Volume 3, Issue 2, February 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Artificial Immune Recognition System (AIRS): A New Approach

N. Priya
Assistant Professor
Department of Computer Science
Ramanujam Centre For Computational Intelligence
Bharath University,
Chennai, Tamilnadu
India

Abstract — In recent years there has been considerable interest in exploring the prospective of Artificial Immune Systems for applications in computer science and engineering. These systems are inspired by various aspects of the immune systems of mammals. Some of these aspects, such as the distinction between self and non-self and the concept of negative selection, have a natural and intuitive fit for applications involving computer security, network intrusion detection. Moreover, research into natural immune systems suggests the existence of learning properties which may be used to advantage in machine learning systems Artificial Immune Systems (AIS) are machine-learning algorithms that embody some of the principles and attempt to take advantages of the benefits of natural immune systems for use in tackling complex problem domains. The Artificial Immune Recognition System (AIRS), is one such supervised learning AIS that has shown significant success on broad range of classification problems. The AIRS algorithm is analysed from the perspective of reasonable design goals for an immune inspired AIS and a number of limitations and areas for improvement are identified. A number of original and borrowed augmentations, simplifications and changes to the AIRS algorithm are then proposed to addresses the identified areas. A professional-level implementation of the AIRS algorithm is produced and is provided as a plug-in for the WEKA machine-learning.

Keywords— Artificial Immune System, AIRS algorithm, Classifier, Classification algorithms

I. INTRODUCTION

An artificial immune system (AIS) is a class of adaptive or learning computer algorithm inspired by function of the biological immune system, designed to apply for difficult problems such us intrusion detection, data clustering, classification and search problems. The field of AIS research is around approximately 15 years and the history mainly concerns with the anomaly identification, such as intrusion detection systems. Nowadays they are applicable to broader domains such as function optimization in the case of AIRS classification. The recent field of approach provides an chance to give design for new specialized algorithms, they are successful in encoding the metaphor. The characteristics of the algorithm contains self regulation, performance, generalization and parameter stability. Mostly AIRS consists number of core principles which give loose abstractions from mammalian immunological research.

II. OVER VIEW

A. Immunology and its Basics

The IS consists of a multitude of cells and molecules which interact in a variety of ways to detect and eliminate infectious agents (pathogens). These interactions are localized because they depend upon chemical bonding—surfaces of immune system cells are covered with receptors, some of which chemically bind to pathogens, and some of which bind to other immune system cells or molecules to enable the complex system of signalling that mediates the immune response. Most IS cells circulate around the body via the blood and lymph systems, forming a dynamic system of distributed detection and response, where there is no centralized control, and little, if any, hierarchical organization. Detection and elimination of pathogens is a consequence of trillions of cells interacting through simple, localized rules. A consequence of this is that the IS is very robust to failure of individual components and attacks on the IS itself.

B. AIRS Algorithm and its parameters

The function of AIRS algorithm is a preparatory pool recognition or memory cells data exemplars) which are representative of the training data the model is exposed to and is suitable for classifying unseen data. The function of the AIRS algorithm is to prepare a pool of recognition or memory cells(data exemplars) which are representative of the training data the model is exposed to,and is suitable for classifying unseen data.

1. ***Initialization***

This step of the algorithm consists of preparing the data for use in the training process, and preparing system variables. The training data is normalized so that the range of each numeric attribute is in the range [0,1]. An affinity measure is required for use through the Training process

2. ***Antigen Training***

The AIRS algorithm is a single-shot algorithm in that only one pass over the training data is required to prepare a classifier. Each antigen is exposed to the memory pool one at a ime. The recognition cells in the memory pool are stimulated by the antigen and each cell is allocated a stimulation value (inverted affinity). The memory cell with the greatest stimulation is then selected as the best match memory cell for use in the affinity maturation process.

3. ***Competition for Limited Resources***

After a number of mutated clones of the best matching memory cell are added to the ARB pool, the process of ARB generation and competition begins The process is quite simple from a high-level. Competition for limited resources is used control the size of the ARB pool and promote those ARBs with greater stimulation (and thus affinity) to the antigen being trained on. The stop condition in the middle of the loop allows the final step of clone generation to be avoided when the ARB pool reaches a desirable state. In this process only ARBs of the same class as the antigen are considered, meaning that the class of an ARB is never adjusted in the mutation process. The final step sees each ARB in the pool have mutated clones generated using the same clonal methods.

4. ***Memory selection***

Once the stop condition for the ARB refinement process is completed, the ARB with the greatest normalised stimulation scoring is selected to become the memory cell candidate.
The ARB is copied into the memory cell pool if the stimulation value for the candidate is better than that of the original best matching memory cell. A check is made to determined if the original best matching memory cell should be removed. This occurs if the affinity between the candidate memory cell and the best matching cell is less than a cut-off. This memory cell replacement cut-off is defined as:
Cut Off =affinity Threshold++affinity Threshold Scalar.

5. ***Classification***

When the training process is completed, the pool of memory recognition cells becomes the core of the AIRS classifier. The data vectors contained within the cells can be de-normalised or left as is for the classification process. Classification occurs using a k-Nearest Neighbour approach where the k best matches to a data instances are located and the class is determined via majority vote.

