UNSUPERVISED CLASSIFICATION BASED NEGATIVE SELECTION ALGORITHM

ESMA BENDIAB*, M. K. KHOLLADI**

*Modeling and Implementation of Complex Systems Laboratory (MISC Laboratory)
**Modeling and Implementation of Complex Systems Laboratory (MISC Laboratory)
Bendiab_e@yahoo.fr,Kholladi@yahoo.fr

Abstract

In the last decade, artificial life has been considered as a promising area for rising challenges to unresolved computational problems. Inspired by natural phenomena, its study focuses on the exploration of complex systems. Neuronal networks, genetic algorithms and more recently artificial immune systems are examples. Artificial Immune Systems (AIS) are one type of intelligent algorithms inspired by the principles and processes of the human immune system. Emulating the discrimination mechanism of the natural system, negative selection algorithm of AIS has been successfully applied on change and anomaly detection.

This paper describes initial investigations in applying negative selection algorithm on pixel classification by maintaining a population of detectors that remove undesired patterns. Its purpose is to find several detectors which do not match to self in the population. We make use of an Euclidian space with an Euclidian performance measure on color images. The experimental show promising results. The obtained classifier is effective and feasible.

Key words: Complex Systems, Artificial Immune Systems, Negative Selection Algorithm, Image Classification.

1 INTRODUCTION

Artificial life (AL) is devoted to a new emerging area that researchers exploit to solve real world problems, using ideas gleaned from natural mechanism. Introduced by Turing and Von Neumann on the models of reaction diffusion and automats, artificial life focuses on the study of emergent phenomena. Its study for those phenomena shows that a system complexity is created from simple interactions either by introduction of knowledge on behalf of the user. Artificial life is based on the observation of interactions between entities. These entities can be simple or complex.

Christopher Langton [1] defines AL as: "the study of systems built by the human being who presents characteristic behaviors of the natural alive systems".

Any artificial life system is typified by two characteristics: evolution and emergence. The fist offers the adaptability to a dynamic environment, when an unforeseen event occurs. Thus the system can evolve. The second is a process where phenomena on a certain level result from low levels interactions. These emergent properties are created when something becomes more than the sum of its parts. The property of emergence is related to complexity, where the increase in the diversity of the elements. The increase of connections number between these elements and the set of nonlinear interactions lead to behaviors not easy to predict.

A complex system can be defined [1] [2] as a system made up of many differentiated elements which interact with each other in a nontrivial way. Neuronal network, genetic algorithms and more recently artificial immune systems are examples of artificial life systems.

Artificial immune systems are relatively new class of meta-heuristics that mimics aspect of the human immune system to solve computational problems [3]. Those methods have shown particular promises for models recognition, machine learning, communication, adaptation, memorization, auto organization and distributed controls [3] [4]. They are massively distributed and parallel, highly adaptive and reactive, evolutionary where learning is native. They are defined as [5]:"the composition of intelligent methodologies, inspired by the natural immune system for the resolution of the problems of the real world".

Growing interests are surrounding those systems due to the fact that natural mechanisms such as recognition, identification, and intruders‟ elimination which allow to the human body to reach its immunity suggest new ideas for computational problems. Artificial immune systems consist of three typical intelligent computational algorithms termed negative selection, clone selection and immune network theory [5][6].

Artificial negative selection is a computational imitation of self non self discrimination [7]. This unfairness is considered as one of the major mechanisms in the complex immune system.

In this paper, we describe our first investigation for solving the problem of the color image classification based on the negative selection algorithm. It is not the preliminarily exploration in this field. Since the artificial immune system was applied generally on the problem of data classification as a problem of discrimination [9] , as well as data analysis and data clustering [10]. For image classification, researchers intend to remote sensing image. We can mention [11-13]. Each one of those works focus on different aspect and different algorithm of AIS. Such as: the clonal selection algorithm [12], network model [13] or moreover combining AIS with other approaches [11].

