A Framework for Pattern Recognition System Based on an Artificial Immune System

Nurul Alam Mohd. Yaakub, Rayner Alfred and Chung Seng Kheau

School of Engineering and Information Technology, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia.

alam.yaakub@gmail.com, {ralfred, kheau}@ums.edu.my

Abstract. The artificial immune system (AIS) has a bright prospect in pattern recognition development. In this paper, a framework to simulate natural immune system is outlined using negative selection algorithm as the pattern detector. An experiment is conducted using bit-strings recognition as a convention to illustrate a pattern recognition problem. The result of comparative preliminary study of applied AIS and SVM for the experiment is discussed. Later discussion of the paper includes suggestions on improving designed framework in pattern recognition domain.

Keywords: Artificial immune systems, negative selection, pattern recognition, Support Vector Machine, Hamming distance method

1 Introduction

Researches over pattern recognition are meant to improve recognition accuracy, scalability in complex data dimension and classification's dispatched time. Current revolutionary developments are those of neural network hybrids, kernel-based learning recognition algorithms particularly with support vector machines (SVM) and mixed ensemble learning models [1].

Identified issues in pattern recognition are over fitting and 'non-cooperative' data [1]. Over fitting problems occurs when too many strict conditions prompt the algorithm to leap categorical steps and miss minimum of error surface. 'Non-cooperative' data is present when inter-classes patterns overlap in feature space causing data dimension complexity and congest the recognition process pipeline. Designed algorithms to counter these problems require complex calculation and architecture which affect performance in terms of computation times and scalability. Due to these reasons, designing an algorithm which is simple and flexible that possesses a certain degree of learning curve is more realistic in performing pattern recognition process.

In this paper, we outline a framework of pattern recognition based on Artificial Immune Systems (AIS). In section 2, we describe the framework of AIS. Then, we describe the framework of pattern recognition system based on AIS, in Section 3. Section 4 provides the experimental design and preliminary results for the evaluation of the proposed framework and Section 5 concludes this paper.

2 Artificial Immune System (AIS)

A new generation of biologically-inspired computation's framework has been discovered in the past 10 years. The introduction of Artificial Immune System (AIS) framework is inspired by the immune system's defence mechanisms found in vertebrates. Castro and Timmis [2] define AIS as follows:

"*AIS is an adaptive system, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving*".

The immune system can be divided into two sub-systems which are innate immune sub-system and adaptive immune sub-system (Fig. 1). The innate immune sub-system deals with the outer layers of physical body defense layer up to sensory or nervous receptors. This first defense layer of our body is inhabited by macrophages, the cells that eat foreign entities like bacteria that intrude the immune system. On the other hand, the adaptive immune sub-system comes into action if the innate immune sub-system fails to filter foreign objects from delving deeper into the defensive layers of our body. The main players of the adaptive immune system include lymphocytes which react towards foreign micro objects entering the body. After a certain protective reaction is completed, lymphocytes genetically remember the traits of previously attacked foreignness, mutated, cloned themselves out and disperse within their own respective zones within the immune system. The illustration of the process is summarized in Fig. 1.

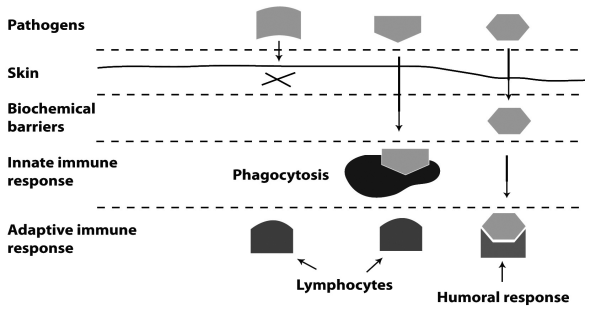

Fig. 1. Immune system's hierarchical diagram. Adapted from [3].

2.1 AIS Framework

AIS practically can be applied in all application domains. To be able to represent in AIS, we need to identify how our application perceives anomaly or threats. We also need to study the application's data process flow that can be represented as threats detector. Suitable mapping of our application domain into the AIS framework determines the flexibility and modularity of resulted AIS engine.

Affinity measuring can be done in binary or real values domain, depending on applications parametric details. Commonly used measuring algorithms by far are those using Euclidean distance algorithm and Hamming method. If input data

parameter satisfies the conditions given by affinity measurement algorithm, it will undergo selection algorithms to determine the identity of the data.