C. Evaluation of the AIRS algorithm

The AIRS algorithm did spring not into existence, in fact, it is a collection of elements and processes developed for other supervised and unsupervised AIS algorithms. This section provides an overview of some of the precursor algorithms to AIRS, specifically the elements and processes borrowed from said algorithms. Also provided is an overview of some of the augmentations and research performed into the AIRS algorithm over the past few years of its history.

III. PRECURSORS TO AIRS

1. ***IMMUNOS-81***

The first approach to using immune system metaphor for a supervised learning and classification system was called Immunos-81 by Carter [9]. The system was designed with the intent of taking advantages of the features of the immune system without adhering too closely to the biological aspects and equations. The exceedingly complex system used T-cell and B-cell equivalents as well as a library of elements currently in the system. It was shown to provide good results in terms of classification accuracy on the Cleveland heart disease datasets

2. ***Immune Network Theory Inspired AIS***

Immune Network Theory (INT) proposes that the immune system maintains a network of cells that learn and maintain memory using feedback mechanisms. A feature of thistheory is that once information is learned by the network, it is then capable of being forgotten unless the information is reinforced. Work in [12] proposes an AIS based onconcepts form INT. The system maintains a population of recognition cells that respond with a stimulus when presented with an antigen, and are connected with links that represent the similarity between the cells.The primary problem with the INT based AIS was that it suffered from a population explosion. the system was extended to support population control mechanisms, and was called a Resource Limited Artificial Immune System (RLAIS). Also introduced with the RLAIS was the concept of artificial recognition balls (ARBs) which represents a number of identical B-cell (recognition cells) in the network. The system is configured to allow a certain number of individual B-cells, which are dividedby and competed for by the ARBs, where the higher the stimulation of the ARB, the more B-cell resource it is allocated. When an ARB losses all of its resources, it is considered not a useful representation of the training data, and is removed from the system. Finally,the last change made was that the NAT was calculated once at the start of the run and kept constant, rather than recalculated each algorithm iteration.

3*. Clonal Selection Inspired AIS*

Clonal selection theory is the idea that those cells that are effective at recognizing pathogenic material are selected (in a Darwinian sense) to survive and propagate. Work in devised a technique called the clonal selection algorithm (CSA), which was revisited in and renamed CLONALG which was based on the clonal selection theory. The technique uses elements of the affinity maturation process of the immune response to maintain a population of cells capable of learning a desired problem. The CLONALG algorithm was shown to be similar to that of evolutionary strategies (ES) from the field of evolutionary computation, though CLONALG searched by blind cumulative variation and selection by cloning, mutating and advancing best match recognition cells.

4. **Expectations and Limitations of AIRS**

1.Easily understood internal representation
2.Ability to generalise from input data
3.Predictable training times
4.Online learning
5.Potential to act as an associative memory
6.Acceptance of continuous and nominal variables
7.Capacity to learn and recall large numbers of patterns
8.Experience-based learning
9.Supervised learning

5. ***Extensions to AIRS***

AIRS had had been shown to be a successful classifier on a broad set of well known classification problems with small numbers of classes. A study was undertaken to investigate the performance of AIRS on problems with many multiple classes and an increased number of features (attributes), compared to the Learning Vector Quantisation (LVQ) algorithm. Artificial problem domains with three, five, eight and 12 classes were evaluated, as were six common problems from the field of machine learning. In just about all cases for the common real-world problems, AIRS out-performed LVQ configured with a similar number of elements as AIRS, as well as an optimised version of the LVQ algorithm. In the case of one of the datasets (credit card classification problem), AIRS achieved results better than the best classification results known at the time for the problem Work in investigated the effect on AIRS of introducing additional irrelevant features to classification datasets. It was speculated that the performance of the technique would decrease given its reliance on the Euclidean distance measure. Results were unexpected and indicated a small drop classification accuracy on the tested problem domains. A comparison of results with LVQ showed that LVQ was capable of out-performing theirs technique when initialized with the same number of codebook vectors discovered by AIRS, though was not capable of doing so when configured independently. A number of augmentations to the AIRS algorithm have been suggested and tested in previous research. Some issues addressed in include the handling of ties during classification, the restructuring of the ARB pool, and resource allocation. While focused on augmentations of version one of the AIRS algorithm, the results are still interesting. Four additional alternative tie handling methods for classification were proposed and tested in addition to the standard AIRS tie handling technique: First labelled first served (traditional AIRS technique) – In the case of a tie, select the class with the lowest assigned identification number (first encountered) Sum of affinities – Select the class with the highest sum affinity Selection based on class proportions – A probabilistic approach that would use the number of instances of each class to determine probabilities Include more memory cells – Continue to add memory cells until the tie was resolved (uses a new parameter k-additional which defines the maximum number of additional cells to include)Interestingly the probabilistic approach "Selection based on proportions" performed the best on the noisy yeast machine learning dataset, and the "sum affinities" approach which logically appear the most sensible performed relatively poorly. The alternative ARB pool structures were centred around the assumption that the fact that ARBs of different classas the antigen in the pool had little impact on training ad the resulting classifier. One alternative resource allocation scheme was also tried where resources were allocated based on the class proportions in the ARB pool (again for AIRS1). Results were inconclusive on the E. coli and yeast machine learning dataset. The issue of using different distance measures for calculating affinity was addressed in. Euclidean distance is traditionally the distance measure used in AIRS algorithm for calculating affinity and stimulation scorings. This requires that all attributes be converted to numeric (real) values, which has no meaning for nominal and potentially less meaning for discrete attributes. Twelve alternative distance measures were proposed for use in AIRS and evaluated, each capable of producing values in the require range of [0,1]. The measures used specifically catered to the nature of attributes, and were tested on common machine learning datasets with a range of numeric, nominal and discrete attribute types. Results indicated that by using more natural and arguably useful measure of comparison, AIRS can achieve better classification accuracy.

