On Parameter Adjustment of the Immune Inspired Machine Learning Algorithm AINE

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

Abstract

A machine-learning algorithm based on the natural immune system metaphor has been developed, AINE (Artificial Immune Network). AINE developed from initial work on Artificial Immune Systems for data analysis, for which detailed experimentation was undertaken as to the affect of altering algorithm parameters had on the behaviour of the system. Two of the parameters, the network affinity threshold and mutation rate have been carried over into the new version, AINE. A third parameter the number of resources has been introduced into AINE as a means by which to control network size and create a stable network structure.

This paper provides details of experiments, which alter these three parameters in AINE. It was expected that the two parameters taken from the AIS would, when altered, exhibit the same behaviour in AINE, that being the NAT scalar affecting network connectivity and mutation rate affecting network size and connectivity. Indeed, this was found to be the case. The third parameter, was designed to create stable network and previous work has shown this to be the case. It is shown in this paper, that the resource parameter can be used to control population size within the network.

Introduction

The natural immune system has been used as inspiration to create a novel unsupervised machine-learning algorithm called AINE (Artificial Immune Network) data analysis technique (Timmis & Neal, 2000). The work presented in that paper is an extension of the artificial immune system (AIS) for data analysis first proposed in (Timmis *et al*, 2000). Within the AIS there were a number of parameters that could be altered to affect the performance. A detailed investigation of the affects of these parameters was undertaken (Timmis, 2000b). This paper presents a detailed examination of the affect of those parameters on the behaviour of AINE, to ensure that the new mechanisms employed by AINE, such as resource limitation, have not caused any undue affects.

Exhaustive experiments were conducted with two significantly different data sets, a simulated data set, which contains linearly separable data and the Iris data set, which contains non-linearly separable data. A variety of values were used on each parameter and the effect of their changing on the resulting networks and behaviour AINE was observed. For consistency, each experiment was conducted five times with the results from each plotted. Where necessary, varying points of the training cycle have been extracted to show consistent behaviour throughout the learning process.

This report is broken down into the following: (i) altering the network affinity scalar (NAT) and its effect on Network size and linkage; (ii) altering the number of resources within AINE and noting its affect on network size and linkage and (iii) investigating the affect of mutation rates is then examined.

Conclusions are then drawn as to the overall affect of the parameters and how they should be used effectively in the learning process.

Previous Observations from AIS

A detailed examination of the effects of the NAT scalar and the mutation rate was undertaken for the AIS, with respect to network connectivity and network size. These results are presented in (Timmis, 2000b).

The NAT could be used to control, to some degree, the connectivity within the network. The NAT scalar can be varied between the values 0 and 1. This scalr value was used to reduce the NAT value within the network, which was used to define if two B Cells were connectd or not. By reducing the value of the NAT, the result was a lower average connectivity within the network. This observation was noticabel for both the data sets used in experimentation. It was also observed, that altering the NAT scalar did not have a significant effct on the resulting size of the network. More over, it was observed that the network acted as a support emchansmim for the evolving patterns and did not distort the results.

Additionaly, it was observed that the mutation rate could be alterd to affect the diversity within the network. This was apparnt by the fact when the mutation rate was increased, this lead to a decrease in the network connectivity. This was to be expected, as if highly muated B Cells were introduced, it stands to reason that there was less chance that connections would be made between B cells. It was also observed that the mutation rate affected the network size, in such that as the mutation rate increased the average connectivity within the network decreased. This stands to reason, as the purpose of the mutation operation is to introduce diversity into the network. As the diversity increases, less B cells can be connected (as defined by the NAT), therefore, a drop in connectivity was expected.

The third parameter within the AIS was the number of times that the antigen training set was presented t o the network. It was found that there were an optimal number of times that this should be done, but that was dependant on the data set. IF the data were presented too many times, then large networks were evolved that yield little useful information. However, if the data was not presented sufficient times, then the patterns in the data were not captured. This lead to the conclusion, that the AIS was really not that usable as an effective technique and another mechanism was required, by which a termination condition could be introduced, once the system has learnt the bets representation of the data. This was achieved with the introduction of the resource limited mechanism and the introduction of Artificial recognition Balls (ARBs), (Timmis, 2000).

Therefore, this paper examines firstly, the affect of the NAT on the network, it was expected that the results should be the same, likewise for the mutation rate. With the introduction of the resource-limited mechanism, experimentation was needed to investigate the effect of using different levels of resources in AINE and how that affected overall performance.

Results

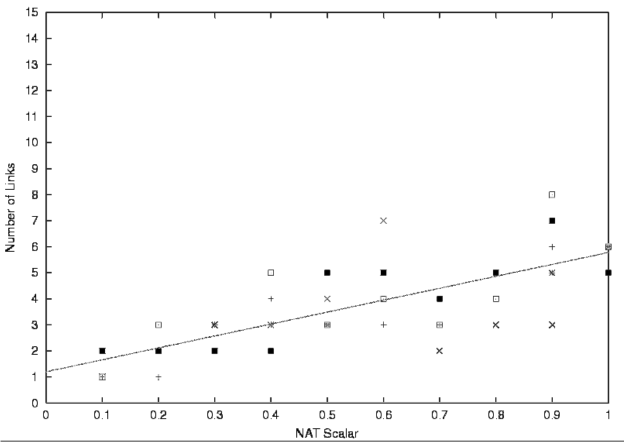
The Affect of the Network Affinity Threshold Scalar

The NAT scalar was introduced to allow a degree of control over the connectivity within the network.
In the original AIS, the NAT scalar was calculated by taking the average distance between each data item in the data set to be learnt and then recalculated after each training iteration. In AINE, the NAT is no longer recalculated (see \cite{Timmis2000c}). In the original AIS it was shown that the NAT scalar had a significant affect on the connectivity of the resulting network. It was shown that the lower the scalar, the lower the resulting affinity and therefore, more stringent and less connected the resulting network. It was also identified that the NAT scalar had little impact on the overall network size, as the dominant factor in the original AIS was the number of training iterations.

