Chapter 4

The Artificial Immune System

The artificial immune system (AIS) is a computational system which is applied to problem domains. The AIS mimics the models of the natural immune system's immune functions and principles. In chapter 2 of this dissertation, the natural immune system (NIS) was discussed. The NIS is a very complex system that is capable of learning the structure of normal cells (self patterns) and classifying foreign cells (non-self patterns). The NIS also builds up a 'memory' of frequently seen non-self patterns to ensure a faster secondary immune response to non-self patterns with identical or similar structure. Although the NIS is not yet fully understood, research has shown that the NIS consists of mature T-Cells and B-Cells that co-operate to detect any foreign cell in the body (as explained in section 2.2.1). The B-Cells produce antibodies through a process known as clonal selection (as discussed in section 2.3).

The capabilities of the NIS to distinguish between normal cells and foreign cells (with only having knowledge on what is normal) and learning the non-self cells' structure, inspired the modeling of the NIS into an AIS for application in non-biological environments. A major advantage of the AIS is that the model only needs to be trained on positive examples (knowledge on the self patterns) to detect or classify non-self patterns in a non-biological environment. A drawback is that a limited knowledge of positive examples or a bad representation of positive examples can lead to misclassification of non-self patterns.

This chapter provides some background information on the different existing artificial immune system models, as well as models inspired by immunology and applications of the artificial immune system.

4.1 Natural to Artificial

There are a few theories surrounding the structure and functioning of the NIS due to its biological complexity. The NIS has many different lymphocytes that form part of the immune response to a foreign cell. Two of these lymphocytes - the T-Cell and B-Cell - have been clearly defined and their functioning in the NIS identified (as explained in section 2.2.1). In the NIS it is mainly the inner working and co-operation between the mature T-Cells and B-Cells that is responsible for the secretion of antibodies as an immune response to antigens. The T-Cell becomes mature in the thymus. A mature T-Cell is self-tolerant, i.e. the T-Cell does not bind to self cells. The mature T-Cell's ability to discriminate between self cells and non-self cells makes the NIS capable of detecting non-self cells. When a receptor of the B-Cell binds to an antigen, the antigen is partitioned and then brought to the surface with an MHC-molecule. The receptor of the T-Cell binds with a certain affinity to the MHC-molecule on the surface of the B-Cell. The affinity can be seen as a measurement to the number of lymphokines that must be secreted by the T-Cell to clonally proliferate the B-Cell into a plasma cell that can produce antibodies. The memory of the NIS on frequently detected antigen is built-up by the B-Cells that frequently proliferates into plasma cells. Thus, to model the proposed artificial immune system, there are a few basic concepts that must be considered:

- There are trained detectors (artificial lymphocytes) that detect non-self patterns with a certain affinity.
- The artificial immune system needs a good repository of self patterns or self and non-self patterns to train the artificial lymphocytes (ALCs) to be self-tolerant.
- The affinity between an ALC and a pattern needs to be measured. The measured affinity can indicate to what degree an ALC detects a pattern.
- To be able to measure affinity, the representation of the patterns and the ALCs need to have the same structure.
- The artificial immune system has memory that is built-up by the artificial lymphocytes that frequently detect non-self patterns.
- When an ALC detects non-self patterns, it can be *cloned* and the clones can be mutated to have more diversity in the search space.

In [29, 39, 40] similar architectures for an AIS are proposed and used for security.

The measured degree of affinity between an ALC and a pattern needs to exceed a certain degree for the pattern to be detected by an ALC. There are different approaches to measure affinity and represent patterns in the AIS. Timmis *et al.* [79] represented patterns as vectors of floating points. The Euclidean distance between an ALC and a pattern was used as a measure of affinity. De castro and Von Zuben [21] used binary strings to represent patterns. The affinity between an ALC and a pattern was measured using the hamming distance. The above mentioned measurements indicate the similarity between an ALC and the pattern. A lower value calculated by the above measurements indicates a stronger affinity between an ALC and the pattern. Hunt and Cooke [44] used the same representation of binary strings, but like Forrest *et al.* [30] the measurement of affinity was done with the *r*-continuous matching rule. The *r*-continuous matching rule is a partial matching rule: An ALC detects a pattern if there are *r*-continuous or more matches in the corresponding positions. *r* is the degree of affinity for an ALC to detect a pattern. A higher value of *r* indicates a stronger affinity between an ALC and the pattern.

