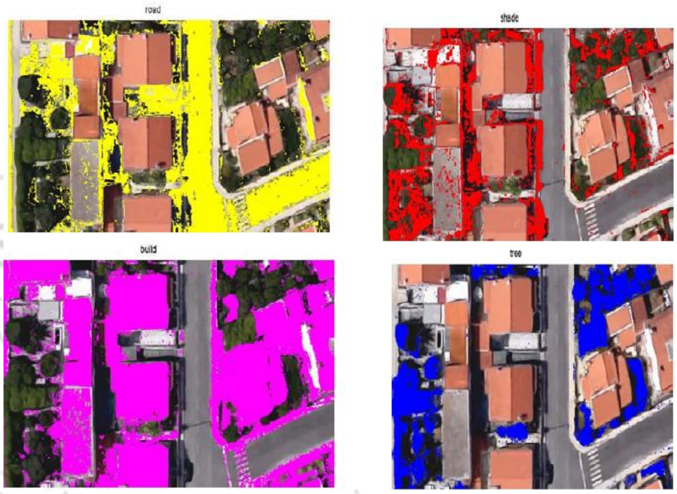

Figure 5(b) Classified regions

C. Performance Based on Accuracy

The Fig.6 shows the accuracy obtained for proposed technique compared to the existing technique using image taken for experimentation.

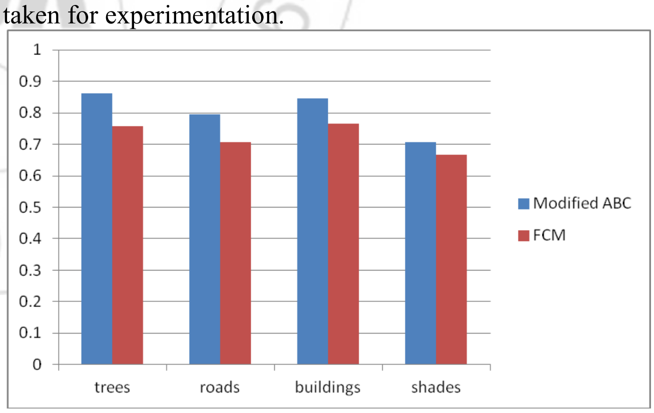

Figure 6(a) Accuracy

In Fig.6(a) the accuracy obtained for the proposed clustering algorithm. It is compared with the existing techniques using taken for experimentation. As shown proposed technique performed better than all the existing algorithms taken for comparison.

4. Conclusion

In this paper, a Modified ABC optimization algorithm for segmentation is proposed with the intention of improving the segmentation in satellite images using feed-forward neural network classifier. Classification accuracy of the proposed algorithm in satellite image classification is calculated and the performance is compared with existing ABC clustering algorithm.

References

[1] H. H Mohammed, "Unsupervised fuzzy clustering and image segmentation using weighted neural networks", CIAP03308-313. 2003
[2] N.R Pal, and S.K Pal," A review on image segmentation techniques", Pattern Recognition,926:1227-1294,1993
[3] L.A. Zadeh, Fuzzy sets, Inform. Control 8 338–353,1965
[4] D.W. Kim, K.H. Lee, and D. Lee, A novel initialization scheme for the fuzzy c-means algorithm for Color clustering Pattern Recognition letters , 25,No.2,pp.227-237.January2004
[5] Lavika Goel,"Land Cover Feature Extraction using Hybrid Swarm Intelligence Techniques - A Remote Sensing Perspective," ACEEE Int. J. on Signal & Image Processing, Vol. 01, No. 03, Dec 2010.
[6] Navdeep Kaur Johal, Samandeep Singh, Harish Kundra, "Hybrid FPAB/BBO Algorithm for Satellite Image Classification," International Journal of Computer Applications, pp.0975- 8887, Vol.6, No.5, September 2010.
[7] Parminder Singh, Navdeep Kaur, Loveleen Kaur,"Satellite Image Classification by Hybridization of FPAB Algorithm and Bacterial Chemotaxis,"International Journal of ComputerTechnology and Electronics Engineering (IJCTEE), Vol.1, No.3, 2011.
[8] O. Gordo, E. Martínez, C. Gonzalo and A. Arquero,"Classification of Satellite Images by means of Fuzzy Rules generated by a Genetic Algorithm,"IEEE Latin America Transactions, Vol. 9, No. 1, March 2011
[9] M. Ganesh, V. Palanisamy,"Multiple-Kernel Fuzzy C-Means Algorithm for Satellite Image Segmentation, "European Journal of Scientific Research, ISSN 1450-216X,Vol.83 No.2, pp.255-263,2012.