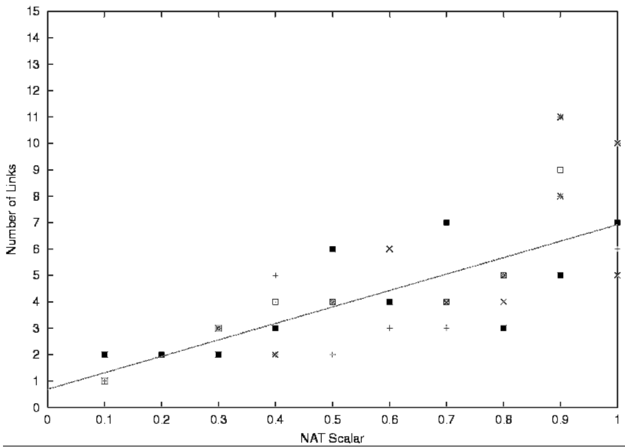
Therefore, it was expected that within AINE the same behaviour would be observed

Connectivity of the Network – Simulated Data

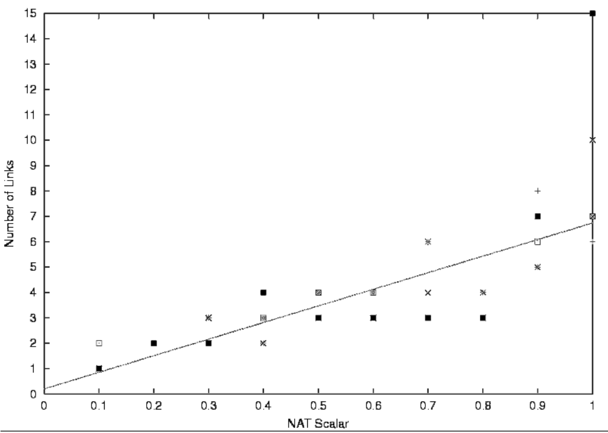
Figure 1 shows the affect of increasing NAT scalar on the network connectivity. The average numbers of links were plotted for five different runs, for a range of possible NAT scalar values. Three different time points were selected to illustrate that the affect is the same at any point of the learning process. As can be seen there is a clear increase in the connectivity of the network as the NAT scalar increases to unity. This means that the higher the NAT scalar, then the higher the connectivity and vice versa. Therefore, the NAT scalar is performing it envisaged task, of being a mechanism by which it is possible to control to some degree network connectivity. This also confirms the behaviour seen in the original AIS.



(a) Second iteration. A steady increase in connectivity can be observed as the NAT scalar is increased



(b) Fifth iteration. Again, a steady increase in network connectivity with respect to NAT scalar.



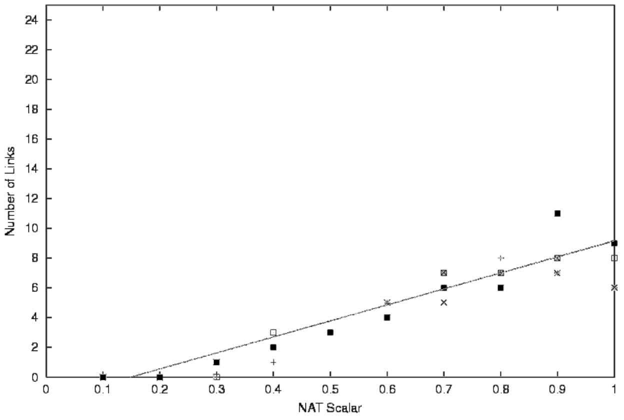
(c) Tenth iteration. Still a steady increase in connectivity can be observed.

Figure 1: NAT scalar affect on network connectivity: simulated data

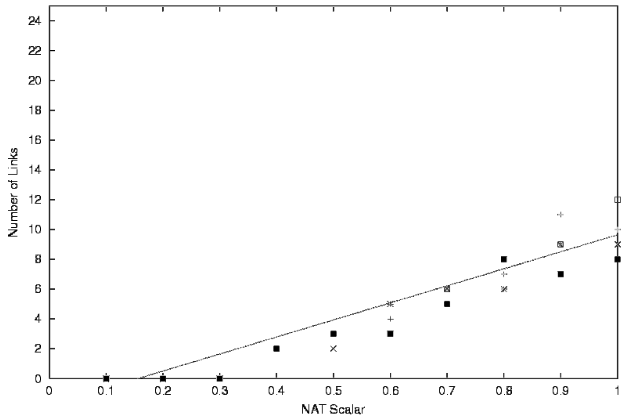
Connectivity of the Network - Iris Data

Figure 2 shows the results obtained from the Iris data set. Here again, a clear correlation can be observed that as the NAT scalar increases to unity, so the connectivity of the network increases. Figures 2(a) through to figure 2(c) show how for three different iterations of the training data the affect of the NAT scalar is the same. For the first three values of the NAT scalar, 0.1 through to 0.3 the average connectivity is zero. There are a few links in the network, indicating that the NAT scalar has been set to stringently to allow any patterns to emerge. As the NAT scalar is increased, then the correlation becomes clear, with the steady rising of network connectivity with respect to NAT scalar.

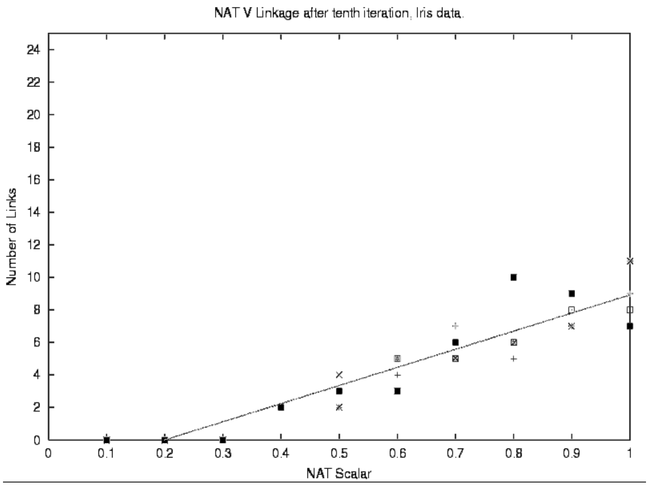
From these results, it can be concluded that the affect of increasing the NAT scalar increases the average connectivity of the network.



