A Scalable Artificial Immune System Model for Dynamic Unsupervised Learning

Olfa Nasraoui[1], Fabio Gonzalez[2], Cesar Cardona[1], Carlos Rojas[1], and Dipankar Dasgupta[2]

[1] Department of Electrical and Computer Engineering, The University of Memphis
Memphis, TN 38152
{onasraou, ccardona, crojas}@memphis.edu
[2] Division of Computer Sciences, The University of Memphis
Memphis, TN 38152
{fgonzalz, ddasgupt}@memphis.edu

Abstract. Artificial Immune System (AIS) models offer a promising approach to data analysis and pattern recognition. However, in order to achieve a desired learning capability (for example detecting all clusters in a dat set), current models require the *storage and manipulation* of a large network of B Cells (with a number often exceeding the number of data points in addition to all the pairwise links between these B Cells). Hence, current AIS models are far from being scalable, which makes them of limited use, even for medium size data sets.
We propose a new scalable AIS learning approach that exhibits superior learning abilities, while at the same time, requiring *modest* memory and computational costs. Like the natural immune system, the strongest advantage of immune based learning compared to current approaches is expected to be its ease of adaptation in dynamic environments. We illustrate the ability of the proposed approach in detecting clusters in noisy data.

Keywords. Artificial immune systems, scalability, clustering, evolutionary computation, dynamic learning

clustering, that addresses the shortcomings of current AIS models. Our approach exhibits improved learning abilities and *modest* complexity. The rest of the paper is organized as follows. In Section 2, we review some current artificial immune system models that have been used for clustering. In Section 3, we present a new dynamic AIS model and

more convenient to think of the antigen index, j, as monotonically increasing with time. That is, the antigens are presented in the following chronological order: $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$. The Dynamic Weighted B-Cell (*D-W-B-cell*) represents an influence zone over the domain of discourse consisting of the training data set. However, since data is dynamic in nature, and has a temporal aspect, data that is more current will have higher influence compared to data that is less current/older. Quantitatively, the influence zone is defined in terms of a weight function that decreases not only with distance from the antigen/data location to the D-W-B-cell prototype / best exemplar as in [11], but also with the time since the antigen has been presented to the immune network. It is convenient to think of time as an additional dimension that is added to the D-W-B-Cell compared to the classical B-Cell, traditionally statically defined in antigen space only. For the i^{th} D-W-B-cell, DWB_i, we define the following weight/membership function after J antigens have been presented:

$$w_{ij} = w_i\left(d_{ij}^2\right) = e^{-\left(\frac{d_{ij}^2}{2\sigma_i^2} + \frac{(J-j)}{\tau}\right)} \tag{1}$$

where d_{ij}^2 is the distance from antigen $\mathbf{x}_j$ (j^{th} antigen encountered by the immune network) to D-W-B-cell, DWB_i. The stimulation level, after J antigens have been presented to DWB_i, is defined as the density of the antigen population around DWB_i:

$$s_{ai,j} = \frac{\sum_{j=1}^{J} w_{ij}}{\sigma_i^2}, \tag{2}$$

The scale update equations are found by setting $\frac{\partial s_{ai,j}}{\partial \sigma_i^2} = 0$, and deriving incremental update equations, to obtain the following approximate incremental equations for stimulation and scale, after J antigens have been presented to DWB_i.

$$s_{ai,J} = \frac{e^{-\frac{1}{\tau}} W_{i,J-1} + w_{iJ}}{\sigma_{i,J}^2}, \tag{3}$$

$$\sigma_{i,J}^2 = \frac{e^{-\frac{1}{\tau}} \sigma_{i,J-1}^2 W_{i,J-1} + w_{iJ} d_{iJ}^2}{2\left(e^{-\frac{1}{\tau}} W_{i,J-1} + w_{iJ}\right)}. \tag{4}$$

where $W_{i,J-1} = \sum_{j=1}^{J-1} w_{ij}$ is the sum of the contributions from the $(J-1)$ previous antigens, $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{J-1}$, to D-W-B-Cell i, and $\sigma_{i,J-1}^2$ is its previous scale value.

