Crop Stage Classification of Hyperspectral Data using Unsupervised Techniques

J. Senthilnath, S.N. Omkar, V. Mani, Nitin Karnwal and Shreyas P.B

Abstract—The presence of a large number of spectral bands in the hyperspectral images increases the capability to distinguish between various physical structures. However, they suffer from the high dimensionality of the data. Hence, the processing of hyperspectral images is applied in two stages: dimensionality reduction and unsupervised classification techniques. The high dimensionality of the data has been reduced with the help of Principal Component Analysis (PCA). The selected dimensions are classified using Niche Hierarchical Artificial Immune System (NHAIS). The NHAIS combines the splitting method to search for the optimal cluster centers using niching procedure and the merging method is used to group the data points based on major physical structures. Results are presented for two hyperspectral images namely EO-1 Hyperion image and Indian pines image. A performance comparison of this proposed hierarchical clustering algorithm with the earlier three unsupervised algorithms is presented. From the results obtained, we deduce that the NHAIS is efficient.

Index Terms—Niche hierarchical artificial immune system, hyperspectral images, principal component analysis.

I. INTRODUCTION

The remote sensing images require accurate classification for many practical applications, such as precision agriculture, monitoring and management of the environment, and security and defense issues. The advent of high resolution sensors and high speed data processing devices has prompted the use of hyperspectral images for image analysis and classification [1]. They cover a wide range of spectral channels and spatial resolution. Thus, every pixel in a hyperspectral image contains values that correspond to the detailed spectrum of reflected light [2]. This rich spectral information in every spatial location increases the capability to distinguish different physical structures, leading to the potential of a more accurate image classification.

In supervised technique, artificial neural networks have been used for land cover mapping problem and have been reported to perform better when compared with statistical classifiers [3]. Support vector machines (SVMs) have also shown good performances for classifying high-dimensional data when a limited number of training samples are available [4, 5]. In unsupervised technique, the knowledge extracted from the data set is in the form of optimal cluster centers. Different unsupervised techniques like minimum spanning forest [6], watershed method [7] have been used to perform pixel-wise classification of hyperspectral imagery.

Hyperspectral image data has been used for various applications such as mapping of several soil properties [8], crop stage identification [9], etc. However, they suffer from the curse of high dimensionality [10]. Various dimensionality reduction techniques have been used in the past to overcome this problem. The large number of dimensions is reduced to first few principal components (PCs) by the Principal Component Analysis (PCA) technique [11].

One of the applications of hyperspectral images is crop stage classification [9]. It is seen that a change in growth stage of a crop is characterized by changes in leaf pigments such as chlorophyll and carotenoids. These changes in turn affect the reflectance spectrum of the crop for intensities of different wavelengths which is used to extract stage information. This fact has been utilized to attempt crop stage classification using the intensity information of the bands [9, 10]. The problem of crop growth classification has been addressed in [12].

In the literature, a popular cluster splitting and merging method namely Iterative Self-Organizing Data Analysis Technique Clustering Algorithm (ISODATA) [13] is used for grouping the data set. However, the main drawback of ISODATA is that it converges to local minima in case of inappropriate choice of initial cluster centers which may lead to poor classification efficiency [14]. In order to overcome local optima problems, population based methods are used [15]. The Artificial Immune System (AIS) is a relatively new population based partitional clustering method. In [16], AIS is directly applied on hyperspectral data without reducing the dimension and also used to extract exact number of cluster centers for classification. Further, many researchers have shown clustering problems can be analyzed efficiently using hierarchical methods [17, 18]. In [19], Hierarchical AIS (HAIS) is applied by changing the suppression threshold of
AIS to capture multiple cluster centers. In this study we have presented a new approach called Niche HAIS (NHAIS). For a given multi-modal data distribution, the agents are explored and exploited by capturing multiple local optima, this is known as niching technique [20]. Here to generate cluster centers niche is combined with AIS. It can be observed from the literature that this approach to capture multiple local optima has not yet been explored.