Unlike those works, [14] proposed an AIS based negative selection algorithm for aerial image segmentation. The authors use hyper spheres detectors in a Euclidian shape space and the sum of squared differences as an affinity metric for grey levels images.

With regards to our approach, we propose the use of the negative selection algorithm to color image classification with a Euclidian shape space and the Euclidian distance as affinity measure.

The rest of the paper is organized as follows. Section 2 contains relevant background information and motivation regarding the negative selection algorithm. Section 3 describes the problem of the image classification via the segmentation. This is followed by the proposed approach in section 5. Section 6 includes experimentations and parameters analysis. The paper ends with a conclusion and future works

2 THE NEGATIVE SELECTION ALGORITHM

The negative selection algorithm is a supervised learning algorithm, based population, introduced by Forrest and al [7] to computer security, network security and anomalies detection problems. It is based on the discriminatory mechanism of the natural system. The goal of the negative selection algorithm is to classify a bit or string representations of real-world data (termed “antigen”) as normal or anomalous (antigens: entities witch invade an organism) [7][8]. It operates in two steps training and testing phases. The basic idea of the negative selection algorithm is to generate a number of detectors in the complementary space and then to apply these detectors to classify new, unseen, data as self or non self. The algorithm can be is summarized in the following steps [7]:

- Define self as a set of S elements of length l over a finite alphabet, a collection that needs to be protected or monitored.
- Then generate a set of D detectors, so that each detector fails to match any element in S. Instead of exact or perfect matching, the method uses a partial matching rule, in which two strings match if and only if they are identical at least at r contiguous positions, where r is suitably chosen parameter.
- Monitor S form changes by continually matching the detectors in D against S.

A schematically representation of the algorithm can be found in figure 1 [4]. In the original description of the algorithm, candidate detectors are generated randomly and then tested (censored) to see if they match any self string. If a match is found, the candidate is rejected .This process is repeated until a desired number of detectors are generated.

A probabilistic analysis is used to estimate the number of detectors that are required to provide a certain level of reliability. The major limitation of the random generation approach appears to be computational difficultly to generate valid detectors, which increase exponentially with the size of self (S).

The algorithm is based on the four following principles [5]:

- Each copy of the detection algorithm is unique;
- detection is probabilistic;
- A robust system should detect, probabilistically, any foreign activity rather than to seek the specific patterns of known intrusions.

No previous knowledge is required

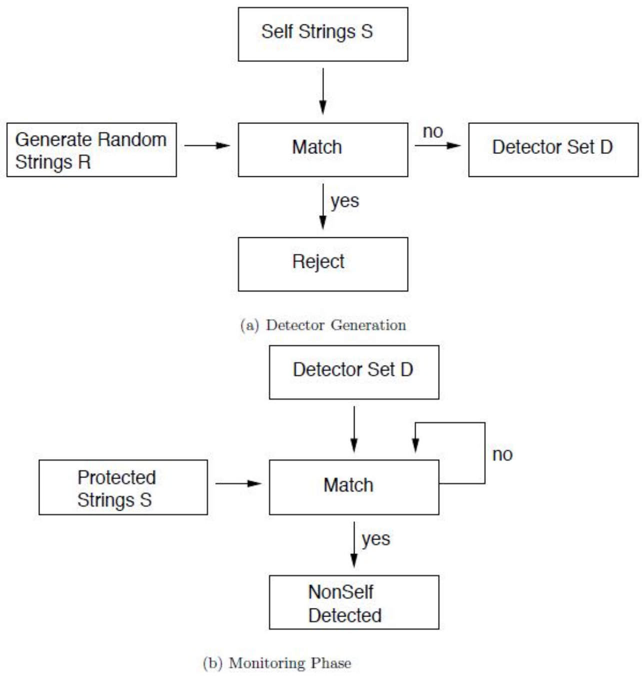

Figure 1: Negative Selection Algorithm by Forrest [7].

3 THE CLASSIFICATION PROBLEM

An image is generally defined as a function *f*:

$$f : \xi \to T$$

Which associate with each pixel $p = (x, y)$ from the space $\xi \subset N^2$ the value *v* in *T*.

