Negative Selection Algorithm :A Survey

Delona C Johny,Haripriya P V
Department of Information Technology
Govt.Engineering college
Bartonhill, Trivandrum

Anju J S Assistant
Professor
Department of Information Technology
Govt.Engineering college Bartonhill,
Trivandrum

Abstract

Interest in the use of biology as a source of inspiration for solving computational problems is greatly increasing now a days.This area of research is referred to Natural Computing. This paper presents the Artificial immune system algorithms such as NSA,Positive selection algorithm,Clonal selection algorithm etc. Negative Selection Algorithm (NSA) is a key concept of Artificial Immune System (AIS) . The Negative Selection algorithm is inspired by the self-nonself discrimination behavior observed in the mammalian acquired immune system.The starting point of this algorithm is to produce a set of self strings, that define the normal state of the system. The task then is to generate a set of detectors, that only bind/recognize the complement of self string. These detectors can then be applied to new data in order to classify them as being self or nonself .The Negative Selection Algorithm was designed for change detection, novelty detection, intrusion detection and similar pattern recognition and two-class classification problem domains.The novelty of this survey is to study the NSA, Data representation in NSA ,Matching techniques,Matching rule for string and real valued representation.

KeyWords-AIS,NSA

I. Introduction

Artificial Immune Systems (AIS) are adaptive systems that is inspired by theoretical immunology and immune functions, principles and models, which are applied to problem solving[1]. Among various mechanisms in the immune system that are explored for AIS, negative selection, immune network model, and clonal selection are still the most discussed models. In the complex immune system,one of the major mechanisms is discrimination between self and nonself. Computational imitation of self/nonself discrimination is Artificial negative Selection.It first designed as a change detection method.It is modeled off the T -cell maturing process that happens in the thymus. T -cells are first assembled with a pseudo-random genetic rearrangement process. T Cells that recognize self cells are eliminated before the rest are deployed into the immune system to recognize and attack outside pathogens.

II. Biological Immune System

The Biological Immune System (BIS), an integral part of the vertebrate immunity, is a dynamic, powerful, intelligent, and interconnection of different components of the body, working in totality to fight, defend, and prevent pathogenic organisms entrance into the body. BIS functions are protection from foreign invaders, and maintaining homeostasis.A role of the immune system is to protect our bodies from infectious agents such as viruses, bacteria, fungi and other parasites.to detect and eliminate pathogens efficiently,the immune system possesses a multi layered protection,detection and elimination architecture.There are two basic types of immunity, innate and adaptive. The static system that identifies and eliminates definite harmful organisms is known as innate immune system.The system that remembers unknown foreign cells and react with them is known as adaptive immune system. The adaptive immune system is a combination of atom cells spread all over the body.There are two type of lymphocytes are present in the cells , that is T-cells and B-cells.These cells recognize and destroy specific substances which are entered into our body. Any molecule or substance that is capable of eliciting an immune response from the lymphocytes is called an antigen or immunogen. A primary response is produced when the human immune system encounters an antigen for the first time. Large numbers of antibodies are created by the immune system

in response to the antigen. These antibodies eliminates the antigen from human body.when the same antigen is encountered again after a period of days ,secondary response is produced by the immune system . Secondary immune response is specific to the antigen that first initiated the immune response and causes a very rapid growth in the quantity of Bcells and antibodies.When similar antigen encountered, a new immunity does not need to be built up, it is already occur.At that time a faster response is attributed to memory cells remaining in the immune system.

III. Artificial immune system

Artificial Immune Systems (AIS) are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving[1]. Our body maintains a large number of immune cells-called lymphocytes, which circulate throughout the body. T cells and B cells are the two types of lymphocytes. T cells and B cells stimulate and suppress each other in certain ways that lead to the stabilization of the network.When affinities exceed a certain threshold,these two cells are connected and the strength of the connection is directly proportional to the affinity they share. AISs have been solve real-world complex problems in the areas of cyber security, robotics, fraud detection, and anomaly detection.Artificial Immune System provides high performance on these areas.Uniqueness,noise tolerance, recognition of foreigners, reinforcement learning and memory, and pattern recognition are the main qualities of AIS. Because of these qualities AIS provides good solutions for complex problems.

