aiVIS – Artificial Immune Network Visualisation

Jon Timmis

Computing Laboratory
University of Kent at Canterbury
Canterbury
Kent. UK.
CT2 7NF
J.Timmis@ukc.ac.uk
http://www.cs.ukc.ac.uk/people/staff/jt6

Abstract

The field of Artificial Immune Systems (AIS) is the use of the natural immune system as a metaphor for solving computational problems. A novel unsupervised machine-learning algorithm, inspired by the immune system, has been developed called AINE. Using various immunological metaphors, AINE evolves a network of objects, known as an Artificial Immune Network (AIN) that is a diverse representation of the data set being learnt.

The results of AINE are visualised in a specially developed tool (aiVIS), which allows the user to interact with the network to perform exploratory analysis. aiVIS presents AINs in such as way as to build up an understanding of the make up of the data set, learning about subtle patterns and clusters within the data set and links between clusters. Unclassified items can then be introduced into the network so to further enhance the exploratory nature of the AIN.

This paper provides an overview of the learning algorithm, but concentrates on the visualisation aspect of the work. The usefulness of using AIN for exploratory visualisation is investigated and an explanation of how aiVIS operates is presented.

Keywords: immune networks, self organising maps, exploratory data analysis, artificial immune systems, cluster analysis, visualisation

1 Introduction

Over the years, biology has been a rich source of inspiration to Computer Scientists (Paton, 1994). Recently, increased use has been made of the natural immune system as a metaphor to create *immune algorithms*; this field is known as Artificial Immune Systems (AIS). What is of interest to researchers in this field is not the modelling of the biological systems, but the abstraction of simple metaphors to help solve problems where traditional techniques are unable to cope. With biologicaly inspired techniques, it is very common to see gross simplification of the biological system used for inspiration. For example, you see a bird, you build an aircraft, you see a spiders web, you build a bullet proof vest. It is quite common to see final biologicaly inspired systems that are quite different to the system that acted as inspiration, due to high levels of abstractions employed.

This paper introduces a visualisation technique that allows the presentation of Artificial Immune Networks (AINs), that are themeselves produced by

an unsupervised machine learning algorithm inspired by immunology. It is proposed that AINs can be used for effective exploratory data analysis. AINs are an effective way of representing meaningful relationships in data via the tool aiVIS, presented in this paper.

Section 2 provides a brief overview of relevant immunology and introduces the concept of using the immune system as a metaphor for machine learning. Section 3 detials the aiVIS tool, which is used to visualise the AINs produced by the learning algorithm. Section 4 shows a comparison of similar techniques from the machine learning community with aiVIS and Section 5 concludes with remarks for future research.

2 The Natural Immune System

2.1 Background Immunology

The natural immune system protects our bodies from infection. This is achived by a complex interaction of white blood cells called B Cells and T Cells. Upon encountering an antigen (an infecting item), B Cells are stimulated by interacting with the antigen and with the help of T Cells undergo rapid cloning mutation. This is an attempt by the immune system to kill of the invading antigen and prepare the immune system for another infection from that antigen (or similar antigen). The immune system maintains a memory of the infection so that if ever exposed to the same antigen a quicker response can be elicited against the infection.

Several theories as to how immune memory works have been proposed. The Immune Network Theory, first proposed in (Jerne 1974) and reviewed in (Perelson 1989), proposes that a network dynamically maintains the immune memory using feedback mechanisms. Thus if something has been learnt, it can be forgotten unless it is reinforced by other parts of the network.

It has been proposed that the immune network can be thought of as being *cognitive* (Varela *et al* 1988) and exhibiting learning capabilities. The authors proposed four reasons as to why they consider immune systems to be cognitive: (i) they can recognise molecular shapes; (ii) they remember history of encounters; (iii) they define the boundaries of molecular self, and (iv) they can make inferences about molecular species they have yet to encounter. Taking these points, the authors explore cognitive mechanisms of the immune system and propose that

the immune network can be thought of as a cognitive network, in a similar way to a neural network.