In this paper, two stages are applied for processing hyperspectral image. Stage 1 uses dimensionality reduction technique namely PCA in order to reduce the dimension of hyperspectral image. In stage 2, NHAIS is applied on reduced dimension to generate optimal cluster centers and merging the centers to their respective classes using majority voting [1, 21]. The main challenge here is to obtain possible combinations of clusters that can be used to split the data set and efficiently merge the data set to their respective groups. Experimental results are demonstrated on two hyperspectral images acquired by the Hyperion and AVIRIS sensor. The proposed NHAIS is compared with the ISODATA, AIS and HAIS methods. The performance measures used are classification efficiency and kappa coefficient.

II. METHODOLOGY

The flow of this study is as shown in Fig. 1. An input \( d \)-band hyperspectral image using dimensionality reduction method is reduced to \( b \)-band hyperspectral image which is considered as a set of \( n \) pixel vectors \( X = \{x_j \in \mathbb{R}^b, j = 1, 2, \ldots, n\} \). Let \( \Omega = \{c_1, c_2, \ldots, c_k\} \) be a set of classes. Unsupervised classification methods assign each hyperspectral data to one of the \( k \)-classes of interest.

A. Dimensionality Reduction

In hyperspectral image, we often encounter high-dimensional data. Though the data are lying in a high-dimensional space, only a few dimensions are actually important for the analysis. Principal Component Analysis (PCA) [11] is one of the most widely used dimensionality reduction technique. It identifies patterns in the data and express data in such a way as to highlight similarities and differences without too much loss of information.

In literature various methods are available to analyze principal components of hyperspectral images like virtual dimensionality, second moment linear dimensionality, modified broken-stick rule etc. Virtual Dimensionality, because of some undesirable properties, produces unreasonable results in case of hyperspectral images. Second moment linear dimensionality technique avoids the pitfalls of virtual dimensionality and is successful in identifying a certain number of principal components. It locates exceptionally large gaps in eigen values and gives a unique solution if the recommended \( a \)-level is used [22].

To ascertain principal components, different methods require a user-defined threshold [23]. The results will depend upon the user-defined threshold, which in all cases may not be optimum, but Modified Broken-Stick Rule (MBSR) avoids it. In MSBR method, \( k \) is the number of principal components out of total \( p \) dimension and ‘\( \lambda \)’ are eigen values of various dimension. The value of \( k \) is defined such that

\[
\sum_{i=1}^{k} \lambda_i > b_j \quad \text{for} \quad j = 1, \ldots, k \quad \text{and} \quad \lambda_{k+1} \leq b_{k+1},
\]

where

\[
b_j = \frac{1}{p-j+1} \sum_{i=j}^{p} \frac{1}{i}
\]

is a fair share of total variability represented by \( \lambda_i \) within \( \lambda_i \ldots \lambda_p \) [24].

B. Niche Hierarchical Artificial Immune System

Artificial immune system (AIS) algorithm is derived from natural process and principles of vertebrate immune system. It takes its inspiration from the human immune system which protects the body from antigens (Ag) with the help of antibodies (Ab). It usually works in two mechanisms namely: innate and adaptive immunity. Innate immunity is for general antigens whereas adaptive immunity is for unknown invader [25, 26]. Clonal selection theory [27] explains the response of adaptive immune system to antigens. There are many algorithms based on clonal selection theory. CLUNAG [28] generates \( N \) antibodies and at the end of each iteration it clones and hyper-mutates them by selecting best antibodies.

Recent research is in the direction of capturing multiple local optima for a given multi-modal functions. In genetic algorithm (GA) niching is adopted to obtain multiple local optima [20]. Inspired from this technique we have developed a new algorithm by binding niching process and AIS. This is known as niche hierarchical artificial immune system (NHAIS). Niche is created wherever density of antibodies (agents) is more. This will make sure that all antibodies within that niche will converge to optimum value.

Initially antibodies are randomly distributed within the data set. Affinity value between antibody (agent) and antigen (data set) is measured by

\[
factor = \frac{d}{\sqrt{\sum_{i=1}^{n} (\max(a_i) - \min(a_i))^2}}
\]

(1)

\[
p = \frac{\sum p \forall \text{dist}(Ab, a(p))}{p=1}
\]

(2)

where \( d \) is number of dimensions, \( n \) is number of samples, \( a \) is data samples and dist function gives Euclidean distance.

\[
\text{affinity} = e^{-p/\text{factor}}
\]

(3)

High value of affinity indicates that antibody is nearer to antigen. Distances between all antibodies are calculated and antibodies which are within specified radius are grouped to
form one niche. Cloning (duplication) of all antibodies are done. Number of clones to be generated is user specified. Clones are hyper-mutated with specified probability. Hyper-mutation is very useful to search the optimal value within the neighborhood of clone. If the antibody is in any niche then we need to make sure that the hyper-mutated clones do not cross their niche radius.

