



AN ARTIFICIAL IMMUNE NETWORK CLUSTERING ALGORITHM FOR MANGROVES REMOTE SENSING IMAGE

Yanmin LUO¹, Peizhong LIU^{2*}, Minghong LIAO³

¹ College of Computer Science & Technology, Huaqiao University, Xiamen, 361021, China

² College of Engineering, Huaqiao University, Fujian, Quanzhou 362021, China

³ Software School, Xiamen University, Xiamen, 361005, China

Emails: pzliu@hqu.edu.cn

Submitted: Oct. 4, 2013

Accepted: Feb. 5, 2014

Published: Mar. 1, 2014

Abstract- Because of not needing training samples and performing the classification just according to the inherent similarity of data in the multidimensional space, unsupervised classification method now gets more and more attention by remote sensing data analyst. Duo to the special growth environment of mangroves, field measurements is difficult to be done to obtain training samples. Therefore unsupervised method provides a good adjunct way for the classification of mangroves remote sensing image. This paper presents an immune network based unsupervised classification method, which is not necessary to define complex objective function. By arbitrarily selecting a certain number of data to be training samples, the proposed algorithm mines the prior knowledge of the samples and selects a few samples to constitute the initial nodes of immune network. After the evolutionary of the immune network, the clustering results are obtained, combined with nearest neighbor classification mechanism, the classification is performed. The experiment results show that the proposed algorithm has better overall stability and can get better clustering result for mangroves remote sensing image than traditional clustering methods.

Index terms: Artificial immune network, unsupervised, remote sensing image, mangroves, classification.

I. INTRODUCTION

Mangroves ecosystem is regarded as one of most productive ecosystems in the world due to its significant ecological value and economic value. Mangroves are valuable ecological and economic resources, being important nursery grounds and breeding sites for birds, fish, crustaceans, shellfish, reptiles and mammals; renewable sources of wood; accumulation sites for sediments, contaminants, carbons and nutrients; and offer protection against coastal erosion [1]. Knowledge of the dynamics of mangroves ecosystem is important in the context of global change. To obtain this knowledge, remote sensing is an indispensable means, it now has become a major technology for detecting and monitoring mangroves [2][3]. Supervised classifications are the primary methods for the interpretation of remote sensing image such as decision-tree [4] and SVM [5]. But the accuracy of supervised methods are affected by the quantity and quality of training samples. In addition, mangroves are mainly grown in intertidal shallows, conventional field survey work is very difficult, which would lead to heavy workload and high cost. As a result it is difficult to accurately locate and sketch the distribution of mangroves. Therefore exploring and researching new technologies and methods [6][7], especially intelligent approaches [8][9] to improve classification accuracy for remote sensing image of mangroves has important significance and value for detection and protection of mangroves ecological environment. Unsupervised methods execute the classification by the inherent similarity of the data classes in multidimensional space. For requiring no prior knowledge of training samples, unsupervised methods gain more and more attention by remote sensing data analysts. For the condition that it is difficult or impossible to collect the data training samples, or it is necessary to take a lot of cost and time to get those data, unsupervised classifications play an important role in remote sensing image, they allow people to be able to get the initial characteristics of all kinds of land cover types and provide the basis for the next design of classifiers and further classification. The aims of this work is to apply the principle of immune network to the cluster of mangroves remote sensing image. We proposed an artificial immune network based unsupervised classification method (AINUC, in short), which is not necessary to define complex objective function. Firstly, the proposed algorithm obtained a certain number of learning samples from data set to be classified randomly, on which the prior knowledge of the training samples was mining. Secondly,

the initial nodes of immune network were established, and the immune network was running to achieve the clustering results. Finally, combined with nearest neighbor mechanism, the classification was performed. The experiment results show that the algorithm can get better clustering result for mangroves remote sensing image than traditional clustering methods.

II. ARTIFICIAL IMMUNE NETWORK

Various aspects of biology have always been the inspiration in developing computational models and problem solving methods. The immune system is one such system that has recently drawn significant attention. There has been considerable interest in exploring and exploiting the potential of artificial immune systems for applications in computer science and engineering. As an artificial immune systems, artificial immune network is based on idiotypic immune network theory [10] which has been proposed by Jerne in the mid-seventies.

The immune network theory suggests that the immune system maintains an idiotypic network of interconnected B cells for antigen recognition. These cells both stimulate and suppress each other in certain ways that lead to the stabilization of the network. Two B cells are connected if the affinity they shared exceeds a certain threshold. The strength of the connection is directly proportional to the affinity they shared [11]. In artificial immune network (AIN) models, the B cell population is made of two sub-populations: the initial population and the cloned population. The initial set is generated from a subset of raw training data to create the B-cell network. The remainders are used as antigen training items. During the evolution of immune network, antigens are selected randomly from the training set and presented to the areas of the B-cell network. If the binding is successful, then the B-cell is cloned and mutated [12]. The mutation yields a diverse set of antibodies that can be used in the classification procedure. Once a new B cell is created, an attempt is made to integrate it into the network at the closest B Cells. If the new B cell can not be integrated, it is removed from the population. If no bind is successful, then a B-cell is generated using the antigen as a template and is then incorporated into the network.

