

An Efficient and Effective Immune Based Classifier

Shahram Golzari, Shyamala Doraisamy,
Md Nasir Sulaiman and Nur Izura Udzir
Faculty of Computer Science and Information Technology,
University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Abstract: Problem statement: Artificial Immune Recognition System (AIRS) is most popular and effective immune inspired classifier. Resource competition is one stage of AIRS. Resource competition is done based on the number of allocated resources. AIRS uses a linear method to allocate resources. The linear resource allocation increases the training time of classifier. **Approach:** In this study, a new nonlinear resource allocation method is proposed to make AIRS more efficient. New algorithm, AIRS with proposed nonlinear method, is tested on benchmark datasets from UCI machine learning repository. **Results:** Based on the results of experiments, using proposed nonlinear resource allocation method decreases the training time and number of memory cells and doesn't reduce the accuracy of AIRS. **Conclusion:** The proposed classifier is an efficient and effective classifier.

Key words: Artificial Immune Recognition System (AIRS), resource allocation, Artificial Immune System (AIS), clonal selection, Artificial Recognition Ball (ARB), nonlinear resource allocation, EXPAIRS generates, feature vector

INTRODUCTION

Artificial Immune System (AIS) is a computational method inspired by the biology immune system. It is progressing slowly and steadily as a new branch of computational intelligence and soft computing (de Castro and Timmis, 2002; de Castro and Timmis, 2003; Golzari *et al.*, 2009). It has been used in several applications such as machine learning, pattern recognition, computer virus detection, anomaly detection, optimization and genre classification (de Castro and Timmis, 2002; de France *et al.*, 2005; Igawa and Ohashi, 2007; Kim *et al.*, 2007; Doraisamy and Golzari, 2010). One of AIS based algorithms is Artificial Immune Recognition System (AIRS). AIRS is a supervised immune-inspired classification system capable of assigning data items unseen during training to one of any number of classes based on previous training experience. AIRS is probably the first and best known AIS for classification, having been developed 2001 Watkins and Boggess (2002).

AIRS has four main steps: Initialization, ARB generation, Competition for resources and nomination of candidate memory cell and finally promotion of candidate memory cell into memory pool. AIRS uses linear method for resource allocation. This linearity increases the number of generations needed to produce the memory cells and the training time of the algorithm.

The aim of this study is applying a new nonlinear approach in the resource allocation of the algorithm to tackle this drawback. The new immune based classifier, AIRS with nonlinear method, is tested against benchmark data to determine the affects of the proposed improvements on AIRS characteristics such as training time, number of memory cells and accuracy.

The rest of the study is organized as follows: Material and Methods section provides a brief description of the AIRS classifier and proposed resource allocation method. Experiment setup section explains the experiments conduction. The results of experiments are presented and discussed in the results section.

MATERIALS AND METHODS

AIRS: Artificial Immune Recognition System (AIRS) was developed by Watkins and Boggess (2002). To show the capability of AIS to do the classification was the initial objective of developing AIRS, but results shown that AIRS is comparable with famous classifiers. Before AIRS, most artificial immune system researches concerned unsupervised learning and clustering. AIRS uses the clonal selection, affinity maturation and memory cell production concepts of the immune system together with the resource limited artificial immune

Corresponding Author: Shahram Golzari, Faculty of Computer Science and Information Technology, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia

system concept introduced by Timmis and Neal (2001). In fact, AIRS is a hybrid algorithm that uses concepts of different immune system theories.

As AIRS is resource limited artificial immune system, this concept i.e., “resource limited” is explained before the focus on details of AIRS.