A more in depth review of immunology can be found in (Timmis 2000).

2.2 The Immune System as a Metaphor

The immune system is an excellent metaphor for solving computational problems. The immune system provides such things as recognition, feature extraction, diversity, learning, memory, inherently distributed and self-regulatory. The past few years have seen an increase in the application of immune algorithms to problems such as robotic control (Ishiguro et al 1998), simulating behaviour in robots (Lee & Sim 1997), network intrusion detection (Kim & Bentley 1998), fault diagnosis (Ishida 1996) and machine learning (Timmis *et al*, 2000). Work in (Timmis & Neal, 2000) builds on (Timmis *et al*, 2000) and uses the immune network as a metaphor to create an unsupervised learning algorithm. This creates a network of objects (analogous to B cells), in which clusters or patterns emerge and can then be visualised.

Using ideas from immunology presented in section 2.1 gave rise to the development of an unsupervised learning algorithm (AINE) (Timmis & Neal 2000), which is briefly examined in this paper. This is done to lay the foundations for the work presented in this paper.

2.2.1 AINE Learning Algorithm

To summarize work in (Timmis & Neal 2000), AINE is initialised as a network of ARB objects (Artificial Recognition Balls); T Cells are currently ignored. ARBs are objects that are representaitve of a single piece of n-dimensional data. Links between ARBs are created if a match between two ARBs, in Eucledian space, is below the Network Affinity Threshold (NAT); this is the average Euclidean distance between each item in the data set. Therefore, if two ARBs are closer together in Euclidean space they can be said to be sore similar, or share a greater affinity. An initial network is created which is a cross section of the data set to be learnt; the remainder makes up the antigen training set. Each member of this set is matched against each ARB in the network, again, with the similarity being calculated on Euclidean distance. An ARB has the concept of being stimulated, as a B cell is in the natural immune system. This stimualation is a bais on deciding how good a match a paraticualr ARB is

for not only the data in the training set, but also adjoining ARBs in the network. The stimulation level of an ARB determines the survival of the ARB in the network. The stimulation level also indicates if the ARB should be cloned and the number of clones that are produced for that ARB. Clones undergo a stochastic process of mutation in order to create a diverse network that can represent the antigen that caused the cloning as well as slight variations. These clones are then integrated into the AIN and if their match value with ARBs in the network is below the NAT. There exist a number of parameters to the algorithm, those being: network affinity scalar (a scalar value to reduce the actual value of the network affinity threshold, thus reducing network connectivity); mutation rate (which affects the diversity of the network) and number of resources that AINE is allocated, this value is approximatly 5 times the number of resources to training items on data sets tested to date. Each one of these can be used to alter algorithm performance.

A population control mechanism forces ARBs to compete for survival based on a finite number of resources that AINE contains; the more stimulated an ARB, the more resources it can claim. This gives rise to a meta-dynamical system that eventually stabilises into a diverse core network structure that captures the main patterns contained within the data.

The output from AINE can be considered as a disconnected graph, which is called an Artificial Immune Network (AIN). Data regarding the contents of each ARB is recorded, along with a list of all links in the network and the affinity between the connected ARBs.

3 Visualisation Using Immune Networks

The AINs evolved by AINE are ideally suited to visualisation. By visualising these networks, the user can build up a good impression of the topological make up of the data, enabling them to identify areas of similarity within the data and perform a more informed exploration of the results.

The visualisation technique employed is an algorithm that uses the idea of allowing the network to self-organise into sensible layout position. The attraction is between the ARBs in the network that are connected. The amount by which there are connected contributes to moving the ARBs closer together in aiVIS panel. However, each ARB also repels against every other ARB in the network. Therefore, what

starts of as a seemingly random network collection of ARBs and links, self-organises into a meaningful and structured network that the user can examine. In effect, an anneling process is observed. Therefore, it can be said that ARBs that are connected to each other reach a dynamic equilibrium. It should be stressed that the representations of the generated AIN are not meaningful as a conventional two-dimensional plot; they represent the information in the AIN as a topological structure. The attraction and repulsion algorithm simply spreads out the cells into a less cluttered shape.