(a) Second iteration. A steady increase in connectivity can be observed as the NAT scalar is increased.



(b) Fifth iteration. Again, a steady increase in network connectivity is observed.

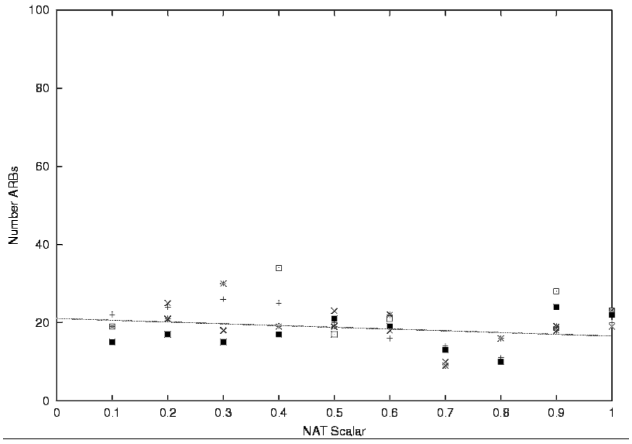


(c) Tenth iteration. Still a steady increase in connectivity can be observed.

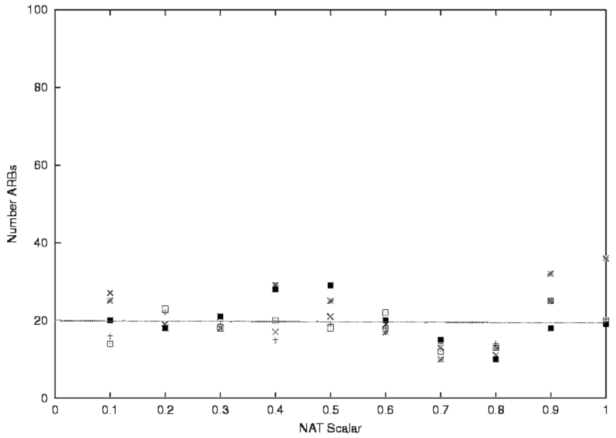
Figure 2: NAT scalar affect on network connectivity: Iris data

Network Size – Simulated Data

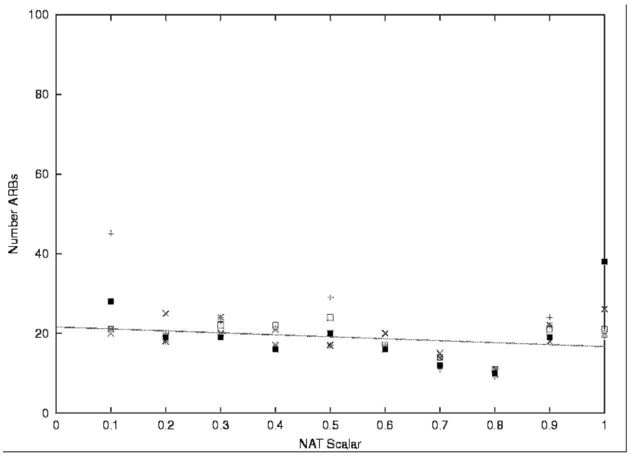
Figure 3 plots the number of ARBs as a function of the NAT scalar for the simulated data. Again, three different time points were selected to illustrate that any effect is not dependant on the time period. Figures 3(a) through to 3(c) all show no correlation between increasing the NAT scalar with network size. The average size of the network remains quite similar regardless of NAT scalar used. Therefore, this confirms the observations with the original AIS and it can be concluded that the affect of the NAT scalar is to control connectivity and altering the NAT scalar has little consequence on the resulting network size.



(a) Second iteration. No significant effect of the NAT scalar can be seen



(b) Fifth iteration. Again, no effect of the NAT scalar can be seen

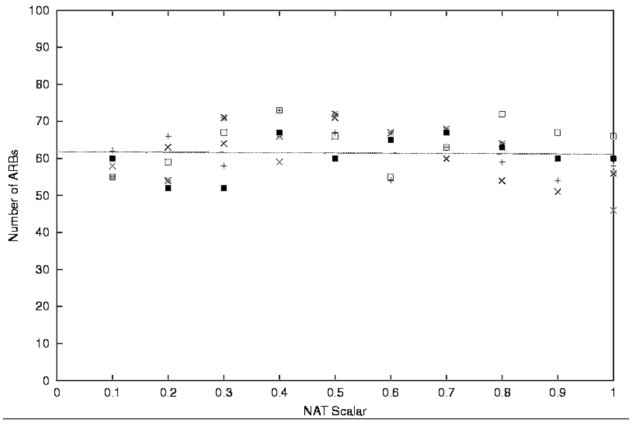


(c) Tenth iteration. No major effect can be noted, although a slight decline is present

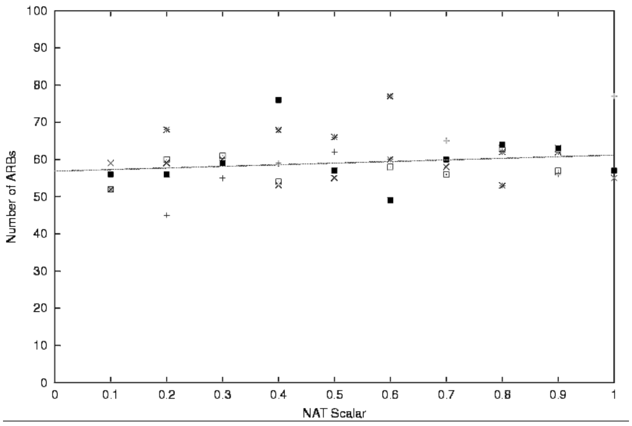
Figure 3: NAT scalar affect on network size: simulated data

