A Danger Theory Inspired Approach to Web Mining

Andrew Secker, Alex A. Freitas and Jon Timmis

Computing Laboratory, University of Kent, Canterbury, Kent, UK, CT2 7NF
{ads3,aaf,jt6}@kent.ac.uk

Abstract. Within immunology, new theories are constantly being proposed that challenge current ways of thinking. These include new theories regarding how the immune system responds to pathogenic material. This conceptual paper takes one relatively new such theory: the Danger theory, and explores the relevance of this theory to the application domain of web mining. Central to the idea of Danger theory is that of a context dependant response to invading pathogens. This paper argues that this context dependency could be utilised as powerful metaphor for applications in web mining. An illustrative example adaptive mailbox filter is presented that exploits properties of the immune system, including the Danger theory. This is essentially a dynamical classification task: a task that this paper argues is well suited to the field of artificial immune systems, particularly when drawing inspiration from the Danger theory.

1 Introduction

Over the last few years, Artificial Immune Systems (AIS) have become an increasingly popular computational intelligence paradigm. Inspired by the mammalian immune system, AIS seek to use observed immune components and processes as metaphors to produce systems that encapsulate a number of desirable properties of the natural immune system. These systems are then applied to solve problems in a wide variety of domains [1]. There are a number of motivations for using the immune system as inspiration for data mining; these include recognition, diversity, memory, self-regulation, dynamic protection and learning [2].

Although some may assume AIS are only of use in computer security for virus detection and suchlike, AIS lend themselves particularly well to data mining, a strength that is in part due to a pattern matching process thought to trigger the natural immune response. In the past AIS have been turned to clustering and classification tasks with encouraging results. Hunt and Cooke [3] created an immune inspired algorithm to classify DNA sequences and the immune inspired classifier AIRS has been benchmarked on a number of standard datasets with impressive results [4, 5]. Timmis and Neal developed an AIS for clustering, AINE [6]. For a summary of immune inspired algorithms for data mining the reader is directed to the review chapter [7].

There are certain challenges associated with web mining, which we believe an AIS is particularly well suited to tacking. Baeza-Yates and Ribeiro-Neto [8] list a number of data-centric problems associated with web mining. These include the following, to which we add our thoughts on immune solutions:

- **Distributed data** – The data on the web spans countless computers. The immune system is naturally distributed. Just like the internet this distribution provides the system with disposability and diversity.
- **High percentage of volatile data** – New computers and data can be added or removed from the internet easily; likewise immune cells are constantly undergoing cell death and reproduction. The ability for both to cope with this situation shows both are adaptive, resilient and robust.
- **Large volume** – The size of the web is incredible and is constantly growing, making it difficult for systems to mine. The immune system too is made from countless numbers of cells. As each type of cell has a specialised function and works independently the system functions efficiently.
- **Quality of data** – The ease with which anyone may publish to the web can raise questions regarding its quality. Errors and omissions are common. The immune system however is noise tolerant, such that absolute matching is not required to trigger a response. Such noise tolerance is essential to an algorithm mining low quality data. The learning characteristics of the immune system are invaluable in this case. The immune system quickly learns the new characteristics of invaders when they mutate, likewise a web mining system may learn to correctly classify documents even with errors, which may be thought of as mutated words.
- **Heterogeneous data** – There exists on the internet many different types of data, in many different languages and formats. The immune system too contains a huge number of different cells each with its own specialised function and is capable of recognising a very large number of different types of antigen.

We propose to extend the field of web mining with AIS by taking inspiration from an immunological theory called the Danger theory [9]. The application of this theory to AIS was identified and discussed in depth in [10]. In this paper the authors state, "it is the authors' intention that this paper stimulates discussion [about the Danger theory] in the research community" (p. 141). The ideas presented in this paper have indeed inspired discussion and as a result significantly influenced this work in which we try to identify the benefits of a Danger theory inspired system and relate these to a practical application. We continue discussion in this area by putting forward some more practical ideas pertinent to the production of such a system, and suggest some implementation details. We believe that by harnessing Danger theory principles a new strand of artificial immune systems may emerge. We believe that because of the difficulties associated with mining such a vast and ever-changing domain, the advantages of Danger theory may be most pronounced in the field of web mining and so concentrate on this topic area. With the Danger theory principles at its core, the final system may harness the principles of context dependent activation and automatic adaptation to changing user actions and preferences.