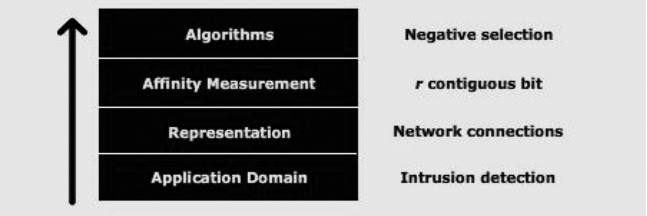

Fig. 2. Intrusion detection system in AIS framework.

For instance, Fig. 2 illustrates the intrusion detection system in AIS framework [4] which detects anomalies within incoming network connections. Good and bad connections are labelled accordingly which also represent self and non-self cells in the immune system. Each network connection is evaluated by *r* numbers of contiguous bits. These bits need to undergo affinity measuring to be translated into comprehensible context to be evaluated by a selection algorithm, such as negative selection algorithm. In general, negative selection algorithm will select bad connection bits as clues to be stored in detector's repository for future detection use. More discussion of negative selection algorithm will be discussed further in this paper.

2.2 Self and Non-self Discrimination

The classical theory of the inner workings of immune system is based on the system's ability to identify and discriminate self cells apart from foreign or non-self cells in the system [5]. The immune system has pattern recognition receptors which cooperate with lymphocytes genetic traits memories, enabling it to discriminate whether an intruding cell is self or foreign. Immune system performs negative and positive selections in deciding its next response towards intruding cells. Negative selection involves memorizing foreign characteristics while positive selection deals with storing self cell's attributes. In this paper, we will focus more on the negative selection as it is safer and more economical to eliminate or detect prospective anomalies before the next computation proceeds.

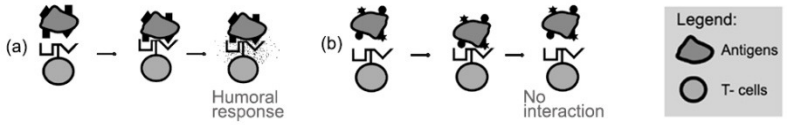

Fig. 3. Self non-self discrimination in adaptive immune system. Diagram in (a) shows matching pair of antigen and T-cell activates response and vice versa in (b). Adapted from [6].

2.3 AIS and pattern recognition

The immune system has a special ability to detect patterns in cells or external particles that is harmful towards the body such as diseases, viruses and cancerous cells. Intelligent molecules in immune system that are responsible for pattern recognition are known as receptors. These receptors, often found in the surface of immune system cells, made it possible to detect almost infinite number of patterns entering the system and are important to trigger immune response.

In AIS, we are modelling the theoretical mechanism on how receptors understand patterns from early detection by lymphocytes. AIS will compute the binding ability of conventionally extracted pattern features before performing classification. Designed AIS framework must be able to discriminate whether entering pattern or data is self, depending on application objectives, or non-self.

AIS is a highly flexible and consistently maturing system. Co-operations from independent sub-modules of AIS suggests high elasticity if parametric adjustment is needed without severely affecting the entire system. AIS is dynamic in such a way that it could avoid redundancy of using similar patterns in matching process and refreshes its patterns repository when stability occurs.

3 Pattern Recognition System Based on AIS

In this section, we outline the framework of an AIS-based pattern recognition algorithm. We will illustrate a bit pattern detection system based on the AIS system described in Section 2. In our study case for pattern recognition problem, we will evaluate the proposed framework by comparing the predictive accuracy of a support vector machine (SVM) algorithm for both AIS-based and non AIS-based detection system.

3.1 The Framework of Pattern Recognition System Based on AIS

In this framework, we outline the designed system in accordance to the Useful Criteria Assessment (UCA) [7]. Based on the UCA, the proposed framework is evaluated based on its distinctiveness and effectiveness. In order to ensure the distinctiveness of the proposed framework, it should be able to replicate or replace current available methods without losing the dynamicity of the immune system. In order to maintain the effectiveness of the proposed system, the new proposed method should be unique and provide better pace and results.

In this work, the feature space for our pattern recognition framework is represented by randomly pre-programmed 12-bits bit-strings. The affinity for sets of binary values will be checked with pre-determined self 12-bits bit-strings and non-self bit-strings using Hamming distance measures. We also incorporate negative selection algorithm from the AIS in order to classify input bit-strings.

Table 1. AIS Attributes for Immune and Pattern Recognition Systems

Attribute	*Immune System*	*Pattern Recognition System*
Self, *S*	Self feature	Positive instance
Non-self, *N*	Non-self feature	Negative instance
Lymphocytes, *L*	Small white blood cells that recognize and react towards APCs	Detectors to identify match
Antigens, *A*	New entity entering the immune system	Random bit-strings input

We are introducing AIS-associated attributes such as Self, Non-self, Lymphocytes (or Detectors) and Antigens. An analogy of the framework of the pattern recognition and the framework of the AIS can be illustrated through the properties outlined in Table 1. Fig. 4 illustrates sequential process of the internal reaction in the immune system. The negative selection algorithm that enables AIS to dynamically detect patterns is discussed in the next section.