Network Size - Iris Data

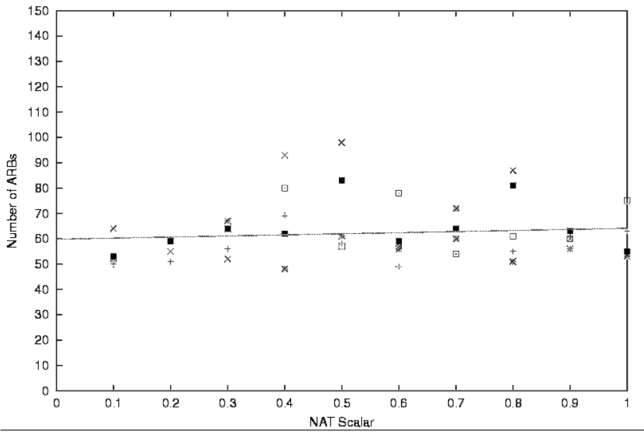
Figure 4 shows the results obtained for the Iris data set. Again, it can be clearly seen that there is no significant correlation between the NAT scalar and network size. A slight increase can possibly be noticed, but this is by no means significant and would not overtly affect the outcome of the network.



(a) Second iteration. No significant effect of the NAT scalar can be seen



(b) Fifth iteration. Again, no effect of the NAT scalar can be seen



(c) Tenth iteration. No major effect can be noted, although a slight decline is present.

Figure 4: NAT scalar effect on network size: Iris data

Summary of the Affect of NAT Scalar

There are two clear observations from these experiments:

- Altering the NAT scalar has a significant impact on the network connectivity. By reducing the NAT scalar towards zero, the connectivity of the network can be reduced. This could be potentially useful if the data is dense in nature and separation out of potential clusters seems difficult.
- Altering the NAT scalar has an insignificant impact on the resulting size of the network.

These observations go to support those found with the original AIS and it can therefore be concluded, that with the given data, the introduction of the resource limited mechanism has not altered the functionality of the NAT scalar, and within AINE, the NAT scalar performs the same function as it was designed for.

The Affect of the Number of Resources

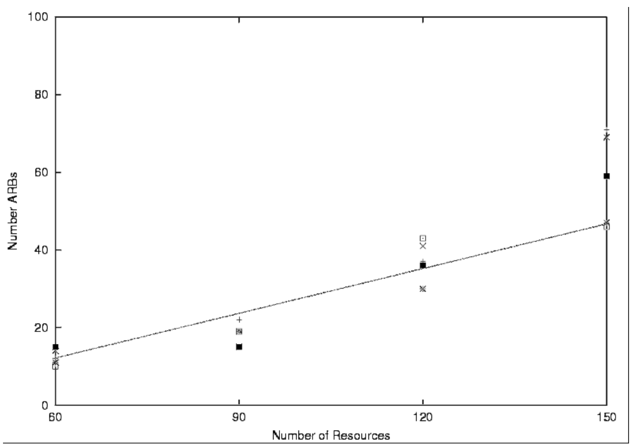
The resource limiting mechanism was introduced into the AIS in order to combat the exponential like growth in the network population. ARBs compete for resources based on their stimulation level and once an ARB could no longer claim resources it was removed from the network. This method has been shown to be effective in creating a stable network structure and it was expected that the more resources AINE was allowed, the larger the networks would be \cite{Timmis2000e}.

To that end, the number of ARBs was plotted as a function of resources to show the effect of increasing resource levels on network population. For consistency, a number of NAT scalars were selected to show the behaviour consistent over different scalars.

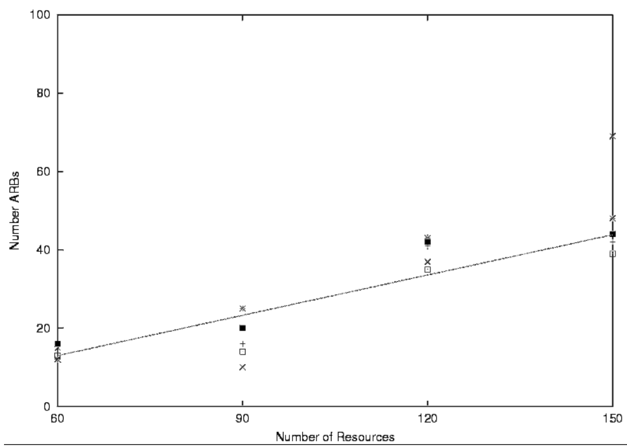
The Affect on Network Size – Simulated Data

Figure 6 shows the affect of altering the resources on the network size. Again, three different time points were selected to check for consistency. As can be clearly seen from all figures there is a steady rise in the size of the networks, when more resources are allocated.

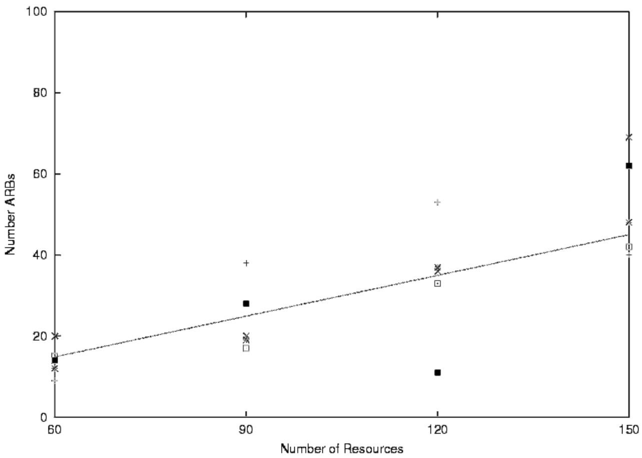
These are as expected, as with more resources it enables for weaker ARBs to claim some resources and thus survive. There is clearly a need to set a reasonable amount of resources; too many and the results will be to general, as weaker ARBs will survive; too few and then only the very strong patterns will survive and weaker patterns maybe lost. There is clearly a balance to be struck. It has been observed that as a general rule of thumb, twice the number of resources should be allocated when compared to the amount of data to be learnt.