Resource limited concept was incorporated to AIS in the study by Timmis and Neal (2001) to control the population and avoid the exponential growth of B-cells in the system. The recognition ball idea from the immune system was used to introduce the Artificial Recognition Ball (ARB) concept. Each ARB in the system is the representative of a number of identical B-cells and the system would have a maximum number of B-cells. The number of B-cells claims by each ARB depends on its affinity. If the total number of B-cells claimed by ARBs is greater than the maximum number of B-cells allowed, then additional B-cells must be removed. In fact, ARBs compete together. In order to remove additional B-cells, the weakest ARB, i.e., ARB that was least awarded B-cells, is selected and a sufficient number of its B-cells are removed. Once the number of its B-cells becomes zero then the ARB is removed from the system. This process is repeated until the allocated B-cells are equal to the maximum number allowed.

In AIRS, feature vectors (labeled data) presented during training and test phases are named as antigens and the system units are called ARBs or B-cells. In theory, similar B-cells are represented with an ARB and ARBs compete with each other for a fixed number of B-cells. AIRS adapts these concepts. With AIRS, ARB and B-cell are the same and ARBs (or B-cells in this case) compete for a fixed number of resources. In this competition, the weaker ARBs are removed from the system and stronger ones stay in the system to generate better ARBs in following step. The algorithm generates new instances as memory cells from the population of ARBs, that would be used in the classification task finally. Therefore memory cells are actually the best ARBs. These ARBs have highest affinities to training antigens.

AIRS has four stages: The first stage is performed once at the beginning of the algorithm. This stage includes normalization and initialization. The following three stages are performed repeatedly for each antigen in the training set. These stages are ARB generation, resource competition and insertion of candidate memory cell into memory cell pool. With normalization, the algorithm puts the distances between two data in the [0,1] interval. Following this, the algorithm initializes the memory cell pool and ARB pool from randomly selected training data. This

prepares the algorithm to generate memory cells repeatedly from antigens. The steps are as follows (Watkins and Bogges, 2002; Watkins *et al.*, 2004):

- A training antigen is compared with all the memory cells in the memory cell pool that have the same class as the antigen. The memory cell most stimulated by the antigen, named MC match, is selected and cloned. The new clones are then mutated. The memory cell and all generated clones are put into the ARB pool. The number of generated clones depends on the affinity between the memory cell and antigen. This affinity is determined by Euclidean distance between the feature vectors of the memory cell and the training antigen. The smaller Euclidean distance means there is higher affinity, generating a large number of clones
- In the next step, the training antigen is presented to all ARBs in the ARB pool. All ARBs are rewarded based on the affinity between the ARB and the antigen and its class. If an ARB and antigen are belonging to the same class, the ARB is rewarded highly for high affinity with the antigen; otherwise, the ARB is rewarded highly for a low value of affinity measure. The rewards are in the form of number of resources (resource allocation). More rewards cause more resources. When the number of resources is calculated for all ARBs, the sum of allocated resources in the system typically exceeds the maximum number allowed for the system and the excess number of resources held by ARBs must be removed from the system. The algorithm finds the ARB with lowest resources and removes its resources and repeats this task until some of the allocated resources do not exceed the number of the allowed resources. Then, ARBs with zero resources are removed from the ARB pool. This procedure is named resource competition. The remaining ARBs are tested for their affinities towards the training antigen. If for any class, the ARBs does not meet a user defined stimulation threshold, then the ARBs are mutated and cloned again. This step is repeated until the affinity for all classes meet the stimulation threshold
- After all classes have passed the stimulation threshold, the ARB with highest affinity of the same class as the antigen, named MC candidate, is chosen as a candidate memory cell. If its affinity for the training antigen is greater than the affinity of MC match then the candidate memory cell is placed in the memory cell pool. if the difference in affinity of these two memory cells is smaller than a