In the next section we discuss the background to the Danger theory including Danger theory immunology and a small literature review concerning the use of Danger theory and AIS. We then discuss web mining and briefly discuss why we believe our system may be productive in this domain. Section 3 continues by using an

example of an adaptive mailbox as an illustration as to how Danger theory may be practically applied. Finally in section 4 we provide some concluding remarks about the Danger theory approach to AIS and web mining and the system we have described.

2 Background

2.1 The Danger Theory

It is acknowledged that the Danger theory is a relatively new area in the realm of artificial immune systems, so to aid the reader's understanding of this paper we would like to discuss some details of this theory. These details have been simplified and so for a more comprehensive review of this field the reader is directed towards the literature, such as [9, 11, 12, 13].

The Danger theory attempts to explain the nature and workings of an immune response in a way different to the more traditional and widely held self/nonself viewpoint. This view states that cells of the adaptive immune system are incapable of attacking their host because any cells capable of doing so are deleted during their maturation process. This view, although seemingly elegant and generally easy to understand has come under criticism as it fails to explain a number of observations. Examples of such may be the lack of immune response to injections of inert but foreign proteins, or the failure of the immune system to reject tumours even though nonself proteins are expressed. Matzinger argues a more plausible way to describe the immune response is as a reaction to a stimulus the body considers harmful, not a simple reaction to nonself. This model allows foreign and immune cells to exist together, a situation impossible in the traditional standpoint. Matzinger hypothesises that cells dying unnaturally may release an alarm signal which disperses to cover a small area around that cell, Antigen Presenting Cells (APCs) receiving this signal will become stimulated and in turn stimulate cells of the adaptive immune system. The term "danger area" was coined by Aickelin and Cayzer in [10] to describe this area, in which the alarm signal may be received by APCs. This simple explanation may provide reasons for the two anomalous observations cited. Foreign proteins in the injection are not harmful and so are ignored, likewise tumour cells are not undergoing necrotic cell death and therefore not releasing alarm signals, hence no immune reaction. The nature of these alarm signals is still under discussion but some possibilities have been empirically revealed. These include elements usually found inside a cell which are encountered outside (pre-packaged signals) or chemicals such as heat shock protein, which are synthesised by cells under stress (inducible signals) [14].

As these danger signals only activate APCs, these cells in turn stimulate B and T-cells into action according to the following rules:

- **Signal one** is the binding of an immune cell to an antigenic pattern or an antigen fragment which is presented by an APC.
- **Signal two** is either a "help" signal given by a T-helper cell to activate a B-cell, or a co-stimulation signal given by an APC to activate a T-cell.

This co-stimulation signal does not fit well in the traditional self/nonself view and also leads to the question "if a B-cell requires stimulation from a T-helper cell to become activated, what activates a T-helper cell?" As Matzinger in [11] states "perhaps for this reason co-stimulation was ignored from its creation by Laferty and Cunningham in 1974, until its accidental rediscovery by Jenkins and Schwartz in 1986" (p. 400). The danger model answers this question (often referred to as the *primer problem*) by stating that T-cells receive their co-stimulation signals from APCs, which in turn have been activated by alarm signals.

There is a criticism of the self/nonself view, which states the thymus is responsible for the negative selection of all autoreactive cells, that as the thymus provides an incomplete description of self, the selection process will impart only a thymus/nonthymus distinction on T-cells. With a few simple laws concerning the described two signal activation mechanisms applied to T-cells we may provide a simple yet plausible explanation of why autoreactive cells are found in the body, yet autoimmune disease is rare:

- A resting T-cell needs two signals to be activated (as described before).
- If a T-cell receives the first signal (a binding of its receptor to an antigen) without the second signal (co-stimulation from an APC) the T-cell is assumed to have matched a host antigen and will die by apoptosis.