(a) Second iteration. A clear increase in network size is observed



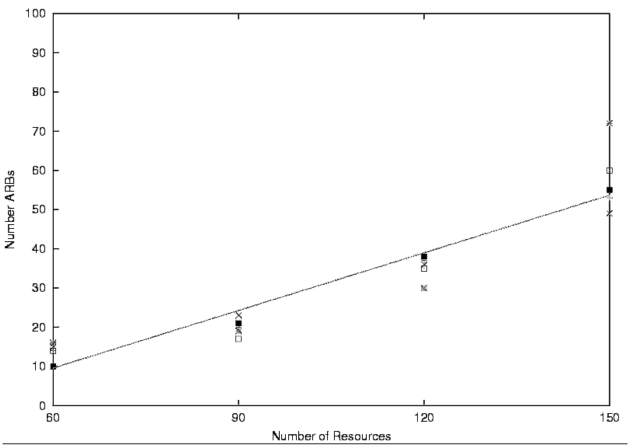
(b) Fifth iteration. The increasing network size is still apparent



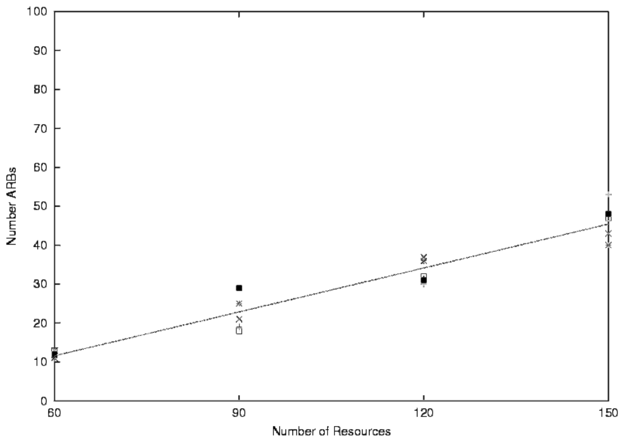
(c) Tenth iteration. Increasing network size still present

Figure 5 - Showing increase in network size as a function of resources: Simulated data, NAT 0.1

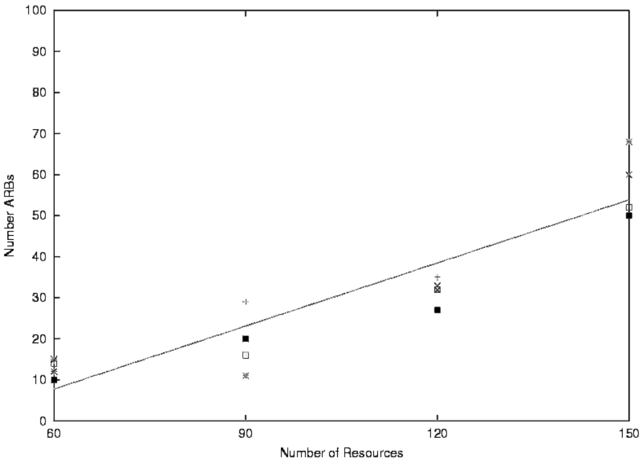
To ensure that these results are consistent throughout all NAT scalars Figure 6 plots the same three time periods as shown in Figure 5. Again, clear rises in network sizes are apparent when more resources are allocated to AINE. These results confirm that the more resources allocated to AINE, the larger the resulting networks.



(a) Second iteration. A clear increase in network size is observed



(b) Fifth iteration. The increasing network size is still apparent



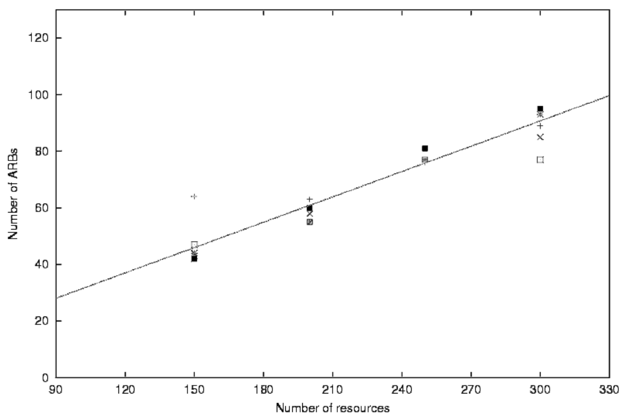
(c) Tenth iteration. Increasing network size still present

Figure 6- Showing increase in network size as a function of resources: Simulated data, NAT 0.5

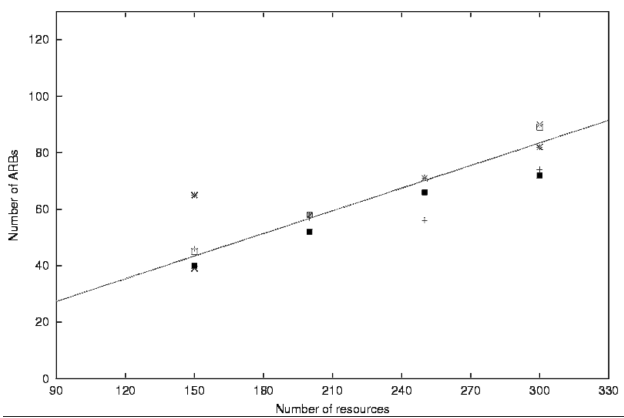
The Affect on Network Size - Iris Data

Again, the Iris was used to confirm the observations using a different data set. Figure 7 shows the affect of increasing the number of resources for a NAT scalar of 0.1, on the network size. It is clear from these figures that a steady increase in the population is observed and that there is a correlation between increasing the number of resources to the resulting network size.   Figure 8 shows similar results obtained for when the NAT scalar is set to 0.5. Again, a clear correlation can be seen between increasing the number of resources with the resulting network size.