D. AIRS and Design Goals

The mentioned set of design goal seem reasonable for an AIRS-like supervised learning system, therefore it is beneficial to evaluate AIRS against each goal to aid in identifying potential limitations.

1. ***Easily understood internal representation***

The benefit of having an easily understood representation is that it allows a user to look at the classification decisions made by the system and understand directly why a particular action was taken. An opaque system creates a situation where the user has little or no understanding of how the system arrived at a decision. Some artificial neural network architectures suffer from this problem. AIRS uses natural or real vectors to represent data in the same manner as data is represented in the domain. AIRS clearly meets this design goal, though it is common during the initialisation process in AIRS to normalise the data vectors, though this may not necessary.

2 ***Ability to generalise from input data***

To generalise is to draw a specific case from a more general case. In terms of classification, AIRS uses exemplars to represent the general case and determines the specific case (classification) based on a best match or majority vote from the top k best matches. Moreover, the exemplars are representative of one or more training instances, meaning that the number of exemplars is less than that of training instances. This data reduction feature has been observed to be up to approximately 50% or more for a rangefor standard machine learning datasets. The minimum requirement for an AIRS like system is to have as many or less exemplars as there are training instances.

3 .Predictable training times

This design goal is the first design goal for which AIRS may have a problem. Although the computational complexity for AIRS has not been described, it can be estimated to be of reasonably high complexity given the repeated number of similarity (affinity and stimulation) elements involved for each antigen presented during training. Further, the algorithm consists of loops within loops within loops (resource allocation and pruning ARB refinement antigen presentation), which are dependent on the antigen in relation to the memory pool at the time in the training schedule. Even though computational improvements have been made on the original AIRS algorithm (AIRS2), this remains an area that needs both further investigation and improvement.

4. *Online learning*

Online learning or continuous learning refers to the algorithms ability to improve or perform further training after delivery of the classifier. LVQ and many artificial neural network algorithms are examples of algorithms that exhibit this feature. This feature has not been explored in the AIRS algorithm to date (to the author's knowledge).

5. *Experience-based learning and Supervised Learning*

The last two design goals are quite vague and similar, in fact you can't have one without the other. The first: experience based learning, indicates that the algorithm learn from its experience with the problem domain (training data), and the second: supervised learning, indicates that the learning it performs is supervised, meaning both input and desired output are known at training time. AIRS is a supervised learning algorithm that learns from experience.

IV. CONCLUSIONS

In this paper elaborately discuss about the artificial immune recognition system (AIRS) algorithm. it used to explain about the competence with the AIRS technique both in terms of history, algorithmic function and application. Further, given some competence, the intent included proposing potential extensions to the technique. These intentions of this work were achieved successfully. A complete history of the AIRS algorithm was provided with a focus on research that was of fundamental or of interest.

REFERENCES

[1] Andrew B. Watkins, Exploiting Immunological Metaphors in the Development of Serial, Parallel, and Distributed Learning Algorithms. 2004. University of Kent.

[2] Julie Greensmith, New Frontiers For An Artificial Immune System 2003. University of Leeds, Hewlett Packard Labs Technical Report number HPL.

[3] Witten and Eibe Frank. Data Mining: Practical machine learning tools with Java implementations, San Francisco: Morgan Kaufmann, 2000

[4] Leandro N. de Castro and Fernando J. Von Zuben, Learning and Optimization Using the Clonal Selection Principle IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, vol. 6, pp. 239-M 251, 2002.

[5] Andrew Watkins and Jon Timmis, "Artificial Immune Recognition System (AIRS): Revisions and Refinements," 1st International Conference on Artificial Immune Systems (ICARIS2002), University of Kent at Canterbury, pp. 173-181, 2002.

[6] Andrew Watkins Home Page http://www.cs.kent.ac.uk/people/rpg/abw5/home.html

[7] Jonathan Timmis Home Page http://www.cs.kent.ac.uk/people/staff/jt6/

[8] Donald Goodman Home Page http://artificial-science.org/

[9] Julie Greensmith Home Page http://www.cs.nott.ac.uk/~jqg

[10] Leandro N. de Castro Home Page http://www.dca.fee.unicamp.br/~lnunes/

[11] Jonathan Timmis Home Page http://www.cs.kent.ac.uk/people/staff/jt6/