Thus Danger theory allows autoreactive cells to roam the body, but if that cell is to bind to a host antigen in the absence of correct antigenic presentation by an APC, the cell will die instead of becoming activated. A number of functions originally the responsibility of the immune system are, under the danger model, actually the responsibility of the tissues. Firstly, by the second law above, simply by existing and expressing their own proteins, tissue cells induce immune tolerance towards themselves. Secondly, as an immune response is initiated by the tissues, the nature and strength of this response may also be dictated by the tissues. Thus different types of alarm signal may result in different types of response. It has long been known that in a certain part of the body an immune response of one class may be efficient, but the same class of response in another may severely harm the host. This gives rise to a notion that tissues protect themselves and use the immune system to do so, a proposition which is in stark contrast to the traditional viewpoint whereby it is the immune system's role to protect tissues.

There is still much debate in the immunological world as to whether the Danger theory is a plausible explanation for observed immune function, however we are not concerned with the plausibility of the biology. If the Danger theory is a good metaphor on which to base an artificial immune system then it can be exploited. The first steps to this are the identification of useful concepts and the application of these concepts to a suitable problem domain. It is these actions we wish to illustrate throughout the rest of this paper.

2.2 Danger Theory and Artificial Immune Systems

There are currently few AIS publications concerned with Danger theory, although the authors believe this is set to change in the coming years. Notable exceptions currently available are as follows. In the review paper [15] the author mentions

Danger theory in a small section (p. 5). In this section it is stated that danger signals may ground the immune response but gives little further detail as to how or why this may occur. In [16] the author raises a number of issues concerning an immune-inspired fault detection system for a computer. In this conceptual paper the author identifies that once again a response can be grounded by the interception of a danger signal. It is suggested that these signals would be raised by dying computer processes and some interesting parallels are drawn between a cell dying by necrosis and a thread terminating abnormally with an error such as a segmentation fault (p. 288). However, we do believe that Danger theory and therefore an implementation is about more than reacting to a threat. We do not believe this paper to propose a danger inspired system, as there is no notion of a danger area surrounding this dying process, nor is there a notion of a context dependent or localised response.

We believe it is a potentially useful area to investigate for a number of reasons. Not least of which may be the increased scalability of the AIS. In the past, the scalability of such systems has been called into question as antigens are compared with all antibodies [17]. The danger area may be one possible solution to this. In the body the immune system may only become activated within the danger area, likewise we may only need to activate antibodies within some area of interest in an AIS. One other advantage may be to harness the role of the tissues and assign different responses to different contexts. We have given thought to the implementation of such ideas, although with no such system having been produced to date we have no literature to refer to for guidance. For example, unlike most AIS algorithms, the tissue cells play a large part in a Danger theory inspired system but how should the behaviour of these cells be implemented? For example, it may be helpful to implement a set of tissue cells in addition to the set of antibodies. Each individual cell may then react to a slightly different stimulus. Furthermore we may also ask how the signal released by these cells should be interpreted. Should a signal from one cell be enough to stimulate an immune response or should activation occur only after a number of cells have been stimulated? If this latter approach is chosen we may then consider an activation function for the immune system such that a certain concentration of signal over a given space or time will initiate a response.

Based on the biology of the Danger theory, we may identify a set of Danger theory characteristics. We believe that the implementation of the majority of characteristics in this set is what may set a true Danger theory inspired AIS apart from other immune inspired systems. This set would include a context dependent danger signal, a notion of a danger area and a localised response within that area.

3 A Practical Application of Danger Theory to Web Mining

Web mining [18, 19] is an umbrella term used to describe three different types of data mining where the data source is the web. These three activities can be summarised as: **Mining structure -** studying the topology of the web made by hyperlinks; **Mining usage -** discovering knowledge from the data users leave behind once they have visited a site in that site's weblogs and **Mining content -** extracting useful information from the content of web pages and the area in which we concentrate our

studies. The strand of web mining we describe in section 3.1 is web content mining, which Linoff and Berry in [19] describe as "the process of extracting useful information from the text, images and other forms of content that make up the pages" (p. 22). We would extend this definition to cover classification of e-mail, as this process extracts information from the e-mail's text and structure for the purpose of class assignment, and e-mails are part of the internet environment. Section 3.2 briefly describes ideas involving the other two strands of web mining,

To our knowledge only a few AIS have been turned towards true web mining tasks. One is a web usage mining system as described in [20]. This aims to mine web logs to detect different user access patterns for a web server. Another is the only reference we have been able to find in the literature to an immune inspired system for text mining and therefore could be turned to web content mining. The papers [21, 22] detail an AIS concept learner for classification of HTML documents into two classes: those which were on a given topic or not on that topic. The algorithm was first tested on the UCI's 1984 Congressional voting records and then turned to classifying pages taken from the Syskill and Webert Web Page Ratings also from the UCI data repository [23]. This dataset consists of HTML pages, each on one of four different topics. The task was for this immune inspired system to predict if an unseen page was on a given topic or not when the system was trained using a number of example pages. The system was compared with a naïve Bayesian classifier and achieved a higher predictive accuracy in three out of four domains. The results showed that the system was relatively insensitive to the size of the training set which was in contrast to the Bayesian system with which it was compared.

