

Hybrid Systems and Artificial Immune Systems: Performances and Applications to Biomedical Research

Vitoantonio Bevilacqua, Cosimo G. de Musso, Filippo Menolascina,
Giuseppe Mastronardi, and Antonio Pedone

Department of Electronics and Electrical Engineering, Polytechnic of Bari,
Via E. Orabona, 4
70125 – Bari, Italy
bevilacqua@poliba.it

Abstract. In this paper we propose a comparative study of Artificial Neural Networks (ANN) and Artificial Immune Systems. Artificial Immune Systems (AIS) represent a novel paradigm in the field of computational intelligence based on the mechanisms that allow vertebrate immune systems to face attacks from foreign agents (called antigens). Several similarities as well as differences have been shown by Dasgupta in [1]. Here we present a comparative study of these two approaches considering evolutions of the concepts of ANN and AIS, respectively hybrid neural systems, Artificial Immune Recognition Systems (AIRS) and aiNet. We tried to establish a comparison among these three methods using a well known dataset, namely the Wisconsin Breast Cancer Database. We observed interesting trends in systems' performances and capabilities. Peculiarities of these systems have been analyzed, possible strength points and ideal contexts of application suggested. These and other considerations will be addressed in the rest of this manuscript.

1 Introduction

The nervous system and the immune one are probably the most complex systems in the vertebrates. Both have been shown to be necessary components for adaptability and then survival to the environment. Learning, memory and associative retrieval are the keywords for these systems and on these aspects researchers have focused their interests in order to replicate such behaviors in artificial systems. Artificial Neural Networks grew in this context and they are nowadays one of the most useful and powerful tools for data classification, clustering and prediction. Starting from the '40s several other bio-inspired models have been successfully proposed including Genetic Algorithms (GA) [2] and Swarm Intelligence [3]. Artificial Immune Systems (AIS) [4] followed this scientific trend. Proposed for the first time by Farmer et al in 1986 [5] AIS field of research underwent a noticeable boost in the mid '90s with the research carried out by Dasgupta [1][6] and then with the pioneering work of de Castro, Timmis [7] and Hunt[8]. AIS systems are knowing a remarkable spread in the scientific community because of their flexibility and potentialities. Someone could

argue that there's a great emphasis on these systems and it is not for a chance. AIS has been shown to be a good choice in several fields of application and one of the most versatile tools researchers are able to use. On the other hand, in their most common implementations, they are characterized by low computational costs; a critical aspect in delicate contexts like e.g. real time computing. Just like Artificial Neural Networks, even Artificial Immune Systems have known different implementations and optimizations through the years; supervised and unsupervised flavors being the former first and resource competition and negative selection the remaining. Hybrid Neural Systems are, probably, among the most famous evolution of Neural Networks being combination of Neural Networks with other Intelligent Systems that are able to conjugate strength points of all of the constituent systems. A Neural-Genetic approach for the problem of breast cancer classification has been described in [9]. However different approaches have been shown to be able to perform quite well on the same problem and, in particular, in this paper, we will focus on the comparison of novel immune systems models and advanced hybrid neural systems. The common platform selected for this comparative study is the Wisconsin Breast Cancer Database (WBCD). The WBCD is fundamentally based on the flattening principle. Composed by 699 cases, each defined by 11 fields, this dataset collects breast cancer cases observed by W.H. Wolberg in the late '80s [5]. Sixteen cases lack of one parameter. Database entries are characterized by the following structure:

(ID, Parameters, Diagnosis)

where ID is the primary key, parameters fields contain numerical values associated to 10; the last field in the entries contains medical diagnosis associated to the cases, it is a binary value representing malignant/benign tumor.

The former first ten indicators are extracted analyzing images obtained through Fine Needle Aspiration (FNA), a fast and easily repeatable breast biopsy exhaustively described in [10].

The systems selected for this comparison are a Hybrid-Neural approach and, for AIS, Artificial Immune Recognition Systems (AIRS, [11]) and aiNet [12]. All of these systems have been described in the following sections of this paper. The comparison among the presented solutions has been assessed using global accuracy metric. Interesting trends have been observed and reported; they mainly concern specific capabilities of Artificial Immune Systems to perform better under certain conditions and algorithms' computational costs.

These and other peculiarities of the systems under investigation have been addressed in the next paragraphs that are organized as follows: firstly an overview of the benchmark dataset is given with preliminary statistical analysis, then the Neural-Genetic approach is explored. Descriptions of AIRS and aiNet implementation follow. "Comparative study" collects all the results of the three systems and "Conclusions and Further Works" paragraphs ends the manuscript with considerations and interpretations of the results giving further cues of research in this field.