This section details the aiVIS system developed for the exploratory analysis of these AINs.

3.1 aiVIS System

In order to view the AIN, the results from AINE are imported into aiVIS. AiVIS applies a simple attraction-repulsion algorithm to layout the network. Each link in the AIN represents the affinity between two ARBs. This distance needs to be drawn by aiVIS in order to represent the relationship between those two ARBs. Initially, all ARBs and associated links are placed in the x, y plane of aiVIS. The objective of aiVIS is to adjust the positions of each ARB and associated links, to such a point where all the similar ARBs are located in the same area of the viewing tool, and the links that are drawn between them match the affinity between the ARBs. The algorithm follows a simple rule, links between ARBs act as an attractor and help to bring ARBs together and ARBs act to repel each other, within a defined radius. Figures 1a-1c show a simple example of how this operates, using a network evolved on the Iris data set (Fisher 1936). In Figure 1a ARBs are placed at a random location. The links between the ARBs have an associated affinity value and the aim is to make the line that is drawn the same (or as near to) as that affinity value. The algorithm continually calculates an amount by which to adjust the x and y position of the associated ARBs. This is achieved by calculating a small shift towards the actual display length required (the affinity between the two ARBs) and the actual length of the link being drawn and is defined as:

$$df = \frac{dl - al}{al * k}$$

Where df is the amount by which to adjust the x and y position of the node, dl is display length required to

be drawn, *al* is the actual length of the line being drawn, and *k* is a scalar constant. Additionally, there is also a repulsion calculation performed, which calculates the amount by which each ARB in the network acts against another ARB in the network, within a given radius: this can be defined in the follwing equations:

$$mdx = \frac{dvx}{dvx^2 + dvy^2}$$

$$mdy = \frac{dvy}{dvx^2 + dvy^2}$$

$$ac = \sqrt{\frac{mdx + mdy^2}{2}}$$

$$nx = \frac{mdx}{ac}, ny = \frac{mdy}{ac}$$

where *dvx* is the distance between two ARBs in the x plane, *dvy* is the distance between two ARBs in the y plane and *ac* is the amount by which to repel each ARB , nx and ny are the new x and y postions of the ARB.

Figure 1b shows the next stage, where separate clusters are starting to emerge from the initial random layout. One cluster has clearly separated from the other larger cluster. Finally, Figure 1c shows three distinct and separate clusters. aiVIS has allowed the AIN to find it's own layout by a process of recalculating ARBs position within the area. The user can click on any of the visible ARBs and drag them around the screen. When this happens the AIN will start to recalculate the new positions and ARBs will move accordingly to maintain the clusters within the network. This is a useful feature of aiVIS as it allows the use to separate out even more clusters and examine them a distance away from other clusters being visualised in aiVIS.

Eventually the AIN come to equilibrium and the affect of the repulsion between all ARBs will cease, as they will be outside the range of influence of each other. This allows a very interactive session between the user and the AIN.
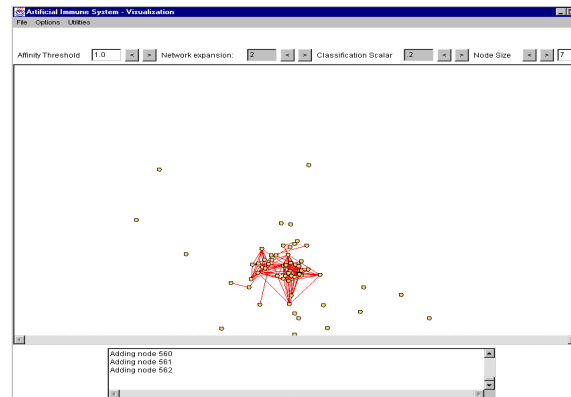