Evolutionary and dynamic network are the two major characteristics of the immune network. Immune network has good generalization ability, memory capacity and tolerance to noise; it is capable of concentrating for arbitrarily shaped data. Artificial immune network and network-based artificial immune system model are the important parts in the research of artificial immune

systems. Some excellent models have been widely used in engineering and data analysis, such as resource limited artificial immune system (RLAIS) [13], aiNet [14] and so on. Relative to the clonal selection algorithm, artificial immune network pays more emphasis on the role of memory cells and interaction among the antibodies in antibody population. The memory ability of the immune network not only relies on memory cells, but also the network structure. Thus it can show better population diversity and more comprehensive understanding on the solved problem. The specific implementation details of variety of artificial immune network algorithm may vary, but they would have common characteristics and similar functions. Artificial immune network has a good application prospect on data clustering, data compression and data purification. It can be used in the researches of engineering technology and information technology to design various scientific and effective algorithms or models [15][16].

III. ARTIFICIAL IMMUNE NETWORK CLUSTERING ALGORITHM FOR MANGROVES REMOTE SENSING IMAGE

The processing of artificial immune network clustering algorithm can be developed into 4 steps.

Step 1: Preparing for the sample set and performing data pro-processing.

Step 2: Mining the prior knowledge of the data to be classified and getting the initial antibody nodes of immune network.

Step 3: Calling the artificial immune network clustering algorithm to get the cluster center for each data type.

Step 4: Performing the classification combined with nearest neighbor mechanism.

a. Preparing for the sample set and performing data pro-processing

For the proposed algorithm is one of unsupervised classification method, it is not necessary to know the category of each training sample. The algorithm randomly selects N_T data to constitute the training sample set from the data set to be classified. N_T should take a appropriate value to guarantee the accuracy and efficiency of the algorithm. Let N be the number of the data to be classified, N_C be the number of categories of data to be classified. The relationship of N_T , N and N_C can be shown by formula (1).

$$N \gg N_T > k \times N_C \quad (1)$$

where $k \in [200, 10]$, and the value of k is proportional to the value of N .

The mainly task of data pro-processing is data standardization. In the classification, different features for decision-making have different ranges. Standardization processing normalizes the value of feature to (0, 1], which can avoid that the feature with large value has a greater influence on classification, and let all the features for decision-making have the same weight to classification. Standardized formula is shown by formula (2).

$$X' = (X - \min(X)) / (\max(X) - \min(X)) \quad (2)$$

where X is value the feature to be normalized, and X' is the result. $\max()$ and $\min()$ are the functions to get the maximum value and minimum value respectively.

b. Mining the prior knowledge of the data to be classified and getting the initial antibody nodes of immune network

In the clustering of immune network, the quality of the initial network nodes has a great impact on efficiency and effectiveness of the algorithm. If the initial network nodes can contain most samples of different categories, then the immune network not only can evolve fast, but also can make the results more precise. Currently, most clustering algorithms do not take full advantage of the prior knowledge of the data to be classified. In this paper, combining with the stochastic method, the fuzzy k-means algorithm (FCM, in short) was used to mining the prior knowledge of the data to be classified. On that the initial nodes of immune network were determined, so that the initial network node can contain various training samples of different categories as much as possible. The processing can be described as following steps:

Step1: Selecting N_{ran} ($N_{ran} > N_{ini}$) data randomly from the data to be classified. where N_{ini} is the number of initial nodes of immune network.

Step2: Calling the FCM algorithm to cluster the N_{ran} data

Step3: Repeat step1~step2 for $N_{cluster}$ times and get the $N_{cluster}$ results.

Step4: Selecting randomly a certain number of data for each category which getting by step 3 to form the initial antibody nodes of immune network.

For the N_{ran} data were selected randomly, which can not ensure that the selected data can contain all categories. By repeating step1 and step2, we can greatly improve the likelihood that the initial nodes of immune network contain various training samples of different categories. The

repeating time $N_cluster$ is proportional to the size of the data to be classified N , and the experiments show the appropriate range of $N_cluster$ is [10,100].

c. Calling the artificial immune network clustering algorithm to get the cluster center for each data category

After getting the initial nodes of immune network, the algorithm calls the artificial immune network based clustering algorithm to get the cluster center for each data category. Immune network achieves the data clustering and purifying by a series of immune mechanisms of the antibodies in immune network which includes antigen recognition, cloning, mutation, immune suppression and network suppression. By those mechanisms, the data to be clustered (antigens) are mapped to memory antibody set. Antibodies memory cells are the refining of the original data, and can well reflect the distribution of the original data. The antibodies memory cells (cluster center) can substitute the raw data, which can achieve the purpose of data compression and purifying.