2 WBCD Dataset and Preliminary Data Analysis

The WBCD dataset, as described above, is composed by 699 cases each of them defined by 11 parameters. The classification process could be faced like a function analysis problem:

$$x_{11} = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$$

with:

x_1 = Radius
 x_2 = Area
 x_3 = Perimeter
 x_4 = Texture
 x_5 = Smoothness
 x_6 = Compactness
 x_7 = Concavity
 x_8 = Concave points
 x_9 = Symmetry
 x_{10} = Fractal dimensions

The attention is then focused on the analysis of the multidimensional space defined by n-tuple associated to each case. Several statistical analyses were then carried out in order to improve knowledge about this space and create the set of information to be used in following steps of development. PCA, PFA and ICA were preliminary carried out in order to gain an adequate model; these results have been previously described [9].

3 Hybrid Neural-Genetic Systems

IDEST (Intelligent Diagnosis Expert System) is a Genetic Optimized Neural System for integrated diagnosis of breast cancer. This system is mainly based on a Feed-Forward Artificial Neural Network whose topology has been optimized using the approach illustrated in [13].

An ANN setup based on the results obtained in previous steps was trained using Back-Propagation algorithm in the variation that updates weights and biases values according to gradient descent momentum. The starting learning rate chosen was 0.3. This choice avoided the occurrence of step-back phenomena in learning phase and it gave the network sufficient energy to exit from "local minima": all this resulted in ANN's good aptitude for convergence. Stop criterion was set to SSE equal to 7E-3 or to the limit of 50000 training epochs. The system was able to complete the training phase in 13000 epochs reaching, therefore, the SSE target. The relatively contained number of epochs needed to accomplish the training step confirms the correctness of the results obtained via linear and non linear analysis and, in particular the accuracy of GA search. System validation was carried out submitting to the network the 228 cases of the validation set and calculating the number misclassified ones. The results returned by this analysis proved to be quite good: no misleading prediction was made on the 228 cases analyzed. Comparison between error distribution in training and validation sets shows low variability and confirms high precision on both classes. The highest obtainable system accuracy was then reached: it is evidently an indicative

result but potentialities of similar systems seem, now, to be supported by more concrete elements. In an *a priori* analysis we have tried to estimate the impact of most significant choices on the accuracy of the ANN's predictions. In this phase we observed how particular decisions contributed to the achievement of such a result. Leaving unchanged the phases of process described until now, but employing training and validation sets assembled ignoring results of linear and non linear analysis (i.e. using sets obtained simply dividing the original dataset in two subsets), it is possible to observe an error on the validation set equal to 4 cases on the 228. This corresponds to an accuracy of 98.6%, a competitive value indeed, which shows, the importance of the observation mostly on intrinsic variance of cases. Error oscillations, furthermore, become more evident right by the cases characterized, by a high variance. Another interesting observation can be made leaving unchanged the process described but eliminating the hybrid ranking inspired by elitist method typical of "evolutionary algorithms" contexts. Suppressing this step we can incur an error, on the validations set, approximately equal to 0.9-0.4%. Results obtained highlight the contribution to the accuracy of learning that choices in data pre-processing phase [13] have generated. Adopted devices allowed obtaining a system capable of taking full advantage of peculiar characteristics of the datasets and of the distribution of information in it.

4 Artificial Immune Recognition System

AIRS (Artificial Immune Recognition System) is a supervised learning algorithm inspired by the function of the biological immune system designed to resolve problems such as intrusion detection, data clustering and classification problems. For this work, we have been employed AIRS1, described in [14], with the goal of developing a binary classification system. In the AIRS environment the feature that should be recognized is represented by the antigens, instead the recognizer features is a pool of antibodies, called memory cells. All the memory cells are created during the training stage and are representative of the training data. The lifecycle [16] of the AIRS system is represented in figure 1.

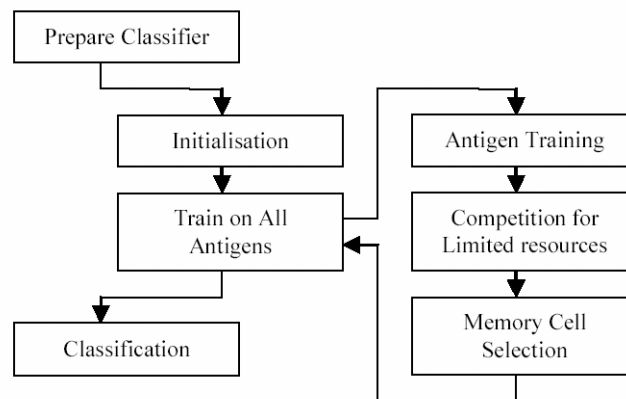


Fig. 1. An overview of the AIRS algorithm

Classification

To classify an unknown antigen the affinity between this feature and all memory cells is calculated; the class of the best match memory cell is the class of the antigen presented.