Figure 1a – Initial random position of the AIN. No clusters are apparant
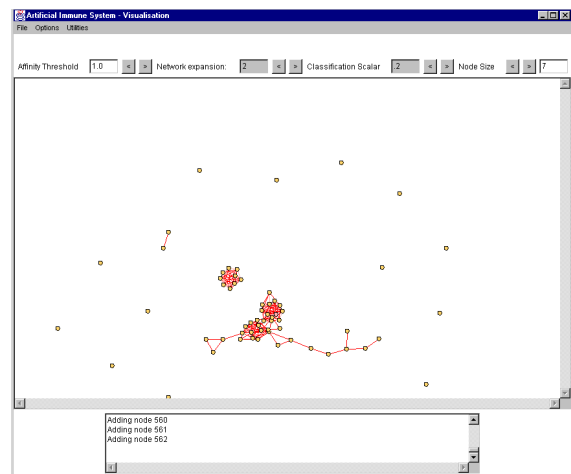

Figure 1b. The clusters within the AIN have started to separate away from each other

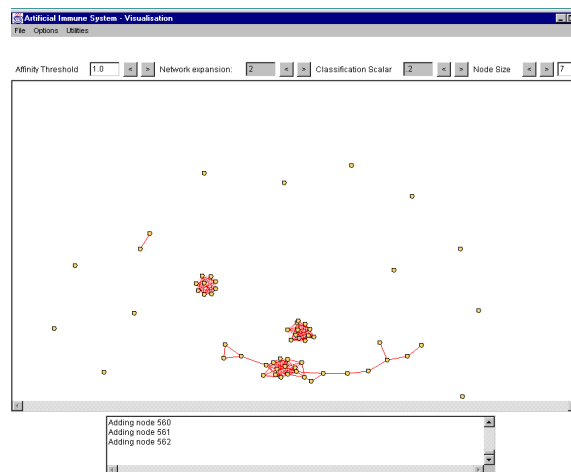Figure 1c. The clusters continue to separate out until a point of equilibrium is reached.

3.2 Exploring AINs using aiVIS

This section details results obtained via the aiVIS system. In order to be able to fully appreciate the visualisation in aiVIS, the data sets used are first explained and presented in visual format.

3.2.1 Data Used

In order to be able to see the benefit of aiVIS, simple plots of the data used are shown Figure 2 and Figure 3. Figure 2 shows the simulated data set. It is a two-dimensional data set with two linearly separable clusters.
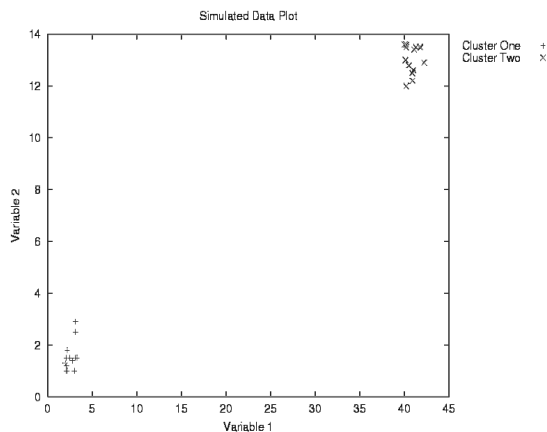
Figure 2. Scatter plot of simulated data. Two clearly separated clusters are visible

The second data set, the Iris data, is more complex, shown in Figure 3. The Iris data is measurements of Iris petal lengths and width; is four dimensional in nature, with each dimension representing a measurement of petal length, width and sepal length and width. There are three different classes of plant with two of those classes being non-linearly separable, i.e a straight line could not be drawn through the data to separate the two classes, there is much cross over of items between classes. This makes the data a good benchmark in attempting to separate out the varying classes of data.
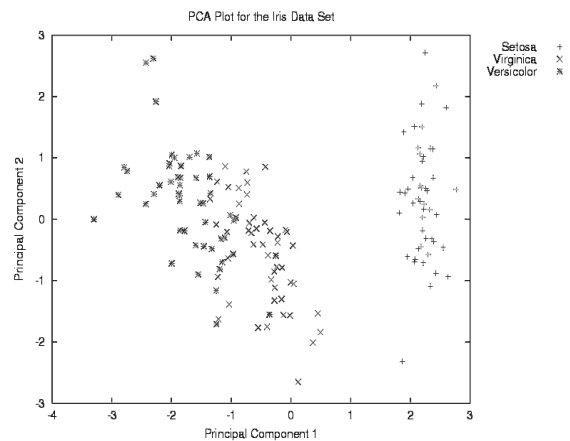
Figure 3. PCA of Iris Data. One cluster is linearly separable from the other two

