AN ALGORITHM FOR REMOTE SENSING IMAGE CLASSIFICATION BASED ON ARTIFICIAL IMMUNE B –CELL NETWORK

Shizhen Xu[a, *], Yundong Wu[b, c]

[a]Insitute of Surveying and Mapping, Information Engineering University – 66 Middle Longhai Road, Zhangzhou, Henan 450052 China - xshizhen@163.com
[b]College of Sciences, Jimei University, 361021 Amoy, China - wuyundong@yahoo.com.cn
[c]Institute of Automation, Chinese Academy of Sciences (CASIA), 100080 Beijing, China

Youth Forum

KEY WORDS: Pattern recognition, Classifier, Image Understanding, Feature Recognition, Land Cover

ABSTRACT:

In this paper a novel supervised classification algorithm (AIBN) for remote sensing image classification based on artificial immune B-Cell network is proposed. Even though some effective classifiers have been proposed in the field of Artificial Immune System, there are still some deficiencies in them. In AIBN, clone selection and immune network theories are used as mechanism for data training in order to gain reduction form of the data. After training the distribution and density information of B cells is learned by AIBN and B cells can directly be used as a criterion for classification. Then, the classification task is carried out. The experiment results show AIBN is superior to Maximum Likelihood classifier and Artificial Immune Recognition System. Some analyses of user defined parameters are made. AIBN provide an alternative way to perform remote sensing image classification task.

1. INTRODUTION

The human immune system is a complex of cells, molecules and organs, which keeps foreign invaders such as viruses, bacteria far away from us. Without immune system, humans can hardly survive. Like other biologically-motivated intelligent computation, Artificial Immune System (AIS) which takes inspiration form natural immune system has been applied to solve many problems, including search and optimization, classification and clustering, anomaly detection, automatic control and so on.

Through the effort of researchers in the field of remote sensing, we could see AIS have been used as an alternative tool to solve remote sensing problems, such as Remote Sensing Image Classification(Zhong, 2006), image registration(Yin, 2003.) and image segmentation(McCoy, 1997).

In this paper we propose a novel algorithm for supervised remote sensing image classification based on immune B-Cell network, we call it AIBN for short. AIBN uses mechanisms of clonal expansion and immune network suppression together to train data set and use K-Nearest- Neighbours (KNN) algorithm to classify data set. Unlike other supervised classification algorithms in the fields of AIS, AIBN takes relationship among antigens into consideration and could reserve the density information of the data set by calculating adaptive radius of each antigen.

The rest of the paper is organized as follows. In section 2, we present an overview of the natural immune systems. In section 3, we review some of the current classification algorithms based on AIS. In section 4, we describe our AIBN algorithm. In section 5, we present its performance on remote sensing image classification problems and make some analysis. Finally, in section 6, we present our conclusions.

2. THE NATURAL IMMUNE SYSTEM

The immune system could perform many tasks, such as learning, memory, pattern recognition, optimization, noise tolerance, generalization, and distribution detection. The basic of those performances are built on the ability of immune system that can recognize all cells within the body and categorize them as self or non-self. Then it removes those non-self substances through some immune mechanisms and immune processes. There are two type of immunity, innate and adaptive. Only adaptive one is normally concerned with. The adaptive immune response is mainly based on the behaviours of two types of lymphocytes: B cells and T cells.

When a pathogen invades the body, some B cells recognize the antigens on the surface of antigen presenting cells with different affinity. With the help of T cells, those B cells begin to proliferate and mutate to produce antibodies for matching the antigen better. It is said memory cells transiting from high affinity B cells retain in the body for a long periods of time after the response. And these memory cells will response more rapidly and powerfully to a similar pathogen in the future response.

Above mentioned content is from the viewpoint of clonal selection. There is a controversial theory, immune network theory, which holds that B cells are interconnected through their idiotopes and parotopes. Then the recognition between B cells could results positive or negative affect to the response to the

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

antigens. Although these two theories are controversial, both of them can used for AIS inspiration.

3. CLASSIFICATION BASED ON AIS

As AIS emerged in the1990s, classification has been become an important application area of AIS. Classification systems based on AIS have attractive features inherited form biological immune system. We could see a lot of works for applying immune metaphors to classification have been reported. Jerome and Carter's Immunos-81 is the first supervised classification system in the field of AIS (Jerome, 2000). Their model consists of T cells, B cells, antibodies, and their interactions. It's very complicated classification system with modest success. Later, Watkins proposed AIRS (Watkins, 2004) based on the principle of resource limited artificial immune systems and it has been applied in many areas. Although AIRS receives success, researches on classification based on AIS are still moving on. Application areas range form document classification (Greensmith, 2002), e-mail classification (Andrew, 2003) to classification of Petroleum Well Drilling operations (Adriane, 2007). Other mechanisms are been used for inspiration for building new algorithms (Grazziela, 2007).