A. Negative Selection Algorithm

Negative Selection Algorithm: Inspired by the positive and negative selection processes that occur in the thymus during the maturation of the cells . Negative selection refers to the identification and deletion (apoptosis) of self-reacting cells, that is T cells that may select for and attack self tissues. This algorithm are used for classification and pattern recognition problem domains . For example, in the case of an anomaly detection domain the algorithm prepares a set of exemplar pattern detectors trained on normal (non-anomalous) patterns that model and detect unseen or anomalous patterns. The NSA got its motivation from the negative selection process in natural immune system . In thymus, if a T-cell detects any self-cell, it is removed and then immune functionality is performed in a T-cell maturation process. Mainly, it is utilized for anomaly detection. This algorithm creates a set of detectors, which includes self-strings only. Later, this detector set is used for the anomaly detection. There are two main steps in NSA algorithm . The first step is censoring in which self-strings and randomly generated strings are matched.The strings that are matched are rejected.The strings that do not get matched are moved to the detector set. In the second step, protected strings are matched with those in the detector set. The strings that get matched are identified as nonself and the rest are matched again.

B. Positive Selection Algorithm

Positive selection is an area in immunology where T cells are selected for their lack of recognition of self .The T cell receptor (TCR) is similar to that of the B cell,except that it is not secreted as antibody and does not bind directly to antigens. Rather it binds to peptides presented in a complex with a specialised self molecule called the Major Histocompatibility Complex (MHC). Recognition of the self-MHC:peptide complex means that T cells undergo a slightly different process to B cells. While T cells are selected for their lack of recognition of self (negative selection as in B cells), they must also be able to recognise the self MHC molecule.

C. Clonal Selection Algorithm

Clonal selection algorithms is an algorithm inspired by the clonal selection theory of acquired immunity that explains how B and T lymphocytes improve their response to antigens over time called affinity maturation.Clonal selection is the method of antigen recognition, cell propagation, and discrimination into the memory cell. Clonal immune characteristic and feature are used for creating many AIS algorithms.The theory specifies that the organism have a pre-existing pool of heterogeneous (individually unique) antibodies. This antibodies can recognise all antigens with some level of specificity. The cells can replicate and produce more cells,When an antigen is matched with an antibody,. During the cell proliferation stage, genetic mutations occur in the clone of cells . This allows the binding ability of the cells to improve with time and exposure to the antigen. According to Darwinian microcosm,the best fittest cells are selected for survival, and genetic mutation provides cell variation.

C-cells basic model and its resulting antibodies could perform as a good primary metaphor. B-cells create specially configured antibodies that are diverse and get excited and stimulated when they meet a foreign antigen.Moreover, resulting clones of B-cells differ in their receptor configuration for performing local biological search to locate the best fitting receptor.The Clonal Selection Algorithm, also called CSA in [10], and renamed to CLONALG in [11] is said to be inspired by the elements of the clonal selection theory such as:

- Maintenance of a memory set
- Selection and cloning of antibodies
- Affinity maturation (mutation)
- Re-selection of clones
- Generation and maintenance of diversity

To develop a memory pool of antibodies is the goal of this algorithm .An antibody and antigen represents an element of a solution and an element or evaluation of the problem space respectively.The algorithm provides two mechanisms for searching for the desired final pool of memory antibodies. The first is a local search provided via affinity maturation of cloned antibodies. More clones are produced for better matched (selected) antibodies, though the scope of the local search is inversely proportional to the selected antibodies rank. This allows the antibodies with low specificity with the antigen to maturate. The second search mechanism provides a global scope. This mechanism involves randomly generated antibodies which are inserted into the population to further increase the diversity .

Steps of clonal Selection Algorithm:

- Initialisation:The first step of the CLONALG technique is initialisation, which involves preparing an antibody pool of fixed size N. This pool is then partitioned into two components, a memory antibody section m that eventually becomes representative of the algorithms solution and a remaining antibody pool r used for introducing additional diversity into the system.
- Loop:The algorithm then proceeds by executing a number of iterations . A single round of iteration is referred to as a generation. The number of generations G the system executes is user configurable, though the system can use a problem specific stopping condition.
  - Select Antigen : A single antigen is selected at random without replacement from the pool of antigens
  - Exposure : The system is exposed to the selected antigen. Affinity values are calculated for all antibodies against the antigen. Affinity is a measure of similarity, and is problem dependent. It is common to use Hamming distance.
  - Selection : A set of n antibodies are selected from the entire antibody pool that have the highest affinity with the antigen.
  - Cloning: The set of selected antibodies are then cloned in proportion to their affinity (rank based).
  - Affinity Maturation (mutation): The clone (set of duplicate antigens) are then subjected to an affinity maturation process to better match the antigen in question. Here, the degree of maturation is inversely proportional to their parent?s affinity (rank based), meaning that the greater the affinity, the lower the mutation.
  - Clone Exposure : The clone is then exposed to the antigen, and affinity measures are calculated.
  - Candidature :The antibody or antibodies with the highest affinity in the clone are then selected as candidate memory antibodies for placement into m. If the affinity of a candidate memory cell is higher than that of the highest stimulated antigen from the memory pool m, then it replaces said antigen. Group replacements occur in a similar or in a batched manner.
  - Replacement: Finally, the d individuals in the remaining r antigen pool with the lowest affinity are replaced with new random antibodies.
- Finish:After the completion of the training regime, the memory m component of the antigen pool is then taken as the algorithms solution. Depending on the problem domain, the solution may be a single best individual antigen or the collective of all antigens in the pool.

IV. Negative Selection algorithm

Negative Selection Algorithm (NSA), is one of the most popular AIS models that has grabbed the eyeballs of researchers. Forrest et al. proposed NSA, based on the principles of self/non-self discrimination in the immune system .It drew motivation from the fact that negative selection of T cells in the thymus and worked around the immune systems philosophy to identify unknown antigens/non-self, while not reacting to self-cells. It builds a self prole by recognizing only normal net-

work patterns as self while other patterns as non-self. This built prole very easily detects non-self patterns and marked them as non-self/anomalous. If any sample matches any self-sample then it is removed so that it does not become a detector. Those patterns who fail to match self pattern becomes detectors and signies non-self. Later on, these detectors are used to detect anomalies. These detectors monitor incoming patterns and if detector matches with any new pattern, then this symbolize detection of an anomaly.

The NSA was inspired by the negative selection process occurring within the NIS and is conceptually illustrated in Figure 1. The main concept behind the NSA is to generate a set of candidate detectors.

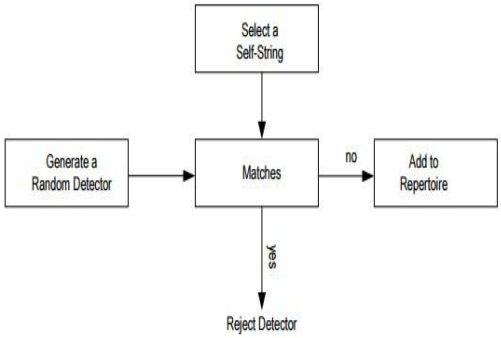

Fig. 1. Negative Selection Algorithm

A. Data Representation and Matching Techniques

For negative selection algorithm, strings (or binary) representation and real-valued representation has been widely used. Also, there is hybrid of both data representations which consist of different data types such as integer, real value, categorical information, boolean value, text information, etc. Since, any data is eventually converted in binary bits, therefore, researches focused on binary representation and one of the most commonly adopted coding scheme in AIS. Forrest et al. introduced rst theory with binary strings because it constituted a nite space that made problem space analysis easy (Forrest et al., 1994). NSA breaks 32-bit string into eight substrings as antigen and antibody. Although not many immune features were employed in this scheme, however it showed the feasibility.

B. Matching Rule for Strings Representation

- R-Contiguous Matching Rule

In r-contiguous matching rule, break string of arbitrary length into shorter segments of predefined length.Matching requirement is defined as r contiguous matching symbols in corresponding positions.

- R-chunk Matching Rule

The r-chunk matching rule was first proposed by Balthrop and is an improved variant of the r-contiguous matching rule developed by Percus . The r-contiguous matching rule was one of the earliest rules which focused on the biological immune system as a model and abstracts the binding between antibody and antigen.Informally, two elements, with the same length, match under r-contiguous rule, if at least r contiguous characters are identical. The r-chunk matching rule achieves the highest matching performance compared with the other matching rules over the binary alphabet. Since matching performance is strongly influenced by the number of generatable detectors, we focus on the number of generatable r-chunk detectors over arbitrary alphabet sizes.

- Hamming Distance

Hamming distance is a matching rule for string representation. The hamming distance of two strings,s1 and s2,is defined as the minimum number of point mutations required to change s1 into s2,where a point mutation is to change a letter,to insert a letter,or to delete a letter. Hamming distance, also called Levenshtein distance,is a generalization of hamming distance.