Given these explanations we can now discuss our reasons for believing a Danger theory based immune algorithm is particularly suitable for a web content mining application. AIS have been shown to be an adaptive and robust computational paradigm. We feel this would be particularly suitable for a *dynamic* environment such as the internet. Internet content is ever-changing and so too are users' expectations. The Danger theory offers us the ability to initiate a response based on this context. To take an example suggested in [10] we could be searching for interesting documents on the internet. A danger zone may arise around an interesting document, possibly a spatial zone incorporating all pages within one hyperlink depth, such that all other documents within this zone are now considered more interesting – thus giving the notion of a localised response. As the users opinions on document interestingness change, so too will the danger signal. Some days we may release a stronger danger signal than others as users' preferences change about the data. This variable danger signal gives a perception of context dependency. Finally, using the notion that tissues may dictate the effector class of response the system may offer different responses to different types of data. Interesting pages on an academic site may be responded to in one way, whereas interesting pages on a commercial site may be responded to in quite another.

3.1 An Adaptive Mailbox

To illustrate the use of a Danger theory inspired approach to artificial immune systems, we propose an adaptive mailbox system which will accept or temporarily

ignore incoming e-mail depending on a measure of predicted interestingness to the user at that moment in time. This is essentially a classification task, but one in which the class boundaries will change both as the user's preferences change and as the status of the mailbox itself changes. During this section we use a number of small sections of pseudocode to illustrate our example. In this pseudocode we use AB to refer to an initially empty set of artificial antibodies.

The system we propose works over two distinct stages. The first will be an initialisation and training stage with the second as a running stage. During this second stage the system will both sort incoming e-mail and use feedback from the user to drive the evolutionary processes natural to an AIS. During the first stage, a summary of which is given in Pseudocode 1, the system must generate an initial collection of antibodies, and so for a given amount of time the system may observe user actions when confronted with a new e-mail. If the user is to delete a message after viewing it for less than a set period of time (normalised by length) or without reading it at all, the e-mail is deemed uninteresting, the content of the e-mail will be processed and an antibody produced. The feature vector of the antibody may include a set of words, a set of features or a combination of the two. This choice may prove critical to the performance of the system, discussion of this may be found in [24]. At this point the system will clone and mutate the antibody with the aim of generalising this antibody set. This process will continue until the system contains an appropriate repertoire of antibodies, each representing a generalised example of an uninteresting e-mail. Interesting e-mails are ignored in this initialisation phase because the goal is to produce only antibodies for classifying uninteresting e-mail.

```
PROCEDURE Initialise_Train()
  WHILE (size of AB < a threshold)
    IF(user expresses disinterest in an e-mail)
      new_ab ← create antibody from e-mail
      add new_ab to AB
      FOREACH (ab ∈ AB)
        clone and mutate ab to maximise affinity
          with new_ab
        add best n clones to AB
```

Pseudocode 1. Initialisation and Training

When the repertoire of immune cells has reached a given size the system may run on new data as described in Pseudocode 2. The AIS will convert all incoming e-mail into a format such that affinity between it and antibodies can be evaluated. Conceptually therefore the e-mail is an antigen. The system makes a distinction between the terms antigen and e-mail, as follows. Antigen is the name used to refer to a processed e-mail which contains just a generalized representation of the original e-mail and the class assigned to the e-mail. The system starts initialising an antigen count to 0. This is the count of all e-mails that have been duly processed. When this count reaches a certain number (K) of antigens, we use the latest K antigens to perform clonal selection and mutation with the set of antibodies as described in the `Update_Population` procedure. Note that it is important that this procedure is performed only after we have a reasonable number of duly processed antigens, to avoid antibodies adapted to just one antigen and preserving generality. One of the