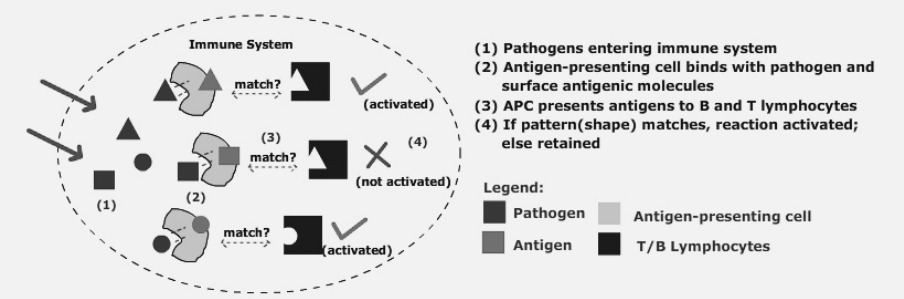

Fig. 4. Diagram showing sequential process of pattern recognition in the immune system.

4 Experimental Design and Results

4.1 AIS vs. Non-AIS (SVM)

In this experiment, we compare the predictive accuracy of the proposed pattern recognition based on AIS with the predictive accuracy of SVM. We propose to use support vector machine (SVM) [8] algorithm for its ability to learn new classes over time and detection precision in pattern recognition researches especially when concerning with real-time processes.

We will run similar repository of feature vectors extracted from methods explained in Section 5, for both SVM and AIS. A set of labelled trained bit-strings will be fed as inputs to the SVM algorithm. The derived support vector or class divider will determine the class of a new set of vectors by comparing the distance of the new vector set with the support vector. For AIS, a set of self bit-strings will be compared with random set of bit-strings to create a stack of detector or antibody. These detectors then will be compared against a new set of test data set using the affinity matching method and will be classified using negative selection algorithm. Fig. 5 illustrates the generic outlook of the framework.

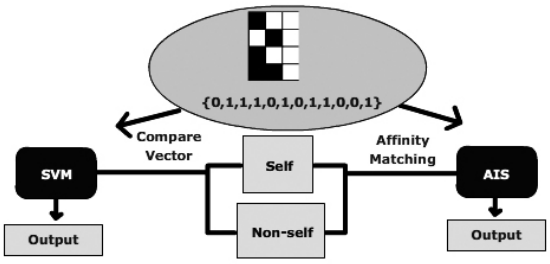

Fig. 5. Workflow to evaluate face detection in both AIS and non-AIS framework

In binary classes level, SVM works as pattern classifier by finding maximum marginal distance between support vectors which are the closest points in a training set. Two classes are separated by a general separator, or known as optimal separating hyperplane (OSH) of which the distance of support vectors can be maximized. To train a basic linear SVM classifier, we find an OSH as:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i x_i \cdot x + b \ . \tag{1}$$

where $i = 1, 2, \dots, N$, x_i belongs to one of two classes identified by the label $y_i \in \{-1, 1\}$ and, α_i and b are coefficients derived from solving quadratic programming problem [8][9]. Suppose a new data point x is found, it can be classified by solving Eq. (1) and using *d(x)* as defined in Eq. (2)

$$d(x) = \frac{\sum_{i=1}^{N} \alpha_i y_i x_i \cdot x + b}{\left\| \sum_{i=1}^{N} \alpha_i y_i x_i \right\|} . \tag{2}$$

to classify x in solving a multi-classification problem. The classification result is more reliable if the distance of x point, $|d|$ from the hyperplane is larger. In solving a binary classification problem, given q classes, that are trained on each own SVMs, the class label y of a pattern x is computed as,

$$y = \begin{cases} 1, & if \ d_{max} + t > 0 \\ -1, & if \ d_{max} + t \leq 0 \end{cases} \quad with \ d_{max} = \max \{d_i(x)\}_{c=1}^{q} \ . \tag{3}$$

where $d_i(x)$ is calculated from Eq. (2) for an SVM to recognize class c and t represents the value of the classification threshold.

In order to represent AIS, we use negative selection algorithm. In the algorithm, there are two stages involved which are censoring and matching stages [10].