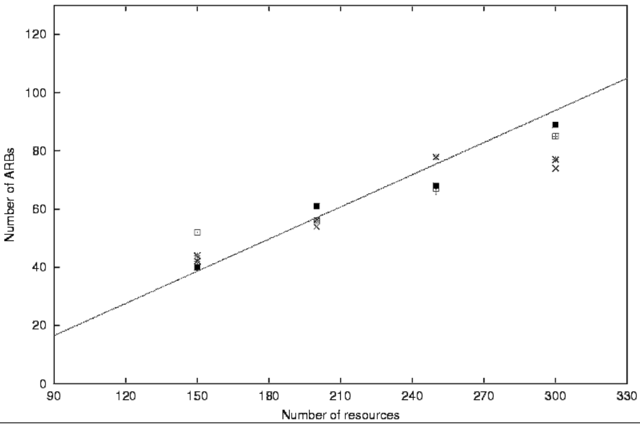
These results confirm that the main role of the resource parameter is to control the size of the network. It was shown in (timmis,xx) that large networks do not make for easy interpretation. AINE in reality acts as a compression tool, the main relationships are maintained between data items, but represented by a reduced number of data items.



(a) Second iteration. A clear increase in network size is observed as number of resources is increased

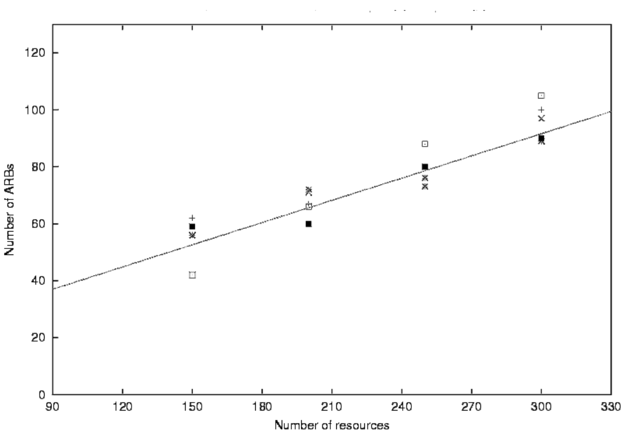


(b) Fifth iteration. The steady rise in network size is still observed.

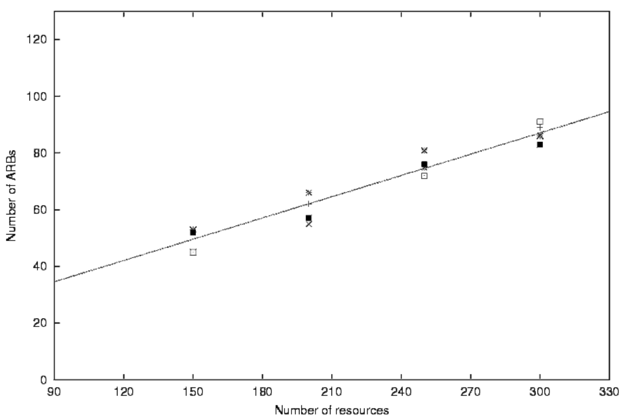


(c) Tenth iteration. Still an increase in network size with respect to number of resources.

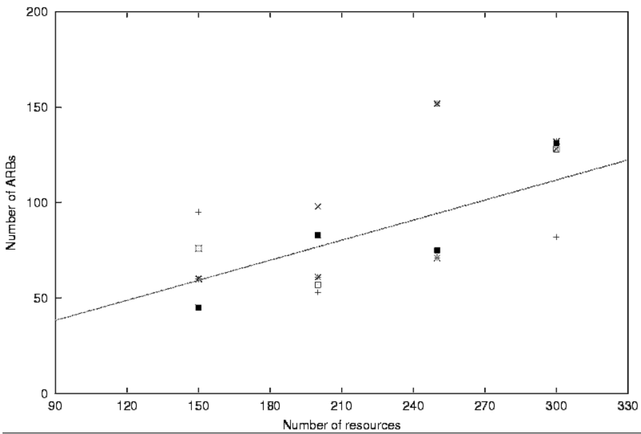
Figure 7- Showing increase in network size as a function of resources. Iris data, NAT = 0.1



(a) Second iteration. The number of resources available affects network size.



(b) Fifth iteration. Network size, again, affected by increasing the number of resources.

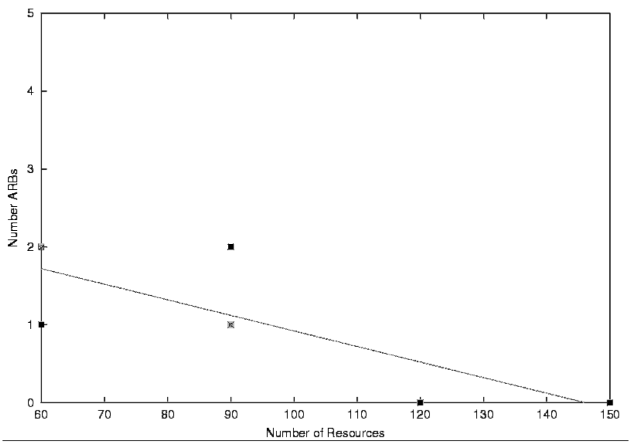


(c) Tenth iteration. Still an increase in network size can be observed.

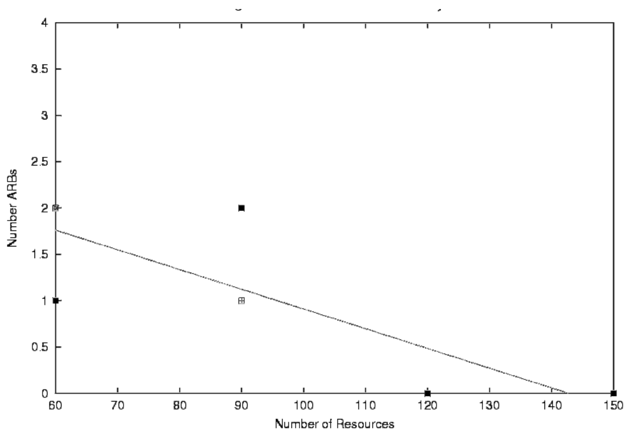
Figure 8- The affect of increasing resources on the network size with NAT set to 0.5: Iris data