(1) The coding way of the antibody and antigen

The coding way of antibody or antigen is one of the important issues to AIS. Compared with binary code, real encoding way is intuitive, more effective and can avoid repeating the operation of encoding and decoding. Also real code is suitable for remote sensing data. Therefore, the proposed algorithm adopts the real encoding way.

In the immune network algorithm, antibody and antigen share the same coding way, which is composed of characteristic values of all the decision-making features of the training sample. The antigen represents the original data set, while antibody corresponds to the B-cell which is one of node or data item in the immune network, and can represent the characteristic of each class pattern and indicate the network after immune evolutionary.

Let N_ag be the number of antigen, N_ab be the number of antibody. Then the antigen set can be represented by $AG = \{ag_1, ag_2, \dots, ag_{N_ag}\}$, $AG \in R^{N_ag}$, where ag_i is the i^{th} individuality of antigen set, and it corresponds to the i^{th} sample of training samples set. Supporting that each sample has N_f features, thus $ag_i = \{ag_{i1}, ag_{i2}, \dots, ag_{iN_f}\}$, where ag_{ij} ($i=1 \dots N_ag, j=1 \dots N_f$) is the value of j^{th} feature of antigen ag_i . The antigen set AG can be described by formula (3).

$$AG = (ag_1, \dots, ag_{N_ag})^T = \begin{bmatrix} ag_{11}, \dots, ag_{1N_f} \\ \vdots & \ddots & \vdots \\ ag_{N_ag1}, \dots, ag_{N_agN_f} \end{bmatrix} \quad (3)$$

Similarly, the antibody set AB can be described by formula (4).

$$AB = (ab_1, \dots, ab_{N_{ab}})^T = \begin{bmatrix} ab_{11}, \dots, ab_{1N_f} \\ \vdots \quad \ddots \quad \vdots \\ ab_{N_{ab}1}, \dots, ab_{N_{ab}N_f} \end{bmatrix} \quad (4)$$

where $ab_i = \{ab_{i1}, ab_{i2}, \dots, ab_{iN_f}\}$ is the i^{th} individuality of antibody set, and ab_{ij} ($i=1 \dots N_{ab}$, $j=1 \dots N_f$) is the value of the j^{th} feature of antibody ab_i .

By the processing of learning and evolution, artificial immune network recognizes the pattern of the data to be processed, which essentially are the processes that antibodies recognize antigens and evolve themselves to the pattern of the data to be processed constantly. The processing of immune response can be formally described by formula (5).

$$AG = \begin{bmatrix} ag_1 \\ ag_2 \\ \vdots \\ ag_i \\ \vdots \\ ag_{N_{ag}} \end{bmatrix} \xrightarrow{f} AB = \begin{bmatrix} ab_1 \\ ab_2 \\ \vdots \\ ab_l \\ \vdots \\ ab_{N_{ab}} \end{bmatrix} \quad (5)$$

If $N_{ag} > N_{ab}$ is satisfied in formula (5), then the scale of antibody is less than the scale of antigen. which means by the processing of immune response the scale of antibody had been compressed.

(2) Antigen recognition

After the initial antibody immune network are established, the antigens of antigen set come into the network to be recognized by antibody cells one by one. In the artificial immune network system the process of antibody recognizing antigen is performed by calculating the affinity between antibody and antigen. For antibody and antigen share the same encoding ways, the affinity is measured by their Euclidean distance. Also the similarity between antibody and antibody or the similarity between antigen and antigen is measured by Euclidean distance. The affinity is inversely proportion to Euclidean distance, which means the higher the affinity the smaller the distance. The affinity of antibody ab_i and antigen ag_j can be calculated by formula (6).

$$S_{ij} = S(ab_i, ag_j) = \frac{1}{D(ab_i, ag_j)} = \frac{1}{\sqrt{\sum_{k=1}^{N-f} (ab_{ik} - ag_{jk})^2}} \quad (6)$$

where $D(ab_i, ag_j)$ are the Euclidean distance of ab_i and ag_j .

(3) Clone proliferation

In immune algorithm, clone proliferation can not only ensure the diversity of antibodies but also ensure the antibody with high affinity has better chance to enter the next generation.

After antigen recognition, the antibodies which have higher affinity will conduct clone proliferation. The antibody with greater affinity will be cloned more. According to the value of affinity, the n highest affinity antibodies are selected to be the subset of select population $Ab_{\{select\}}$. Then the selected antibodies are submitted to perform the clone operation. The clone size is an increasing function of the antibody affinity. Formula (7) shows the number for the anyone antibody ab in $Ab_{\{select\}}$ to be cloned.

$$N_{clone}(ab) = round(N_{ab} - N_{ab} \cdot D(ab, ag_j)) \quad (7)$$

where N_{ab} is the total amount of antibodies in antibody set, $D(ab, ag_j)$ are the Euclidean distance of ab and ag_j , and $round(\cdot)$ is the operator that rounds the value on parenthesis towards its closest integer. Supporting the antibodies population after clone is $Ab_{\{clone\}}$.