3.3 Results on Test Data with aiVIS

This section presents the results from running AINE with the two test data sets. The usefulness of aiVIS for the visualisation of AINs is examined.

3.3.1 Initial Exploratory Analysis

Once the network has been evolved by AINE, it can be loaded directly into aiVIS. Figure 4 illustrates results from the simulated data. By this stage, the network has found equilibrium. First glance at the AIN shows that there are clearly two different clusters within the data set. It is worthy of note, that the AIN being visualised consists of approximately 17 ARBs, whereas the data set being learnt consists of 30 items. Significant relationships between data have been maintained being displayed in the AIN. This shows that the AIN is effective in reducing the complexity of the data, whilst still providing useful information. Compared to Figure 2, initially the same basic information can be gleaned. However, with aiVIS topological information from the AIN can be gleaned, whilst also allowing the user to interact with the AIN. Using this idea, the user can interrogate each ARB for the contents and clusters of ARBs to allow for the development of a more informed understanding of the patterns or clusters contained within the data. For example, investigation of the AIN can lead to the understanding of why certain ARBs are connected; with a connection indicating an affinity and hence relationship between items.
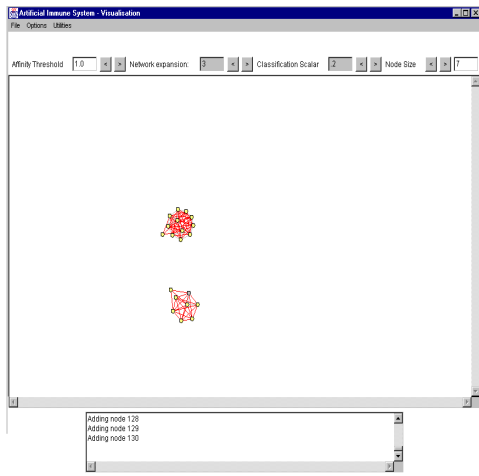
Figure 4. Initial exploration of simulated data. Two clusters are clearly visible as are relations (or links) between ARBs

Figure 5 shows the results from AINE on the Iris data set. Here the AIN shows three distinct classes within the data. When compared Figure 5 more information can be gleaned, as from Figure 3 it is not clear there are three distinct classes. The AIN has separated out this non-linearly separable data set and maintained the three classes of plants
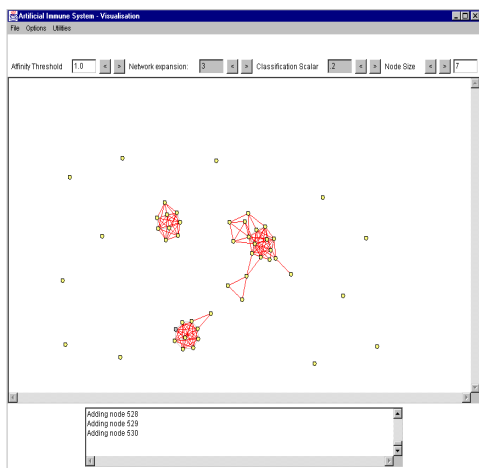

Figure 5. Initial exploration of Iris data. Three distinct clusters have been identified, compared to just two from a simple PCA

Should the network be very highly connected; then the user can increase the length of the displayed links in order to spread out AIN and make viewing the network easier. The user can also attach the

labels on each node (should they be provided by the data) and the affinity value on the link between each ARB. All of these features make aiVIS combined with AINs a very useful and powerful visualisation mechanism.

