RABNET: A Real-Valued Antibody Network for Data Clustering

Helder Knidel
LBiC/DCA/FEEC/UNICAMP
PO Box 6101
Campinas/SP, Brazil, 13083-852
+55 19 3788-3885
knidel@dca.fee.unicamp.br

Leandro N. de Castro
Graduate Program in Informatics
Carvalho de Mendonça, 144
Vila Mathias, Santos/SP, Brazil
+55 13 3226 0500
lnunes@unisantos.br

Fernando J. Von Zuben
LBiC/DCA/FEEC/UNICAMP
PO Box 6101
Campinas/SP, Brazil, 13083-852
+55 19 3788-3885
vonzuben@dca.fee.unicamp.br

ABSTRACT
This paper proposes a novel constructive learning algorithm for a competitive neural network. The proposed algorithm is developed by taking ideas from the immune system and demonstrates robustness in the initial experiments reported here for a benchmark problem. Comparisons with results from the literature are also provided. To automatically segment the resultant neurons at the output, a tool from graph theory was used with promising results. General discussions and avenues for future works are also provided.

Categories and Subject Descriptors
I.2.6 [**Artificial Intelligence**]: Learning - connectionism and neural nets, I.5.3 [**Pattern Recognition**]: Clustering – Algorithms;

General Terms
Algorithms, Design.

Keywords
Artificial Immune Systems, Data Clustering, Artificial Neural Networks

1. INTRODUCTION
This paper extends the work proposed in [3] on the use of features from the human immune system, and also artificial immune systems [5], to design novel artificial neural network learning algorithms. In particular, the artificial neural network to be developed is modeled as a competitive and constructive (i.e., with network growing and pruning phases) antibody network. The version presented here uses real-valued vectors to represent the weights and is, thus, termed RABNET (Real-valued Antibody Network). The resultant hybrid system has a typical competitive neural network architecture [7] similar to a one-dimensional self-organizing map [8]. In order to adapt to the environmental stimulation (input patterns), the network makes use of several features of an immune response, such as the clonal expansion of the most stimulated cells, death of the non-stimulated cells (apoptosis) and the affinity maturation of the repertoire. The network does not have a predefined number of neurons, which will be determined dynamically based on immune principles. Finally, a minimal spanning tree [2] is used to automatically determine the final number of clusters in the neural network after learning, and thus in the input data.

GECCO'05, June 25-29, 2005, Washington, DC, USA.
ACM 1-59593-010-8/05/0006.

2. RABNET ALGORITHM
Inspired by ideas from immunology, the RABNET development assumes an antigen population (**Ag**) that should be recognized by an antibody repertoire (**Ab**). At the beginning of the adaptation process, RABNET contains a single antibody (neuron) in the network and grows when required.

The RABNET algorithm is summarized in the pseudocode presented below.

```
1. Initialize randomly a single antibody in the network and define
the parameters: α, β, γ and σ;
2. While not reached the convergence criterion do:
   2.1. For each input pattern do:
      2.1.1. Present a random antigen to the network;
      2.1.2. Calculate the Euclidean distance between the antigen
      presented and the antibodies in the network;
      2.1.3. Find the winner antibody;
      2.1.4. Increase the concentration level of the winner;
      2.1.5. Update the weights of the winner;
   2.2. If (Iteration > γ) then α = σ * α;
   2.3. If (iteration is multiple of β) then Clone if necessary;
   2.4. If the concentration level of a given antibody is zero, then it
   is pruned from the network.
3. Use the MST criterion proposed to automatically segment the
neurons at the output of the network.
```

The followings steps of the algorithm can be stressed:

Competitive Phase (2.1.1-2.1.4): The competitive phase involves finding the most similar antibody $\mathbf{Ab}_K$ to a given antigen **Ag**; i.e., to find the winner neuron to a given input pattern. This antibody is said to have the highest affinity with the antigen.

Network Growing (2.3): Network growing is inspired by clonal selection, where the most stimulated cell is selected for cloning. Two bio-inspired parameters control the network dynamics and metadynamics: one related to the concentration of antigens (τ) recognized by a given antibody, and the other related to the affinity threshold (ε) between an antigen and an antibody.

Network Pruning (2.4): The strategy adopted here to perform network pruning is based on the concentration level of each antibody. If the concentration level of a given antibody $\mathbf{Ab}_i$ is zero ($\tau_i = 0$), it means that antibody i was not stimulated by any of the antigens. In this case, antibody i can be pruned. The process of network pruning is performed every iteration.