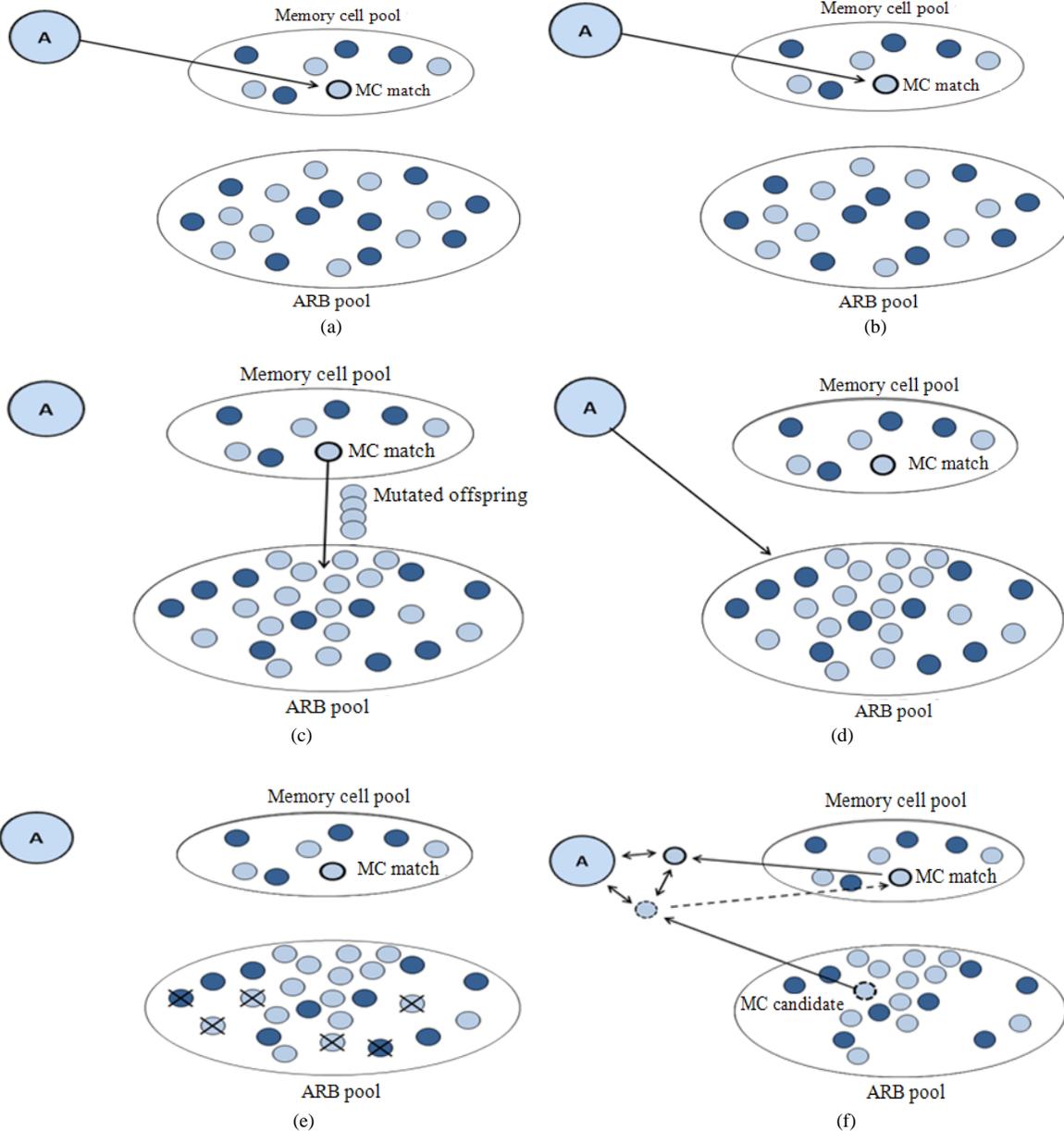


Fig. 1: Processes of Algorithm for a Given Antigen (A) (a) Memory Cell Identification ;(b) MC match Finding; (c) ARB Generation;(d) Recourse Allocation;(e) Resource Competition;(f) Memory Cell Introduction

user defined threshold, the MC match memory cell is removed from the memory cell pool

These steps are repeated for each training antigen. After completion of training, memory cells are used for classification. The KNN is used to classify test data. The class of a test data is determined by majority voting among the k most stimulated memory cells. The detailed description of AIRS can be found in (Watkins and

Bogges, 2002; Watkins *et al.*, 2004). Fig. 1 presents the processes of algorithm for a given antigen A.

