Investigating an Artificial Immune System to Strengthen the Promoter Region Structure Prediction and Promoter Region Identification using Cellular Automata Classifier

P. Kiran Sree[1], I. Ramesh Babu[2], N.S.S.N. Usha Devi[3]

Abstract

Genes carry the instructions for making Promoter Regions that are found in a cell as a specific sequence of nucleotides that are found in DNA molecules. But, the regions of these genes that code for Promoter Regions may occupy only a small region of the sequence. Identifying the promoter regions play a vital role in understanding these genes. In this paper we have explored an artificial immune system can be used to strengthen and identify the Promoter Region coding regions in genomic DNA system in changing environments, and Cellular Automata (CA) technique for Promoter Region structure prediction of small alpha/beta Promoter Regions using Rosetta. From an initial round of Rosetta sampling, we learn properties of the energy landscape that guide a subsequent round of sampling toward lower-energy structures. Three different approaches to improve tertiary fold prediction using the genetic algorithm are discussed: (i) Refinement of the search strategy, (ii) combination of prediction and experiment and (iii) inclusion of experimental data as selection criteria into the genetic algorithm. It has been developed using a slight variant of genetic algorithm. Good classifier can be produced especially when the number of the antigens is increased. However, an increase in the range of the antigens had somehow affected the fitness of the immune system. Experimental results confirm the scalability of the proposed Artificial Immune System Fuzzy Multiple Attractor Cellular Automate (AIS FMACA) based classifier to handle large volume of datasets irrespective of the number of classes, tuples and attributes. We note an increase in accuracy of more than 5.2%, over any existing standard algorithms for addressing this problem. This was the first algorithm to identify Promoter Region coding regions in mixed and non overlapping exon-inton boundary DNA sequences also. The accuracy of predicting the structure of Promoter Regions was also found comparable.

Keywords : Cellular Automata (CA), unsupervised learning Classifier, Genetic Algorithm (GA), artificial immune system, Coding Regions, Fuzzy Multiple Attractor Cellular Automata (FMACA), Pattern Classifier, Promoter Region Structure Prediction

1. Introduction

Many of the challenges in biology are now challenges in computing. Bioinformatics, the application of computational techniques to analyze the information associated with bimolecules on a large scale, has now firmly established itself as a discipline in molecular biology. Bioinformatics is a management information system for molecular biology. Bioinformatics encompasses everything from data storage and retrieval to the identification and presentation of features within data, such as finding genes within DNA sequence, finding similarities between sequences, structural predictions. Analyzing the coding regions is not the scope of the project.

[1]Associate Professor, Department of CSE, SRKIT, Vijayawada, email : profkiran@yahoo.com.

[2]Professor, CSE, Acharya Nagarjuna University, Guntur.

[3]Graduate Student of JNTU, Hyderabad.

For better understanding of the specified objectives, we presented CA, FCA, AIS fundamentals in Section II and Section III .Section IV presents the design of AIS FMACA based pattern classifier [3], [7] as well as rule formation and chromosome representation. In Section V, we address the problem of Promoter Region coding region identification [11], [12] in DNA sequences. In order to validate the design of proposed model, experimental results are also reported in this section.

2. Cellular Automata (CA) and Fuzzy Cellular Automata (FCA)

A CA [4], [5], [6], consists of a number of cells organized in the form of a lattice. It evolves in discrete space and time. The next state of a cell depends on its own state and the states of its neighboring cells. In a 3-neighborhood dependency, the next state qi (t + 1) of a cell is assumed to be dependent only on itself and on its two neighbors (left and right), and is denoted as

$$qi(t+1) = f(q^{i-1}, qi(t), qi+1(t)) \quad E(1)$$

where qi (t) represents the state of the i^{th} cell at t^{th} instant of time, f is the next state function and referred to as the rule of the automata. The decimal equivalent of the next state function, as introduced by Wolfram, is the rule number of the CA cell [9],[10],[11]. In a 2-state 3-neighborhood CA, there are total 256 distinct next state functions.

2.1 FCA Fundamentals

AFCA [2], [6] is a linear array of cells which evolves in time. Each cell of the array assumes a state qi, a rational value in the interval [0, 1] (fuzzy states) and changes its state according to a local evolution function on its own state and the states of its two neighbors. The degree to which a cell is in fuzzy states 1 and 0 can be calculated with the membership functions. This gives more accuracy in finding the coding regions.In a FCA, the conventional Boolean functions are AND , OR, NOT.

2.2 Dependency Matrix for FCA

Rules defined in equations 1, 2 should be represented as a local transition function of FCA cell. That rules are converted into matrix form for easier representation of chromosomes [16].