3.2 Dynamic Stimulation and Suppression

We propose incorporating a dynamic stimulation factor, $\alpha(t)$, in the computation of the D-W-B-cell stimulation level. The static version of this factor is a classical way to simulate memory in an immune network by adding a compensation term that depends on other D-W-B-cells in the network [3]. In other words, a group of intra-stimulated D-W-B-cells can self-sustain themselves in the immune network, even after the antigen that caused their creation disappears from the environment. However, we need to put a limit on the time span of this memory so that truly outdated patterns do not impose an

additional superfluous (computational and storage) burden on the immune network. We propose to do this by an annealing schedule on the stimulation factor. This is done by allowing each group of D-W-B-cells to have their own stimulation coefficient, and to have this stimulation coefficient decrease with the age of the sub-net). In the absence of a recent antigen that succeeds in stimulating a given subnet, the age of the D-W-B-cell increases by 1 with each antigen presented to the immune system. However, if a new antigen succeeds in stimulating a given subnet, then the age calculation is modifed by refreshing the age back to zero. This makes extremely old sub-nets die gradually, if not re-stimulated by more recent relevent antigens. Incorporating a dynamic suppression factor in the computation of the D-W-B-cell stimulation level is also a more sensible way to take into account internal interactions. The suppression factor is not intended for memory management, but rather to control the proliferation and redundancy of the D-W-B-cell population. In order to understand the combined effect of the proposed stimulation and suppression mechanism, we consider the following two extreme cases: **(i)** When there is positive suppression (competition), but no stimulation. This results in good population control and no redundancy. However, there is no memory, and the immune network will forget past encounters. **(ii)** When there is positive stimulation, but no suppression, there is good memory but no competition. This will cause the proliferation of the D-W-B-cell population or maximum redundancy. Hence, there is a natural tradeoff between redundancy/memory and competition/reduced costs.

3.3 Organization and Compression of the Immune Network

We define *external interactions* as those occuring between an antigen (external agent) and the D-W-B-cell in the immune network. We define *internal interactions* as those occuring between one D-W-B-cell and all other D-W-B-cells in the immune network. Figure 1(a) illustrates internal (relative to D-W-B-cell$_k$) and external interactions (caused by an external agent called “Antigen"). Note that the number of possible interactions is immense, and this is a serious bottleneck in the face of all existing immune network based learning techniques [3,9,11]. Suppose that the immune network is compressed by clustering the D-W-B-cells using a linear complexity approach such as K Means. Then the immune network can be divided into several *subnetworks* that form a parsimonious view of the entire network. For global low resolution interactions, such as the ones between D-W-B-cells that are very different, only the *inter-subnetwork interactions* are germane. For higher resolution interactions such as the ones between similar D-W-B-cells, we can drill down inside the corresponding subnetwork and afford to consider all the *intra-subnetwork interactions*. Similarly, the external interactions can be compressed by considering interactions between the antigen and the subnetworks instead of all the D-W-B-cells in the immune network. Note that the centroid of the D-W-B-cells in a given subnetwork/cluster is used to summarize this subnetwork, and hence to compute the distance values that contribute in the internal and external interaction terms. This divide and conquer strategy can have significant impact on the number of interactions that need to be processed in the immune network. Assuming that the network is divided into roughly K equal sized subnetworks, then the number of internal interactions in an immune network of N_B D-W-B-cells, can drop from N_B^2 in the uncompressed net-

work, to $\left(\frac{N_B}{K}\right)^2$ *intra-subnetwork interactions* and $K-1$ *inter-subnetwork interactions* in the compressed immune network. This clearly can approach linear complexity as $K \rightarrow \sqrt{N_B}$. Figure 1(c) illustrates the reduced internal (relative to D-W-B-cell$_k$) interactions in a compressed immune network. Similarly the number of external interactions relative to each antigen can drop from N_B in the uncompressed network to K in the compressed network. Figure 1(b) illustrates the reduced external (relative to external agent "Antigen") interactions. Furthermore, the compression rate can be modulated by choosing the appropriate number of clusters, $K \approx \sqrt{N_B}$, when clustering the D-W-B-cell population, to maintain linear complexity, $O(N_B)$.