main advantages of the immune inspired approach is this built in ability to adapt. The use of the clonal selection principle here has the effect of allowing our antibody set to change over time reflecting users changing preferences and the changing nature of the e-mail received. The clonal selection procedure will lead to an increase in the size of set AB over time and so to counter this the final line of this procedure will remove the *w* most unhelpful antibodies in the set AB.

```
PROCEDURE Continuous_adaptation()
  antigen_count ← 0
  LOOP
    receive incoming e-mail
    Process_User_Feedback(email)
    ag ← preprocess e-mail into antigen
    antigen_count ← antigen_count + 1
    IF(antigen_count = K)
      AG ← last K antigens
      Update_Population(AG)
      antigen_count ← 0
    compute degree of danger()
    WHILE(danger is high)
      compute temporal danger zone
      AG ← all e-mails in the danger zone
      FOREACH(ag ∈ AG)
        FOREACH(ab ∈ AB)
          compute affinity (ab,ag)
        high_aff ← highest affinity value
        IF(high_aff > a threshold)
          move ag to temporary store

PROCEDURE Process_User_Feedback(email)
  wait for feedback from user
  IF (user considers email uninteresting)
    assign class uninteresting to email
  ELSE
    assign class interesting to email

PROCEDURE Update_Population(AG)
  FOREACH(ab ∈ AB)
    FOREACH(ag ∈ AG)
      compute affinity(ab,ag)
      U ← {ag | ag's class is uninteresting}
      I ← {ag | ag's class is interesting}
      quality_ab ← ∑_{i∈U} aff(ab,ag_i) − ∑_{j∈I} aff(ab,ag_j)
    clone and mutate ab in proportion to
     quality_ab
  remove from AB the w antibodies with the
   lowest value of quality_ab
```

Pseudocode 2. Continuous adaptation

The procedure `Process_User_Feedback` has the main goal of assigning a class (uninteresting or interesting) to the e-mail, based on the user's feedback, so that the affinity maturation of the antibodies is based on the class of the last *K* antigens.

Note that, at any given moment, there might be several e-mails waiting for the feedback from the user to be classified and duly processed into an antigen. The system keeps working in an asynchronous way, while those e-mails are waiting feedback from the user.

In the next part of the pseudocode, the system computes the degree of danger at the moment, which depends on the current status of the mailbox. So what may the nature of this danger signal be? The danger signal should signal something is wrong but there is still time to recover, and should come from something the antigens have little or no control over. There is also no reason why the danger signal must indicate danger in the real world; it could be a positive signal as long as it signals that something of significance to the system is taking place in a particular area. In this instance we considered several possibilities for such a danger signal, such as abnormal frequency and/or size of e-mail messages and a high number of unread messages or an abnormally full inbox. Although the system might work with a combination of danger signals, each of them requiring a somewhat different response, in this paper we focus on a single danger signal based on the idea that the mailbox has too many messages. Hence, this part of the pseudocode works as follows. First, the system computes the degree of danger. If the user has no messages waiting to be read then we may not have a danger signal at all. If however there are many unread messages a danger signal may be raised. Every time a new e-mail arrives the danger area should be re-evaluated based on the current state of the mailbox. Although in the natural immune system the danger area is spatial, in our system we are not so constrained. The nature of this danger area must be decided upon based on what we want it to signal or how we want to react to it. Hence in this example we propose a temporal danger zone, the size of which will vary according to a measured value of the danger signal. Thus unread messages in the users inbox which may have been let through previously may become candidates for removal on receipt of a danger signal. This temporal danger area will therefore stretch into the inbox's past and make the system even more adaptive: the larger the degree of danger, the larger the size of the danger area.

Once the danger area has been computed, the system has to decide, for each e-mail in that area, whether the e-mail is interesting or not. This decision consists of predicting the class of the e-mail without user feedback, based solely on the affinity between each email and the antibodies. Hence, the system computes the affinity between each e-mail in the danger area and each antibody. For each e-mail in the danger area, if the affinity between that e-mail and the most similar antibody is greater than a threshold, then the e-mail is considered uninteresting (since all antibodies represent uninteresting e-mails) and the e-mail is moved into temporary storage or otherwise hidden from the user. Otherwise the e-mail is considered interesting and e-mail remains in the inbox. This is an analogy to the natural situation as a recognition of an antigen by an antibody is signal one. There is a correspondence in the natural system between the release of a danger signal and the activation of APCs which in return supply the co-stimulatory signal to activate T-cells. Just the presence of danger in this context may therefore take the place of signal two.