The proposed AIS has a set of ALCs that is used to detect non-self patterns. The ALC set is trained with a training set that consists of self patterns, non-self patterns or self and non-self patterns. Each pattern in the training set has been labeled as self or non-self. The ALCs can be trained to become self-tolerant like a mature T-Cell or to detect non-self patterns with a higher affinity. The trained ALC set is then used to detect non-self patterns. Hightower *et al.* [38] and Oprea [60] used a training set that consisted of non-self patterns. A genetic algorithm was used to evolve ALC sets with a larger detection ratio of non-self patterns in the training set. Forrest *et al.* [30] used a technique known as negative selection to train ALCs to become self-tolerant. The training set consisted of self patterns represented by nominal valued attributes or binary strings. The ALCs are randomly generated and tested against the training set of self patterns. If the ALC does not detect any of the self patterns in the training set, it is added to the ALC set. The training set is monitored by continually testing the ALC set against the training set for changes. A negative selection method to train an ALC with continuously-valued self patterns are presented by Gonzalez *et al.* [37]. The continuously-valued negative selection method evolves ALCs that are the furthest away from the training set of self patterns and that are separated to maximise the non-self space coverage. A randomly generated ALC that is trained with negative selection does represent a pattern in non-self space, but not necessarily an antigen. A different approach is proposed by Kim [49] where ALCs are not randomly generated and tested with negative selection, but an evolutionary process is used to evolve ALCs towards non-self and to maintain diversity and generality among the ALCs. The model by Potter and De Jong [64] applies a co-evolutionary

genetic algorithm to evolve ALCs towards the selected class of non-self patterns in the training set and further away from the selected class of self patterns. Once the fitness of the ALC set evolves to a point where all the non-self patterns and none of the self patterns are detected, the ALCs represent a description of the concept. If the training set of self and non-self patterns is noisy, the ALC set will be evolved until most of the non-self patterns are detected and as few as possible self patterns are detected. The evolved ALCs can discriminate between examples and counter-examples of a given concept. Each class of patterns in the training set is selected in turn as self and all other classes as non-self to evolve the different concept in the training set. The negative selection method used in the proposed AIS, evolves ALCs that are binary-valued to cover the maximum non-self space with the furthest distance from the training set of self patterns and the least overlap among the evolved ALCs.

The clonal selection of the immune system was modeled by an algorithm proposed by de Castro [21, 24]. The presented algorithm, CLONALG, performs machine-learning and pattern recognition tasks. The training set consists of non-self patterns. The ALCs in the ALC set are randomly initialised and the ALCs that recognised the selected non-self pattern in the training set with the highest affinity, are cloned. The clones are mutated and then a new population of ALCs are selected from the mutated clones and the previous ALCs, according to their affinity with the selected non-self pattern in the training set. A selection of ALCs with high affinity go into the memory pool. New randomly initialised ALCs are inserted into the set of ALCs for the next non-self training pattern to be recognised. The algorithm can also be used to solve complex problems, eg. multi-modal function optimisation.

Timmis introduced the concept of artificial recognition balls (ARBs) in a resource limited artificial immune system [78]. An ARB has the same representation as an ALC, but stands for a number of identical ALCs. Thus, each ARB allocates a number of resources based on its stimulation level. The total number of resources of the system is bounded. Watkins adopted the concept of ARB's [81]. To be able to identify a memory cell, a training set of antigens (non-self patterns) is presented to the system. A subset of the training set is used to create an initial batch of memory cells. Each of the remaining non-self training patterns are then matched against the memory cells. The memory cell with the closest match to a specific non-self training pattern is then cloned to form an ARB. The level of cloning is determined by the strength of the affinity between the non-self pattern and the memory cell. The newly created ARB is then added to a pool of existing ARBs of the same class. Resources are allocated to an ARB depending on the affinity

between the ARB and the presented non-self pattern as well as the class of the non-self pattern. The ARBs are clonally expanded until the average stimulation level of the ARBs is above a certain threshold. When the limit for available resources has been reached by the ARBs, resources are removed from the ARBs with the lowest affinity until the limit is no longer exceeded. The ARBs with bad performance will have no resources allocated to them and are removed from the pool. The ARBs in the pool are evaluated on their affinity with a non-self pattern. If the best matching ARBs in the pool have a higher affinity with the presented non-self pattern than the best matching memory cell in the memory pool, then the ARBs are added to the memory pool. Only the memory pool is used to classify non-self patterns in the test set.

