Chapter 1

Artificial Immune System

1.1 INTRODUCTION

People have had a keen interest in the biosphere since ancient times and have gotten inspiration from the structures and functions of biological systems and their regulatory mechanisms continuously. Since the mid-twentieth century, researchers have focused on the simulation of biological systems, especially the structures and functions of human beings. For example, artificial neural network simulates the structure of the nerve system of the human brain, fuzzy control is very similar to the fuzzy thinking and inaccurate reasoning of human beings, and evolutionary computation algorithms are the direct simulations of the evolved processes of natural creatures.

In recent years, the biological immune system has become an emerging bioinformatics research area. The immune system is a complex system consisting of organs, cells, and molecules. The immune system is able to recognize the stimulation of "self" and "nonself," make a precise response, and retain the memory. It turns out from much research that the immune system consists of a variety of functions such as pattern recognition, learning, memory acquisition, diversity, fault-tolerant, distributed detection, and so on.

These attractive properties of the biological immune system have drawn extensive attention of engineering researchers who have proposed many novel algorithms and techniques based on those principles of immunology. After introducing the concept of immunity, many researchers in engineering have obtained more and more promising results, such as computer network security, intelligent robots, intelligent control, and pattern recognition and fault diagnosis. These research efforts and applications not only can help us to further understand the immune system itself, but also to re-examine and solve practical engineering problems from the perspective of information processing way in biological immune system.

Building a computer security system using the principles of the immune system opens a new research field of information security. Many structures, functions, and mechanisms of the immune system are very helpful and referential to the research into computer security, such as antibody diversity, dynamic coverage, and

Artificial Immune System: Applications in Computer Security, First Edition. Ying Tan.

distribution. We believe that the features of the immune system are the roots and original springs for us to build perfect computer security systems.

1.2 BIOLOGICAL IMMUNE SYSTEM

1.2.1 Overview

Biological immune system (BIS) is a highly complex, distributed, and parallel natural system with multiple levels, which can identify the self, exclude the nonself, for maintaining security and stability in the biological environment. It makes use of innate immunity and adaptive immunity to generate accurate immune response against the invading antigens. The BIS is robust to noise, distributed, self-organized, noncentral control, and has enhanced memory [1]. The original substance in an organism is called the self such as normal cells. The non-original substance in the organism is called the nonself like the invading antigens.

Biological immune systems consists of innate immunity (also known as nonspecific immune) system and adaptive immunity (also known as specific immune) system. The two systems mutually cooperate to resist the invasion of external antigens. Specifically, innate immune response starts the adaptive immune response, influences the type of adaptive immune responses, and assists adaptive immune to work. Adaptive immune response provides a more intense specific immune response [2].

Innate immune system is an inherent defense system that comes from a long-term evolutionary process. It is the first line of defense against antigens, which provides the innate immune function of the body. Usually, the innate immune system makes use of innate immune cells to recognize the common pattern formed by a variety of nonself. Therefore it can identify a variety of antigens, effectively preventing the invasion of most antigens. If an antigen breaks up the body's innate immune defense barrier, the adaptive immune system of the human body will be invoked and becomes responsible for the immune response to that specific antigen.

An adaptive immune system mainly has the following three functions:

1. Identifies specific antigens.
2. Provides the specific immune response to clear the corresponding antigen.
3. Provides a mechanism for immune memory.

Specific memory cells are able to remember the corresponding antigens. When the same antigen invades the body again, the memory cells will propagate and divide rapidly, providing a more intense immune response to it.

1.2.2 Adaptive Immune Process

Lymphocytes are the main effective immune substances in the adaptive immune system, which consists of T lymphocytes and B lymphocytes. The generation process of the lymphocytes is shown in Fig. 1.1. After negative selection, bone marrow stem

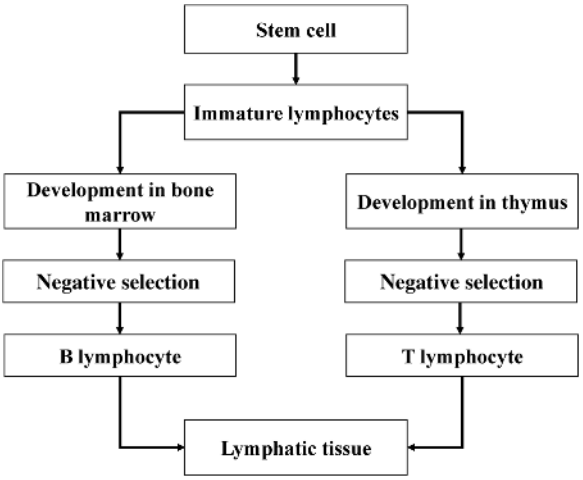

Figure 1.1 The generation and differentiation of lymphocytes.

cells grow to the B cells and T cells in the bone marrow and thymus. Other cells involved in the adaptive immune response include phagocytic cells, dendritic cells, and so on.

In the generation process of lymphocytes, they are affected by a large number of self. The lymphocytes which react with self will apoptosis, and the remaining lymphocytes will go to lymphoid organs and tissues, cycling in the organism with the lymphatic blood. This process in the biological immune system is called the negative selection process [3]. Based on the negative selection process, the biological immune system is able to successfully identify self and nonself, without the need of any nonself information.

In the first time adaptive immune response, T cells and B cells will proliferate and differentiate into effector T cells and effector B cells, respectively. The effector T cells are able to specifically recognize invading antigens and eliminate the antigens directly through cell lysis. This immune process is called cellular immunity. Different from effector T cells, the effector B cells specifically recognize and destroy the antigens by secreting antibodies, which is a kind of immunoglobulin. This process is called humoral immunity. In such a process, a few effector cells will differentiate into memory cells, achieving the immunological memory that is able to remember the antigens for a long time.

When the antigens invade the organism again, the adaptive immune system will produce the secondary immune response. In case of the secondary immune response, the memory cell is capable of proliferation and differentiation quickly, producing a large number of effector cells and providing a more intense immune response.

The proliferation and differentiation process of the lymphocytes cloning process are actually the processes of cloning and mutation of lymphocytes, respectively. Such clone and mutation processes result in the diversity of immune cells in the biological immune system. It is this kind of diversity that

gives the biological immune system the ability to identify unknown antigens and new variants of known antigens.

1.3 CHARACTERISTICS OF A BIOLOGICAL IMMUNE SYSTEM

Artificial immune system (AIS) is a bionic system inspired from immunology principles of the biological immune system. The key to designing the artificial immune system is to take full advantage of the immunology principles and to replicate the effectiveness and capability of the biological immune system in computer systems. A biological immune system has a number of inspirational characteristics that artificial immune system can borrow from, including:

- Distributivity: Lymphocytes in the biological immune system are able to detect abnormity independently, with the control of a center, which means that they constitute a highly distributed system. When designing the artificial immune system, this feature is very helpful to the self protection and robustness of AIS. The architecture based on agents has been proposed to simulate the distributivity of the immune system.
- Multi-layered: The biological immune system has a multi-layer structure. A single layer of the biological immune system cannot protect the organism from all invasions, but the cooperation of multiple layers is able to achieve the security protection of the system. Although this feature is not unique to the biological immune system, it is a very important feature of the biological immune system. Studies and implementations of the multi-layered feature in the artificial immune system for computer systems can greatly enhance security of computer systems.
- Diversity: In nature, although the bodies protected by the biological immune system are the same on the whole, each body has its own differences. The diversity of different bodies is also very helpful to the protection against invasions. Diversity is from two aspects: one is the body's own diversity, the other is the diversity of the biological immune system. The combination of the two aspects increases the "diversity" greatly and is very important to the protection of our body. In the field of computer system security, the implementation of the diversity can be also achieved in two aspects—the diversity of computer operating systems and the diversity of the artificial immune system.
- Disposability: No immune cells in the biological immune system are indispensable. Every immune cell has a lifecycle. In the study of artificial immune systems, we can borrow the mechanism to achieve the lifecycle of immune antibodies.
- Autonomy: The biological immune system does not require a central control node. They can automatically recognize and destroy invading antigens and unitize the illness and death of immune cells to update themselves, achieving the immunologic function on their own.