(4) Immune mutation

Mutation operator is very important to artificial immune algorithm. It enables the new antibodies to match the antigen more closely, and avoid falling into local optimal solution. The set $Ab_{\{select\}}$ is submitted to a directed immune mutation process which means affinity maturation. Then a mutated set $Ab_{\{mutation\}}$ is generated. Each antibody ab_k from $Ab_{\{clone\}}$ will suffer a mutation with a rate α_k which is inversely proportional to its affinity (S_{kj}) to the antigen ag_j , the higher the affinity, the smaller the mutation rate. Supporting the antibody after mutation in $Ab_{\{mutation\}}$ is ab_k^* , which parent antibody in $Ab_{\{clone\}}$ is ab_k . The mutation operator can be showed by formula (8).

$$ab_k^* = ab_k + \alpha_k (ab_j - ab_k); \quad \alpha_k \propto 1/S_{kj} \quad (8)$$

(5) Antibody apoptosis

By cloning selection and mutation, the training samples (which are antibodies) are expanded. This expansion of the training samples is based on the premise of the antigen recognized. Therefore, the antibodies obtained after clone proliferation and immune mutation should ensure a

certain degree of affinity with the corresponding antigen ag_j . The purpose of antibody apoptosis is to remove the antibodies which are generated by clone proliferation and immune mutation but has low affinity with antigen ag_j . The antibody apoptosis is performed by setting the corresponding threshold value T_{aa} . If the affinity between the antibody in $Ab_{\{mutation\}}$ and antigen ag_j is greater than T_{aa} , then the antibody will be cleared. Let $Ab_{\{aa\}}$ be the antibody population collection after antibody apoptosis.

(6) Clonal suppression

The clone proliferation and mutation of antibody is to ensure the diversity of antibody. If any two antibodies after clone and mutation have great similarity, which means one of them is redundant. For the immune network, the similar antibodies will bring redundant nodes which would affect the evolution speed of immune network. The process of clonal suppression is to eliminate the similar antibodies in $Ab_{\{aa\}}$, in order to avoid antibody repetition and improve the efficiency of the algorithm. The clonal suppression is performed by setting the corresponding threshold value T_{cs} . If the affinity of two antibodies is less than T_{cs} , then one of them will be cleared. Let $Ab_{\{cs\}}$ to be the antibody population collection after performing the clonal suppression on $Ab_{\{aa\}}$. The antibodies in $Ab_{\{cs\}}$ will be entered into the immune network and participate in the network competition, which can make them come into being maturation.

(7) Network suppression

Network suppression make the cloning antibodies come into being affinity maturation. Its principle is similar with the natural selection process of species. During the evolution of immune network, the clonal suppression antibody population $Ab_{\{cs\}}$ will enter the network memory cells collection and participate in online competition. Then the mutual recognitions inner the antibody cells will occur. The mutual recognition between antibody cells will causes their autoimmune response. Network suppression can eliminate the autoimmune response. When the clonal suppression antibody population $Ab_{\{cs\}}$ enter the immune network, the antibody which is similar to one of network memory antibody cells will be cleared. The similarity between antibodies is measured by their affinity. The network suppression is performed by setting the corresponding threshold value T_{ns} . If the affinity of antibody in $Ab_{\{cs\}}$ and anyone antibody in immune network memory cells set M is less than T_{ns} , then the antibody of $Ab_{\{cs\}}$ will be eliminated.

After network suppression, the immune network completed a recognition and evolution for a antigen ag_j , and the immune network memory cells set M completes an update and achieve

affinity maturation and become the new antibody memory cells set M' . The antibodies in new set M' will successful entry into the next generation for the next evolution.

(8) Algorithm description

The artificial immune network based unsupervised classification algorithm (AINUC) can be described as follows:

step1: Getting N_{ag} training samples from the data to be classified as the antigen set AG .

step2: Getting the initial antibody nodes(N_{ab}) of immune network which is also the initial memory cells set of immune network M .

step3: For each antigen $ag_j, (j=1, \dots, N_{ag})$ in AG do

3.1 Antigen recognition: calculate the affinities between ag_j and all the antibodies in immune network memory cells set M based on formula (6).

3.2 Select the n antibodies which have the highest affinity with ag_j to be the select population $Ab_{\{select\}}$.

3.3 Clone proliferation: clone the n antibodies to get the clonal population $Ab_{\{clone\}}$. The clone size is an increasing function of the antibody affinity.

3.4 Immune mutation: perform mutation on the antibodies in $Ab_{\{clone\}}$ and generate the mutated set $Ab_{\{mutation\}}$. The mutation rate is inversely proportional to the affinity.

3.5 Antibody apoptosis: eliminate the antibody in $Ab_{\{mutation\}}$ which the affinity of the antibody and antigen ag_j is less than T_{aa} , and generate the antibody population $Ab_{\{aa\}}$.

3.6 Clonal suppression from antibody: eliminate one of the any two antibodies in $Ab_{\{aa\}}$, whose affinity is greater than the threshold T_{cs} , and generate the clonal suppression antibody population $Ab_{\{cs\}}$.