Parameter set for this algorithm is presented below:

Seed: 1
 Clonal rate: 10.0
 Hyper clonal rate: 2.0
 Mutation rate: 0.1
 Total resources: 650
 Affinity threshold scalar: 0.1
 Stimulation threshold: 0.99

Seed is the number of antigens selected to seed the initial pool of memory cell.

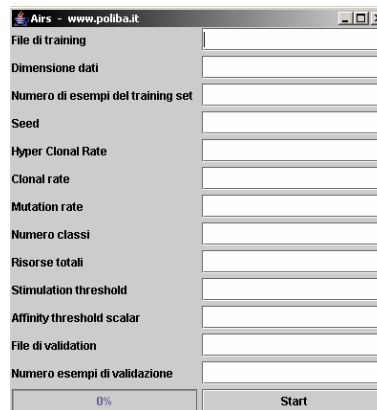


Fig. 2. Java application GUI

5 aiNet

aiNet (Artificial Immune Network) was proposed to the scientific community by de Castro and Von Zuben in the 2001 [12]. AiNet combines compression techniques with the applications of graph theory, yielding an unsupervised classifier. The centroids extracted by the algorithm are analyzed by minimizing a tree with MST (Minimal Spanning Tree - MST) where function cost is the distance among the centroids. AiNet performs well as filtering of dataset of great dimensionalities, describing the distribution of data. The cells of the immune net are represented in a space of the same dimension as the input data. The dimension of the net, that is the number of cells that composes it, is defined by a mechanisms inspired by the dynamics of the biological immune system. aiNet is based on two principles of the biological immune system: Clonal Selection and Immune Network Theory [18]. Clonal Selection defines how the system reacts to antigens' invasion: when an antigen is recognized by the system a subset of the antibodies that recognized the antigen undergoes cloning and changing by

introducing diversity in the population and then adapting itself to the invaders. This principle allows us extracting repetitive pattern in a dataset, because all the antigens with a particular sequence of values will be recognized by the same antibody. All the cells that compose the biological immune system interact with each other in total absence of external stimuli. This gave the idea of the existence of the model of such interactions, with a communication net that connects various elements. In the biological system chemical messages are exchanged determining the antibodies' survival or death. In the computational model the interaction between two antibodies is given by their relative distance; moreover nearest antibodies would recognize similar antigens, vanishing all the net. The model of de Castro & Von Zuben is based on these two principles. The recognition of the pattern influences the clone, mutation and selection task according to the principle of Clonal Selection, while the recognition of the elements of a same net is determined by the network suppression, eliminating its redundancy.

The stop criterion proposed is a max number of iterations/generations. These properties turn out to be important in the analysis of biomedical dataset where several information are available for each patient but reduced sample numbers is usual. At the end of each run of aiNet the extracted antibodies represent an internal representation of the system for the spatial distribution of antigens. In aiNet the affinity between Ab and Ag is given by their relative distance: the use of Euclidean distance is very common in this context, especially in case of real valued data. Hamming distance is preferred in case of binary strings. Classification is then unsupervised and this is a critical factor in the analysis of biomedical data where information extraction tends most often to be an explorative rather than confirmative one.

6 Results

The aim of this paper is the comparison of three different approaches to solve the same problem. Two of these techniques are supervised, IDEST and AIRS, while aiNet is an unsupervised learning algorithm. Results are reported in the table below:

Table 1. For each technique, the rate per cent of features correctly classified

	IDEST	AIRS (reducing training items)	AIRS	AINET
Training	100,00%	99,00%	98,50%	95,00%
Validation	100,00%	94,00%	100,00%	

IDEST uses an ANN (Artificial Neural Network) to classify and to recognise the data. Learning phase for this system is slower than the AIRS learning process, but performs better results both on the training and validation set. ANN explores the space of the features better than AIRS does because the latter is a single shot algorithm and for this reason is faster. Using the same training and validation set (502 and 185 items) also for the AIRS, this algorithm presents a rate of features correctly recognised for the validation set greater than for the training set. It could be argued

that the training set is greater than the double of the validation set (see Tab. 3). With a 200 entries training set we obtained that the accuracy is better for the training set. In conclusion, reducing learning set, AIRS maintains good results where IDEST starts to fail. For the classification, IDEST needs to train a lot of feature because a neural network must process a minimum set of items to learn its space, while AIRS needs to process a reduced data set, for its nature of unsupervised system, because AIRS is the extension of AIS (Artificial Immune System – unsupervised algorithm). For these investigations we consider that AIRS is better than a system based on ANN (as IDEST) when it needs to implement an on-line learning process or when the number of the features is not so large to train a neural network. Finally we have used an unsupervised system, aiNet, to understand how all the features are distributed in the space and with this technique we have calculated two clusters that classify correctly the 95% of the all items.