- Adaptability: The biological immune system is able to learn newfound invading pathogens, and form the memory. The speed of response to the same pathogen invasion will be accelerated. Learning mechanisms of the biological immune system are very important to the artificial immune system. The artificial immune system should not only remember the abnormal immune information found in the past, but also dynamically learn the immune rules to handle the emerging unknown anomalies.
- No secure layer: In the biological immune system, any cell can be the invaded by pathogens, including lymphocytes. But other lymphocytes can kill the invading pathogen. The mutual help between the lymphocytes forms the basis of the security of the biological immune system.
- Dynamically changing coverage: The biological immune system can maintain a good balance between the space and time of the detector set. The biological immune system cannot form a large detector set to contain all the invasion information. At any time, the detector set flowed into the body is just a small portion of the entire detector set. The flowed detector set will update itself over time and the lifecycle. Such a mechanism has great benefits for enhancing the portability and coverage of the biological immune system.
- Identity via behavior: In the field of encryption, the encryption algorithm is used for identification. However, the biological immune system uses the representations of antibody and antigen for identification. In the field of computer systems, any representation is based on "0" and "1" at the bottom. Finding a reasonable representation will result in good recognition effect.
- Anomaly detection: The biological immune system is able to recognize the pathogen that is never seen. This phenomenon is called anomaly detection. This feature is conducive to the artificial immune system for achieving the function to detect unknown anomalies or to find new viruses in the field of computer security.
- Incomplete detection: Any match between antibodies and antigens is not a complete match. This feature can enhance the diversity and generalization of detectors. Just a few antibodies are able to detect a large number of antigens.
- Numbers game: The numbers game mainly refers to the time of the invasion and the protective response. Immune response must be faster than the speed of invasion, otherwise the immune protection will be overwhelmed by the invasion. Researchers of artificial immune system indicate that more attention should be paid to the lightweight of the system.

1.4 ARTIFICIAL IMMUNE SYSTEM

Artificial Immune System (AIS) is a computational intelligence system inspired by the working mechanism and principles of the biological immune system. Based on the concept and idea of "getting wisdom from nature," and by simulating the working mechanism of biological immune systems, artificial immune systems

successfully achieve many advantages of biological immune systems, including noise patience to learn without a teacher, distributed, self-organized, no center control, and strengthening memory, and other features [4]. Artificial immune systems have developed into a hotspot research field of computational intelligence [5], and attracted many interested researchers.

There are a variety of immune algorithms and models in a artificial immune system. Among them, most algorithms try to utilize the mechanisms of learning and memory of biological immune systems for problem solving. Most algorithms and models achieved great success. In the artificial immune algorithm, the antigen corresponds to the objective function for solving problems and constraints, the antibody corresponds to candidate solution, and antigen and antibody affinity matching degree corresponds to candidate solution with objective function. The general steps of the artificial immune system algorithm 1 is shown next. In Algorithm 1, when suspended, it was the best match with the antigen-antibody, which has been optimized to the solution that solves the problem successfully.

Algorithm 1 General Steps of Artificial Immune Algorithm (AIS)

Step 1 Input antigen,
Step 2 Initialize antibody populations,
Step 3 Calculate affinity for each antibody,
Step 4 Check the lifecycle of each antibody and update the antibody,
Step 5 If the abort condition, then go to **Step 6;** otherwise steer for **Step 3**,
Step 6 Output antibodies.

Based on the negative selection mechanism of biological immune systems, Forrest and associates first proposed a negative selection algorithm [6], as shown in Algorithm 2, for anomaly detection in computer systems. This algorithm is one of the most important AIS algorithms, and is of very good robustness in identifying self and "variant," without reference to information of variant, which can be used to detect unknown antigens. It is especially suitable for unknown computer security monitoring, fault diagnosis under changing environments, computer malware detection, anomaly detection, intrusion detection, and so on.

Algorithm 2 Negative Selection Algorithm (NSA)

Step 1 Define self as a category of detector set.
Step 2 Generate a detector randomly; this detector undergoes "autologous" match. If a match occurs, then this detector is removed, otherwise it is added to the variant detector concentration.
Step 3 Abort condition judgment; if the variant does not contain a sufficient concentration detector detector, steer for **Step 2**, otherwise abort.

According to the clonal selection theory proposed by the Australian immunologist Burnet [7], Castro and Zuben proposed a clonal selection algorithm inspired by

the clonal selection mechanism of the artificial immune system [8,9]. In the B cell cloning process, according to the affinity, clonal selection algorithm in the vicinity of the candidate to produce a variation of individual clones as a population of individuals for expanding the search. In such a way, the clonal selection algorithm can help prevent an evolutionary algorithm from premature, i.e., avoid falling into local minima [10], then leads to improving the optimization speed of the algorithm [11].

Application of biological diversity mechanism in the immune system helps to improve the global search ability of optimization algorithms and accelerates their convergence speeds. Negative expression mechanism of biological immune system, self-organization, and unsupervised learning may provide us useful mechanisms to cognise unknown environments by use of the known information. The mechanism of immune memory that can save the previous knowledge learned is very important and vital for many intelligent systems. Other artificial immune models and algorithms such as artificial immune network models, dendritic cells algorithms, and so on, are not introduced here due to space limitations. In the midst of the rapid development of artificial immune algorithms, many people continue to put forward a variety of novel artificial immune models and algorithms for many real-world problems.

Artificial immune systems have been successfully applied to many practical fields, including computer security, optimization, fault diagnosis, and pattern recognition, to name a few. In particular, computer malware detection based on immune principles has been developed rapidly and achieved many fruitful results and achievements, attracting more and more researchers. Nevertheless, these artificial immune systems and malware detection methods are not perfect. Most of them have some deficiencies and shortcomings, which stimulates researchers to explore more efficient models and algorithms.

1.5 ARTIFICIAL IMMUNE SYSTEM MODELS AND ALGORITHMS

1.5.1 Negative Selection Algorithm

Inspired by the generation process of T cells in immune systems, Forrest and associates [6] proposed a negative selection algorithm, which has become one of the most famous AIS algorithms. Biological immune systems have the ability to distinguish between self cells and nonself cells, which makes it able to recognize invading antigens. T cells play a key role in this process. The generation of T cells includes two stages: the initial generation stage and the negative selection stages. First, T-cell receptors are generated by a random combination of genes. In order to avoid the erroneous recognition of self, T cells are filtered in the thymus (i.e., negative selection process). The T cells that can recognize the self cells will be removed, while others that are approved by the T cells are able to participate in the immune response. Forrest and associates applied the same principle to the distinction of self and nonself in computer systems. They generated the detector set by a negative selection process to recognize the nonself that invaded the computers.

The process of negative selection algorithm is shown in Algorithm 3. The negative selection algorithm includes the detector set generation stage and the nonself detection stage. In the detector set generation process, the self gene library is constructed from the self files. Then the detector set is randomly generated. The detectors that match the self gene library are removed from the detector set according to the negative selection principle. The main role of the detector set is that it can fully cover the nonself data space. Therefore, the number of detectors tend to be more substantial. In the stage of nonself detection, the algorithm conducts the *r*—contiguous bits match between the sample and the detectors in the detector set one by one. Once a match occurs, the sample will be labeled as "nonself."

The key to the negative selection algorithm is to design the detector representation and matching functions. Regarding these two aspects, researchers have carried out a lot of work on negative selection algorithms [12].