3.7 Network suppression: add $Ab_{\{cs\}}$ to immune network memory cells set M , and eliminate the antibody of $Ab_{\{cs\}}$ whose affinity with anyone antibody of M is less than the threshold T_{ns} , and generate the new memory cells set M' which will entry into the next generation.

step 4: Repeat step 3 until a termination criterion is satisfied or the predetermined generation number is reached.

The algorithm executes the loop from step 3.1 to step 3.7. After each loop, one generation of antibody is generated. Steps 3.1 to 3.6 describe the clonal selection and affinity maturation processes. Step 3.7 simulates the immune network activity. In AINUC, the clonal immune response is elicited by each presented antigenic pattern. There are two suppressive steps in this

algorithm: the clonal suppression is to eliminate intra-clonal self-recognizing antibodies, while the network suppression is to search for similarities between different sets of clones. The loop will terminate when a criterion is satisfied or a predetermined generation number is reached. Finally the network antibodies in memory set will represent internal images of the antigens presented to the network.

d. Performing the classification combined with nearest neighbor mechanism

After the learning of AINUC algorithm, the clusters will serve as internal mirrors responsible for mapping existing clusters in the data set onto network clusters. Figure 1 is the illustration of the AINUC algorithm. Figure 1(a) shows the data set (antigens set) composed of three regions with a high density of data. Figure 1(b) shows the network architecture (antibodies in network memory set) generated by AINUC algorithm which can well represent internal images of the antigens in Figure 1(a).

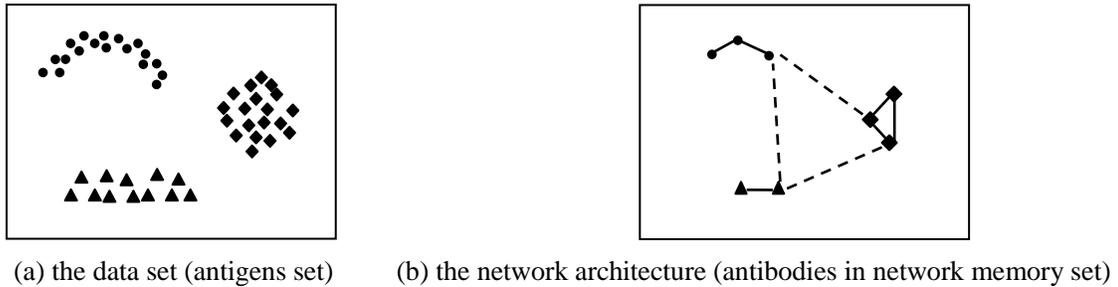


Figure 1. The illustration of the AINUC algorithm

There are three clusters in Figure 1(b), the solid lines indicate the inner connections of each cluster, and the dashed lines suggest connections to be pruned. By the solid lines and dashed lines the final network structure can be defined and detected.

Based on the network architecture (antibodies in network memory set, M) generated by AINUC algorithm, the Minimum Distance method is used to classify the samples to be classified. For an unknown sample $TS=(t_1, t_2, \dots, t_{N_f})$, calculating the distances between TS and all the antibodies in M . If the antibody ab_i in M satisfies the condition shown in formula (9), then TS is classified into the same category as the antibody ab_i .

$$ab_i = \arg \min_{ab_j} (dis(ab_j, TS)) \quad (9)$$

IV. EXPERIMENT AND ANALYSIS

a. Data source and decision-making features

The Zhangjiang estuary in southeastern China, a national mangroves reserve, was chosen as the region for algorithm experiment. The reserve is bounded by 22°53'45"N ~ 23°56'00"N and 117°24'07"E ~ 117°30'00"E, with an area of 2360 ha. The remote sensing data used is the 30-m resolution multispectral Landsat Thematic Mapper(TM) image which was acquired on Sep 25, 2006. Fig.2 (a) shows the study area. According to field investigation in the area, the land-cover type is defined by seven categories. Table 1 lists all the seven land cover types and their descriptions.

Table 1: Land cover types in classification

Class Name	Land-cover Class	Description
C1	Mangroves	Mangrove forests
C2	Agricultural land	Crop fields, paddy fields and grasslands
C3	Upland vegetation	Deciduous or evergreen forest land, orchards, and tree groves
C4	Water	Permanent open water, lakes reservoirs, bays, and estuaries
C5	Urban area	Residential, commercial, industrial and other developed land
C6	Littoral zones	Land in the intertidal zone or the transitional zone between land and sea
C7	Fallow land	Fields no longer under cultivation

In the experiments, spectral features, geographical features and image texture features are integrated to be the decision-making features, which have been proved to be efficient for classification of remote sensing image [6]. Twelve features were extracted for each sample, which are composed by the spectrums of TM image:band1~band7, Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM) and 3 texture features based on Gray-Level Co-occurrence Matrix (GLCM). Mangrove forests are constrained by topographic factors. Field investigations validate that mangrove forests were not distributed above 8 meter elevation. Therefore, the digital elevation model (DEM) can be used to exclude non-mangrove pixels that had similar spectral attributes with mangrove pixels, but were above the elevation limiting line.