3.3.2 Presenting Unseen Data

aiVIS allows the possibility of introducing items of data that were not used as part of the learning process, called *unseen data,* or *antigen* node. Figure 7 shows two examples of presenting unseen items to both the simulated data and Iris data AINs. The affinity between the unseen item and each ARB is calculated and if the affinity falls below the NAT threshold, as calculated by AINE, the unseen item and ARB are connected. Figure 6a, shows a square *antigen* node connected to the lower of the two clusters. Figure 6b, shows an unseen *Setosa* class of plant being connected to the top cluster. The user can then click on the *antigen* node and a table is presented to the user that lists the connected ARBs and their associated affinity and data. This provides the user the ability to classify the unseen item and provide the means by which to undertake further exploratory analysis of the data.

This feature makes aiVIS a simple and easy to use classification tool that can be used for detailed exploratory data analysis.
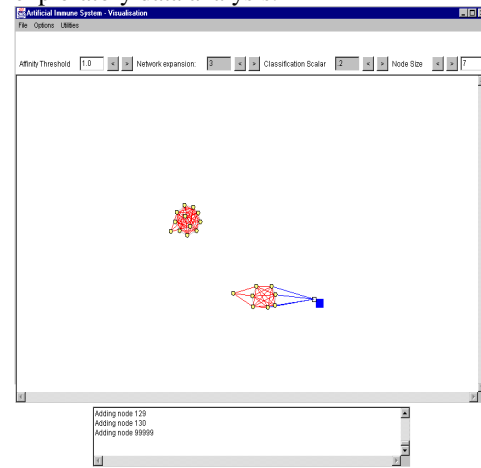

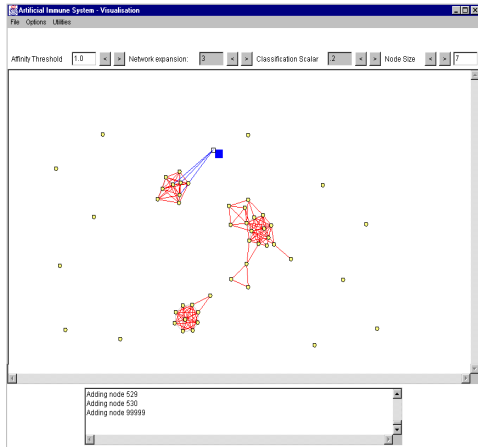Figure 6a. Presenting an unseen data item with the simulated data

Figure 6b. Presenting an unseen data item with the Iris data

3.3.3 Interacting with AINs

The aiVIS system provides the ability for a user to explore the evolved network to gain an understanding as to the inherent patterns/clusters within that data.

Refering to Figure 6a aiVIS contains a number of features that allow effective exploratory analysis; these are located in the top panel of aiVIS. Starting from left to right, the *Network Affinity* setting allows the user to affect what connections are shown within the AIN. All links between ARBs that have a value less than is shown will be displayed on screen. This is a very useful feature if the networks are highly connected and its usefulness is discussed later. The *Network Expansion* value allows the user to increase or decrease the size of the links between ARBs, thus spreading out or bringing together clusters of B cells for easier viewing. The value shown in the *Classification Scalar* is for use when unseen antigens are presented to the system; the lower the value, the better the match has to be to create a link. Finally, the *Node Size* value allows for the increasing and decreasing of the physical size of the ARBs being drawn.

4.0 Related Work

The results obtained from aiVIS can be compared with algorithms such as Kohenen networks, or Self-Orgainsing feature maps (Kohonen, 1988) and simple single linkage cluster analysis (Everit, 1974). aiVIS however shows something different in terms of results and interpretation. Figure 7 shows a

dendrogram of the Iris data. The dendogram is a useful exploratory tool, quick to generate and identifies clusters and sub-clusters within a data set. However, dendrograms are difficult to interpret correctly as it is always hard to ignore horizontal proximity when interpreting them. In addition, single-linkage clustering did not manage to successfully separate out into three distinct clusters, this is due to the nature of the data, but would point to limitations with this technique and its inability to separate out data as effectively as might be required.