Dasgupta and González [13] represented the detector as a rectangular function of the real number space, which is able to measure the degree of "abnormal."

González, Gupta, and Gómez [14] analyzed the limitations of the binary string representation and its matching process. They discussed the experimental performance of binary-type detectors and analyzed distribution of such types of detectors set in the data space. They pointed out that the binary type of detector is not able to characterize the data spatial structure well of certain issues.

Balachandran and associates [15] conducted the investigation on multiple shapes of detectors in the real value space: super rectangle, super sphericity, super spheroidicity, and so on. In addition, they gave a uniform negative selection model.

Balthrop and associates [16] proposed the *r*—block matching function. The matching process measures the match status of the detectors and the text character block. The matching method can reduce the vulnerabilities of detection and improve the detection range of the detector set.

Algorithm 3 Negative Selection Algorithm

Input: The self set $SELF = self_i$.
Output: The detector set $D = d_i$.

1. $D = \phi$.
2. **While** termination condition does not meet **do**.
3. Randomly generate a detector set N.
4. **For all** detector d in the detector set N.
5. **For all** self $sel\,f_i$ in the self set $SELF$ **do**
6. **If** $Affinity\,(d, sel\,f_i) < \theta$ **then**
7. Remove d from the detector set N.
8. Continue.
9. **End if**
10. **End for**
11. **End for**
12. $D = D \cup N$.
13. **End while.**

Ji and Dasgupta [17] adopted the Euclidean distance as the detector matching function in the real value space, and dynamically adjusted the matching threshold value according to the length of detectors.

In the traditional negative selection algorithm, the detector set for nonself is generated randomly. This random method without wizard will consume a lot of resources. Furthermore, the traditional negative selection algorithm is more concerned about nonself characteristic of the samples, while what the biological immune system really cares about is the danger of antigens. The concern of anomaly detection in computer systems is the risk of the sample. Therefore, how to improve the negative selection algorithm's concern about the risk of the sample becomes a valuable work.

1.5.2 Clonal Selection Algorithm

In biological immune systems, each B cell produces a kind of antibody, in order to identify the corresponding antigen. When the antibody and antigen match (i.e., binding) and receive a stimulus signal emitted by the helper T cells, the corresponding B cells of antibodies are activated and cloned and differentiated into plasma cells and memory B cells. When the memory B cells encounter the same antigen again, they will generate a lot of antibodies with high affinities. Burnet [18] proposed the biological clonal selection theory to explain the process of cloning and the relationship between proliferation and differentiation of the immune cells and the affinities.

Inspired by this theory, De Castro and Von Zuben [8] proposed the clonal selection algorithm. The core idea of the algorithm is to select and clone the cells with high affinities and clear the cells with low affinities, while cloning and mutating the cells based on affinities of antigens and antibodies.

Algorithm 4 gives the pseudo-code of the clonal selection algorithm. First, the initial solution set is is regarded as the set of immune cells, and n solutions with the highest affinities are selected from the set. Then the selected n solutions are cloned. The amount of offspring is proportional to the affinity, and the degree of mutation is inversely proportional to the affinity. According to the affinity, immune cells with low affinity in the collection will be replaced with a certain probability. If an optimal solution is not found, then the algorithm goes to the next iteration. It can be seen that the algorithm gradually approaches the optimal affinity set in the iterative process, like the reinforcement learning. Furthermore, the random replacement in the algorithm is able to effectively maintain the diversity of the immune cell set.

De Castro and Von Zuben [9] further analyzed and discussed this algorithm and applied it to learning and optimization problems. They used the binary encoding for solutions.

Algorithm 4 Clonal Selection Algorithm

Input: The pattern set S.
Input: n, the number of antibodies to be cloned.
Input: d, the number of new antibodies at each iteration.

Output: The memory detector set M.

1. Randomly generate the candidate antibody set P_r.
2. Randomly generate the memory detector set M.
3. **While** termination condition does not meet **do**
4. $P = P_r \cup M$.
5. Select n antibodies with best affinities from P, denote the set as P_n.
6. $C = \phi$
7. **For all** antibody $a \in P_n$ **do**
8. Clone a to get new antibodies, denote the set as A, the size of the set A is proportional to the affinity of a.
9. $C = C \cup A$
10. **End for** $C^* = \phi$
11. **For all** antibody $a \in C$ **do**
12. Mutate the antibody a to a^*. The degree of mutation is proportional to the affinity of a.
13. $C^* = C^* \cup a^*$
14. **End for**
15. Select the best antibodies from C^*; replace M with the best set.
16. Randomly generate the candidate antibody set N_d with size d.
17. Replace the d antibodies with lowest affinities in P_r and M with antibodies in N_d.
18. **End while.**

On the basis of De Castro and Von Zuben's work, researchers have proposed a number of clonal selection algorithm variants. The clonal selection algorithm gradually becomes an important branch of artificial immune system.

Cutello and Nicosia [19] gave a new strategy to maintain diversity. For each B cell, they defined the probabilistic half-life period to control the cycle of the B cell, and updated the immune cell set according the lifecycle.

Garrett [20] added the parameters of the clonal selection algorithm to the representation of solutions, and used the real value to encode the solution. The parameters of the algorithm can be automatically adjusted in an iterative process. This method avoids the process of parameter selection, which is very useful to problems with uncertainty parameters.

Watkins and associates [21] studied the distributed nature of clonal selection algorithms and gave the parallel implementation of this algorithm. This method divided memory B cells into multiple independent groups, and each group evolves independently. At last the solutions from all of the groups are integrated to obtain the final result.

Cruz and associates [22] discussed different variants of clonal selection algorithm. The binary encoding strategy and the real value encoding strategy were compared. They also analyzed the affect of Cauchy mutation and Gaussian mutation to the performance of clonal selection algorithm.

Brownlee [23] made a comprehensive analysis of the development of the clonal selection algorithm. He pointed out that the common character of clonal selection algorithm variants in aspect of operators and the framework, and compared clonal selection algorithms with evolutionary computation algorithms.

1.5.3 Immune Network Model

Immune Network Theory [24] explains the relationship between the immune system B cells: no matter the presence or absence of the antigen, B cells in the immune system have excitation and inhibition affect with each other. The mutual excitation and inhibition of B cell-cell make the B-cell network stable. The excitation of a B cell is not only affected by the antigen, but also affected by the excitation and inhibition from other B cells in the immune network.

Inspired by the ideology of the immune network theory, Hunt and Cooke [25] proposed an artificial immune network method and applied it to the DNA sequence recognition. In this method, B cells are correlated according to the degree of affinity and inhibition. The population of B cells includes two subgroups: the initial population and the cloned population. In the training phase, the training set is divided into two parts, one for generating the initial B-cell network, and the other part is used as antigens to stimulate B-cell network. When the affinity between an antigen and a B cell exceeds a predetermined threshold, the B cell is excited and will be cloned and mutated. The generated B cells then join the network and will be dynamically adjusted by the excited state of the network.

This work found the basic features of the immune network theory. Regarding the mechanisms of the immune network and the representation method of B cells, researchers have proposed a variety of artificial immune network approaches [26].

Timmis and Neal [27] proposed the idea of artificial recognition balls. Each artificial recognition ball represent a group of similar B cells. There exist excitation and inhibition among the artificial recognition balls to maintain the stability of the immune network. This method assumes that the network resources are limited; the overall number of B cells represented by the artificial recognition balls B cells is limited.

Neal [28] proposed self-stable artificial recognition balls that are controlled distributively. Each artificial recognition ball automatically controls its own resources.

Nasaroui and associates [29] applied the fuzzy theory to the artificial immune network. The artificial recognition balls are represented as fuzzy sets in the data space. The method also proposed to merge artificial recognition balls according to affinity, which is similar to the crossover operator in evolutionary computation.