The procedure `Process_User_Feedback()` works as follows. The expression of disinterest by the user may be measured by the time the user views the e-mail – normalized by the e-mail's length – or user's actions upon receipt of the e-mail, such as deleting it immediately or leaving it to read later. The exact details of how the system will interpret user actions are irrelevant for the purposes of this paper.

To end this section we refer back to the end of section 2.2 in which we identified a number of points we believe would set a Danger theory inspired AIS apart from immune inspired systems. For clarification and to illustrate that by our definition this system is truly danger inspired we may draw the parallels described in Table 1.

Table 1 – Comparison between Danger theory characteristics and the adaptive mailbox

Danger Theory	Adaptive mailbox
Tissue	Mailbox
Signal 1	High affinity between antibody and antigen (e-mail)
Signal 2	Receipt of danger signal
Source of danger	High number of unread messages
Danger area	Temporal (last K e-mails)
Immune response to danger	Move emails to temporary storage
Localised response	Only emails in the danger area are considered to be moved to temporary storage

3.2 Other web mining applications

We can further identify a number of areas in which such a system may be of use to a user whilst they are having dealings with the internet covering the remaining two strands of web mining. In the above example the mailbox is one tissue, the danger signal released by which provoke one type of immune response, but these ideas can be extended to different tissues we may encounter when using the internet. One such example could be turning the system to web usage mining. Consider a system used to mine the logs generated by user accesses to an academic webpage. A significant drop in accesses could trigger a danger signal. This may prompt the system to identify page accesses that have disappeared. The classes of these may point to developing situations such as a declining quality of publication or general disinterest in a given subject area. In this situation, the danger signal may only be raised if, other webpages hosted by the same department are not suffering such a fate, or the author has updated his site recently and so should be generating interest. This gives the danger signal the required context dependency.

Similarly a web usage mining system may be implemented with an e-commerce site in mind. A drop in the frequency of access to pages detailing certain products or services may point to a change in customer preferences similar to the manner above. In this case the immune response will be triggered and may begin to look for interesting information from competitors/associates websites, a principle described in [25]. This may give rise to information on new products or services as yet not detailed on the monitored site. The danger signal here has carried the information concerning

what is going wrong. The different type of danger signal triggers a different response to that described in the previous section.

One final possibility is for the web mining system covers the domain of web structure mining. The system may monitor the local internet topology in which an academic or business related page is embedded. If the structure-mining system detects fewer and fewer pages or internet sites are linking to the one being monitored, the danger signal may once again be raised and in a similar manner to above. The adaptive AIS may then search for information on the internet which may enable the user to reverse this decline.

4 Conclusions and Further Work

In this conceptual paper we have discussed how immune inspired algorithms exhibit a similar set of desirable features to the natural immune system. We have also discussed the relatively new Danger theory and given examples as to how Danger theory principles may be used in the field of AIS. We believe that a Danger theory inspired approach to web mining could lead to the production of new and effective algorithms for knowledge discovery on the web. We believe the scalability of immune algorithms may be enhanced by initiating an immune response only when a local danger signal is present which may also yield an increase in result quality as the danger signal may be released in a context dependent manner.

Although the potential is clear, until a danger inspired AIS system is realized no firm claims can be made with regard to improvements over more traditional algorithms. However the area of danger inspired AIS algorithms is an exciting one. Our ultimate idea, combining artificial tissues capable of releasing artificial danger signals, is a significant paradigm shift for the field of artificial immune systems and with a relevant application already identified, one we are keen to pursue.

References

1. deCastro, L. N., & Timmis, J. (2002). *Artificial Immune Systems: A New Computational Intelligence Approach*: Springer.
2. Dasgupta, D. (1999). An overview of Artificial Immune Systems. In D. Dasgupta (Ed.), *Artificial Immune Systems and Their Applications* (pp. 3-21): Springer
3. Hunt, J. E., & Cooke, D. E. (1996). Learning using an artificial immune system. *Journal of Network and Computer Applications, 19*(2), 189-212.
4. Watkins, A. (2001). AIRS: A resource limited artificial immune classifier. Masters Thesis, Mississippi State University.
5. Watkins, A., & Timmis, J. (2002). Artificial Immune Recognition System (AIRS): Revisions and Refinements. *In proceedings of The First International Conference on Artificial Immune Systems (ICARIS 2002)* (pp 173-181), Canterbury, UK.
6. Timmis, J., & Neal, M. (2001). A resource limited artificial immune system for data analysis. *Knowledge Based Systems, 14*(3-4), 121-130.