Classifiers based on AIS are almost on assumption that the training set constitutes the initial antibodies' population of the system, and a suppression mechanism that tries to reduce this training set into a smaller subset. This subset is supposed to contain the most significative samples, without losing much capability of generalization (George, 2005). Let us take AIRS classifier for example. AIRS is concerned with developing a set of memory cells that give a representation of the training data. When an antigen invades in, firstly AIRS evolves a candidate memory cell through the process of clone, mutation and resource competition, and then determines whether this candidate cell should be added to the pool of memory cells or not.

Without discussing the attractive virtue of AIRS, we take some problems of AIRS or other classifier based on AIS into consideration. We notice that after initialization these algorithms ignore the relationships among antigens. For example, AIRS employs one-shot learning on each antigen. However in reality, there are different antigens in the body at the same time and they interact with each other. We also notice the purpose of training process is to eliminate redundancy within the training data. In AIRS, this information reduction process is controlled by some parameters. Whether a candidate memory cell can be added in the pool of memory cells or not depends on the one parameter-the affinity threshold scalar. By this way, all the memory cells are positioned uniformly in antigens' space. But the densities of antigens in antigens' space may be vary form one to another. The acceptation or rejection of the candidate memory cell is determined by one same parameter. This will lead to lose the density information in the data set, and finally results in that the borders different types of memory cells might be fuzzy and overlap; the presentation of original data set might be distortion; inaccuracy emergences using KNN to classify at the last stage.
Keeping aforementioned words in mind, we attempt to build a new algorithm based on immune B-cell network for remote sensing image classification to overcome those deficiencies in next section.

4. THE AIBN ALGORITHM FOR REMOTE SENSING IAMGE CLASSIFICATION

Although several features could be used in remote sensing image classification, we only concern with the spectral features. In multi-spectral remote sensing images, every pixel could be presented as a vector by grouping gray levels of *N* bands, such as $x = \{x_1, x_2, \cdots, x_N\}$. Classification is a process of assigning each pixel to one of *nc* classes.

In our algorithm only immune B cells are under our discussion. We don't distinguish between notion of B cell and antibody. Then our immune network is composed by many antibodies or B cells and their interaction. In AIBN, a training spectral vector is presented by an antigen; a B cells in the network after training which used in KNN algorithm is referred as a memory cell in AIRS. We take Euclidean distance as the measurement of the similarity in our algorithm.

AIBN is an iterative procedure that can be summarized into three main phases:
1. Clonal expansion and affinity maturation: each antigen is presented to antibodies of same type; the most simulated antibody within some range goes through clone and mutation processes.
2. Network suppression: the interaction between the antibodies of same type is quantified and if one recognizes another, then one of them must be removed from the network.
3. KNN classification: after training KNN algorithm is run to determine the type of each pixel.
Now we list the pseudocode of AIBN firstly in figure 1, and then present the description of the parameters and details about AIBN.

```
1 Initialize parameters: G, a, b, calculate E, aff1, aff2, ⋯, affnc
2 For iteration 1 to G do:
  2.1 For each antigen Ag do:
    2.1.1 Select the best matching antibody Ab;
    2.1.2 If Ab is within Ag's recognition radius R
          Clone and mutate Ab;
        end;
    end
  2.2 Calculate the local density den for each Ab within the
      average radius;
  2.3 Calculate the suppression radius r of each Ab;
  2.4 Suppress antibodies giving survival priority for those
      with smaller suppression radius r.
End
3 Perform the KNN algorithm.
```

Figure1. The pseudocode of AIBN

In AIBN, there are four user defined parameters *G*, *a*, *b* and *K*.
1. *G*: iteration times;
2. *a*: adjustable radius scalar;
3. *b*: minimum radius scalar;
4. *K*: an integer used in KNN algorithm.

The parameter *a* is within the range of [0, 1], so does the parameter *b*. And *a* is larger than *b*. The parameter *a* is used in two occasions. One is in determining recognition radius *R* and another is in determining the largest suppression radius. The parameter *b* is used to determine the smaller suppression radius.

There are still some unique notions in AIBN.

1. Normal radius E: average distance among all the antigens;

$$E = \frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\|Ag_i - Ag_j\right\|}{(n(n-1)/2} \quad (1)$$

2. Recognition radius *R*: determine the best matching antibody cloning and mutating or not in step 2.1.2;

$$R = a \times E \quad (2)$$

3. Average radius aff_i (i=1,2, ⋯,nc): average distance among the antigens of the same type;
4. Local density *den*: the number of antibodies located within average radius aff_i of a certain antibody;
5. Suppression radius *r*: a certain antibody has a suppression effect on other antibodies within the suppression radius.