De Castro and Von Zuben [30] combined the clonal selection algorithm and the immune network theory. In the adjustment process of the network, it conducted clonal selection and suppression to the immune cells based on the affinity.

1.5.4 Danger Theory

Matzinger [31] analyzed the limitations of self and nonself theory, and proposed the immune danger theory on this basis. According to the traditional immune theory, the function of biological immune system is to distinguish between self and nonself. However, some harmless variant, such as food, embryos, and transplanted organs will not trigger an immune response. Therefore, Matzinger pointed out that the

immune system's function is to detect danger, rather than detect nonself. Danger signals are generally released by injured cells before death and can synergisticly stimulate the antigen-presenting cells.

From the perspective of artificial immune system, Aickelin and Cayzer [32] analyzed the danger theory and discussed how to build the corresponding artificial immune models. They proposed the concept of the danger zone. The core of the danger theory is the cooperative stimulation of danger signals, and scope for danger signals is the local area of the injured cells (i.e., danger zone). The activation of B cells requires two conditions: one is the match of corresponding antibody and antigen, the other one is locating in a danger zone and being stimulated by the danger signal.

The key to build a danger model is to define a reasonable danger signal and the danger zone based on the original matching principle. In practical problems, danger signals can be dangerous independent mechanism and can be regarded as the information representation of a problem. For the definition of danger zones, the similarity in the space or the time can be used, and the correlation between the data can also be used. The danger theory has the potential to be used in anomaly detection and data mining.

On the basis of Aickelin and Cayzer's work, researchers have proposed a number of artificial immune models based on the danger theory [33–35].

Secker and associates [36] explored how to apply the danger theory to web mining. The definition of danger signals is based on the user's behavior and interests. The danger zone is defined according to the distance in time and space of the documents. This work mainly discussed ideas and model's framework, without giving a specific implement algorithm.

Aickelin and associates [37] analyzed the relationship between the danger-theory based artificial immune system and the intrusion detection system. They discussed how to defined the danger signals and danger zones based on the intrusion behavior in order to build a more robust intrusion detection system.

Prieto and associates [33] applied the danger theory to the control strategy of robot soccer goalkeeper. When football is located in our region, a first immune signal is generated. When an opposing player comes into the penalty area with the ball (danger zone), the dangerous signal will be generated.

Chao and associates [35] detected the anomaly in the software system based on the danger theory. In the running process of the software, the abnormal changes of the system resources will result in a danger signal, indicating the anomaly of the software.

1.5.5 Immune Concentration

Immune concentration is an immune inspired algorithm for feature extraction. In this section, I will take spam detection [38] as an example to introduce the concept of immune concentration.

The essence of the feature extraction method lies in the construction of concentration feature vectors. Tan and associates [39,40] presented global concentration

(GC) based feature extraction methods for spam filtering. Zhu and Tan [41,42] proposed local concentration (LC) based feature extraction methods. In these methods, statistical term selection methods [43] are utilized to remove uninformative terms. Then a tendency function is well designed to generate two detector sets [41,42,44,45]. The tendency of a term t_i is defined in Eq. 1.1. $T(t_i)$ measures the difference between the term's occurrence frequency in two types of messages. Terms are added to corresponding detector sets according to their tendency. Detector concentration, which corresponds to antibody concentration in BIS, are then extracted from messages by using the detector sets. In addition, a sliding window is utilized to slide over a message to extract position-correlated information from messages. By using a sliding window, a message is divided into local parts. At each movement of the window, a spam detector concentration S_i and a legitimate detector concentration L_i are calculated with respect to the two detector sets and the terms in the window according to Eqs. 1.2 and 1.3.

$$T(t_i) = P(t_i|c_l) - P(t_i|c_s) \tag{1.1}$$

where $P(t_i|c_l)$ denotes the probability of t_i's occurrence, given messages are legitimate emails, and $P(t_i|c_s)$ denotes the probability of t_i"s occurrence estimated in spam.

$$S_i = \frac{\sum_{j=1}^{w_n} M(t_j, D_s)}{N_t} \tag{1.2}$$

$$L_i = \frac{\sum_{j=1}^{w_n} M(t_j, D_l)}{N_t} \tag{1.3}$$

where N_t is the number of distinct terms in the window, D_s denotes the spam detector set, D_l denotes the legitimate email detector set, and $M($ denotes the match function, which measures the number of terms in the window matched by detectors.

Each sliding window defines a specific local area in a message. To explore the effects of a sliding window, we design two strategies using a sliding window with fixed-length (FL) and using a sliding window with variable-length (VL). When a fixed-length sliding window is utilized, messages may have different number of local areas (corresponding to different number of feature dimensionality), as messages vary in length. To handle this problem, we may either expand a short message by reproduce the existing features, or reduce the dimensionality of long messages by discarding uninformative features. In VL strategy, the length of a sliding window is designed to be proportional to the length of a message, and there is no need for specific process of feature dimensionality. Preliminary experiments showed that both the two strategies are effective in extracting discriminative features. In the circumstance that the size of a window is set to infinite, a message is taken as a whole for getting concentration features, GC feature vectors are extracted. When the window size is smaller than the message length, the window divide a message into individual local parts, and LC features are extracted from each window.

1.5.6 Other Methods

Dasgupta and associates [46] made a comprehensive analysis of a variety of biological immune model, and pointed out that the biological immune system is a highly complex network composed of biological tissue, immune cells, chemical molecules, and other parts. On this basis, they proposed a multi-layer multi-resolution immune learning model, which integrated a variety of immunization strategies, including dynamical detector generation, clonal selection, and the interactions of immune cells. The method is able to make full use of the function of various immune cells; helper T cells, suppressor T cells, B cells, and antigen-presenting cells will synergistically interact information to detect anomalies.

Wang and associates [47] presented a complex immune system to simulate the representation and process of antigens. This method also used the interactions among a variety of immune cells. It comprised five immune processes, and mainly concerned the processing and representation of the antigens, and the interactions between the antigen presenting cells, T cells, and B cells. Experiments show that the system has a good memory and noise immunity.

Zhang and Hou [48] combined the niche strategy and clonal selection algorithm and proposed a hybrid immunological method. This method combined the negative selection, clonal selection, mutation, and niche strategies, which is able to effectively reduce the number of detectors.

Li and associates [49] proposed an efficient artificial immune network that combined a artificial immune network and particle swarm optimization algorithm. In the particle swarm optimization, particles' behavior can be affected by optimal particle in the population. The interacting with the optimal particle the swarm is able to speed up the convergence of the particle swarm. As far as the immune network, they introduced the interaction between immune cells and the optimal immune cell, making the immune network converging to a stable state with a faster speed.

De Castro and Von Zuben [50] proposed the Bayesian artificial immune system which replaces the basic clone and mutation operator with a probabilistic model for solving complex optimization problems.

1.6 CHARACTERISTICS OF THE ARTIFICIAL IMMUNE SYSTEM

The biological immune system has been evolving for hundreds of millions of years and plays a very important role in the protection of the body from bacterial invasion. Although the immune system may encounter problems, generally speaking, we can see its unique protective effect. The working principles of the biological immune system will have some inspiration and reference meaning on the research job of security protection technologies of computer systems, providing a brand new thinking of computer security, if the computer systems are seen as human bodies and the external intrusions as harmful viruses.

Immunity refers to the ability of the body to identify self or nonself and exclude nonself. The biological immune system is the body's natural system with functions of resistance to the disease itself and prevention of invasion from harmful bacteria. This system itself has many characteristics, some of which get certain significance on the research of computer system security.

1.6.1 Distributed Detection

The immune system works in a way of distributed detection, in which the detector to detect the bacterial invasion is very small but with high detection efficiency; centralized control center and collaboration are not required. Computer security systems are not equipped with the function of distributed detection and the use of the control center has actually reduced the factor of safety protection of the system.

