IMAGE CLUSTERING AND CLASSIFICATION

Dr. S. Praveena
E.C.E, Mahatma Gandhi Institute of Technology, Hyderabad,India
veenasureshb@gmail.com

Abstract—

This paper presents a hybrid clustering algorithm and feed-forward neural network classifier for region clustering of trees, shade, building, grass and road. The image is clustered using the hybrid - Artificial Bee Colony (ABC-KFCM) algorithm that is developed by hybridizing the ABC and KFCM to obtain the effective clusteirng in satellite image and classified using neural network. The accuracy and MSE of the proposed hybrid algorithm is compared with the algorithm like KFCM.

Keywords—ABC-KFCM; Segmentation Algorithm;Neural Network; Feature Extraction; Satellite Image Classification.

I. 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].

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]. This paper related to a study of their hybrid intelligent classifier along with other current Soft Computing classifier like: 1) Ant Colony Approach, 2) Hybrid Particle Swarm Optimization-cAntMiner (PSO-ACO2), 3) Fuzzy sets, 4) Rough-Fuzzy Tie up; the Semantic Web Based Classifiers and the traditional probabilistic classifiers such as the Minimum Distance to Mean Classifier (MDMC) and the Maximum Likelihood Classifier (MLC). This paper proposes a hybrid clustering algorithm and neural network classifier for satellite image classification.

In this work:
ABC algorithm and KFCM are combined to improve the clustering of images.

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

II. PROPOSED ABC-KFCM AND NEURAL NETWORK

This section explains the proposed satellite image classification based on hybrid ABC-KFCM algorithm and Feed-Forward Neural Network.

A.Training Phase

In training phase different colors of building , road, shade ,grass and trees are taken. For each different layer are extracted. For each layer histogram, maximum value of histogram and mean values are calculated. These features are given to neural network classifier to train the different regions in the image.

B. Testing Phase:

The input satellite image is given for pre-processing using median filtering technique and the output of median filter is used for segmentation. In segmentation process, initially the H, T and L layers are extracted from the pre-processed image and the layers are given separately to the ABC-KFCM 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 satellite image.

C.Clustering

In this process, the input pre-processed satellite image is converted different layers. i.e H, T and L layers are extracted from it. Thereafter, the ABC-KFCM algorithm is applied on each layer (H, T and L) separately to cluster the pixels. Here, the KFCM operator is incorporated in the ABC algorithm to segment the satellite image effectively. The segmentation process is done based on ABC-KFCM algorithm. Consider the ABC-KFCM 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 would contain 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.

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. We can alter two or more centroids based on the above equation and we would get a new solution. From the above formula the perturbation on the position Sij decreases as the difference between the parameters Sij and Skj decreases. Therefore the step length is adaptively reduced as the search approaches to the optimum solution in the search space. 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.

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.

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 KFCM operation.

D. *Classification*

The clustered regions features are calculated and compared with neural network classifier features. The best clustering algorithm is decided by the following parameters.

III. 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. The internal metric is Mean Square Error.

Evaluation Metrics

The metrics used for evaluation are accuracy, MSE. The calculations of metrics are as follows:

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

Mean Square Error

The Mean Square Error (MSE) for a clustering is done by summing the least distance value between the centroid and data of each cluster. The best cluster would have the nearest data of the centroid in it. It is defined as follows:

$$MSE = \sum_{n=1}^{N} \min dis(X_b, C_n) \qquad where,\ b = 1,2,....,T$$

In the above equation N is the total number of clusters, X_b is the b^{th} data in n^{th} cluster, C_n is the centroid of n^{th} cluster and T is the total number of data in the n^{th} cluster.

A .Performance Based on External Metrics

The table 1 shows the accuracy obtained for proposed technique compared to the existing techniques using image taken for experimentation.

In table the specificity obtained for the proposed segmentation algorithm ABC-KFCM is compared with the existing techniques\ using taken for experimentation. As shown proposed technique performed better than existing algorithms,

Accuracy	ABC-KFCM	KFCM
Road	0.845	0.75
Building	0.815	0.83
shade	0.889	0.8106
tree	0.893	0.7507
grass	0.88	0.72

TABLE 1

B.Performance Based on Internal Metrics

This section shows the performance comparisons by means of DB-index for satellite image taken for experimentation. The better performances of these indices are judged based on less value. The table 2 shows MSE, value of proposed technique compared to the existing techniques.

Clustering Technique	MSE
ABC-KFCM	0.04786602
KFCM	0.0546602

TABLE 2

It shows that ABC-KFCM is has very less MSE which indicates it does better clustering.

IV. CONCLUSION

In this paper, a new optimization algorithms for clusteirng is proposed with the intention of improving the segmentation in satellite images using feed-forward neural network classifier.. The overall steps involved in the proposed technique in three steps such as, i) Pre-processing, ii) segmentation using ABC-KFCM algorithm, and iii) classification 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 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.