Sufficient summary statistics for each cluster of D-W-B-cells are computed, and can later be used as approximations in lieu of repeating the computation of the entire suppression/stimulation sum. The summary statistics are in the form of average dissimilarity within the group, cardinality of the group (number of D-W-B-cells in the group), and density of the group.

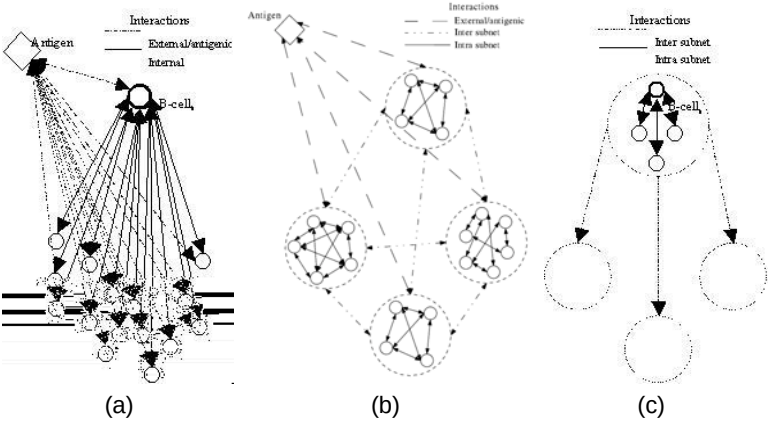


Fig. 1. Immune network interactions: (a) without compression, (b) with compression, (c) Internal Immune network interactions with compression

3.4 Effect of the Network Compression on Interaction Terms

The D-W-B-cell specific computations can be replaced by subnet computations in a compressed immune network. The stimulation and scale values become

$$s_i = s_{ai,J} + \alpha(t)\frac{\sum_{l=1}^{N_B^i} w_{il}}{\sigma_{i,J}^2} - \beta(t)\frac{\sum_{l=1}^{N_B^i} w_{il}}{\sigma_{i,J}^2}, \tag{5}$$

where $s_{ai,J}$ is the pure antigen stimulation given by (3) for D-W-B-cell$_i$; and N_B^i is the number of B-cells in the subnetwork that is closest to the J^{th} antigen. This will modify the D-W-B-cell scale update equations to become

$$\sigma_{i,J}^2 = \frac{1}{2}\frac{e^{-\frac{1}{\tau}}\sigma_{i,J-1}^2 W_{i,J-1} + w_{iJ}d_{iJ}^2 + \alpha(t)\sum_{l=1}^{N_B^i} w_{il}d_{il}^2 - \beta\sum_{l=1}^{N_B^i} w_{il}d_{il}^2}{2\left(e^{-\frac{1}{\tau}}W_{i,J-1} + w_{iJ}\right) + \alpha(t)\sum_{l=1}^{N_B^i} w_{il} - \beta\sum_{l=1}^{N_B^i} w_{il}}. \quad (6)$$

3.5 Cloning in the Dynamic Immune System

The D-W-B-cells are cloned (i..e, duplicated together with all their intrinsic properties such as scale value) in proportion to their stimulation levels relative to the average stimulation in the immune network. However, to avoid preliminary proliferation of good B-Cells, and to encourage a diverse repertoire, new B-Cells do not clone before they are mature (their age, t_i exceeds a lower limit t_{min}). They are also not removed from the immune network regardless of their stimulation level. Similarly, B-cells with age $t_i > t_{max}$ are frozen, or prevented from cloning, to give a fair chance to newer B-Cells. This means that

$$N_{clones_i} = K_{clone}\frac{s_i}{\sum_{k=1}^{N_{D-W-B-cell}} s_k} \text{ if } t_{min} \leq t_i \leq t_{max}. \quad (7)$$