All of the above models are supervised training methods except the work by de Castro and the work by Timmis, which are unsupervised. CLONALG and AINE were both algorithms for clustering. The training sets consist either of self patterns, non-self patterns or both. The following section introduces the network theory of interconnected B-Cells in the natural immune system.

4.2 The Network Theory

The theory of Jerne is that the B-Cells are interconnected to form a network of cells [45, 62]. When a B-Cell in the network respond to a foreign cell, the activated B-Cell stimulates all the other B-Cells to which it is connected in the network. Work that has been done in AISs on the network theory of B-Cells can be found in [77, 79, 80]. Timmis [77] implemented the network theory with interconnected ARBs. When two ARBs have a high affinity between them, a link is established between them. Therefor, a network of ARBs is formed based on the similarity and affinity among the ARBs. In the network of ARBs, closely related ARBs form clusters that represent clusters in the data set. This model resulted in a successful unsupervised training method to visualise data. Another unsupervised approach to cluster data was done by de Castro and Von Zuben [19]. In this model, named aiNet, the B-Cells formed part of an edge-weighted graph. Some of the B-Cells were connected with edges, with a weight (connection strength) assigned to each edge. Thus, the graph is not necessarily fully connected. The network is trained by representing non-self patterns to the interconnected B-Cells. The distance between B-Cells with the highest affinity for the non-self pattern are then decreased. The immune network theory has also been successfully developed and applied to optimise multi-modal functions [17].

4.3 The Genetic Artificial Immune System

The algorithm proposed in this dissertation, named GAIS - Genetic Artificial Immune System, represents all patterns in space as binary vectors and uses the hamming distance as affinity measurement. A GA is used to evolve ALCs with maximum non-self space coverage and minimum overlap among existing ALCs (as explained in chapter 6). The ALCs are trained with an adopted negative selection method (as explained in section 5.2.2.1) or with positive selection (as explained in section 5.2.2.2). The affinity threshold of an ALC is used to determine a match with a non-self pattern. With the adopted negative selection method the affinity threshold is determined by the distance to the closest self pattern from the ALC. The algorithm is supervised with a training set that consists of self patterns. GAIS is different from existing AIS models in that the GA does not evolve ALCs towards non-self patterns (as in the model of Kim [49]), neither does it evolve ALCs with a larger detection ratio of non-self patterns in the training set (as in the models of Hightower *et al.* [38] and Oprea [60]). The GA in GAIS evolves ALCs with a maximum non-self space coverage and a minimum overlap with existing ALCs. The main goal is thus to evolve mature ALCs to detect non-self patterns that have not been presented to GAIS during training. Surely all evolved ALCs will cover non-self space, but not all ALCs will detect non-self patterns. Therefor, a proposed transition function, the life counter function (as explained in section 5.3), determine an ALC's status (as defined in section 5.2.1). ALCs with annihilated status are removed in an attempt only to have mature and memory ALCs with optimum classification of non-self patterns. GAIS has the advantage to classify patterns in a problem space where only positive patterns are available for training.

4.4 Applications and Other Models

The artificial immune system has been successfully applied to many problem domains. Some of these domains range from network intrusion and anomaly detection [16, 28, 29, 37, 39, 40, 49, 50, 74, 75] to data classification models [65, 82], (the model of Pramanik *et al.* [65] bridges the models of [59] and [48]), virus detection [30], concept learning [64], data clustering [19], robotics [46, 80], pattern recognition and data mining [11, 44, 77, 79]. The AIS has also been applied to the initialisation of feed-forward neural network weights [23], the initialisation of centers of a radial basis function neural network [22] and the optimisation of multi-modal functions [17, 33]. The interested reader is referred to [15, 18, 20] for more information on AIS applications.

4.5 Conclusion

The chapter gave a background overview on the existing theories of the functioning of the natural immune system and the modeling thereof, as well as the successful applications of the AIS in non-biological environments. The next chapter will explain the modeling of the T-Cell and B-Cell as artificial lymphocytes in the AIS. The different states in the life cycle of an artificial lymphocyte is discussed. The next chapter also presents a threshold function that is used to determine the state of an artificial lymphocyte.