1.6.2 Detection of Abnormality

The immune system is able to identify the invading bacteria that the system has never seen and take corresponding measures. The specific targets of the current computer security protection system are generally decided by the protective strategies or the protection system itself, without automatic intrusion detection of the latest way of invasion.

1.6.3 Learning and Memory

The immune system is able to automatically learn the structure of invading bacteria, and memorize this information in order to reply to this type of bacteria faster and more timely subsequently. Current computer security systems do not have the ability of self-learning.

1.6.4 Diversity

Different biological bodies have different immune systems. A certain weakness of one immune system is not the weakness of another. A virus might be able to break through one protective immune system, but the possibility of breaking through other immune systems is very small. Thus, the immune systems have strong ability to protect the overall population. While for computer systems, the security systems are always the same. Once a loophole is found, any computer system using this kind of security system will suffer the threat of invasion through this loophole.

1.6.5 Incomplete Detection

The immune system does not require making nonself test on every invading cell. It has great flexibility and may sacrifice a portion of the body functions or resources in

order to ensure the normal functions of the body in general. Computer security systems generally do not have the ability of overall analysis of the system and its functions are generally specific and fixed.

1.7 APPLICATIONS OF ARTIFICIAL IMMUNE SYSTEM

1.7.1 Virus Detection

According to the ability of distinguishing self and nonself of the immune system, Forrest proposes principles and laws of BIS that AIS can take inspiration from and he has done a lot of research work to support this. By taking inspiration from the mechanism of BIS resisting and destroying unknown biological virus, T. Okamolo proposed a distributed agent-based anti-virus system. It consists of two parts: the immune system and the recovery system. The function of the immune system is identifying the nonself information (computer virus) by grasping the self information; the recovery system copies files from the non-infected computer to the computer which has been infected through the network to cover the files on it [45,51–53]. Based on the same principles, AIS is also used for hacking prevention, network security maintenance, and system maintenance.

1.7.2 Spam Filtering

Spam filtering is an important and typical pattern recognition problem because spam causes many problems to our daily communication life. In solving the problem, both classical statistical methods and AIS methods have been presented, and most of them focus on studying feature extraction methods and design of classifiers. The main function of feature extraction is to extract discriminative information from messages and transform messages into feature vectors. The statistical feature extraction methods try to collect and analyze numerical characteristics of messages, such as term frequencies, and relation between terms and email categories. Some prevalent ones are Bag-of-Words (BoW) [54], Sparse Binary Polynomial Hashing (SBPH), and Orthogonal Sparse Bigrams (OSB) [55]. Different from the statistical ones, the AIS methods [56] construct feature vectors by mimicking the process of antibody creation in BIS. In design of classifiers, classical pattern recognition methods, e.g. Naive Bayes(NB) [57,58], Support Vector Machine (SVM) [59–61], k-Nearest Neighbor (k-NN) [62,63], and Artificial Neural Network (ANN) [64,65] were proposed on the basis of statistical theory. On the contrast, AIS models were inspired by natural functions and mechanisms of BIS [56,66].

These statistical and AIS methods are quite different in terms of both origins and principles, which endow them with quite distinct properties. Combining the strength of both approaches may help achieve better performance. We will introduce and discuss several recent works of our laboratory [39–42,67–69], which applied mixed principles to feature attraction, classifier combination, and classifier updating, so as to demonstrate the rationality of combining statistical and AIS methods for

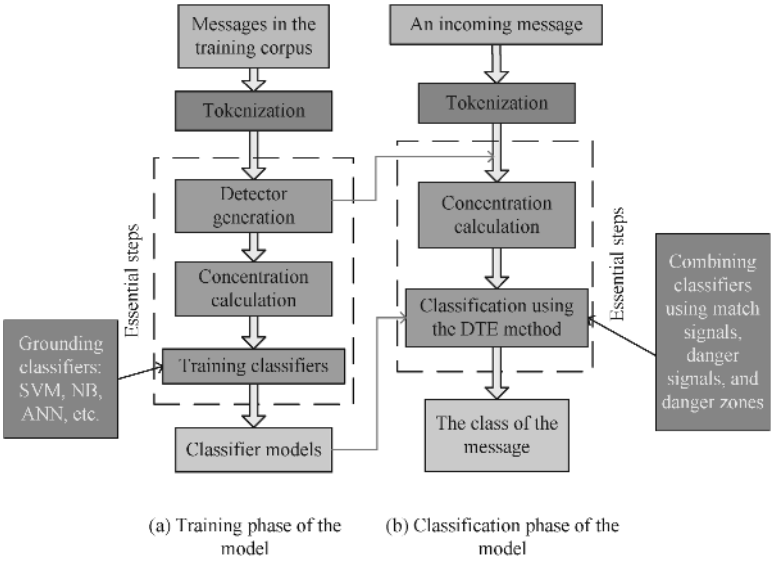

Figure 1.2 Training and classification phases of the immune-based model.

spam filtering. In addition, we present a generic framework of an immune-based model for spam filtering, and online implementation strategies are given to demonstrate how to build an immune-based intelligent email server.

1.7.2.1 Concentration-Based Feature Representation

Based on these previous works, we present a generic framework of an immune-based spam filtering model, as depicted in Fig. 1.2. According to the model, concentration-based feature vectors are extracted from messages by computing match concentration of detections. Classifiers are then built on the concentration vectors of training corpus. Finally, incoming messages can be classified by using the Danger Theory-based Ensemble (DTE) method. In addition, classifiers are updated at all times based on the drift of messages and classification performance. In the following subsections, we briefly introduce and discuss the principles of these methods [70,71], and analyze the rationality of combining statistical principles with AIS ones.

Experiments were conducted on real-word corpora Ling, PU1, PU2, PU3, PUA, and Enron-Spam[1] using cross validation to investigate the performance of the concentration-based method. Meanwhile, four benchmark criteria, namely spam precision, spam recall, accuracy, and F_1 measure were adopted in analyzing the results. Among them, accuracy and F_1 were more important as they indicated the overall performance of approaches. From these experimental results, it can be seen that the combination of statistical information and immune characteristics helps achieve the

[1] The PU corpora and Enron-Spam are available from the web site: http://www.aueb.gr/users/ion/publications.html.

best discriminative performance. The success lies in the following aspects: (1) By using term selection methods, noise and uninformative terms can be removed, which reduce computational complexity and enhance effectiveness of detectors. (2) The concentration principle helps obtain feature vectors with lower dimensionality. (3) The sliding window strategies provide effective ways of defining local area in messages, and extracting position-correlated information.

1.7.2.2 Danger Theory Inspired Ensemble Method

Mimicking the DT theory, we defined artificial signals and danger zones, and classifiers were combined using them [67]. First, two types of artificial signals, namely, signal 1 (match signals) and danger signals, were respectively generated using two independent classifiers. Depending on the classification results, negative or positive signals would be generated. After the production of the signals, the two classifiers were interacted through the transmission of the signals. Mimicking the DT mechanism, the transmission of the signals was designed to be different. An activated signal 1 would be sent only to the specific sample, upon which the signal was arisen. However, an activated danger signal would be sent to all the test samples within the danger zone, besides the specific sample. Finally, the result was acquired based on the interaction among classifiers.

The framework of the DTE method is depicted in Fig. 1.3. A test sample gets labeled by the first two classifiers if the two signals agree with each other. Otherwise, a third classifier (self-trigger process) is utilized to solve the conflict and get the test sample classified. According to the method, three classifiers are combined in a cascade way. Similar to other cascade method, the order of classifiers can be determined according to classifier performance on training corpus. The characteristics of the DTE method lie in the interaction among classifiers by using the danger zone and the signals.