7 Conclusions and Further Works

In this paper we compared different and evolved approaches from different fields of research. ANNs and Artificial Immune Systems derived approaches have been compared to each other. Although small differences in accuracy levels have been observed being IDEST the top performing algorithm, some observations should be made. As shown in Tab. 2 computational costs of these three algorithms are markedly different.

Table 2. Time (in seconds) needed to complete the training phase

	IDEST	AIRS	AINET
Training set	1223	96	135

Execution times have been computed flooring the mean of 100 runs of each system. It is evident that the Neural-Genetic system is more accurate but is most time consuming. On the other hand, AIRS and aiNet show smaller computational resources, the cost being a little degradation in accuracy. However it seems that a well planned fine tuning of the parameters for AIS based systems can lead to a remarkable improvement of the results. For their characteristics in terms of computational times and levels of accuracy, immune based approach seems to be a good alternative to well-established paradigms. Questions about sensitivity analysis of parameters and optimal feature sets are currently being investigated. Some interesting behaviours of Immune based systems are even under investigation: they mainly refer to the ability of such systems, under certain conditions to outperform classical approaches (like ANN or SVM). We are trying to model these behaviours (as reported in Tab. 1, columns 3 and 4) and to understand the inner mechanisms that lead to them; it is obvious, in fact, that this aspect could turn out to be a strength point of top performing AIS based systems in biomedical field. Maintaining a low sample size can allow containing experimental costs of the research pipeline.

References

1. Dasgupta, D.: Artificial Neural Networks and Artificial Immune Systems: Similarities and Differences, *Proc. of the IEEE SMC* **1** (1997) 873-878
2. Holland, J.H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor. (1975)
3. Bonabeau, E., Dorigo, M., Theraulaz, G.: *Swarm Intelligence: From Natural to Artificial Systems* (1999)
4. de Castro, L.N., Timmis, J.I.: *Artificial Immune Systems: A New Computational Intelligence Approach*, Springer-Verlag, London, September (2002) 357 p.
5. Farmer, J.D., Packard N., Perelson, A.: The immune system, adaptation and machine learning, *Physica D* **22** (1986) 187--204
6. Dasgupta, D.: *Artificial Immune Systems and Their Applications*, Springer-Verlag, Inc. Berlin, January (1999)
7. DeCastro, L., Timmis, J.: *Artificial Immune Systems: A New Computational Intelligence Approach* (2001)
8. Timmis, J., Neal, M., Hunt, J.: An Artificial Immune System for Data Analysis **55** *Biosystems* (2000) 143--150
9. Bevilacqua, V., Mastronardi, G., Menolascina, F.: Intelligent information structure investigation in biomedical databases: the breast cancer diagnosis problem, *ISC 2005*
10. Wolberg, W.H., Street, W.N., Heisey, D.M., Mangasarian, O.L.: Computer derived nuclear features distinguish malignant from benign breast cytology. *Cancer Cytopathology* **81** (1997) 172-179
11. Watkins, A., Timmis, J., Boggess, L.: Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Machine Learning Algorithm, *Genetic Programming and Evolvable Machines*, 5-3 (2004) 291--317
12. de Castro, L.N., Von Zuben, F.J.: aiNet: An Artificial Immune Network for Data Analysis, (full version, pre-print), Book Chapter in *Data Mining: A Heuristic Approach*, H. A. Abbass, R. A. Sarker, and C. S. Newton (eds.), Idea Group Publishing (2001), USA, Chapter XII, pp. 231-259.
13. Bevilacqua, V., Mastronardi, G., Menolascina, F., Pannarale, P., Pedone, A.: A Novel Multi-Objective Genetic Algorithm Approach to Artificial Neural Network Topology Optimisation: The Breast Cancer Classification Problem, *IJCNN 2006*
14. Watkins, A.: AIRS: A resource limited artificial immune classifier - Department of Computer Science, Mississippi State University (2001)
15. Timmis, J., M. Neal, et al., An Artificial Immune System for Data Analysis - *BioSystems* **55** (1/3) (2000) 143-150
16. Brownlee, J.: Artificial Immune Recognition System (AIRS) a review and analysis- Centre for Intelligent Systems and Complex Processes (CISCP), Swinburne University of Technology (SUT) (2005)
17. Jerne N.K.: Towards a Network Theory of Immune System, *Annals of Immunology* (1973)
18. de Castro L.N.: <http://www.dca.fee.unicamp.br/~lnunes/>