Texture contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. The texture of mangroves is being fine and smooth, while other green vegetations are coarse and rippled. Leen-Kiat Soh et al. formulated 10 features of GLCM in ref. [17]. This paper take 3 of them (Variance, second moment and dissimilarity) to be the texture features which are proved to be useful for mangroves classification by experiments.

b. Parameters analysis

T_{aa} , T_{cs} and T_{ns} are the main parameters to AINUC algorithm, their values will affect the performance and the convergence speed of the algorithm. A too large value of T_{aa} will generate a lot of antibodies with low affinities to antigen, which may lead to generate some new categories that do not actually exist. While too small value of T_{aa} will influence the effect of clonal proliferation.

The purpose of cloning is to remove the similar antibody generated by clones proliferation. If the value of T_{cs} is too small, many similar antibodies will enter the memory network, which would increase the burden of network suppression. On the other hand, too greater value of T_{cs} leads to lower speed of immune network evolution, and may affect global property of the algorithm.

The value of T_{ns} will affect the final result of the immune network directly. If T_{ns} is too small, then the final form of network structure will be complex, which is not conducive to subsequent classification. if T_{ns} is too large, the categories generating by immune network are likely to be less which leading to miss some categories. Reasonable value of these three parameters is important to AINUC algorithm, and also their values have a relationship with the specific problems to be solved. After a number of experimental analyses, we found that for mangroves remote sensing images, the best combination of these three parameters is: $T_{aa}=0.3$, $T_{cs} =0.5$ and $T_{ns} =0.1$. Therefore we adopt these three values in the next experiments.

c. Experimental results and analysis

In order to evaluate the efficiency of the proposed algorithm (AINUC), we compare the classification results of AINUC algorithm with K-Means and ISODATA. The AINUC algorithm was programming and running in Matlab. The pre-processing of remote sensing image and the two other clustering algorithms were all done under the ESRI ENVI 4.5. In the experiment, each

algorithm was repeated 50 times, and the better one was taken as the results. The images of the classification results are shown in Figure 2(b)~(e).

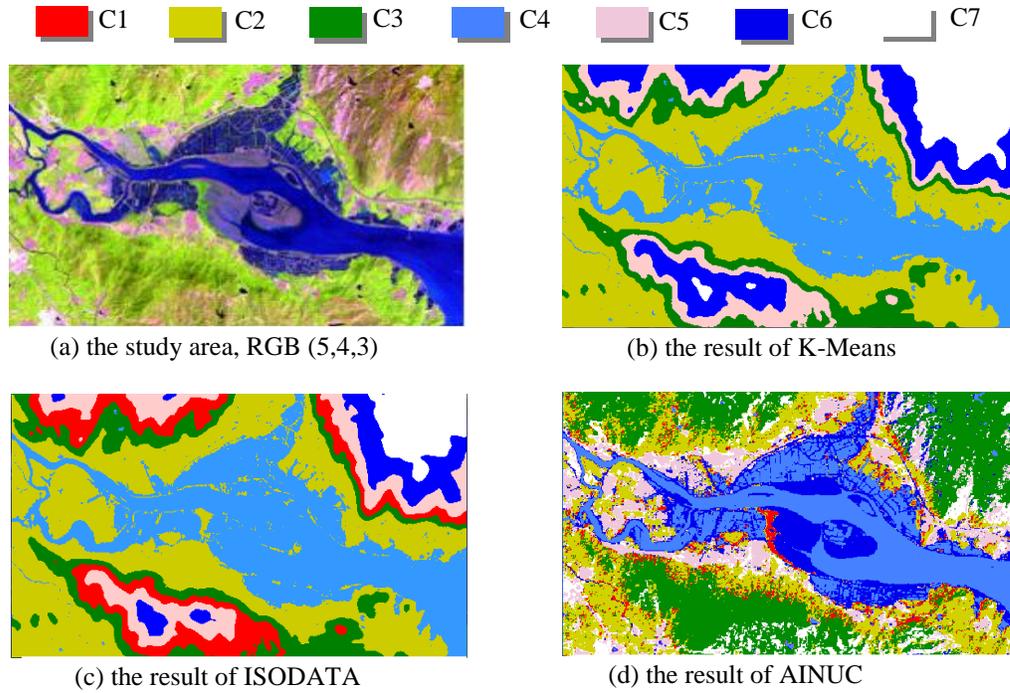


Figure 2. The study area and the classification result images

As shown in Figure 2, the clustering result of AINUC algorithm is much better than the other two algorithms. AINUC algorithm generally is able to identify the main areas of various land-cover types, while K-Means and ISODATA just can identify the main areas of water and agricultural land, as for the other land-cover types they almost can not correctly distinguish. Although AINUC algorithm misclassify many agricultural land and upland vegetation to mangroves, but it can basically identify the major mangroves areas. On the other hand, K-Means and ISODATA are completely unable to correctly identify the mangroves areas and both K-Means and ISODATA misclassify all the mangroves into agricultural land. K-Means can not identify any mangroves, and the mangroves which indentified by ISODATA are almost agricultural land and upland vegetation actually. As for littoral zones, only AINUC can indentify correctly, K-Means and ISODATA almost can not correctly distinguish.