This space can be composed of: a gray level value ($T = \{0,\ldots\ldots, 255\}$), a tri-chromatic representation ($T = [0, 1]^3$ or $T = \{0,\ldots\ldots\ldots, 255\}^3$) or a spectral response ($T \subset R^n$) for an image of *n* spectral bands).

The segmentation aims to partition the pixel space ξ of an image *f* into a set of *K* homogeneous regions $\{R_k\}$, $1<k<K$ under certain criterion (e.g. the pixels values *v* in each region).

It is therefore a function $\pi : \xi \rightarrow \Re$ that associates each pixel *p* having index *k*, with the region R_k to which it belongs. Each region R_k is constructed as a related component, in term of a set of adjacent pixels of value *k*.

More formally [15], we defined the concept of a discreet way P_{pq} of a pixel *p* to a pixel *q* as the set of pixels:

$P_{pq} = \{pi;\ i = 0; \ldots\ldots ; m\}$ with $p_0 = p$, $p_m = q$, $\forall\ i = 0; \ldots\ldots\ldots ;(m-1)$; p_i is adjacent to p_{i+1}.

Two pixels $p=(x_p\ ,y_p)$ and $q=(x_q\ ,y_q)$ are adjacent if their distance is aqual to 1. The pixel neighborhood of a pixel *p* denoted by *N (p)* is the set of pixels *q* adjacent to *p*.

Finally, a region R_k is a related component if it checks: $\forall p, q \in R_k, \exists P_{pq} avec \forall p_i \in p_{pq}, p_i \in R_k$ which tantamount to $\pi(p_i) = k$

The unsupervised classification [16] also called data partitioning or clustering is to gather data within homogeneous groups called classes. Applied to an image, the classification can be represented as a function $\pi : \xi \rightarrow C$ which associates to each pixel *p* having the index *k* with the class C_k to which it belongs. As well for the regions R_k, the content of the classes C_k must be homogeneous (for example, pixels compose a class should have similar values *v*). However, regions R_k produced by the segmentation, class C_k are not a related component. To relieve this limitation, a classification algorithm must assure the spatial connectivity of the produced classes. Therefore, the spatial information is present in images via connectivity and adjacency notions between neighborhood pixels. Moreover, the application of the classification must be in the pixel space ξ rather than in the characteristic space *T*.

4 THE PROPOSED APPROACH

4.1 Description of the algorithm elements in the context of the classification problem

4.1.1 *Antibodies*

Elements which are responsible of antigens identification in nature, they compose the set of detectors. They can be of two types:

- Relevant detectors: (solution) set of elements which are capable of identifying non-self.
- Not relevant detectors: set of elements which are able to identify self.

4.1.2 *Antigens*

They represent the set of segmented pixels. They can be of two forms:

- The Self: represent the desired segments.
- The non-self: represent the unwanted segments.

A schematically representation of antibody and antigen in the context of image classification can be shown in figure 2.

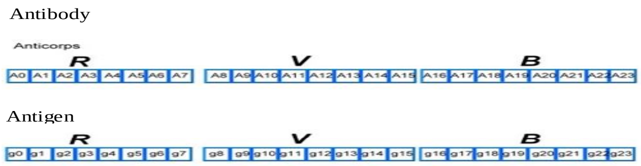

Figure 2: Representation of the AIS elements in the context of image classification by an Euclidian Shape pace

4.1.3 *Affinity measure*

The interaction between the specific self detectors and antigens is measured by the affinity measure. It represents the relationship between antibody and antigen with a considered distance (Ab-Ag).

For this purpose, we used the Euclidean distance in Euclidean space between the average values of RGB color values and RGB detectors. This distance is defined by equation Eq.1.

$$\text{Affinity} = | Ab - Ag | \qquad 1$$

4.2 Description of the approach

The approach describes a population of M individuals, detectors, which identify the shape of a set P (antigen). In this context, the negative selection algorithm must define a set of detectors which match to the non self shape.

