International Journal of Engineering, Science and Mathematics
Vol. 7 Issue 3, March 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: **http://www.ijesm.co.in**, **Email: ijesmj@gmail.com**
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Reduction of Davies Bouldin index using hybrid clustering algorithm and Naive Bayes Classifier

Dr.S.Praveena
E.C.E , MGIT
Hyderabad,India

Abstract—**This paper presents a hybrid clustering algorithm and Naïve Bayes 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 Kernelized Fuzzy C Means to obtain the effective clustering in satellite image and classified using Naïve Classifier. The Davies Bouldin -index of the proposed hybrid algorithm is compared with the algorithm like FCM.**

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

I. Introduction

Swarm intelligence (SI) is artificial intelligence based on the collective behavior of decentralized, self-organized systems. SI systems are typically made up of a population of simple agents interacting locally with one another and with their environment. In order to have clarity in the satellite images [1] used Particle Swarm Optimization technique to segment it. When incorporated with traditional clustering algorithms, problems such as local optima and sensitivity to initialization are reduced, thus exploring a greater area using global search. This segmented image is further classified using Kappa coefficient.

Tie Qi Chen and Yi Lu [2] developed a fuzzy clustering algorithm that iteratively generates color clusters using a uniquely defined fuzzy membership function and an objective function for clustering optimization. The region segmentation algorithm merges clusters in the image domain based on color similarity and spatial adjacency.

Recently, tremendous works focus on using kernel method, which first maps the data into high dimension space to gain high discriminant capability, and then calculates the measure of the samples in their original data space with kernel function. Kernel fuzzy C-Means (KFCM) is proposed by substituting the Euclidean distance with kernel function. KFCM not only to certain extent overcomes limitation of data intrinsic shape dependence and can correctly clustering, but also overcome sensitivity to initialization and noise data and improve the algorithm robustness. However, like FCM algorithm, KFCM algorithm still exists some drawbacks, such as the sensitivity to initialization and the tendency to get trapped in local minima. Therefore, an improved kernel fuzzy C-Means based on artificial bee colony (ABC-KFCM) is put forward.

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

This section explains the proposed image classification based on hybrid ABC-KFCM algorithm and Naïve Bayes Classifier. The proposed technique is discussed in two phases which are training and testing phases.

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

The input is given for pre-processing using median filtering technique and the output of median filter is used for clustering. In clustering, 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 using Navie Bayes classifer to classify the building, road, shade, grass and tree regions of the given input satellite image.

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:

$$fit_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\left(1-k(x_{j,}o_j)\right)$$

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.

The algorithm of ABC-KFCM is as follows:

Step 1: Initialize the parameters of ABC and KFCM including population size SN, maximum cycle number MCN, limit, clustering number c, m and ε;

Step 2: Initialize the kernel metrics and membership function for KFCM.

Step 3: Generate the initial population (cluster center) cij , and evaluate the fitness function.

Step 4:ABC algorithm

4.1 Set cycle to 1

4.2 Set s to 1

4.3 FOR each employed bee { Produce new solution vij by KFCM ceters.

$$c_i = \frac{\sum_{l=1}^{n} u_{il}^{m} k(x_l, c_i) x_l}{\sum_{l=1}^{n} u_{il}^{m} k(x_l, c_i)}$$

Calculate the value fiti Apply greedy selection process}

4.4 Calculate the probability values pi for the solutions (c_{ij})

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

4.5 FOR each onlooker bee { Select a solution cij depending on pi Produce new solution vij Calculate the value fiti Apply greedy selection process}

4.6 If the searching times surrounding an employed bee s exceeds a certain threshold limit, but still could not find better solutions, then the location vector can be reinitialized randomly according to Eq. go to step 4.2

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

4.7 If the iteration value is larger than the maximum number of the iteration (that is, cycle> MCN, output the best cluster centers. If not, go to Step 4.1.

Step 5: KFCM algorithm

5.1 Update membership matrix

5.2 Update the cluster centers

5.3Compute E i.e difference between old and new membership functions. If it is less than therehold value then , stop; If not, go to Step 5.1.

Solutions.

C. Classification

The clustered regions features are calculated and compared with Naïve Bayes classifier features. The best clustering algorithm is decided by DB index, and Sensitivity.

III. Results and Discussion

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

A.Evaluation Metrics

The metrics used for evaluation are Sensitviy, DB index. The calculations of metrics are as follows:

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

DB Index

The Davies Bouldin (DB) Index is a metric exploited to evaluate the clustering algorithm. The DB-Index is an internal evaluation scheme that validates how well the cluster is done based on the quantities and features inherent to the dataset. The DB-Index calculation is as follows:

$$DBI = \frac{1}{N}\sum_{n=1}^{N} D_{n,n+1}$$

Where, $D_{n,n+1} = \dfrac{d_n + d_{n+1}}{M}$

$$d_n = \frac{1}{T}\sum_{b=1}^{T}\left|X_b - C_n\right|^2$$

$$M = \sum_{n=1}^{N-1} \sum_{f=n+1}^{N} \sqrt{(C_n - C_f)^2}$$

In the above equations DBI denotes the Davies Bouldin (DB) Index, N denotes total number of clusters, $d_{n,n+1}$ denotes clustering scheme measurement between each cluster, dn denotes the value of distance between each data in the n^{th} cluster and centroid of that cluster, dn+1 denotes the value of distance between each data in the next cluster and the centroid of n^{th} cluster, M denotes sum of the Euclidean distance between each centroid, T is the total number of data in the cluster, X is the data in the n^{th} cluster and C_n is the centroid of n^{th} cluster.

A .Performance Based on External Metrics

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

In table 1 the sensitivity 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 the existing algorithms taken for comparison.

TABLE 1

Sensitivity	ABC-KFCM	FCM
Road	0.80	0.75
Building	0.85	0.83
shade	0.88	0.8106
tree	0.81	0.7507
grass	0.84	0.72

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 the DB-index, performances of proposed technique compared to the existing techniques.

TABLE 2

Clustering Technique	DB-index
ABC-KFCM	0.31602
FCM	0.46602

The TABLE2 shows the DB-index comparison using image taken for experimentation. It shows that ABC-KFCM is has very less XB-index 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 Naïve Bayes Classifer.Classification Sensitivity of the proposed algorithm in satellite image classification is calculated and the performance is compared with existing clustering algorithm.

References

[1] Jagdeep Kaur, “Remote Image Classification Using Particle Swarm Optimization”, International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 7, July 2012.

[2] Chen Tie Qi and Lu Yi. Color image segmentation-aninnovative approach. *Pattern recognition*, vol. 35, 2002, pp. 395-405.

[3] Sanghamitra Bandyopadhyay, Sriparna Saha, “A Validity Index Based on Symmetry: Application to Satellite Image Segmentation”, Springer, 2013, pp 125-163.

[4] J. K. Udupa and S. Samarasekera, “Fuzzy connectedness and object definition: theory,algorithm and applicationsinin Segmentation,”*Graph. Models Image Process.*, vol. 58, no. 3, pp. 246–261, 1996.