In censoring or training stage, suppose that we have a set of face or self feature vectors, $S = \{s_1, s_2, \dots, s_m\}$ where m is the total feature vector population in set S. A random set of inputs, $D_0 = \{d_{01}, d_{02}, \dots, d_{0n}\}$, is generated as the initial detectors in which it has similar format as the self feature vectors in *S*. The population of D_0 can be greater than or less than the value of m.

Assuming that a detector set will accept a pattern when the total of Hamming distance Eq. (4) for all elements in feature vector s and d_0 that exceed the Hamming distance's tolerance against cross-reactivity threshold, r, which is derived in Eq. (5).

$$H = \sum_{i=1}^{L} \delta_i \,. \quad where \quad \delta_i = \begin{cases} 1 & if \quad s_{yi} \neq d_{0xi} \quad y = 1,2,3,\dots m \\ 0 & otherwise \quad x = 1,2,3,\dots n \end{cases} . \tag{4}$$

$$r = (L-1) - \varepsilon \quad where \quad \begin{matrix} L = length\ of\ bit-string \\ \varepsilon = cross-reactivity\ threshold \end{matrix} . \quad . \tag{5}$$

Given that the total of Hamming distance value of each element in feature vector s and d_0 surpasses r, d_0 will be stored as an element in the detector set D while those which do not exceed r will be eliminated. As a result, we will collect all negative patterns (detector set, D). We can represent censoring stage by using the following notation,

$$d_{0x} \,\epsilon\, D \quad when \quad H(s_y, d_{0x}) > r \quad \begin{matrix} y = 1,2,3,\dots m \\ x = 1,2,3,\dots n \end{matrix} . \tag{6}$$

where n represents D_0 size of population and r represents rejection tolerance threshold.

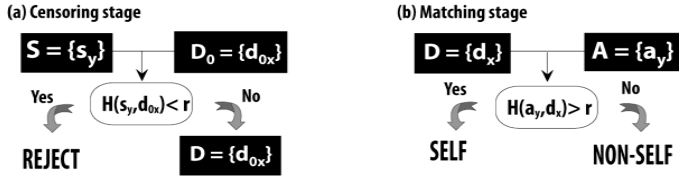

Fig. 6. Diagram in (a) shows censoring of initial detector set and self set, while (b) performs matching between detector and antigen set.

Matching process (Fig 6b) will be performed after censoring stage (Fig 6a) is completed. We will no longer involve self feature vectors as derived set of detectors had been determined earlier. At this point of time, we will compare antigens, A or feature inputs with appointed detectors set D. Affinity rate of antigen and detector set will be the benchmark to test the presence of non-self patterns. Let a_y represent antigen set A and d_x to represent detector set D. We will get the value for self detection flag, c using this following condition,

$$C = \begin{cases} 1\ if\ H(a_y, d_x) > r \\ 0\ if\ H(a_y, d_x) < r \end{cases} \quad when\ \delta_i = \begin{cases} 1\ if\ a_{yi} \neq d_{xi} & y = 1,2,3,\dots p \\ 0 \quad otherwise & x = 1,2,3,\dots q \end{cases} . \tag{7}$$

Input bit-string will be perceived as self if the value of C = 1 and as non-self if C = 0.

4.2 Preliminary Results

In this paper, we design the experimental design in order to show any significant improvement of the SVM algorithm by implementing the AIS-based framework of a pattern recognition system.

In this experiment, 10 data sets of randomly controlled 13-bits binary strings which comprises of 12-bits of feature data and 1-bit data for identification is prepared. We represent self patterns by pre-determining a set of 12-bits binary patterns which also correspond to numerical patterns in 3x4 cells to be recognized by the algorithms. The aim of each implemented algorithm is to detect the self patterns which are represented by the numbers '1' and '4'. Six possible features for '1' and '4' patterns are identified to discretely label the self or non-self binary strings (Fig. 8). Fig. 7 shows the translation of the numerical patterns in 3x4 cells into 12-bits binary strings.

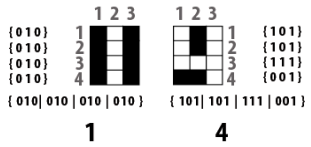

Fig. 7. 12-bits binary string representation for numerical pattern '1' and '4'.

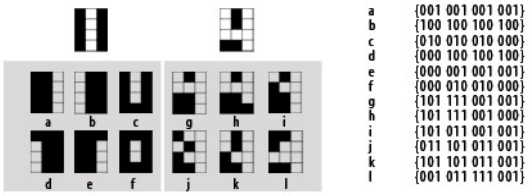

Fig. 8. 6 possible features (3x4 numerical pattern and binary strings) for '1' and '4' patterns.

Table 2. Table showing accuracy results for detecting bitstring patterns using SVM vs. AIS.