Best clone is selected for the antibody by considering their affinity value and it is used for next generation. By doing this we search neighborhood of the antibody using clones and move to best position at each generation until it is converged within the niche. If many clones converge to the same point then that point will be considered as the cluster center.

If the dataset is varying with number of samples in each class, AIS fails to generate the appropriate cluster centers. The non-linearity of the data set results in cluster centers that do not belong to the true classes. Also, an exact number of cluster centers are generated which fails for efficient classification. This number needs to be optimal cluster centers for effective processing and classification [29]. NHAIS overcomes these drawbacks. In this study, NHAIS generates optimal cluster centers by splitting the data set. These cluster centers are further merged with the help of majority voting method to generate centers belonging to their true classes.

The implementation of the NHAIS based on artificial immune system principles is as follows:

Step 1. Initialization – randomly distribute antibodies.
Step 2. Find affinity value of each antibodies using Eq. 3.
Step 3. Sort all the antibodies according to their affinity values.
Step 4. Find out all niches and note all the antibodies within that niche.
Step 5. Generate clones for all antibodies.
Step 6. With specified probability (generally very less) hyper-mutate the clones.
Step 7. If the hyper-mutated clone has moved outside the niche region then that clone is abandoned.
Step 8. Select the best clone to replace corresponding antibody.
Step 9. Repeat step 2 to 8 until they are converged which is selected as cluster centers.
Step 10. Merge data points to closest clusters using majority voting method for each data points belonging to the cluster.
Step 11. Clusters are grouped in agglomerative fashion using labels.
Step 12. Assign each data point to one of the class.
Step 13. Calculate the performance measures of each class.

The important parameters in this algorithm are: \( N \) is the number of \( Ab \) selected for cloning and \( R \) is niche radius.

Fig. 2 summarizes the whole process of clustering. The given 2-dimensional data set has two classes. After applying NHAIS on it, NHAIS has split the data set into three cluster centers instead of two (Fig. 2(a)). This is because of the non-linearity in the data set. The data set is now merged according to the distance from the cluster centers. The data points closer to a cluster center are clubbed into one cluster. So there are 3 clusters now (Fig 2(b)). Then, we label a cluster based on the maximum number of data points (i.e. majority voting method) belonging to a class label. To estimate the effectiveness and efficiency of the clustering algorithm, we require class labels for the data points. The cluster is compared against the class label and if it is found to contain more data points belonging to a particular class, it is classified to that class (Fig 2(c)). This process repeats for all the data points. After this, the performance of each class is evaluated.

![Fig. 2](image)

**Fig. 2**. Scheme illustrating the NHAIS algorithm: (a) Cluster split using NHAIS; (b) Assigning data to the closest center; (c) Majority voting for classification.

### III. PERFORMANCE MEASURES

To classify and evaluate the performance based on individual, average and overall classification accuracy for a given data set, we use four unsupervised techniques namely ISODATA, AIS, HAIS and NHAIS. Initially, the data set is used to arrive at the classification matrix which is of size \( n \times n \), where \( n \) is the number of classes [30]. A typical entry \( q_{ij} \) in the classification matrix shows how many samples belonging to class \( i \) have been classified into class \( j \). For a perfect classifier, the classification matrix is diagonal. However due to misclassification we get off-diagonal elements.

The performance measures considered are: individual class efficiency (\( \eta_i \)), average efficiency (\( \eta_a \)), overall efficiency (\( \eta_o \)) and kappa coefficient (\( \kappa \)). These are defined as:

\[
\eta_i = \frac{q_{ii}}{\sum_j q_{ij}} \quad \eta_a = \frac{1}{n_c} \sum_i \eta_i \quad \eta_o = \frac{1}{N} \sum_{i=1}^{n_c} q_{ii}
\]  

![Fig. 3](image)

**Fig. 3**. Illustrating convergence of algorithm: (a) Standard Rastrigin function; (b) Antibody converged to multiple local maxima.