Resource allocation: Resource competition is one stage of AIRS. The purpose of resource competition in AIRS is improving the selection probability of high-affinity ARBs for next steps. Resource competition is done based on the number of allocated resources for each ARB. The distribution of resources

is done by multiplying stimulation value with clonal rate that shown in Eq. 1. Marwah and Boggess (2002) have used a different resource allocation method. In their method, the antigen classes occurring more frequently get more resources. AIRS and the study by Marwah use the linear resource allocation and the number of allocated resources has a linear relation to affinities. The resource number difference between high-affinity ARBs and low-affinity ARBs is not much here. The low-affinity and high-affinity ARBs would survive and selection pressure is not high. Therefore, generation of memory cell from antigen will be a time consuming process and increases the training time of the AIRS algorithm:

$$\text{Resources}_{\text{Original}} = \text{stimulation value} \times \text{clonal rate} \quad (1)$$

Nonlinear resource allocation: A nonlinear resource allocation method was proposed in the study by Polat and Gunes (2007) to reduce the classification time and number of memory cells. In this method, resource allocation was done nonlinear to affinities. The difference in resources number between high-affinity ARBs and low-affinity ARBs was bigger in this method than the original linear allocation method. The method allocated less number of resources for ARBs with stimulation values between 0 and 0.50 and more for ARBs with stimulation values between 0.50 and 1. However, the training time of the AIRS algorithm was not considered in the resource allocation method determination.

In this study, a new nonlinear resource allocation method for AIRS is proposed with an aim to reduce the training time of the algorithm. Training time of AIRS is the time required to evolve memory cells from antigens. Evolving memory cells is an evolutionary process. The selection pressure of this evolutionary process would affect the training time of AIRS. By increasing the selection pressure, a memory cell is evolved from a given antigen at a faster process and finally the training time of the algorithm would reduce. How resources are assigned to ARBs (resource allocation) has a direct effect on the selection pressure of the evolutionary process. If the resource allocation method allocates more resources for high-affinity ARBs, then the algorithm would tend to select high-affinity ARBs and the selection pressure of the algorithm increases. Therefore, allocating more resources for high-affinity ARBs would reduce the training time of AIRS. Utilizing nonlinear resource allocation methods would allocate more resources for high-affinity ARBs in comparison to linear resource allocation method. Also the difference in resources number between high-affinity ARBs and low-affinity ARBs would be bigger

than linear resource allocation method. The resource allocation method named EXP, given by Eq. 2, would satisfy these conditions:

$$\text{Resources}_{\text{EXP}} = \begin{cases} \text{Exp}\left(\frac{\text{stimulation value} - 0.5}{0.5}\right) \times \\ \text{clonal rate if stimulation value} < 0.5 \\ (2 - \text{Exp}\left(\frac{0.5 - \text{stimulation value}}{0.5}\right)) \times \\ \text{clonal rate if stimulation value} \geq 0.5 \end{cases} \quad (2)$$

Experimental setup: The EXP resource allocation method was utilized in AIRS and the algorithm, named EXPAIRS was developed. The performances of EXPAIRS were evaluated in comparison to AIRS for it's the training time, classification accuracy and data reduction. Experiments were carried out in order to determine how EXPAIRS performed compared to AIRS. For this study, a number of datasets were retrieved from the well-known UCI machine learning repository (Blake and Merz, 1998). We selected datasets with varying number of attributes, instances and classes, from simple toy datasets to difficult real world learning problems.

Stratified ten-fold cross validation approach was used to estimate the predictive accuracy of the algorithms. In addition, since there are some randomness in AIRS and the cross validation method, both algorithms were run ten times on each dataset to achieve more reliable results. The two tail paired t test was performed to compare the mean of each of performance metrics, i.e., accuracy, training time and number of memory cells, of both algorithms and test hypothesis. The commonly used level of significance 0.05 was applied in this study.