- Hamming shape space

The common representation is defined over Hamming shape-space in which a universe consists of all strings of fixed length over a finite alphabet. The original NSA employed binary strings to represent patterns and detectors and an r-contiguous-bit(rcb) matching rule was used to determine the matching between patterns and detectors.Under the rcb matching rule,a pattern is matched by detector if they have the same bits in atleast r contiguous position.

C. Matching Rule for Real-Valued Representation

- Euclidean Distance

The distance defined over some of the elements of the vector. It is equivalent to the distance projected to a lower-dimensional space degraded from the original

space.Euclidean distance used in a matching rule is not calculated over all the elements of the vector. Only some elements are used to calculate Euclidean distance over a lower-dimensional space, similar to partial matching in string representation that only uses some bits.Ji and D. Dasgupta measure can be chosen contiguously as in rcb rule or randomly. In both cases, the chosen positions need to match between the two points whose distance is calculated. Permutation mask and crossover closure concepts from string representation can be extended to the way these elements are chosen.

- Manhattan Distance

Manhattan Distance is the distance between two points measured along axes at right angles. For example, given two points p1 and p2 in a two-dimensional plane at (x1, y1) and (x2, y2) respectively, the Manhattan distance between p1 and p2 is given by —x1 - x2— + —y1 - y2—.

- Minkowski Distance

The Minkowski distance can be considered as a generalization of both the Euclidean distance and the Manhattan distance.It is a metric in a normed vector space.The Minkowski distance of order p between two points

$$X = (x_1, x_2, x_n) \text{and} Y = (y_1, y_2..y_n) \in R^n$$

is defined as:

$$\left(\sum_{i=1}^{n} |x_i - yi|^p\right)^{1/p}$$

V. Conclusion

In surveying on Negative selection algorithm based on Artificial immune system,NSA gaining the popularity and attraction.Majority of works proposed are based on variation of NSA.Its performance is based on the interaction between the detector generation algorithm and matching technique adopted for use.Matching techniques and rules used are different in different works.Relying on the type of data representation, either for strings or real-valued, the proper detection algorithm must be assigned.

References

[1] Praneet Saurabh , Bhupendra Verma, “An efficient proactive artificial immune system based anomaly detection and prevention system,” Expert Systems With Applications ELSEVIER Journal, Vol. 60, pp. 311 320, 2016.

[2] Stephanie Forrest, Alan S. Perelson, Lawrence Allen, Rajesh Cherukur, Self-Nonself Discrimination in a Computer,In Proceedings of IEEE Symposium on Research in Security and Privacy, 1994.

[3] Tamer F. Ghanem , Wail S. Elkilani , Hatem S. Ahmed , and Mohiy M. Hadhoud, “An enhanced Hybrid Anomaly-based Detection Approach ,” ICAIME, Vol. 60, pp. 169-177 , 2014.

[4] Mohammad Reza Abdolahnezhad, Touraj Banirostam, “Improved Negative Selection Algorithm for Email Spam Detection Application ,” International Journal of Advanced Research in Electronics and Communication Engineering, Vol. 5, 2016.

[5] Praneet Saurabh , Bhupendra Verma, “A Novel Immunity Inspired Approach for Anomaly Detection ,” International Journal of Computer Applications, Vol. 94, 2014.

[6] Leandro Nunes de Castro, Fernando Jos Von Zuben, “ARTIFICIAL IMMUNE SYSTEMS: PART I BASIC THEORY AND APPLICATIONS,” Technical Report TR DCA 01/99, 1999.

[7] Dipankar Dasgupta, Senhua Yu, Fernando Nino, “Recent Advances in Articial Immune Systems: Models and Applications,” Applied Soft Computing ELSEVIER Journal, 2011.

[8] Furong Liu, Qiaoling Wang, Xiaozhi Gao, “Survey of Artificial Immune System,” 2005.

[9] Mark Muchmore , “Artificial Immune Systems : A Survey ,” University of Northern Iowa Cedar Falls, IA 50613 , 2005.

[10] Leandro N. de Castro and Fernando J. Von Zuben, ”The Clonal Selection Algorithm with Engineering Applications,” GECCO 2000, Workshop on Artificial Immune Systems and Their Applications, Las Vegas, USA, pp. 36-37, 2000

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