Let us consider a Rastrigin function to illustrate how NHAIS capture all the modes (maxima). Rastrigin function is a standard benchmark optimization problem which is given by [20]:

\[
F(x, y) = 20 + (x^2 - 10 \cos(2\pi x) + y^2 - 10 \cos(2\pi y))
\]  

From Fig. 3(a) we can observe that the distribution of this function for variables \( x \) and \( y \) in the region \([-5 5]\). For this distribution the number of local maxima is 100. Here using NHAIS algorithm (step (1) to step (9)) all the modes (maxima) are captured by setting parameter value \( N \) and \( R \) as 600 and 0.1 respectively as shown in Fig 3(b).
where \( q_{ij} \) is the number of correctly classified samples, \( n_i \) is the number of samples for the class \( i \), and \( N \) is the number of samples in the data set.

The kappa coefficient [1, 31] is a statistical measure of inter-rater agreement for grouping of qualitative items. It is generally considered as a more robust measure to analyze classification matrix.

\[
\kappa = \frac{\text{Observed accuracy} - \varepsilon}{1 - \varepsilon}
\]

\[\varepsilon = \frac{\sum_{i=1}^{N} (\text{classified class } i \text{ total pixels } \times \text{actual class } i \text{ total pixels})}{(\text{total pixels})^2}\]

IV. RESULTS AND DISCUSSIONS

In this section, we present the results obtained for hyperspectral satellite images classification problem. In our study, we consider two hyperspectral satellite images- EO-1 Hyperion satellite image containing 3 crop stages, of a region around Meerut, Uttar Pradesh, India [19] and AVIRIS Indian Pines image [31].

A. Image 1 – EO-1 Hyperion Image

It has 225 bands, with a bandwidth of the order of 10 nm ranging from 300 to 2400 nm. The image resolution is 30 by 30 m. Each image has a width of 7.5 km and length of 100 km. The region of study is around Meerut city in Uttar Pradesh, India. The latitudinal and longitudinal positions of the four corners of the image are 29 15 56.20N 77 38 9.56E, 29 15 49.53N 77 37 41.58E, 29 1 20.77N 77 37 47.23E and 29 1 14.17N 77 38 9.56E.

The data consists of three stages of growth of the wheat crop. It has a total of 352 samples with 225 bands. For this image, ground truth is created using multiple sources like topographical maps, manual analysis and ground verification. The details are given in TABLE I. A plot of wavelength versus change in reflectance for three stages of crop for the complete wavelength range of 300-2400 nm, is as shown in Fig. 4. From the figure, we can observe that reflectance value for a crop in a particular stage increases as the crop grows from stage 1 to stage 3 in the red edge region (band 41 to band 48 in HYPERION image). Similar trend is also visible in the initial portion of the near infrared region (Bands 75-125). While a reverse trend is apparent in the later half (Bands 125-225) of the near infrared region (near infrared region ranges from 750nm to 2500nm).

To reduce the high dimension of the image data, PCA is used. We have employed MBSR method to reduce the number of dimensions. The number of PC obtained by using MBSR method is 6. This is shown in Fig. 5. From this figure, we can observe that the point of intersection for number of principal components (PCs) versus percent of variability gives 6 PCs.

![Fig.4. Effect of reflectance for a selected pixel of three crop stages](Image)

![Fig.5. Number of PCs versus Percent of Variability](Image)
TABLE I. Classification Efficiency for the crop stage data set

<table>
<thead>
<tr>
<th>Class type</th>
<th>Sample size</th>
<th>$\eta$</th>
<th>PCA-ISO (%)</th>
<th>PCA-AIS (%)</th>
<th>PCA-HAIS (%)</th>
<th>PCA-NHAIS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emerging</td>
<td>110</td>
<td>$\eta_1$</td>
<td>73.6</td>
<td>76.4</td>
<td>81.8</td>
<td>86.4</td>
</tr>
<tr>
<td>Mature</td>
<td>109</td>
<td>$\eta_2$</td>
<td>32.1</td>
<td>44.9</td>
<td>73.4</td>
<td>73.4</td>
</tr>
<tr>
<td>Milking</td>
<td>133</td>
<td>$\eta_3$</td>
<td>71.4</td>
<td>65.4</td>
<td>82.7</td>
<td>84.2</td>
</tr>
<tr>
<td>Total</td>
<td>352</td>
<td>$\eta_4$</td>
<td>59.1</td>
<td>62.5</td>
<td>79.3</td>
<td>81.3</td>
</tr>
</tbody>
</table>