For assessing the clustering results, Two hundred test samples were randomly selected from a WORLDVIEW-2 image (1.8-m resolution) and the Google Earth image which had been used

for geometric rectification. Overall accuracy and Kappa coefficient of the classifications were then computed, they are showed in Table 2.

Table 2: The comparing of overall accuracy

	K-Means	ISODATA	AINUC
Overall	33.1%	35.8%	81.4%
Kappa	0.289	0.323	0.778

As shown in Table 2, the overall classification accuracy of AINUC is 81.4%, and the Kappa coefficient is 0.778, both are much higher than K-Means and ISODATA. This indicates that the developed algorithm can improve the overall clustering accuracy effectively for mangroves remote sensing image. The results also confirmed that AINUC algorithm has the characteristics of continuous self-learning, dynamic regulation and immune memory ability. Especially the adaptive capacity can ensure the stability and versatility of AINUC algorithm. It is suitable for remote sensing images and other complex data processing field. Therefore AINUC algorithm is significantly better than K-Means and ISODATA for complex and high-dimensional data processing.

One of the characteristics AINUC algorithms take advantage of a priori knowledge of the sample. By the fuzzy k-means, the distribution of training samples was mining, and thus initial antibodies of B-cell was determined. For validating the effect of prior knowledge to the clustering, we compared AINUC in two cases. They are using prior knowledge and without prior knowledge respectively. For the later, we just randomly selected N_{ini} samples from the initial training sample set to be the initial node of immune network. The experiment results are shown in Figure 3(a) ~ (b).

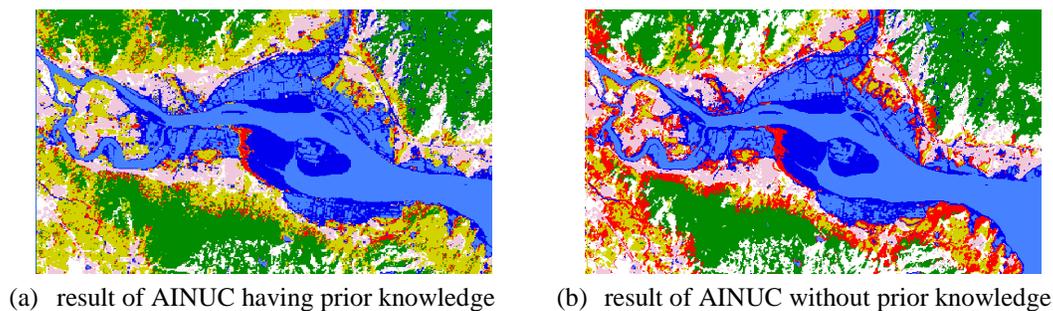


Figure 3. The comparing of clustering results for AINUC on prior knowledge

As shown in Figure 3, this two methods result in good classification for the water, littoral zones and upland vegetation. However, for the left land cover types, the algorithm without prior knowledge had serious misclassification and leakage points. Although it can identify main area of mangroves, but many samples of agricultural land were wrongly identified as mangroves. The AINUC algorithm having prior knowledge seems avoid these problems.

The confusion matrices for clustering results of the two algorithms are listed in table 3, and the Overall accuracy and Kappa coefficient are showed in table 4.

Table 3: The comparing of confusion matrices

AINUC having prior knowledge									AINUC without prior knowledge								
Class	C1	C2	C3	C4	C5	C6	C7	Total	Class	C1	C2	C3	C4	C5	C6	C7	Total
C1	181	55	2	0	0	0	0	238	C1	200	132	25	0	0	0	1	358
C2	15	127	69	0	0	0	8	219	C2	0	61	70	0	0	0	0	131
C3	3	11	129	0	0	0	15	158	C3	0	0	48	0	0	0	5	53
C4	0	1	0	189	0	0	0	190	C4	0	0	0	193	0	0	0	193
C5	0	2	0	0	194	0	54	250	C5	0	2	0	0	197	0	40	239
C6	1	3	0	11	0	200	5	220	C6	0	1	0	7	0	200	0	208
C7	0	1	0	0	6	0	118	125	C7	0	4	57	0	3	0	154	218
Total	200	200	200	200	200	200	200	1400	Total	200	200	200	200	200	200	200	1400

Table 4: The comparing of overall accuracy

	AINUC having prior knowledge	AINUC without prior knowledge
Overall	81.4%	75.2%
Kappa	0.778	0.710

In confusion matrix, the elements of main diagonal are the number of samples which are classified correctly. As shown in Table 3, the sum of main diagonal elements of AINUC having prior knowledge is larger than those of AINUC without prior knowledge, which shows the using of prior knowledge can improve the classification accuracy effectively. Although AINUC without prior knowledge can identify more mangroves points from the 200 samples, but it misclassify many agricultural land and forest samples into mangroves too. The overall classification accuracy of AINUC having prior knowledge is 81.4%, and the Kappa coefficient is 0.778, both higher than those of AINUC without prior knowledge. All of these results

demonstrate that the AINUC algorithm developed in this study is a competent classification method for mangroves satellite images. This is because AINUC is a self-adaptive method which is data-driven and can adjust itself to the data.