The fundamental difference between the negative selection algorithm and other approach for image classification is in the nature of the identification itself. Its purpose is to find several detectors which do not match to self in the population (desired Segments).

The principle of the algorithm, as defined in the previous section, adjusted to the color images classification is presented in the following steps:

1- Initialization.
2- Censoring: generation of detectors.
3- Monotoring (control): classification phase.

A detailed description of the steps is given bellow:

1- Initialization.

- Generate the Self detectors
- Selection of the threshold.
- Image test reading ;
- Randomly generation N candidates detectors.

2- Censoring phase: (Learning phase)

- Affinity assessment: determines the affinity between detectors and the self elements.
- Selection: If a candidate detector matches a self element, then this detector is removed. Otherwise the candidate is added to set of relevant detectors (possible Solutions);

3- Classification

After the censoring phase, all relevant detectors are prepared for the image classification; it is done by the following steps:

- Match the image pixels with each competent detectors.
  - If the pixels match detectors according threshold then
- the pixel belongs to non self set; and it is eliminated;
  - Else
- The pixel belongs to the Self set (segments desired).
- It is classified to the nearest segment, which verifies maximum distance (affinity).

5 EXPERIMENTAL RESULTS AND EVALUATION

To verify the performance of color image classification based negative selection algorithm, we make use of synthesis image shown in figure 3.

We wish to emphasize that the purpose of this section is not to measure the performance of our method compared to other existing methods, but to demonstrate and validate the feasibility and effectiveness of the approach.

The following parameters of the negative selection algorithm were set:

- Number of self detectors;
- The parameter r which determines the perfect or partial matching.

We measure classification success using the well-known criteria producer's accuracy or completeness (the number of pixels that are correctly assigned to a certain class divided by the total number of pixels of that class in the reference data) and user's accuracy or correctness (the number of pixels correctly assigned to a certain class divided by the total numbers of pixels automatically assigned to that class.

Therefore, the overall accuracy, O, is calculated as shown in equation 2, where A is the number of pixels assigned to the correct class and B is the number of pixels that actually belong to that class. It is a good measure of the accuracy of a classification scheme.

$$O = \frac{a}{b} * 100 \qquad (2)$$

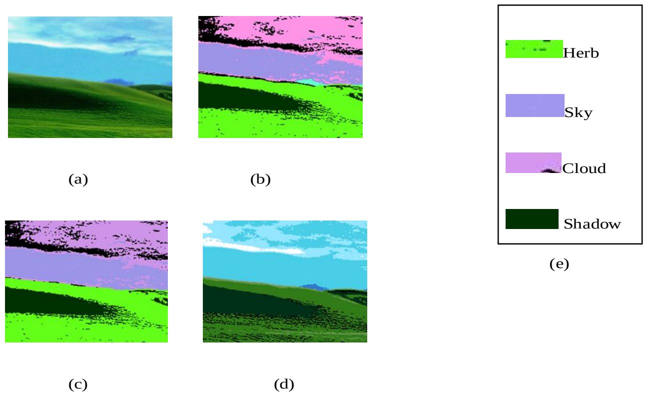

Figure 3: (a) image test used in the experiments. (a) (b), (c), (d),Images resulted with different overall accuracy depending on parameters

With the considered image, in figure fig.3(a), five distinct classes can be found: Herb, Sky, cloud, shadow and *others*. figure3(e) show the results with one color per class: Herb in light green, Sky in blue, Cloud in ,Shadow in dark green, and others in black. The error matrix of the experiment is shown in Table 1.

Compare figure fig.3 (b) with figure fig.3(c) and figure fig.3(d), change in the threshold and the number of self detector influence the classification quality. The overall accuracy decreases from 70,00% figure 3 (b) to 69.8% figure 3 (c). For a number of applications this accuracy is still acceptable.

The overall accuracy calculation of fig 2 (b) is presented on table 1.

Table 1: The overall accuracy.