Weights Updating (2.1.5): Updating the weights in RABNET is similar to the weights updating procedure used in winner-takes-all competitive neural networks [8].

Convergence Criterion (2): The convergence criterion used checks the stability of the network topology (number of neurons in the network), and the weights stability (variation in the weight vectors). It is assumed that the network topology has reached stability if during the last 10.β iterations there was no variation in the number of neurons. Concerning the weights, they are assumed to have stabilized if the sum of their modules does not vary by more than 10^{-4} from the current iteration to the past 10.β iterations.

Defining the Number of Clusters (3): One of the main objectives of RABNET is to cluster unlabelled data. The RABNET adaptation procedure builds a network where the spatial distribution of the input patterns (antigens) is represented by the spatial distribution of the network weight vectors (antibodies). After network learning, to automatically determine the number of clusters detected by the network we propose the use of a minimal spanning tree (MST) [2], associated with an algorithm to determine and remove the inconsistent edges.

3. PERFORMANCE EVALUATION

To assess the performance of the proposed network, RABNET is applied to the Two-Spirals data set [1] and the obtained results are compared with equivalent ones in the literature. The parameters used to run RABNET are: $\alpha = 0.2$, $\beta = 2$, $\gamma = 100$ and $\sigma = 0.95$.

The input patterns are normalized within the [0,1] interval, and an affinity threshold $\varepsilon = 0.025$ is empirically obtained. Some classification results obtained by the application of different unsupervised learning algorithms, including RABNET, are summarized in Table 1. Figure 1 depicts the spatial location of each neuron in the network and the result of the application of the MST criterion to the network generated. Before applying the MST cluster identification procedure, the two spirals in Figure 1 were connected by an edge that was considered inconsistent and thus pruned.

Table 1. Classification results for the Two Spirals data set. (Based on [6] and [4].)

Algorithm	Number of Units	Error Rate (%)
GCS	145	0
DCS-GCS	135	0
LVQ	114	11.9
SOM	144	22.2
ESOM	105	0
aiNet	121	0
RABNET	90	0

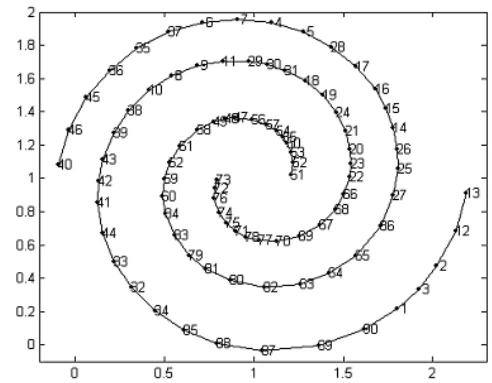

Figure 1. Spatial distribution of neurons and clusters obtained by the application of RABNET to the Two Spirals data set.

4. DISCUSSION AND FUTURE TRENDS

This paper introduced RABNET, a real-valued antibody network, developed by mixing ideas from the immune system with concepts from artificial neural networks. A competitive neural network will be produced by a learning algorithm based on ideas from the immune system. Network weights and structure are determined by the repeated presentation of the input data. After learning, the number of clusters (groups) identified by the network is automatically determined using an MST. So, the neighborhood definition and the cluster discrimination are both performed a posteriori.

5. ACKNOWLEDGMENTS

The authors thank FAPESP and CNPq for their financial support.

6. REFERENCES

[1] B. Fritzke (1994). "Growing Cell Structures - A Self-Organizing Network for Unsupervised and Supervised Learning". Neural Networks 7:9, 1441-1460.

[2] C. T. Zahn, (1971). "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters. IEEE Trans. on Comp. C-20, 68-86.

[3] L. N. de Castro, F. J. Von Zuben, and G. A. de D. Júnior (2003), "The Construction of a Boolean Competitive Neural Network Using Ideas From Immunology", Neurocomputing, 50C, pp. 51-85.

[4] L. N. de Castro and F. J. Von Zuben, (2001), "Immune and Neural Network Models: Theoretical and Empirical Comparisons", International Journal of Computational Intelligence and Applications (IJCIA), 1(3), pp. 239-257.

[5] L. N. de Castro and J. Timmis, (2002). *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer-Verlag.

[6] N. Kasabov, (2002). *Evolving Connectionist Systems: Methods and Applications in Bioinformatics*, Brain Study and Intelligent Machines, Springer Verlag.

[7] S. Haykin, (1999). *Neural Networks – A Comprehensive Foundation*, Prentice Hall, 2nd Ed.

[8] T. Kohonen, (2000). *Self-Organizing Maps*, Springer-Verlag, 3rd Ed.