RESULTS

Table 1 shows the achieved results by the algorithms for experiment datasets. The values in parenthesis are standard deviations. The * indicates that P-value is under 0.05 and difference between the results of the algorithms is significant.

As can be seen, EXPAIRS takes less time than AIRS in all cases. Also the difference between times is significant in all cases; therefore EXPAIRS is more efficient than AIRS.

Based on the results, the classification accuracies of EXPAIRS and AIRS are comparable. EXPAIRS is better in five cases while AIRS is better in six cases. An important point to note is on the significance difference between accuracies. The difference is not significant on

Table 1: Comparison of achieved result by AIRS and EXPAIRS

Dataset	Training time (S)		Accuracy (%)		# Memory cells	
	AIRS	EXPAIRS	AIRS	EXPAIRS	AIRS	EXPAIRS
Balance	8.041 (0.09)	4.710 * (0.02)	86.07 (3.12)	86.32 (3.38)	533.8 (51.58)	462.5* (62.53)
Breast-cancer	5.698 (0.02)	5.102 * (0.02)	97.17 (1.94)	97.10 (2.06)	217.1 (8.21)	208* (7.48)
Credit-crx	9.255 (0.03)	5.823 * (0.02)	85.19 (4.16)	85.22 (4.63)	235.3 (8.00)	220* (14.28)
German	21.269 (0.07)	14.395 * (0.03)	71.17 (3.84)	70.75 (4.03)	601.5 (16.44)	530.2* (11.08)
Glass	2.442 (0.02)	1.598 * (0.01)	61.96 (10.01)	62.97 (9.20)	118.8 (6.16)	118.0 (4.99)
Hepatitis	1.028 (0.01)	0.523 * (0.01)	84.54 (6.62)	83.64 (7.56)	61.4 (2.83)	55.8* (1.81)
Segmentation	3.525 (0.02)	2.498 * (0.02)	82.38 (7.61)	81.10 (7.01)	67.2 (4.36)	66.1 (4.62)
Ionosphere	21.561 (0.06)	8.636 * (0.03)	87.53 (5.00)	87.16 (4.67)	187.9 (6.64)	162.3* (8.76)
Iris	3.780 (0.02)	0.750 * (0.01)	95.60 (5.20)	95.67 (5.47)	50.6 (3.40)	46.7* (3.33)
Pima	8.079 (0.02)	5.286 * (0.02)	73.70 (4.31)	73.26 (5.17)	320.8 (13.99)	313.5 (15.58)
Zoo	1.536 (0.01)	0.897 * (0.01)	96.75 (6.09)	96.95 (5.86)	59.0 (1.41)	58.7 (1.49)

all cases. This confirms that utilizing EXP resource allocation method does not significantly affect the classification accuracy of AIRS. This is due the selection pressure as well. The selection pressure of EXPAIRS is higher than AIRS but not at a level that would generate extreme premature memory cells which would have a significant negative affect on the accuracy of AIRS.

Moreover, EXPAIRS reduces the data more than AIRS. This shows that the replacement between memory cells occurs in EXPAIRS more than AIRS. The results thus far have shown the important ability of EXPAIRS to do more data reduction whilst does not reduce the classification accuracy of AIRS

DISCUSSION

The reason for the efficiency of EXPAIRS is the higher selection pressure of it in contrast to AIRS.

The heart of AIRS is evolving memory cells from ARBs. This is an evolutionary process. Resource competition is the main mechanism in developing memory cells from population of ARBs. Each ARB uses its reward to participate in competition. The reward is the number of allocated resources. More stimulated ARBs obtain more resources.