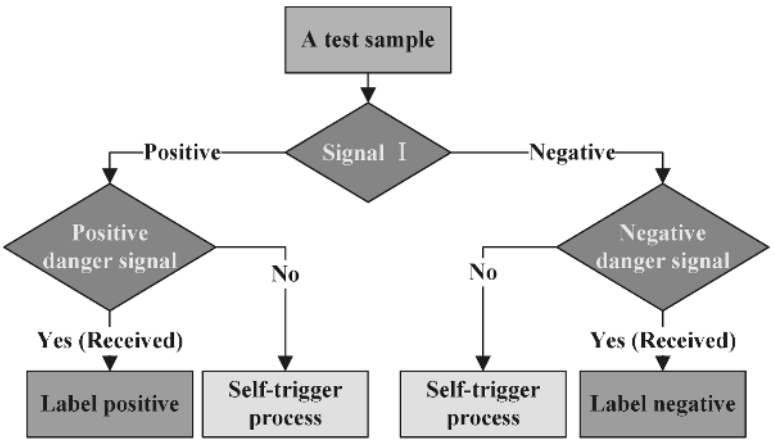

Figure 1.3 The framework of the DTE method.

The interaction between the first two classifiers is expressed as shown in Eq. 1.4.

$$E(x_i) = \sum_{x_j \in D} \delta(c_1(x_i), c_2(x_j)) K(d(x_i, x_j)) \tag{1.4}$$

where x_i and x_j are test samples, D denotes the test set, $c_1(x)$ and $c_2(x)$ are the two classifiers, $d(x_i, x_j) = \|x_i - x_j\|$ is the distance between two samples, $K(z)$ is defined in Eq. 1.5, and $\delta(y_1, y_2) = 1$, if $y_1 = y_2$, and 0 otherwise.

$K(z)$ defines the effect of the danger zone as follows:

$$K(z) = \begin{cases} 1 & if\ z \le \theta \\ 0 & otherwise \end{cases} \tag{1.5}$$

where θ is the size of the danger zone.

After obtaining the weighted result $E(x_i)$, the sample x_i can get its class label using Eq. 1.6.

$$L(x_i) = \begin{cases} c_1(x_i) & if\ E(x_i) \ge 1 \\ f(x_i) & otherwise \end{cases} \tag{1.6}$$

where $f(x)$ denotes the class label given by the third classifier.

The performance of the DTE was investigated on four real-world corpora, namely PU1, PU2, PU3, and PUA using ten-fold cross validation. In the experiments, SVM, NB, and Nearest Neighbor (NN) were utilized as three grounding classifiers. SVM was utilized to generate match signal, NB was utilized to generate danger signal, and NN was utilized in the self-trigger process. The experimental results [67] show that the danger zone provides a well defined interaction between the two types of signals, and classifier are combined through the interaction. By using the DTE method, the performance of classifiers can be effectively improved.

1.7.2.3 Immune-Based Dynamic Updating Strategies

Mimicking dynamic mechanisms of BIS, we proposed several classifier updating strategies [68,69]. The updating process of SVMs is depicted in Figs. 1.4 and 1.5. Support vectors (SVs) of a SVM are used as detectors (antibodies) and SVs are

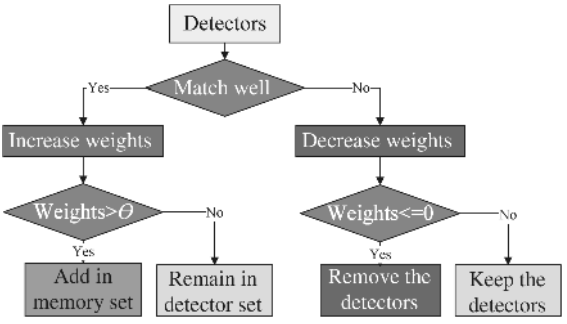

Figure 1.4 Updating SVs according to their performance.

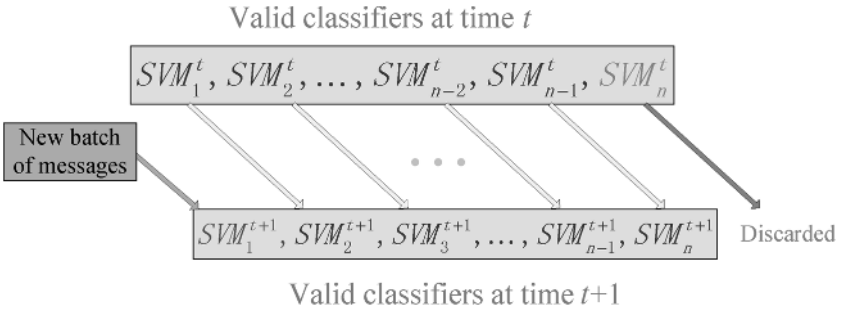

Figure 1.5 Updating SVMs with time according to their lifespan.

updated according to their performance by mimicking the dynamic mechanisms of BIS. In measuring the importance of SVs, we assign weights to SVs, and build up two sets, a Detector Set and a Memory Set. The weight is increased when a SV correctly classifies a sample (according to hamming distance), and vice versa [68,69]. When the weight of a SV is above a pre-defined threshold, the SV will be add to the memory set and the weight will be increased significantly. On the contrary, when the weight of a SV is decreased to zero, the SV will be culled from the detector set. In addition to SVs, the whole SVM is also updated with time. The updating of a SVM is in a greater magnitude as most of the SVs will be changed in this process. In the process, a sliding window strategy is adopted, and the window size controls the life-span of SVMs. When the updating moment is arrived, the oldest SVM is discarded and a new SVM will be built using the new arrival messages. The final classification decision will be made by the majority voting of the SVMs in effect.

1.7.3 Robots

D. W. Lee proposes a controlling method distributed robots based on the principle of homeostasis in the immune system. In this method, each the robot is regarded as a B cell and each environment condition as an antigen, while the behavior strategies adopted by the robots are taken as antibodies and the controlling parameters of the robots as T cells. Under different environment conditions, each robot will first select a set of behavioral strategies that are adapted to the environment conditions of itself. Then this set of behavioral strategies are individually communicated with other robots around one by one, and some behavioral strategies will be stimulated while some others are suppressed. The behavioral strategies that are stimulated more than others will finally be adopted by the robot. Based on the distributed controlling mechanism of the immune system, Lshiguro implements the gait controlling and speed measuring of a six-legged walking robot. The action strategies based on the principle of interaction between B cells in the immune system are used to control the movement of self-regulation robots. The main idea of this strategy is: several basic and different operators of the self-regulation robot are pre-designed and each operator is regarded as an agent that can make action decisions based on its surrounding environment and send controlling commands to the system, and the system

will dynamically determine the robot's actions according to the collaboration and competition status between the agents.

1.7.4 Control Engineering

AIS can be readily identified as a feedback controller based on the principles of fast response and rapid determination of foreign intrusions. It has been applied to the car's rear collision prevention system by comprehensive processing signals transmitted from sensors and controlling each actuator executing corresponding operations quickly and accurately [72,73]. Takahashi designed an immune feedback controller of PID with activation item of controlling the response speed and suppression item of controlling the stabilizing effect. The validity of the controller is verified by simulating a discrete, single-input and single-output system. In addition, AIS is also used in sequence controlling, dynamic and complex controlling and other aspects.

1.7.5 Fault Diagnosis

Distributed diagnosis system combined immune network and learning vector quantization can be used to accurately detect the sensors where failure occurs in controlled object. This system has two modes: training mode and diagnosis mode. In the training mode, data of sensors working normally are trained and achieved through LVQ; in the diagnosis mode, the immune network determine the sensors with faults based on the knowledge acquired by LVQ. Experiments show that the system can automatically identify the failed sensors in the group of working sensors. While in the past, this is implemented by detecting the output of each sensor independently. The self-learning ability of the immune system is also used in the monitoring system of computer hardware, in which the system marks out the area fault occurs in and takes appropriate recovery actions one the computer hardware system goes wrong.