3.6 Learning New Antigens and Relation to Outlier Detection

Somatic hypermutation is a powerfull natural exploration mechanism in the immune system, that allows it to learn how to respond to new antigens that have never been seen before. However, from a *computational* point of view, this is a very costly operation since its complexity is exponential in the number of features. Therefore, we model this operation in the artificial immune system model by an instant antigen duplication whenever an antigen is encountered that fails to activate the entire immune network. A new antigen, $\mathbf{x}_j$ is said to activate the i^{th} B-Cell, if its contribution to this B-Cell, w_{ij} exceeds a minimum threshold w_{min}. Antigen duplication is a simplified rendition of the action of a special class of cells called *dendritic* cells whose main purpose is to *teach other immune cells such as B-cells to recognize new antigens*. Dendritic cells (which have long been mistaken to be part of the nervous system), and their role in the immune system, have only recently been understood. We refer to this new antigen duplication, a *dendritic injection*, since it essentially injects new information in the immune system.

3.7 Proposed Scalable Immune Learning Algorithm for Clustering Evolving Data

Scalable Immune Based Clustering for Evolving Data

Fix the maximal population size N_B*;*
Initialize D-W-B-cell population and $\sigma_i^2 = \sigma_{init}$ *using the first batch of the input antigens/data;*
Compress immune network into K *subnets using 2-3 iterations of K Means;*
Repeat for each incoming antigen $\mathbf{x}_j$ {
  Present antigen to each subnet centroid in network and determine the closest subnet;
  IF antigen activates closest subnet Then {
    Present antigen to each D-W-B-cell, D-W-B-cell$_i$*, in closest immune subnet;*
    Refresh this D-W-B-cell's age ($t = 0$*) and update* w_{ij} *using (1);*
    Update the compressed immune network subnets incrementally;
  }
  ELSE
    Create by dendritic injection a new D-W-B-cell = $\mathbf{x}_j$ *and* $\sigma_i^2 = \sigma_{init}$*;*
  Repeat for each D-W-B-cell$_i$ *in closest subnet only* {
    Increment age (t) for D-W-B-cell$_i$*;*
    Compute D-W-B-cell$_i$*'s stimulation level using (5);*
    Update D-W-B-cell$_i$*'s* σ_i^2 *using (6);*
  }
  Clone and mutate D-W-B-cells;
  IF population size $> N_B$ *Then*
    Kill worst excess D-W-B-cells, or leave only subnetwork representatives of oldest subnetworks in main memory;
  Compress immune network periodically (after every T *antigens), into* K *subnets using 2-3 iterations of K Means;*
}

3.8 Comparison to Other Immune Based Clustering Techniques

Because of paucity of space, we review only some of the most recent and most related methods. The Fuzzy AIS [11] uses a richer knowledge representation for B-cells as provided by fuzzy memberships that not only model different areas of the same cluster differently, but are also robust to noise and outliers, and allow a dynamic estimation of scale unlike all other approaches. The Fuzzy AIS obtains better results than [9] with a reduced immune network size. However, its batch style processing required storing the entire data set and all intra-network interaction affinities.

The Self Stabilizing AIS (SSAIS) algorithm [12] maintains stable immune networks that do not proliferate uncontrollably like in previous versions. However, a single NAT threshold is not realistic for data with clusters of varying size and separation, and SSAIS is rather slow in adapting to new/emerging patterns/clusters. Even though SSAIS does not require storage of the entire data set, it still stores and handles interactions between all the cells in the immune network. Because the size of this network is comparable to that of the data set, this approach is not scalable.

The approach in [13] relies exclusively on the antigen input and not on any internal stimulation or suppression. Hence the immune network has no memory, and would not be

able to adapt in an incremental scenario. Also, the requirement to store the entire dataset (batch style) and the intense computations of all pairwise distances to get the intial NAT value, make this approach unscalable. Furthermore, a single NAT value and a drastic winner-takes-all pruning strategy may impact diversity and robustness on complex and noisy data sets.