How resources are assigned to ARBs has a direct effect on the number of ARBs that survive after the resource competition. EXPAIRS with the EXP resource allocation method tended to select a few super individuals during resource competition in contrast to AIRS; therefore it has a higher selection pressure. The presence of super individuals, which are much better than the average fitness of the population, is the common cause of rapid convergence in evolutionary process. Such super individuals prevent other individuals from contributing in the next generation. Therefore, within the next few generations a super individual can

eliminate other appropriate individuals causing a rapid convergence in result. Hence, each antigen (feature vector) in EXPAIRS generates a memory cell rapidly in contrast to AIRS and evolving memory cells from antigens take less time in EXPAIRS. At the result EXPAIRS will be more efficient than AIRS.

CONCLUSION

AIRS is a very effective immune based classifier. In this study this improved and a novel immune based classifier, EXPAIRS, was introduced by utilising a nonlinear resource allocation method in AIRS. The proposed algorithm required less training time and memory cells in comparison to AIRS and accuracy that is comparable as well.

REFERENCES

- Blake, C.L. and C.J. Merz, 1998. UCI Repository of Machine Learning Databases. <http://www.citeulike.org/user/wkiri/article/1611605>
- de Castro, L.N. and J. Timmis, 2002. Artificial Immune Systems: A New Computational Intelligence Approach. 1st Edn., Springer Verlag, ISBN-10: 1852335947, pp: 380.
- de Castro, L.N. and J. Timmis, 2003. Artificial immune systems as a novel soft computing paradigm. *Soft Comput.*, 7: 526-544. DOI: 10.1007/s00500-002-0237-z
- de France, F.O., F.J. Von Zuben and L.N. de Castro, 2005. An artificial immune network for multimodal function optimization on dynamic environments. *Proceedings of the 2005 Conference on Genetic and Evolutionary Computation*, June 25-29, ACM New York, NY, USA., pp: 289-296. DOI: 10.1145/1068009.1068057

- Doraisamy, S. and S. Golzari, 2010. Automatic musical genre classification and artificial immune recognition system. *Adv. Music Inform. Retrieval.*, 274: 390-402. DOI: 10.1007/978-3-642-11674-2-17
- Golzari, S., S. Doraisamy, M.N. Sulaiman and N.I. Udzir, 2009. A review on concepts, algorithms and recognition based applications of artificial immune system. *Adv. Comput. Sci. Eng.*, 6: 569-576. DOI: 10.1007/978-3-540-89985-3_70
- Igawa, K. and H. Ohashi, 2007. Discrimination-based artificial immune system: Modeling the learning mechanism of self and non-self discrimination for classification. *J. Comput. Sci.*, 4: 204-211. ISSN: 1549-3636
- Kim, J., P.J. Bentley, U. Aickelin, J. Greensmit and G. Tedesco *et al.*, 2007. Immune system approaches to intrusion detection-a review. *Natural Comput.*, 6: 413-466. DOI: 10.1007/s11047-006-9026-4
- Marwah, G. and L. Boggess, 2002. Artificial immune systems for classification: Some issues. *Proceedings of the 1st International Conference in Artificial Immune Systems, (ICAIS'02)*, University of Kent at Canterbury, UK., pp: 149-153.
- Polat, K. and S. Gunes, 2007. An improved approach to medical data sets classification: Artificial immune recognition system with fuzzy resource allocation mechanism. *Exp. Syst.*, 24: 252-270. DOI: 10.1111/j.1468-0394.2007.00432.x
- Timmis, J. and M. Neal, 2001. A resource limited artificial immune system for data analysis. *Knowl-Based Syst.*, 14: 121-130. DOI: 10.1016/S0950-7051(01)00088-0
- Watkins, A. and L. Boggess, 2002. A resource limited artificial immune classifier. *Proceedings of the 2002 IEEE Congress on Evolutionary Computation*, May 12-17, Computer Lab., Kent University, Canterbury, UK., pp: 926-931. DOI: 10.1109/CEC.2002.1007049
- Watkins, A., J. Timmis and L. Boggess, 2004. Artificial Immune Recognition System (AIRS): An immune-inspired supervised learning algorithm. *Genet. Program. Evol. Mach.*, 5: 291-317. DOI: 10.1023/B:GENP.0000030197.83685.94