The Affect on Network Connectivity – Simulated Data

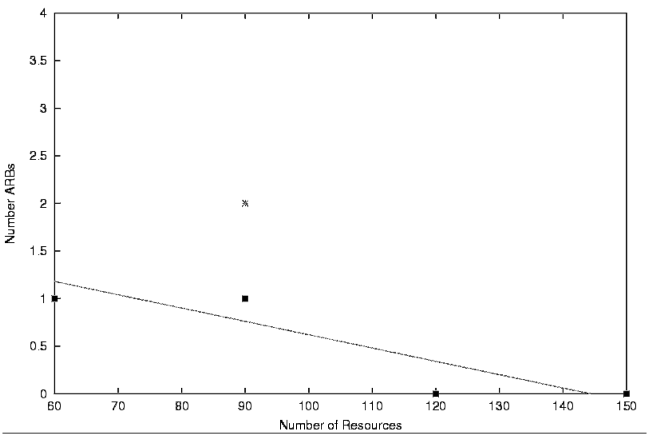
The results of these experiments are not so simplistic. Is was expected that increasing the number of resources would have little if no affect on the average linkage within the network. In general, this appears to be the case. However, Figure 9 shows a decrease in network connectivity for a low NAT scalar. However, Figure 10 shows a different picture. Here, there is no significant effect of the increasing number of resources on connectivity. This could lead to the conclusion that the main effect being seen in Figure 9 is that of a low NAT scalar on what is already a very sparse data set. The original NAT value is already very low and taking into account that the NAT scalar value is 0.1 in this case is affecting the connectivity in the network.



(a) Second Iteration.

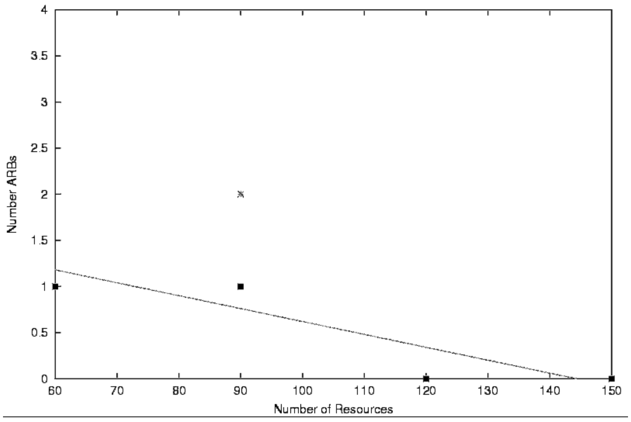


(b) Fifth iteration

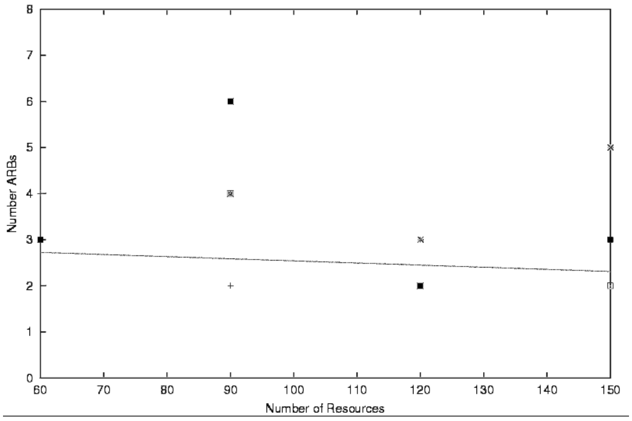


(c) Tenth iteration

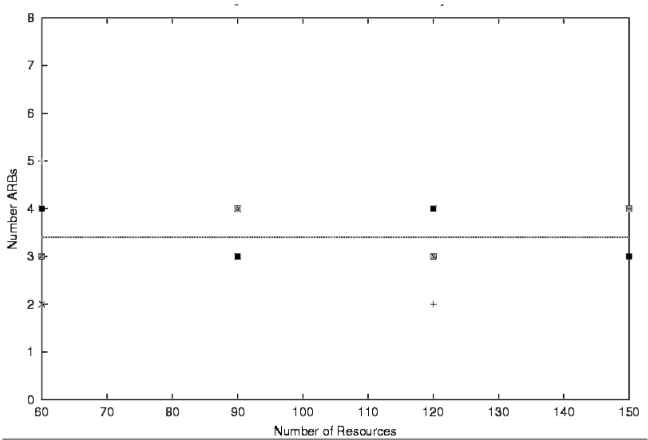
Figure 9 - The affect of increasing resources on the network connectivity NAT 0.1: Simulated data



(a) Second iteration



(b) Fifth iteration



(c) Tenth iteration

Figure 10 - The affect of increasing resources on the network connectivity NAT 0.5: Simulated data

The Affect on Network Connectivity - Iris Data

Figure 11 shows the affect of increasing the number of resources on network connectivity, for the Iris data set. These results show an insignificant affect on the linkage, as the number of resources increases. A very slight increase is observed. It is possible that this is the result of an increasing network size, combined with data that is more dense. These two conditions combined would easily lead to this behaviour being observed. Further investigation of this phenomena is being undertaken at present.