Step 2.1 is corresponding to clonal expansion and affinity maturation phase; step 2.2, 2.3and 2.4 is referred as Network suppression phase; step 3 is corresponding to the final classification phase.

In step 1 of the algorithm, the parameters are set and initial population of antibodies is randomly generated. The number of antibody sets is *nc*, each set only reserve same type of antibodies. Also, normal radius *E* and average radiuses of different classes are calculated.

In step 2.1, all the antigens are presented to the antibody set of same class one by one. That means when one antigen *Ag* is presented, distances between this antigen and all these antibodies is calculated. The best matching antibody *Ab* which is located nearest this antigen is selected. If the distance between *Ag* and *Ab* is larger than recognition radius *R*, One clone of *Ab* is allowed to be generated and then mutation operation is carried out. We use the simple mutation mechanism in aiNet (De Castro, 2000), the formula of the mutated antibody Ab' is:

$$Ab' = Ab - rand \times (Ab - Ag) \quad (3)$$

In the formula, *rand* is a random number uniformly generated between 0 and 1. If the distance between *Ag* and *Ab* is smaller than *R*, it means *Ag* had been recognized by AIBN and redundant learning is unnecessary. For the sake of avoiding over head of antibodies, only one clone is allowed for the selected antibody.

In step 2.2, local density of each antibody is estimated. Its value is the number of antibodies within a hypersphere centred in the antibody and with the average radius aff_i. The calculated densities are used to determine the suppression radiuses of the antibodies.

In step 2.3, suppression radiuses of the antibodies are calculated depending on normal radius *E*, adjustable radius scalar *a*, and minimum radius scalar *b*. Hypothesis that the density of an antibody is presented by *den*, den_{max} is the largest density in the antibodies set and den_{min} presented the smallest one. Then the formula of the suppression radius *r* is:

$$r = [a + \frac{b-a}{den_{\max} - den_{\min}} \times (den - den_{\min})] \times E \quad (4)$$

Note that the suppression radius of an antibody located at a region with the highest density will have its value set to $b \times E$. The others will have a larger radius, as the density decrease.

The largest suppression radius is set to $a \times E$. This mechanism reserved density information of the antibody sets through the alterable suppression radiuses.

After the suppression radiuses of all antibodies are defined, the network suppression operation takes place. If the distance between two antibodies is smaller than the suppression radius of one of them, the one with larger radius is suppressed. Then it is removed form the B cell set. Note that after network suppression, distribution information of the invading antigens is presented by residual antibodies through a compact form.

After training has completed, the residual network B cells are available for classification. In step3, the classification is performed in a K-Nearest Neighbour approach. Each B cell is iteratively presented with each spectral vector. The classification of a spectral vector is determined by using a majority vote of the outputs of the *K* most matched B cells.

5. RESULTS AND DISCUSSION

Then we apply the proposed algorithm above to remote sensing image classification task and compare the results with those of Maximum Likelihood classifier and the well-known AIRS.
The images used for classification are acquired by the Thematic Mapper (TM) sensor in ERDAS software suit. The purpose of the experiment is to classify the remote sensing images into four classes using the six bands of seven. The four classes are supposed to be water, mountain, road and building. According to the composite image, as figure.2 (a) shows, 517 ground reference points are selected as training samples and 599 ground points are chosen as testing samples.
Figure.2 (d) illustrates the classification results using AIBN. The parameters to be specified in AIBN are G=50, a=0.1, b=0.05, k=3. Figure.2 (b), (c) illustrates the classification result using Maximum Likelihood classifier and AIRS2 respectively. Table 1 shows the classification accuracy of four classifiers.

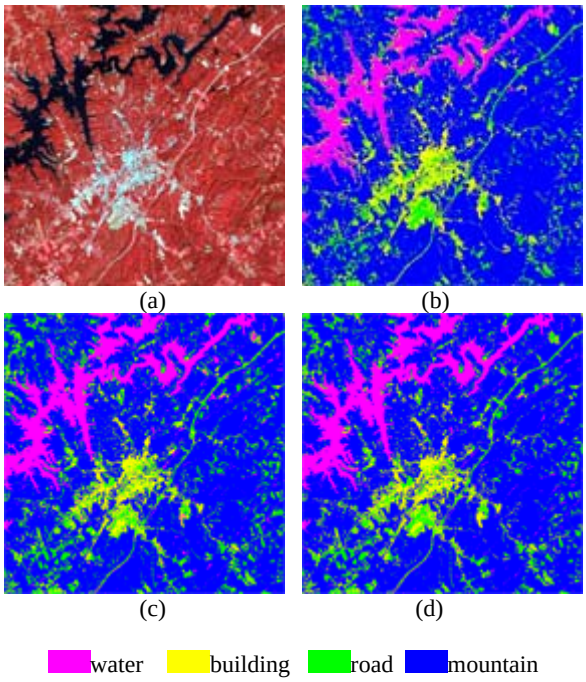

Figure2. (a) Composite image of band 4,3,2; (b) False-color image of the classification map of Maximum Likelihood; (c) False-color image of the classification map of AIRS2; (d) False-color image of the classification map of AIBN;