7. Timmis, J., & Knight, T. (2002). Artificial Immune Systems: Using The Immune System as Inspiration for Data Mining. In H. A. Abbass, R. A. Sarker & C. S. Newton (Eds.), *Data Mining: A Heuristic Approach* (pp. 209-230): Idea Group Publishing.
8. Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*.: Addison Wesley Longman
9. Matzinger, P. (2002a). The Danger Model: A Renewed Sense of Self. *Science, 296*, 301-305.
10. Aickelin, U., & Cayzer, S. (2002). The Danger Theory and Its Application to Artificial Immune Systems. *In proceedings of The First International Conference on Artificial Immune Systems (ICARIS 2002)*(pp. 141-148), Canterbury, UK.
11. Matzinger, P. (1998). An Innate Sense of Danger. *Seminars in Immunology, 10*(5), 399-415.
12. Anderson, C., & Matzinger, P. (2000). Danger: The view from the bottom of the cliff. *Seminars in Immunology, 12*(3), 231-238.
13. Matzinger, P. (2002b). The Real Function of The Immune System or Tolerance and The Four D's. Retrieved 30/10/2002, 2002, from http://cmmg.biosci.wayne.edu/asg/polly.html
14. Gallucci, S., & Matzinger, P. (2001). Danger signals: SOS to the immune system. *Current Opinion in Immunology, 13*(1), 114-119.
15. Williamson, M. M. (2002). Biologically Inspired Approaches to Computer Security (*HP Labs Technical Reports HPL-2002-131*): HP Labs Bristol, UK. Available from: http://www.hpl.hp.com/techreports/2002/HPL-2002-131.html
16. Burgess, M. (1998). Computer Immunology. *In proceedings of The 12th Systems Administration Conference (LISA 1998)*, Boston, USA.
17. Kim, J., & Bentley, P. J. (2001, July 7-11, 2001). An Evaluation of Negative Selection in an Artificial Immune System for Network Intrusion Detection. *In proceedings of The Genetic and Evolutionary Computation Conference 2001 (GECCO 2001)* (pp. 1330-1337), San Francisco, USA.
18. Chakrabarti, S. (2003). *Mining the web (Discovering Knowledge from Hypertext Data)*: Morgan Kaufmann.
19. Linoff, G. S., & Berry, M. J. A. (2001). *Mining the web (Transforming Customer Data into Customer Value):* Wiley.
20. Nasraoui, O., Dasgupta, D., & Gonzalez, F. (2002). The Promise and Challenges of Artificial Immune System Based Web Usage Mining: Preliminary Results. *In proceedings of The SIAM Workshop on Web Analytics* (pp. 29-39), Arlington, VA
21. Twycross, J. (2002). An Immune System Approach to Document Classification (*HP Labs Technical Reports HPL-2002-288*): HP Labs Bristol, UK. Available from: http://www.hpl.hp.com/techreports/2002/HPL-2002-288.html
22. Twycross, J., & Cayzer, S. (2002). An Immune System Approach to Document Classification (*HP Labs Technical Reports HPL-2002-292*): HP Labs Bristol, UK. Available from: http://www.hpl.hp.com/techreports/2002/HPL-2002-292.html
23. Blake, C. L., & Merz, C. J. (1998). *UCI Repository of machine learning databases.* Retrieved 20 May 2003, from http://www.ics.uci.edu/~mlearn/MLRepository.html
24. Diao, Y., Lu, H., & Wu, D. (2000). A comparative study of classification based personal e-mail filtering. *In proceedings of The Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2000)* (pp. 408-419), Kyoto, Japan.
25. Liu, B., Ma, Y., & Yu, P. S. (2001). Discovering unexpected information from your competitors' web sites. *In proceedings of The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2001)* (pp. 144-153), San Francisco, USA.