In [14], an approach is presented that exploits the analogy between immunology and sparse distributed memories. The scope of this approach is different from most other AIS based methods for clustering because it is based on binary strings, and clusters represent different schemas. This approach is scalable, since it has linear complexity, and works in an incremental fashion. Also, the gradual influence of data inputs to all clusters avoids undesirable winner-take-all effects of most other techniques.

Finally, the aiNet algorithm [4] evolves a population of antibodies using clonal selection, hypermutation and apoptosis, and then uses a computationally expensive graph theoretic technique to organize the population into a network of clusters.

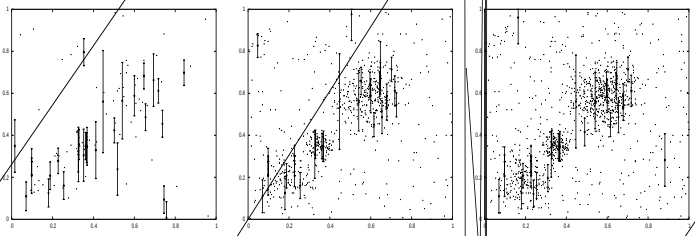
Table 1 summarizes the charateristics of several immune based approaches to clustering, in addition to the K Means algorithm. The last row lists typical values reported in

4 Experimental Results

Clean and noisy 2-dimensional sets, with roughly 1000 to 2000 points, and between 3 and 5 clusters, are used to illustrate the performance of the proposed immune based approach. The implementation parameters were as follow: The first 0.02% of the data are used to create an initial network. The initial value for the scale was $\sigma_{init} = 0.0025$ (an upper radial bound derived based on the range of normalized values in $[0, 1]$). B-cells were only allowed to clone past the age of $t_{min} = 2$, and the cloning coefficient was 0.97. The maximum B-cell population size was 30 (an extremely small number considering the size of the data), the mutation rate was 0.01, $\tau = 1.5$, and the compression rate, K varied between 1 and 7. The network compression was performed after every $T = 40$ antigens have been processed. The evolution of the D-W-B-cell population for 3 noisy clusters, after a *single* pass over the antigens, presented in random order, is shown in Figure 2, superimposed on the original data set. The results for the same data set, but with antigens presented in the order of the clusters is shown in Figure 3, with the results of RLAINE [9] in Fig. 3 (d). This scenario is the most difficult (worst) case for single-pass learning, as it truly tests the ability of the system to memorize the old patterns, adapt to new patterns, and still avoid excessive proliferation. Unlike the proposed approach, RLAINE is unable to adapt to new patterns, given the same amount of resources. Similar experiments are shown for a data set of five clusters in Figure 4 and 5. Since this is an unsupervised clustering problem, it is not important that a cluster is modeled by one or several D-W-B-cells. In fact, merging same-cluster cells is trivial since we have not only their location estimates, bue also their individual robust scale estimates. Finally, we illustrate the effect of the compression of the immune network by showing the final D-W-B-cell population for different compression rates corresponding to $K = 1, 3, 5$ on the data set with 3 clusters, in Fig. 6. In the last case ($K = 5$), the immune interactions have been practically reduced from quadratic to linear complexity by using $K \approx \sqrt{(N_B)}$. It is worth mentioning that despite the dramatic reduction in complexity, the results are virtually indistinguishable in terms of quality. The effect of compression is further illustrated for the data set with 5 clusters, in Fig. 7. The antigens were presented in the most challenging order (one cluster at a time), and in a single pass. In each case, the proposed immune learning approach succeeds in detecting dense areas after a single pass, while remaining robust to noise.