Accuracy	Maximum likelihood	AIRS2	AIBN
Overall	86.89%	87.98%	88.57%
Kappa	0.824	0.839	0.843

Table1. Comparison of four methods of classification

From the classification results, it is observed that the AIBN classifier produces better classification results than Maximum Likelihood and AIRS classifier. When compared AIBN with Maximum Likelihood, AIBN does not need to specify distributional form for the underlying model and compute any parameter of the model. It has some attractive biological features, such as self-regulatory, self-adaptive, and nonlinear and so on. When compared with AIRS, AIBN takes relationship among antigens into consideration and is able to learn the density information of the data. The number of user defined parameter is fewer than that of AIRS. For reason above, AIBN works very well in this experiment.

Then we will discuss the influence of user defined parameters on residual B cells in network in AIBN. Its value influences the number of B cells and accuracy of the classification. In our experiment, we notice when G is larger than 10, the number of left B cells after training and the accuracy of the classification are relatively stable. And the parameter *a* has a strong influence on the number of residual B cells. The influence of *b* is weaker than that of *a*. Table2 shows when *a* increases from 0.1 to 0.3, the number of left B cell decreases quickly, and accuracy is also decreases.

a	0.1	0.15	0.2	0.25	0.3
Number of B cells	137	96	62	43	27
Accuracy (%)	88.75	86.98	84.66	82.87	80.43

Table2. Influence of *a* on number of left B cells

6. COCLUSION

In this paper, a novel algorithm based on artificial immune B-Cell Network, AIBN, is proposed. AIBN is a little different from other classifiers based on AIS, such as, only one clone is permitted; interaction between antibodies take place after the process that all the antigen are presented; local densities of antibodies are need to be estimated and so on. Comparisons between two classifiers and our algorithm are made. The experiment results demonstrated AIBN is an effective method in remote sensing image classification and its average performance is better than Maximum Likelihood classifier and AIRS which is a successful supervised classifier based on AIS. Finally, we believe that AIS will provide an alternative way to solve some of the remote sensing problems well.

REFERENCES

Adriane, S., Jose M., 2007. Artificial Immune Systems for Classification of Petroleum Well Drilling Operations. In: ICARIS 2007, LNCS 4628, PP. 47-58.

Andrew, S. Alex, F. et al., 2003. AISEC: an artificial immune system for e-mail classification. In: Evolutionary Compution, 2003. 1,pp. 131-138.

De Castro, L., 2000. An Evolutionary Immune Network for Data Clustering. In: Proc. of the IEEE SBRN 2000. Brazil，pp. 84-89.

George, B., Tiago, B., 2005. Adaptive Radius Immune algorithm for data clustering. In: ICARIS 2005, LNCS 3627, PP. 290-303.

Grazziela. F., Nelson, E. et al., 2007. The SUPRAIC algorithm: a suppression immune based mechanism to find a representative training set in data classification tasks. In: ICARIS 2007, LNCS 4628, PP. 50-70.

Greensmith , J., Cayzer, S., 2002. An Artificial Immune System Approach to Semantic Document Classification. In: Proc. ICARIS 2002. pp 136-146.

Jerome, H., Carter, M., 2000. The Immune System as a Model for Pattern Recognition and Classification. Journal of the American Medical Informatics Association. 7(1), pp.28-41.

McCoy, D., Devarajan V., 1997. Artificial Immune Systems and Aerial Image Segmentation. In: Systems, Mans, and Cybernetics and Stimulation, 1997 IEEE International Conference on System. 1, pp867-872.

Watkins, A. Timmis, J., 2004. Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm. Genetic Programming and Evolvable Machine. 5,PP. 291-317.

Yin, G. 2003. The Multi-Sensor Fusion: Image Registration Using Artificial Immune Algorithm. In: *Soft Computing Techniques in Instrument, Measurement and Related Applications* 2003, pp. 32-36.

Zhong, Y., Zhang L. et al. 2006. An Unsupervised Artificial Immune Classifier for Multi/Hyperspectral Remote Sensing Imagery. IEEE Trans. On Geosciences and Remote Sensing, 44(2), pp. 420-431.