1.7.6 Optimized Design

For the nonlinear optimization problem with multiple local minima, the general optimization methods are difficult to find the global optimal solution, while genetic mechanism based on diversity of the immune system can be used for optimal search. It can avoid premature convergence for improving the genetic algorithm and dealing with multi-criteria problems. It has been currently used for function testing, the traveling salesman problem, VLSI layout, structure design, parameter correction of permanent magnet synchronous motor and others.

1.7.7 Data Analysis

AIS has the ability of data analysis and classification by combining the advantages of classifiers, neural networks and machine inference [74,75]. Therefore, it has been

used in fields of data mining and information processing. Timmis discussed how to implement an unsupervised and self-learning AIS specifically.

1.8 SUMMARY

Biological immune system (BIS) provides a natural biological defense system for biological creatures to defend against external antigens. Artificial immune system (AIS) is a computational intelligence system inspired by the working mechanism and principle of BIS. The working mechanism simulating the BIS allows the AIS to access the many advantages of BIS. At present, AIS has been widely used in many fields such as pattern recognition, function optimization, computer security, robot control, data analysis, and so on.

REFERENCES

1. Janeway, Charles A. Jr, P.T.M.W. and Shlomchik, M.J. Immunobiology: The Immune System in Health and Disease, no. 2 June 21, 2001.
2. Zhang, X. (2010) *Viral Immunology*. vol. 1, Science Press. (in Chinese)
3. Sun, Z. and Wei, W. (2003) Artificial immune system and its application. *Computer Engineering*, **29** (15). (in Chinese)
4. Mo, H. and Jin, H. (2003) Application of artificial immune system to computer security. *Journal of Harbin Engineering University*, **24** (3), 278–282. (in Chinese)
5. Wang, J., Liu, X.Y., and Wang, X. (2006) Artificial immune system and analysis of its models. *Computer Technology and Development*, **16** (7), 105–107. (in Chinese)
6. Forrest, S., Perelson, A., Allen, L., and Cherukuri, R. (1994) Self-nonself discrimination in a computer, in Research in Security and Privacy, 1994. Proceedings. 1994 IEEE Computer Society Symposium on, IEEE, pp. 202–212.
7. Burnet, S.F.M. et al. (1959) *The Clonal Selection Theory of Acquired Immunity*, Vanderbilt University Press, Nashville.
8. De Castro, L.N. and Von Zuben, F.J. (2000) The clonal selection algorithm with engineering applications, in Proceedings of GECCO, vol. 2000, pp. 36–39.
9. De Castro, L.N. and Von Zuben, F.J. (2002) Learning and optimization using the clonal selection principle. *Evolutionary Computation, IEEE Transactions on*, **6** (3), 239–251.
10. Zhang, X. and Jiao, L. (2004) Feature selection based on immune clonal selection algorithm, *Journal of Fudan University (Natural Science)*, **43** (5), 926–929. (in Chinese)
11. Jiao, L. and Du, H. (2003) Development and prospect of the artificial immune system. *Acta Electronica Sinica*, **31** (10), 1540–1548. (in Chinese)
12. Ji, Z. and Dasgupta, D. (2007) Revisiting negative selection algorithms. *Evolutionary Computation*, **15** (2), 223–251.
13. Dasgupta, D. and González, F. (2002) An immunity-based technique to characterize intrusions in computer networks. *Evolutionary Computation, IEEE Transactions on*, **6** (3), 281–291.
14. González, F., Dasgupta, D., and Gómez, J. (2003) The effect of binary matching rules in negative selection, in *Genetic and Evolutionary Computation—GECCO 2003*, Springer, pp. 195–206.
15. Balachandran, S., Dasgupta, D., Nino, F., and Garrett, D. (2007) A framework for evolving multi-shaped detectors in negative selection, in Foundations of Computational Intelligence, 2007. FOCI 2007. IEEE Symposium on, IEEE, pp. 401–408.
16. Balthrop, J., Esponda, F., Forrest, S., and Glickman, M. (2002) Coverage and generalization in an artificial immune system, in Proceedings of the Genetic and Evolutionary Computation Conference, Citeseer, pp. 3–10.

17. Ji, Z. and Dasgupta, D. (2004) Real-valued negative selection algorithm with variable-sized detectors, in *Genetic and Evolutionary Computation–GECCO 2004*, Springer, pp. 287–298.
18. Burnet, F. (1978) Clonal selection and after. Theoretical Immunology, pp. 63–85.
19. Cutello, V. and Nicosia, G. (2002) Multiple learning using immune algorithms, in Proceedings of 4th International Conference on Recent Advances in Soft Computing, RASC, pp. 102–107.
20. Garrett, S. (2003) A paratope is not an epitope: Implications for immune network models and clonal selection. Artificial Immune Systems, pp. 217–228.
21. Watkins, A., Bi, X., and Phadke, A. (2003) Parallelizing an immune-inspired algorithm for efficient pattern recognition. *Intelligent Engineering Systems through Artificial Neural Networks: Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Complex Systems and Artificial Life*, **13**, 225–230.
22. Cruz-Cortés, N., Trejo-Pérez, D., and Coello, C. (2005) Handling constraints in global optimization using an artificial immune system. Artificial Immune Systems, pp. 234–247.
23. Brownlee, J. (2007) Clonal selection algorithms, *Tech. Rep.,* Complex Intelligent Systems Laboratory (CIS), Centre for Information Technology Research (CITR), Faculty of Information and Communication Technologies (ICT), Swinburne University of Technology, Victoria, Australia, Technical Report ID: 070209A.
24. Jerne, N. (1974) Towards a network theory of the immune system, in Annales D'Immunologie, vol. 125, pp. 373–389.
25. Hunt, J. and Cooke, D. (1996) Learning using an artificial immune system. *Journal of Network and Computer Applications*, **19** (2), 189–212.
26. Galeano, J., Veloza-Suan, A., and González, F. (2005) A comparative analysis of artificial immune network models, in Proceedings of the 2005 Conference on Genetic and Evolutionary Computation, ACM, pp. 361–368.
27. Timmis, J. and Neal, M. (2001) A resource limited artificial immune system for data analysis. *Knowledge-Based Systems*, **14** (3), 121–130.
28. Neal, M. (2002) An artificial immune system for continuous analysis of time-varying data, in Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS), vol. 1, pp. 76–85.
29. Nasaroui, O., Gonzalez, F., and Dasgupta, D. (2002) The fuzzy artificial immune system: Motivations, basic concepts, and application to clustering and web profiling, in Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on, vol. 1, IEEE, pp. 711–716.
30. de Castro, L. and Von Zuben, F. (2001) AINET: An artificial immune network for data analysis. *Data Mining: A Heuristic Approach*, **1**, 231–259.
31. Matzinger, P. (2001) Essay 1: The danger model in its historical context. *Scandinavian Journal of Immunology*, **54** (1–2), 4–9.
32. Aickelin, U. and Cayzer, S. (2002) The danger theory and its application to artificial immune systems. Artificial Immune Systems, pp. 141–148.
33. Prieto, C., Nino, F., and Quintana, G. (2008) A goalkeeper strategy in robot soccer based on danger theory, in Evolutionary Computation, 2008. CEC 2008. (IEEE World Congress on Computational Intelligence). IEEE Congress on, IEEE, pp. 3443–3447.
34. Zhang, C. and Yi, Z. (2010) A danger theory inspired artificial immune algorithm for on-line supervised two-class classification problem. *Neurocomputing*, **73** (7–9), 1244–1255.
35. Chao, Y., Yiwen, L., and Aolin, L. (2011) The danger sensed method by feature changes. *Energy Procedia*, **13**, 4429–4437.
36. Secker, A., Freitas, A., and Timmis, J. (2003) A danger theory inspired approach to web mining. Artificial Immune Systems, pp. 156–167.
37. Aickelin, U., Bentley, P., Cayzer, S., Kim, J., and McLeod, J. (2003) Danger theory: The link between AIS and IDS? Artificial Immune Systems, pp. 147–155.
38. Tan, Y. and Zhu, Y. (2010) Advances in anti-spam techniques. *CAAI Transactions on Intelligent Systems*, **5** (3), 189–201.
39. Tan, Y., Deng, C., and Ruan, G. (2009) Concentration based feature construction approach for spam detection, in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, IEEE, pp. 3088–3093.