The Kohonen network (Figure 8) results present a slightly different picture. The network has identified one larger cluster of data and three smaller ones. The smaller ones are made up of Setosa class items, which is comparable to the dendrogram result in that in the Setosa area of the dendrogram there do seem to be three sub-clusters. The Kohonen Network has failed to separate out the Versicolor and Virginica class of plant, again a similar result to the dendrogram. The interpretation of the Kohonen Network is made easier by the user interface tool, allowing the user to click on each network node to investigate what data items are associated with the node. Once again, although the topological ordering of the data in the network is meaningful it is hard to interpret the strength of relationships within the data.

AINE has evolved an AIN which displays three separate clusters in the data set (Figure 5). A feature of the aiVIS is the ability to manually investigate the evolved network via the visualisation technique that has been developed. Upon investigation of the network by the user it can be seen that each of the three clusters represents each of the three classes in the Iris data set. The network not only represents the clusters in the data, but also relationships between individual cells. This allows the user to build up a fuller picture of the relationships and patterns contained within the data, that would have previously gone unseen. This might be particularly apparant when using single linkage clustering and Kohonen networks.

AINE has created a diverse set of all three class derivatives which enables aiVIS to be an effective classifier on items which vary significantly from the original training items (Timmis, 2000).
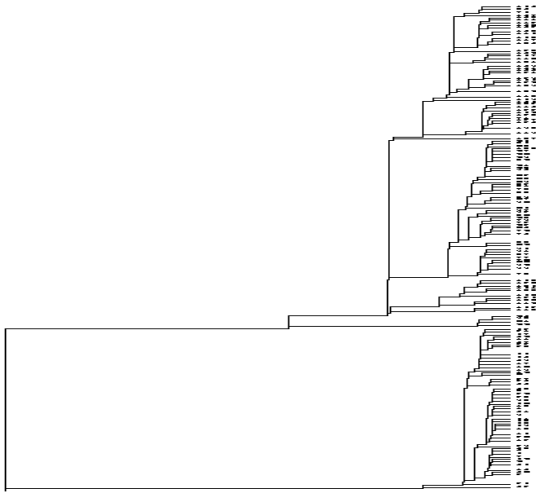
Figure 7. Single linkage clustering of the Iris data. Interpretation of dendrograms is difficult especially when data sets grow large
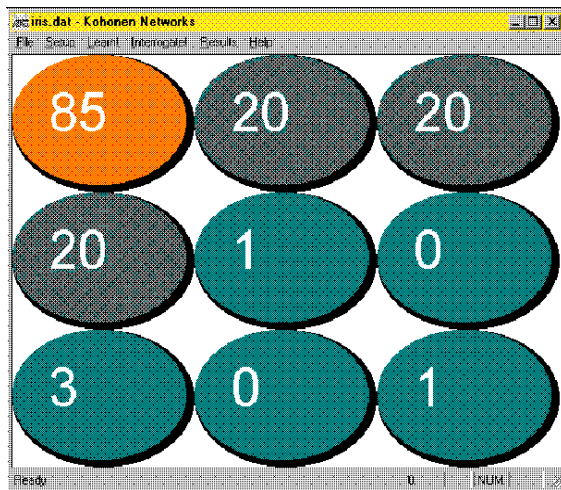

Figure 8. Kohonen representation of the Iris. Software (Neal, 1999)

5.0 Future Work

There is great potential for the work described in this paper. Currently, an investigation of the use of AINs and the extension of aiVIS is being undertaken. This research will apply work described in this paper to document classification and web site navigation. For example, it is proposed to evolve an AIN representative of the pages contained on a web site. This AIN would then be visualised in a tool based on aiVIS. Additional features would be to allow the user

to view a hierarchy of pages on the site, and to follow links between relevant pages and clusters of pages on that site.

It is also proposed to augment aiVIS to include a toolkit of statistical and cluster analysis technique, such as inter cluster analysis and hierarchical cluster analysis.

6.0 Summary

This paper has introduced the metaphor of the natural immune system for unsupervised machine learning and visualisation.