V. CONCLUSIONS

A unsupervised classification algorithm was proposed in this paper, which inspired in the clonal selection principle and affinity maturation of the immune network. The antibodies network represents the candidate clustering solutions, while the antigen represents the data to be classified. A decimal encoding method for antibody and antigen is used which is suited for remote sensing data. The affinity is measured based on Euclidean distance. The clone selection operator, mutation operator and suppression operator are involved in the evolution of immune network. By fuzzy k means algorithm the prior knowledge of training samples was mining, which can improve the classification accuracy effectively. A series of experiments were carried out to test the performances using mangrove remote sensing image. By comparing with traditional clustering algorithm, the algorithm was verified to be a efficient method. For future research, the algorithm will be further explored for more extensive remote sensing applications, and the more self- adaptive method would be carried out.

VI. ACKNOWLEDGEMENT

This work was supported by Quanzhou scientific and technological planning projects of Fujian, China(2013Z28), and new IT platform construction project in Fujian Province(2013H2002).

REFERENCES

- [1] D. M. Alongi, "Present state and future of the world's mangrove forests," *Environmental Conservation*, 2002, 29(3), pp. 331-349.
- [2] B. W. Heumann, "Satellite remote sensing of mangrove forests: Recent advances and future opportunities," *Progress in Physical Geography*, 2011, 35(1), pp. 87-108.

- [3] C. Giri, E. Ochieng, L. Tieszen, Z. Zhu, A. Singh, T. Loveland, J. Masek, N. Duke, "Status and distribution of mangrove forests of the world using earth observation satellite data," *Global Ecology and Biogeography*, 2011, 20(1), pp. 154-159.
- [4] K. Liu, X. Li, X. Shi, S. Wang, "Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning," *Wetlands*, 2008, 28(2), pp. 336-346.
- [5] B. W. Heumann, "An Object-Based Classification of Mangroves Using a Hybrid Decision Tree--Support Vector Machine Approach," *Remote Sensing*, 2011, 3(11), pp. 2440-2460.
- [6] Y. Luo, M. Liao, J. Yan, C. ZHANG, "A multi-features fusion support vector machine method (MF-SVM) for classification of mangrove remote sensing image," *Journal of Computational Information Systems*, 2012, 8(1), pp. 323-334.
- [7] V. Rodríguez-Galiano, F. Abarca-Hernández, B. Ghimire, M. Chica-Olmo, P. Atkinson, C. Jeganathan, "Incorporating spatial variability measures in land-cover classification using Random Forest," *Procedia Environmental Sciences*, 2011, pp. 344-349.
- [8] Y. Zhong, L. Zhang, J. Gong, P. Li, "A supervised artificial immune classifier for remote-sensing imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, 2007, 45(12), pp. 3957-3966.
- [9] Y. Luo, M. Liao, J. Yan, C. Zhang, S. Shang, "Development and demonstration of an artificial immune algorithm for mangrove mapping using landsat TM," *Geoscience and Remote Sensing Letters, IEEE*, 2013, 10(4), pp. 751-755.
- [10] N. K. Jerne, "Towards a network theory of the immune system," *Annales d'immunologie*, 1974.
- [11] D. Dasgupta, Z. Ji, F. Gonzalez, "Artificial immune system (AIS) research in the last five years," *CEC'03. The 2003 Congress on Evolutionary Computation*, 2003.
- [12] J. Hunt, J. Timmis, E. Cooke, M. Neal, C. King. *Jisys, "The Envelopment of an Artificial Immune System for Real World Applications,"*. City: Springer, 1999.
- [13] J. Timmis, M. Neal, "A resource limited artificial immune system for data analysis," *Knowledge-Based Systems*, 2001, 14(3), pp. 121-130.
- [14] L. Nunes de Casto, F. J. Von Zuben, "An evolutionary immune network for data clustering," *Neural Networks, 2000. Proceedings. Sixth Brazilian Symposium on*, 2000.
- [15] G. B. Bezerra, L. N. de Castro, "Bioinformatics data analysis using an artificial immune network," City: Springer, 2003.

[16] F. O. de França, F. J. Von Zuben, L. N. de Castro, "An artificial immune network for multimodal function optimization on dynamic environments," Proceedings of the 2005 conference on Genetic and evolutionary computation, 2005.

[17] R. M. Haralick, K. Shanmugam, I. H. Dinstein, "Textural features for image classification," Systems, Man and Cybernetics, IEEE Transactions on, 1973, 3(6), pp. 610-621.