40. Ruan, G. and Tan, Y. (2010) A three-layer back-propagation neural network for spam detection using artificial immune concentration. *Soft Computing*, **14** (2), 139–150.
41. Zhu, Y. and Tan, Y. (2010) Extracting discriminative information from e-mail for spam detection inspired by immune system, in Evolutionary Computation (CEC), 2010 IEEE Congress on, IEEE, pp. 1–7.
42. Zhu, Y. and Tan, Y. (2011) A local concentration based feature extraction approach for spam filtering. *Information Forensics and Security, IEEE Transactions on*, **6** (2), 486–497.
43. Yang, Y. and Pedersen, J. (1997) A comparative study on feature selection in text categorization. International Conference on Machine Learning, pp. 412–420.
44. Wang, W., Zhang, P., Tan, Y., and He, X. (2011) An immune local concentration based virus detection approach. *Journal of Zhejiang University-Science C*, **12** (6), 443–454.
45. Wang, W., Zhang, P., and Tan, Y. (2010) An immune concentration based virus detection approach using particle swarm optimization, in *Advances in Swarm Intelligence*, Springer, pp. 347–354.
46. Dasgupta, D., Yu, S., and Majumdar, N. (2003) Mila multilevel immune learning algorithm, in *Genetic and Evolutionary Computation, ECCO 2003*, Springer, pp. 183–194.
47. Wang, W., Gao, S., and Tang, Z. (2008) A complex artificial immune system, in Natural Computation, 2008. ICNC'08. Fourth International Conference on, Vol. 6, IEEE, pp. 597–601.
48. Zhang, Y. and Hou, C. (2003) A clone selection algorithm with niching strategy inspiring by biological immune principles for change detection, in Intelligent Control. 2003 IEEE International Symposium on, IEEE, pp. 1000–1005.
49. Li, Z., Zhang, Y., and Tan, H. (2007) An efficient artificial immune network with elite-learning, in Natural Computation, 2007. ICNC 2007. Third International Conference on, Vol. 4, IEEE, pp. 213–217.
50. de Castro, P. and Von Zuben, F. (2009) BAIS: A bayesian artificial immune system for the effective handling of building blocks. *Information Sciences*, **179** (10), 1426–1440.
51. Chao, R. and Tan, Y. (2009) A virus detection system based on artificial immune system, in Computational Intelligence and Security, 2009. CIS'09. International Conference on, IEEE, vol. 1, pp. 6–10.
52. Guo, Z., Liu, Z., and Tan, Y. (2004) An nn-based malicious executables detection algorithm based on immune principles, in *Advances in Neural Networks-ISNN 2004*, Springer, pp. 675–680.
53. Wang, W., Zhang, P., Tan, Y., and He, X. (2009) A hierarchical artificial immune model for virus detection, in Computational Intelligence and Security, 2009. CIS'09. International Conference on, IEEE, vol. 1, pp. 1–5.
54. Androutsopoulos, I., Paliouras, G., and Michelakis, E. (2004) *Learning to filter unsolicited commercial e-mail*, "DEMOKRITOS," National Center for Scientific Research.
55. Siefkes, C., Assis, F., Chhabra, S., and Yerazunis, W.S. (2004) Combining winnow and orthogonal sparse bigrams for incremental spam filtering, in *Knowledge Discovery in Databases: PKDD 2004*, Springer, pp. 410–421.
56. Oda, T. and White, T. (2003) Developing an immunity to spam, in *Genetic and Evolutionary Computation–GECCO 2003*, Springer, pp. 231–242.
57. Sahami, M., Dumais, S., Heckerman, D., and Horvitz, E. (1998) A bayesian approach to filtering junk e-mail, in Learning for Text Categorization: Papers from the 1998 workshop, vol. 62, pp. 98–105.
58. Çıltık, A. and Gungör, T. (2008) Time-efficient spam e-mail filtering using n-gram models. *Pattern Recognition Letters*, **29** (1), 19–33.
59. Drucker, H., Wu, S., and Vapnik, V.N. (1999) Support vector machines for spam categorization. *Neural Networks, IEEE Transactions on*, **10** (5), 1048–1054.
60. Tan, Y. and Wang, J. (2004) A support vector network with hybrid kernel and minimal vapnik-chervonenkis dimension. *IEEE Trans. On Knowledge and Data Engineering*, **26** (2), 385–395.
61. Andreas, J. and Tan, Y. (2011) Swarm intelligence for non-negative matrix factorization. *International Journal of Swarm Intelligence Research*, **2** (4), 12–34.
62. Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Sakkis, G., Spyropoulos, C.D., and Stamatopoulos, P. (2000) Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach. arXiv preprint cs/0009009.

63. Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Spyropoulos, C.D., and Stamatopoulos, P. (2003) A memory-based approach to anti-spam filtering for mailing lists. *Information Retrieval*, **6** (1), 49–73.
64. Clark, J., Koprinska, I., and Poon, J. (2003) A neural network based approach to automated e-mail classification, in Web Intelligence, IEEE/WIC/ACM International Conference on, IEEE Computer Society, pp. 702–702.
65. Wu, C.H. (2009) Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks. *Expert Systems with Applications*, **36** (3), 4321–4330.
66. Guzella, T.S., Mota-Santos, T.A., Uchoa, J.Q., and Caminhas, W.M. (2008) Identification of spam messages using an approach inspired on the immune system. *Biosystems*, **92** (3), 215–225.
67. Zhu, Y. and Tan, Y. (2011) A danger theory inspired learning model and its application to spam detection. Advances in Swarm Intelligence, pp. 382–389.
68. Ruan, G. and Tan, Y. (2007) Intelligent detection approaches for spam, in Natural Computation, 2007. ICNC 2007. Third International Conference on, IEEE, vol. 3, pp. 672–676.
69. Tan, Y. and Ruan, G. (2014) Uninterrupted approaches for spam detection based on svm and ais. *International Journal of Computational Intelligence*, **1** (1), 1–26.
70. Mi, G., Zhang, P., and Tan, Y. (2013) A multi-resolution-concentration based feature construction approach for spam filtering, in The International Joint Conference on Neural Networks (IJCNN 2013), IEEE, vol. 1, pp. 1–8.
71. He, W., Mi, G., and Tan, Y. (2013) Parameter optimization of local-concentration model for spam detection by using fireworks algorithm, in *The Fourth International Conference on Swarm Intelligence (ICSI 2013)*, vol. 1, Springer, pp. 439–450.
72. Huang, X., Tan, Y., and He, X. (2011) An intelligent multi-feature statistical approach for discrimination of driving conditions of hybrid electric vehicle. *IEEE Transactions on Intelligent Transportation Systems*, **12** (2), 453–465.
73. Gu, S., Tan, Y., and He, X. (2013) Recent-biased learning for time series forecast. *Information Science*, **237** (10), 29–38.
74. Gu, S., Tan, Y., and He, X. (2010) Laplacian smoothing transform for face recognition. *Science China (Information Science)*, **53** (12), 2415–2428.
75. Andreas, J. and Tan, Y. (2013) Efficient euclidean distance transform algorithm of binary images in arbitrary dimensions. *Pattern Recognition*, **46** (1), 230–242.