The results in terms of classification efficiency and kappa coefficient are compared among ISODATA, AIS, HAIS and NHAIS. From TABLE I, it is apparent that NHAIS gives better classification efficiency than other algorithms. ISODATA and AIS yields lower individual efficiency of crop stage 2. This is due to reflectance of the samples belonging to stage 2 is getting classified into Stage 1 and Stage 3. The reason for better classification efficiency of HAIS and NHAIS in comparison to that of AIS is that AIS is a parametric method and generates only a fixed number of cluster centers while HAIS and NHAIS are hierarchical in nature and generate many cluster centers. These are local optimal cluster centers and are further merged to improve the classification efficiency. However individual efficiency of stage 1 and stage 3 is better in NHAIS which leads to increase in overall efficiency and kappa coefficient in comparison with HAIS.

B. Image 2—AVIRIS Indian Pines

The Indian Pines image was acquired by the AVIRIS sensor in Northern Indiana. The spatial dimension is 145 by 145 pixels, with a spatial resolution of 20 m by 20 m and 220 spectral channels. In [31], 20 water absorption bands have been removed and 200 bands are used. In our study also we use 200-band image. Sixteen classes of interest are considered, with the number of labeled samples for each class is as shown in TABLE II.

Similarly, to reduce the high dimension of this dataset, PCA is used. Here also the number of dimensions is selected based on the MBSR. Finally the selected dimensions are used for further clustering and classification. The ISODATA, AIS, HAIS and NHAIS are applied on the reduced data set and the classification efficiency is noted. The same analysis is done for the selection of parameters as for the crop stage data set.

TABLE II gives the individual, average, overall efficiencies and kappa coefficient of the ISODATA, AIS, HAIS and the proposed NHAIS. From this table, we can observe that HAIS and NHAIS yield higher average, overall efficiency and kappa coefficient when compared to the ISODATA and AIS. When comparing between HAIS and NHAIS, Overall efficiency is marginally better in HAIS over NHAIS. However individual efficiency of small data sets like “Alfalfa” and “Oats” classes, NHAIS is better than HAIS. Thus average accuracy of NHAIS is better. Since difference in kappa coefficient between HAIS and NHAIS is small, we can conclude that both algorithms are equally good. But in NHAIS we have additional advantage that it extracts small classes with high efficiency.

V. CONCLUSION

A niche hierarchical clustering method for crop stage classification and Indian Pines data set of hyperspectral images is proposed and implemented in this paper. The proposed method could successfully perform data clustering using the principles of splitting and merging the cluster centers along with niching process. NHAIS is used to split the hyperspectral data set to optimal number of cluster centers and majority voting method is used to merge the data set. The high dimensionality of the data is reduced to lower dimensions using PCA by considering Modified Broken-Stick Rule. Parameters are varied to generate optimal cluster centers. The experimental results show that the proposed NHAIS is better in comparison with other unsupervised techniques namely ISODATA and AIS and high efficiency in classifying small classes compared to HAIS. This evinces that the proposed algorithm is applicable for processing of the hyperspectral image.
ACKNOWLEDGMENT

The authors would like to thank the associate editor and reviewers for their comments that helped in improving this paper.

REFERENCES


J.Senthilnath is a research scholar at the Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India. His research interests include nature inspired computational techniques, satellite image processing, machine learning and computer vision.

S.N.Omkar is working as a principal research scientist at the Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India. His research interests include nature inspired computational techniques, satellite image processing and parallel computing.

V. Mani is working as a professor at the Department of Aerospace Engineering, Indian Institute of Science, Bangalore, India. His research interests include distributed computing, queuing networks, evolutionary computing, and neural computing.

Nitin Karnwal completed his undergraduate program in Instrumentation and Control Engineering from National Institute of Technology-Tiruchy, Tamilnadu, India in July 2012.

Shreyas P.B currently pursuing his under-graduation in Electronics and Communication Engineering at National Institute of Technology Karnataka, Surathkal, India.