Table 1: FA Rules

Non-complemented Rules		Complemented Rules	
Rule	Next State	Rule	Next State
0	0	255	1
170	q_{i+1}	85	$\overline{q_{i+1}}$
204	q_i	51	$\overline{q_i}$
238	$q_i + q_{i+1}$	17	$\overline{q_i + q_{i+1}}$
240	q_{i-1}	15	$\overline{q_{i-1}}$
250	$q_{i-1} + q_{i+1}$	5	$\overline{q_{i-1} + q_{i+1}}$
252	$q_{i-1} + q_i$	3	$\overline{q_{i-1} + q_i}$
254	$q_{i-1} + q_i + q_{i+1}$	1	$\overline{q_{i-1} + q_i + q_{i+1}}$

Example 1: A 4-cell null boundary hybrid FCA with the following rule

< 238, 254, 238, 252 > (that is, < (qi+qi+1), (qi-1+qi+qi+1), (qi + qi+1), (qi-1 + qi) >) applied from left to right, may be characterized by the following dependency matrix. While moving from one state to other, the dependency matrix indicates on which neighboring cells the state should depend. So cell 254 depends on its state, left neighbor, and right neighbor fig (1). Now we represented the transition function in the form of matrix. In the case of complement [5[,[6],[8],FMACA we use another vector for representation of chromosome.

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Figure 1 : Matrix Representation

2.3 Transition From One State To Other

Once we formulated the transition function, we can move form one state to other. For the example 1 if initial state is P (0) = (0.80, 0.20, 0.20, 0.00) then the next states will be

P (1) = (1.00 1.00, 0.20, 0.20),
P (2) = (1.00 1.00, 0.40, 0.40),
P (3) = (1.00 1.00, 0.80, 0.80),
P (4) = (1.00 1.00, 1.00, 1.00).

FMACA Based Pattern Classifier

An n-cell FMACA with k-attractor basins can be viewed as a natural classifier. It classifies a given set of patterns into k distinct classes, each class containing the set of states in the attractor basin.

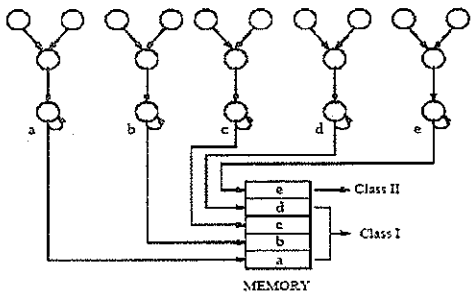


Figure 2 : FMACA Based Classification Strategy with 5 Attractor Basins Classifying the Elements into Two Classes.

Note : (i) An attractor basin covers the elements belonging to one class only.
(ii) Each attractor points to the memory location that stores the class information.

Suppose, we want to design a FMACA based pattern classifier to classify a training set S = {S1, S2, · · · , SK} into K number of classes. First, a FMACA with k ($k \geq K$) number of attractor basins is generated. The training set S gets distributed into k attractor basins (nodes). Let S be the set of elements in an attractor basin. If S belongs to only one class, then label that attractor basin as that class. Otherwise, this process is repeated recursively for each attractor basin (node) until all the patterns in each attractor basin belong to only one class.

3. Artificial immune systems

Artificial immune systems are motivated by the theory of immunology. The biological immune system functions to protect the body against pathogens or antigens that could potentially cause harm. It works by producing antibodies that identify, bind to, and finally eliminate the pathogens. Even though the number of antigens is far larger than the number of antibodies, the biological immune system has evolved to allow it to deal with the antigens. The immune system will learn the criteria of the antigens so that in future it can react both to those antigens it has encountered before as well as to entirely new ones. In 2002, de Castro and Timmis [17], suggested that for a system to be characterized as an artificial immune system, it has to embody at least a basic model of an immune component (e.g. cell, molecule, organ), it has to have been designed using the ideas from theoretical and/or experimental immunology.

The human body is protected against foreign invaders by a multilayered system[15],[16]. The immune system is composed of physical barriers such as the skin and respiratory system, physiological barriers such as destructive enzymes and stomach acids and the actual immune system, which has two complementary parts, the innate and adaptive immune systems. The innate immune system is an unchanging mechanism that detects and destroys certain invading organisms, whilst the adaptive[17],[18] (or acquired) immune system responds to previously unknown foreign cells and builds a response that can remain in the body over a long period of time. Of most interest to us is the adaptive immune system,

which is composed of a number of different agents performing different functions at a number of different locations in the body. The precise interaction of these agents is still a topic for debate [13], [16]. Two of the most important cells in this process are two types of white blood cells, called T-cells and B-cells. Both of these originate in the bone marrow (hence the 'B'), but T-cells pass on to the thymus to mature (hence the 'T'), before they circulate the body in the blood and lymphatic vessels.

T-cells come in three types; T-helper cells which are essential to the activation of Bcells, Killer T-cells which bind to foreign invaders and inject poisonous chemicals into them causing their destruction, and suppressor T-cells which inhibit the action of other immune cells thus preventing allergic reactions and autoimmune diseases. B-cells are responsible for the production and secretion of antibodies, which are specific Promoter Regions that bind to the antigen. Each B-cell can only produce one particular antibody. The antigen is found on the surface of the invading organism and the binding of an antibody to the antigen is a signal to destroy the invading cell. As can be gleaned from the brief explanations above, there is more than one mechanism at work in the human immune system [8], [12], [13], and [16] However let us now concentrate on the essential process exploited in most AIS: The matching between antigen and antibody which subsequently leads to increased concentrations (proliferation) of more closely matched antibodies. In particular, the negative selection mechanism and the 'clonal selection' and 'somatic hyper mutation' theories are primarily used in AIS models.