N	Self	Non-self	SVM				AIS			
			Correctly classified instance		*Incorrect classified instance*		*Correctly classified instance*		*Incorrect classified instance*	
			n	%	N	%	n	%	n	%
1000	**206**	**794**	871	87.1000	129	12.9000	955	95.5000	45	4.5000
1200	**242**	**958**	1068	89.0000	132	11.0000	1107	92.2500	93	7.7500
1400	**283**	**1117**	1243	88.7857	157	11.2143	1302	93.0000	98	7.0000
1600	**324**	**1276**	1428	89.2500	172	10.7500	1469	91.8125	131	8.1875
1800	**366**	**1434**	1595	88.6111	205	11.3889	1753	97.3889	47	2.6111
2000	**404**	**1596**	1777	88.8500	223	11.1500	1798	89.9000	202	10.1000
2200	**445**	**2003**	1755	88.7273	248	11.2727	1995	90.6818	205	9.3182
2400	**492**	**1908**	2122	88.4167	278	11.5833	2164	90.1667	236	9.8333
2600	**529**	**2071**	2312	88.9231	288	11.0769	2290	88.0769	310	11.9231
2800	**568**	**2232**	2465	88.0357	335	11.9643	2465	88.0357	335	11.9643

Table 2 shows that the predictive accuracy percentage of SVM is consistent but somehow decreases as the number of N increases while the AIS shows an explicit decreasing trend with standard deviation of 3.018 compared to 0.614 for SVM as the number of N increases. Insignificant improvement of AIS implementation is shown by the result derived from t-test calculation. However, 80% of AIS accuracy percentages have outdone that of SVM for all the tests.

Obvious decreasing trend in AIS results shows that the implemented abstraction of AIS suffers as the number of input increases, as is the increase of antigens amount to be compared with input. However, another possible cause for this trend might be due to noise created upon random data creation. Extreme difference between AIS against SVM results for the first eight data sets might be another indicator that it is still possible to surpass SVM detection accuracy rate given that more genetic behaviour is added to current implemented AIS algorithm. For instance, a coefficient for growth rate in a population can be incorporated to complement current lack of dynamics in bigger samples.

5 Conclusion

The challenge of applying AIS for pattern recognition system is to determine the most suitable parametric features in order to simulate AIS dynamic behaviour. While natural immune system is triggered by numerous intermolecular chemical communications, we have to carefully simplify the whole complex process without compromising the immune system's dynamicity and adaptability. Thus, we are proposing to perform probabilistic mutation towards detector's feature and infection rate as to add "age" or "experience" for the detector with respect to the values of infection rate for each feature vector or "cell". In order to improve the predictive accuracy for large set of data, we propose to add "growth rate" so that the dynamic property of the immune system is not compromised.

In terms of usability test evaluation [7], we need to improve on the algorithm distinctiveness as the representation of AIS is still using primitive convention of negative selection implementation. We believe that the framework can be distinctively effective given that the mentioned suggestions can be incorporated in future development of the research.

References

1. Polikar, R.: Pattern Recognition. In Wiley Encyclopedia of Biomedial Engineering. John Wiley & Sons, Inc. (2006)
2. de Castro, L. N., & Timmis, J.: Artificial Immune Systems: A new computational intelligence approach. Springer. (2002)
3. de Castro, L. N., & Timmis, J.: Artificial Immune Systems: A Novel Paradigm to Pattern Recognition. In e. L. Alonso J. Corchado and C. Fyfe, Artificial Neural Networks in Pattern Recognition (pp. 67-84). University of Paisley. (2002)

4. Cayzer, S.: Artificial Immune Systems. Bristol, Semantic and Adaptive Systems, HP Lab. (2006)
5. Burnet, F. M., & Fenner, F.: The Production of Antibodies, 2nd Edition. Melbourne: Macmillan and Co. (1949)
6. Aickelin, U., & Cayzer, S.: The Danger Theory and Its Application to Artificial Immune Systems. 1st International Conference on Artificial Immune Systems (ICARIS - 2002). Canterbury. (2002)
7. Garrett, S. M. How do we evaluate artificial immune systems? Evolutionary Computation 13(2) , 145-178. (2005)
8. Vapnik, V. : Statistical Learning Theory. New York: John Wiley and Sons. (1998)
9. Berthold, M., & Hand, D. J.: Intelligent Data Analysis - An Introduction (Second Edition). Heidelberg, Berlin, Germany. (2003)
10. Forrest, S., Perelson, A. S., Allen, L., & Cherukuri, R.: Self-nonself discrimination in a computer. Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy. Los Alamitos, CA: IEEE Computer Society Press. (1994)