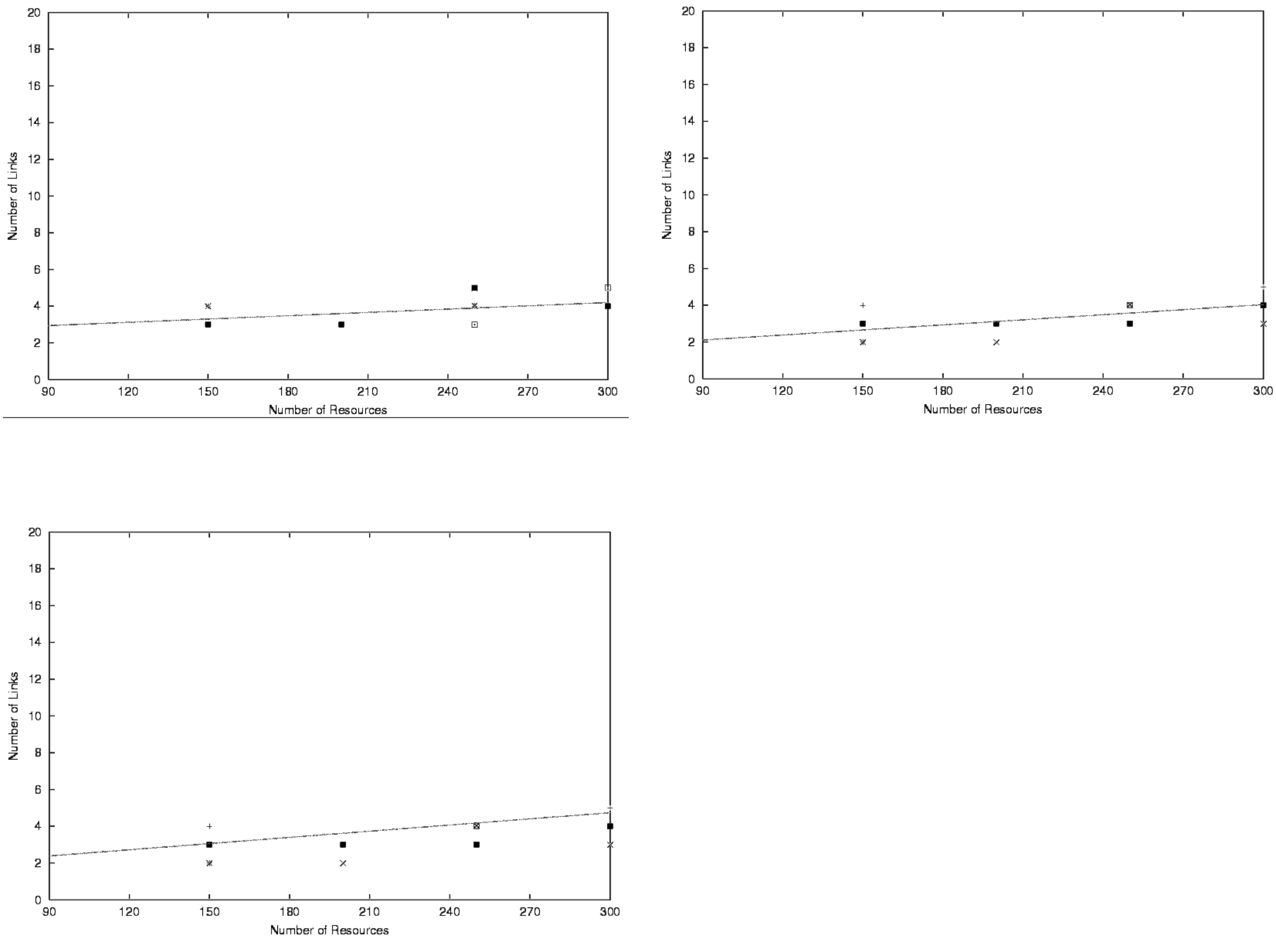


Figure 11 - Showing the affect of resources on network connectivity: Iris Data

The Affect of Mutation Rate

The Affect on Network Size – Simulated Data

The Affect on Connectivity – Simulated Data

The Affect on Network Size – Simulated Data

The Affect on Connectivity – Iris Data

Conclusions

An in-depth investigation was undertaken into examining the affect of parameters on AINE. From these experiments, a number of observations are clear:

- The NAT scalar can be used to affect connectivity within the network. The lower the NAT value, the lower the network connectivity. This is an effective tool in creating more or less stringent networks, depending on the make up of the data set being learnt.
- The NAT scalar has little if no impact on overall network size. It is clear that the network structure acts as a support mechanism for AINE and does not overly affect the resulting network. This means that the patterns that emerge from the network are from the antigen training set and not from the network itself.
- Increasing the number of resources available increases the resulting size of the network. This can be attributed to the fact that with more resources available, weaker patterns will survive within the network, as they are able to claim small amounts of resources with their low stimulation level. Setting the number of resources correctly allows for the chance for weaker patterns to emerge, but untimely if they remain weak, they will be overcome by the stringer patterns; which is exactly the behaviour the system should exhibit.
- Altering the number of resources has an inconsequential affect on network connectivity. This was expected behaviour as the main role for the resource mechanism is to control population size. It was shown that with the Iris data set, a slight increase in connectivity was observed when the number of resources was increased. This was small, but can be attributed to the fact that as more ARBs exist it increase the chance that there are more similar ARBs in the network, thus increasing the average connectivity, albeit by a small amount. This parameter should not be taken as a means to control network connectivity. Probably due to the nature of the data being used and no clear conclusions can be drawn at this stage.

Future Research

So far, tow good benchmark data sets have been experimented with. Future work will include testing other data sets, numeric in the first instance, moving then onto more complex data types, to see if the same observations are made.

Despite the conclusions drawn form this paper, clearly the setting of these parameters is sensitive to the nature of the data being learnt. To this end, research is on going into the investigation of using a Genetic Algorithm (GA) to optimise these parameters (Garret & Timmis, 2000). The GA requires a fitness function that returns the measure of what a good network is; this is not simple. Some mechanism needs to be found, that defines the *goodness* of a network. Usually, a human in the loop, visualising the network and making an informed decision, does this. Clearly, the human needs to be taken out of the loop and the metadynamics of AINE analysed to reveal what makes a good resulting network.

References

Garret, S and Timmis, J. (2000). *GA Parameter Optimisation of an Immune Inspired Machine Learning Algorithm (AINE)*. In preparation.

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

Timmis, J. (2000a). Visualising Artificial Immune Networks. University of Wales, Aberystwyth. Technical report number UWA-DCS-00-034

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

Timmis, J and Neal M. (2000). *A resource limited artificial immune system for data analysis*. Proceedings of ES2000. BCS International Conference on Knowledge Based Systems and Applied Artificial Intelligence.