By using this metaphor, meaningful networks can be evolved that are a diverse representation of data. These AINs can be visualised in such a manner as to allow interactive exploratory data analysis. It has been shown that these Artificial Immune Networks (AINs), via the use of the system aiVIS, can provide an excellent way of representing AINs that contain meaningful relationships in data. AINs are evolved to create a diverse representation of the data being learnt, with the AINs using a reduced number of nodes to represent the main features within the data set; thus acting as a form of compression.

The tool aiVIS has been introduced, that allows the visualisation of these AINs. The AINs are laid out using a simple *attraction-repulsion* algorithm that enables the AIN to find equilibrium, such that identification of patterns is made easy. The user can interact with the AIN, being able to investigate clusters, contents of clusters to enable an understanding of the data to be gained. aiVIS allows the visualisation of *n*-dimensional data in two dimensions, as it is the topological structure of the AIN that is of interest and similarity is based on whole data items, rather than on an inter-dimensional basis, a standard technique for the machine learning community.

When compared to other simialr techniques, aiVIS coupled with AINs provide a useful tool for exploratory analysis offering compression of the data whilst still maintaining important relationships in the data. This research paves the way for future research and exciting opportunities for development of the use of the immune metaphor in exploratory visualisation and data mining.

7.0 References

Everitt, B. 1974. *Cluster Analysis*. Pub. Heinemann

Fisher, R. 1936. The use of multiple measurements in taxonomic problems. *Annual Eugenics*. **7.** II. pp 179-188

Ishiguro, A, Ichikawa, S, Shibat, T and Uchikawa, Y. 1998. Modernationsim in the immune system: Gait acquisition of a legged robot using the metadynamics function. *pp 3827-3832 of: Proceedings of IEEE International Conference on Systems and Man and Cybernetics (SMC).* San Diego, USA. IEEE.

Jerne, N. 1974. Towards a network theory of the immune system. Annals of Immunology (Inst. Pasteur). **125C**, pp. 373-389

Kim, J and Bentley, P. 1998. The human immune system and network intrusion detection. Proceedings of 7th European Congress on Intelligent Techniques - Soft Computing. Aachan, Germany

Kohonen, T. 1997. *Self Organising Maps.* 2nd Edn. Springer

Lee, Dong-Wook and Sim, Kwee-Bo. 1997. Artificial immune network based co-operative control in collective autonomous mobile robots. *pp 58-63 of: proceedings of IEEE International Workshop on Robot and Human Communication. Sendai, Japan.* IEEE.

Neal, M (1999). KOHO: A windows program for self organising feature map generation. Technical Report UWA-DCS-99-033. University of Wales, Aberystwyth.

Perelson, A. 1989. Immune Network Theory Immunological Review. **110,** pp**.** 5-36

Paton, R. 1994 *Computing with Biological Metaphors.* Chp. Introduction to computing with biological metaphors. Publisher Chapman and Hall. pp 1-9

Timmis, J, Neal, M and Hunt, J. 1999. Data Analysis with Artificial Immune Systems and Cluster Analysis and Kohonen Networks : Some Comparisons. *pp. 922-927 of: Proceedings of the IEEE International Conference SMC Tokyo, Japan.* IEEE

Timmis, J, Neal, M and Hunt, J. 2000. An Artificial Immune System for Data Analysis. *Biosystems.* **55**(1/3), pp. 143-150

Timmis, J and Neal, M. 2000. A resource limited artificial immune system. Proceedings of ES2000. BCS International conference on Knowledge Based Systems and Applied Artificial Intelligence. Cambridge. UK.

Timmis, J. 2000. Artificial Immune Systems : A novel data analysis technique inspired by the immune network theory. Ph.D. Thesis. University of Wales, Aberystwyth. 2000.

Varela, F, Coutinho, A, Dupire, B and Vaz, N. 1988. Cognitive Networks : Immune and Neural and Otherwise. *Theoretical Immunology : Part Two, SFI Studies in the Sciences of Complexity*, **2**, pp.359-371