Producer Accuracy(Completeness)	User's Accuracy(Correctness)
Sky=67.46%	Sky =77.05%
Cloud=64.08%	Cloud =67.41%
Shadow=77.11%	Shadow= 73.14%
Other=86.67%	Other= 89.00%
Overall accuracy = 70.0%	

6 CONCLUSION AND FUTURE WORKS

This paper describes the application of the negative selection algorithm of artificial immune systems, to the color image classification problem. According to the experiment's procedures and results of the previous section, the conclusions are as shown below.

The Negative selection algorithm of the artificial immune system can be applied to the color images classification. From the experimental results, the obtained classifier is effective and feasible.

However, we recognize that the proposed method should be compared with other classifiers method in order to evaluate the quality of the method. In future work, we will further investigate the potential influence of the other parameters and also consolidate our results using more test data and alternative indices for measuring the affinity metric, the representations space also the colorimetric space of images. More than, we will apply the proposed method to other classification problems object recognition, particularly to larger problems and other types of problems.

BIBLIOGRAPHIE

[1] M. Sipper, "An introduction to artificial life", Explorations in artificial life (special issue of al expert). Miller freeman. September 1995.

[2] T. Carden, "Image processing with artificial life", Artificial Intelligence with mathematic 2OO1- 2002.

[3] De Castro, L. R. and Timmis, J. , "Artificial Immune Systems: A New Computational Intelligence Paradigm", Springer-Verlag New York, Inc., Secaucus, NJ, 2002

[4] Hart, E. Timmis, J., "Application areas of AIS: The past, the present and the future", Applied Soft Computing, v.8 n.1, p.191-201, January, 2008 .

[5] Dasgupta,D, "Artificial Immune Systems and their applications", Springer-Verlag (Ed.), 1999.

[6] Timmis, J., "Artificial immune systems---today and tomorrow", Natural Computing: an international journal, v.6 n.1, p.1-18, March 2007

[7] Forrest, S. and Al, "Self-Non self Discrimination in a Computer", Proceedings of the IEEE Symposium on Security and Privacy, p.202, May 16-18, 1994

[8] Ji,Z., Dasgupta, D., " Revisiting Negative Selection Algorithms", Evolutionary Computation, Massachusetts Institute of Technology v.15 n.2, p.223-251, 2007

[9] Igawa, K. and Ohashi, H., "A Discrimination Based Artificial Immune System for Classification", Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Vol-2 (CIMCA-IAWTIC'06), p.787-792, November 28-30, 2005

[10] De Castro,L.N. and Von Zuben, F.J., " aiNet: an artificial immune network for data analysis". In: Abbas, H.A., Charles, R.A.S., Newton, S. (Eds.), Data Mining: A Heuristic Approach, Idea Group Publishing, USA. pp. 231-259. 2002

[11] Tong, H., Zhao,M., and LiZ., "Applications of Computational Intelligence in Remote Sensing Image Analysis", (Eds.): ISICA 2009, CCIS 51, pp. 171–179, Springer-Verlag Berlin Heidelberg 2009

[12] Cheng, J., and al, "Image Segmentation Based on Chaos Immune Clone Selection Algorithm ", D.-S. Huang, L. Heutte, and M. Loog (Eds.): ICIC 2007, LNAI 4682, pp. 505–512, 2007. Springer-Verlag Berlin Heidelberg 2007

[13] Xua, S., Wu, Y., «An algorithm for remote sensing image classification based on artificial immune B cell network", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B6b. Beijing 2008.

[14] McCoy, D.F. Devarajan, V., "Artificial immune systems and aerial image segmentation", Proceedings of the IEEE International Conference on Systems Conference on Systems, Man, and Cybernetics: Computational Cybernetics and Simulation, Orlando, FL, USA, pp. 867-872. 1997.

[15] LU, D., and WENG, Q., "A survey of image classification methods and techniques for improving classification performance", International Journal of Remote Sensing Vol. 28, No. 5, 10 , 823–870, March 2007.

[16] Ilea, D. and P. Whelan,"An adaptive unsupervised classification algorithm based on color-texture coherence". IEEE Transactions on Image Processing 17(10), 1926–1939. 2008