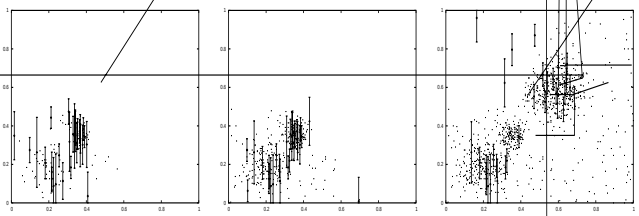
5 Conclusion

We have introduced a new robust and adaptive model for immune cells, and a scalable immune learning process.The D-W-B-cell, modeled by a robust weight function, defines a gradual influence region in the antigen, antibody, and time domains. This is expected to condition the search space. The proposed immune learning approach succeeds in detecting dense areas/clusters, while remaining robust to noise, and with a very modest D-W-B-cell population size. Most existing methods work with B-cell population sizes often exceeding the size of the data set, and can suffer from premature loss of good detected immune cells. The proposed approach is favorable from the points of view of scalability, as well as quality of learning. Quality comes in the form of diversity



(a) (b) (c)

Fig. 2. Single Pass Results on a Noisy antigen set presented one at a time in random order: Location of D-W-B-cells and estimated scales for data set with 3 clusters after processing (a) 100 antigens, (b) 700 antigens, and (c) all 1133 antigens



and continuous adaptation as new patterns emerge. We are currently investigating the use of our scalable immune learning approach to extract patterns from evolving Web clickstream and text data for Web data mining applications.

Acknowledgment. This work is partially supported by a National Science Foundation CAREER Award IIS-0133948 to Olfa Nasraoui and support from Universidad Nacional de Colombia for Fabio Gonzalez.

References

1. D. Dasgupta, *Artificial Immune Systems and Their Applications*, Springer Verlag, 1999.
2. I. Cohen, *Tending Adam's Garden*, Academic Press, 2000.
3. J. Hunt and D. Cooke, "An adaptative, distributed learning system, based on immune system," in *IEEE International Conference on Systems, Man and Cybernetics*, Los Alamitos, CA, 1995, pp. 2494–2499.
4. L. N. De Castro and F. J. Von Zuben, "An evolutionary immune network for data clustering," in *IEEE Brazilian Symposium on Artificial Neural Networks*, Rio de Janeiro, 2000, pp. 84–89.
5. J.D. Farmer and N.H. Packard, "The immune system, adaptation and machne learning," *Physica*, vol. 22, pp. 187–204, 1986.
6. F.J. Varela H. Bersini, "The immune recruitment mechanism: a selective evolutionary strategy," in *Fourth International Conference on Genetic Algorithms*, San Mateo, CA, 1991, pp. 520–526.
7. S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in *IEEE Symposium on Research in Security and Privacy*, Los Alamitos, CA, 1994.
8. D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology," in *5th International Conference on Intelligent Systems*, Reno, Nevada, 1996.
9. J. Timmis and M. Neal, "A resource limited artificial immune system for data analysis," *Knowledge Based Systems*, vol. 14, no. 3, pp. 121–130, 2001.
10. T Knight and J Timmis, "Aine: An immunological approach to data mining," in *IEEE International Conference on Data Mining*, San Jose, CA, 2001, pp. 297–304.
11. O. Nasraoui, D. Dasgupta, and F. Gonzalez, "An artificial immune system approach to robust data mining," in *Genetic and Evolutionary Computation Conference (GECCO) Late breaking papers*, New York, NY, 2002, pp. 356–363.
12. M. Neal, "An artificial immune system for continuous analysis of time-varying data," in *1st International Conference on Artificial Immune Systems*, Canterbury, UK, 2002, pp. 76–85.
13. Wierzchon and U. Kuzelewska, "Stable clusters formation in an artificial immune system," in *1st International Conference on AIS*, Canterbury, UK, 2002, pp. 68–75.
14. E Hart and P Ross, "Exploiting the analogy between immunology and spares distributed memories: A system for clustering non-stationary data," in *1st International Conference on Artificial Immune Systems*, Canterbury, UK, 2002, pp. 49–58.
15. N. K. Jerne, "The immune system," *Scientific American*, vol. 229, no. 1, pp. 52–60, 1973.
16. J. Timmis, M. Neal, and J. Hunt, "An artificial immune system for data analysis," *Biosystems*, vol. 55, no. 1, pp. 143–150, 2000.