3.1. The Human Immune System

The human immune system is a complex system of cells, molecules and organs that represent an identification mechanism capable of perceiving and combating dysfunction from our own cells and the action of exogenous infectious microorganisms. The human immune system protects our bodies from infectious agents such as viruses, bacteria, fungi and other parasites. Any molecule that can be recognized by the adaptive immune system is known as an antigen(Ag).The basic component of the immune system is the lymphocytes or the white blood cells. Lymphocytes exist in two forms, B cells and T cells. These two types of cells are rather similar, but differ with relation to how they recognize antigens and by their functional roles, B-cells are capable of recognizing antigens free in solution, while T cells require antigens to be presented by other accessory cells. Each of this has distinct chemical structures and produces many Y shaped antibodies form its surfaces to kill the antigens.

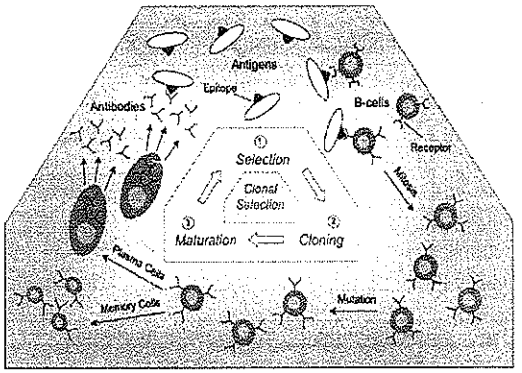


Figure 3 : Immune System

Fig 2 shows a depiction of the immune response when antigens invade the body. B-cells that are able to bind to antigen become stimulated by Helper T-cells (not shown). Then they begin the repeated process of cell division (or mitosis). This leads to the development of clone cells with the same or slightly mutated genetic makeup. B-cells with the same genetic makeup will have identical receptors. However some B-cells will become mutated, and thus have slightly modified receptors. This results in the creation of a new B-cell that might have an increased

affinity for the antigen. This phenomenon is called clonal selection because it is the antigen that essentially selects which B-cells are to be cloned [6]. This will eventually lead to the production of plasma cells and memory cells. Plasma cells mass produce and secrete soluble B-cell receptors that are now called antibodies. These antibodies bind to other antigen to neutralize and mark them for destruction by other cells. Some memory cells can survive for long periods of time by themselves, while other memory cells form a network of similar cells to maintain a stable population. This helps to keep the immune system from extinguishing itself once the antigen has been completely removed.

3.2 Fuzzy logic with Artificial Immune System

This work focuses on one kind of AIS in-spired by the clonal selection principle of the biological immune system. In essence, when an immune system detector (a lymphocyte) has a high affinity (a high degree of matching) with an antigen (invader microorganism), this recognition stimulates the proliferation and differentiation of cells that produce antibodies. This process, called clonal expansion (because new cells are produced by cloning and mutating existing cells), produces a large population of antibodies targeted for that antigen.

Fuzzy systems are very effective in expressing the natural ambiguity and subjectivity of human reasoning. Membership functions determine to which degree a given object belongs to a fuzzy set. In a fuzzy system this degree of membership varies from 0 to 1. Membership functions can take different forms, varying from the simplest ones (triangular functions) to more complex functions (parameterized by the user). n a classification problem with n attributes, fuzzy rules can be written as:, where is an n-dimensional pattern vector, (i=1,…,n) is the i-th attribute's linguistic value (e.g. small or large), C is the class predicted by the rule, and N is the number of fuzzy if-then rules. In addition, it has been suggested that an AIS based on the clonal selection principle, called CLONALG, can be used for classification in the context of pattern recognition [6], although originally proposed for other tasks.

However, unlike the AIS algorithm proposed in this paper, neither AIRS nor CLONALG discovers comprehensible IF-THEN rules. Hence, neither of those two algorithms addresses the data mining goal of discovering comprehensible, interpret able knowledge (see Introduction). Also, they do not discover fuzzy knowledge, unlike the algorithm proposed in this paper. AIS for discovering IF-THEN rules are proposed in [9]. Unlike the algorithm proposed in this paper, that work is based on ex-tending the negative selection algorithm with a genetic algorithm. We have avoided the use of the negative selection algorithm because this kind of AIS method has some conceptual problems in the context of the classification task, as discussed in [10]. Also, again that work does not discover fuzzy rules. Fuzzy AIS is proposed in, however, that work addresses the task of clustering, which is very different from the task of classification addressed in this paper. To the best of our knowledge, the algorithm proposed in this paper is the first AIS for discovering fuzzy classification rules based on the clonal selection principle.

After discussing the clonal selection principle and the affinity maturation process, the development of the clonal selection algorithm (CSA) is straightforward. The main immune aspects taken into account were: maintenance of the memory cells functionally disconnected from the repertoire, selection and cloning of the most stimulated cells, death of non-stimulated cells, affinity maturation

and re-selection of the clones with higher affinity, generation and maintenance of diversity, hyper mutation proportional to the cell affinity.

3.3 The Proposed Algorithm

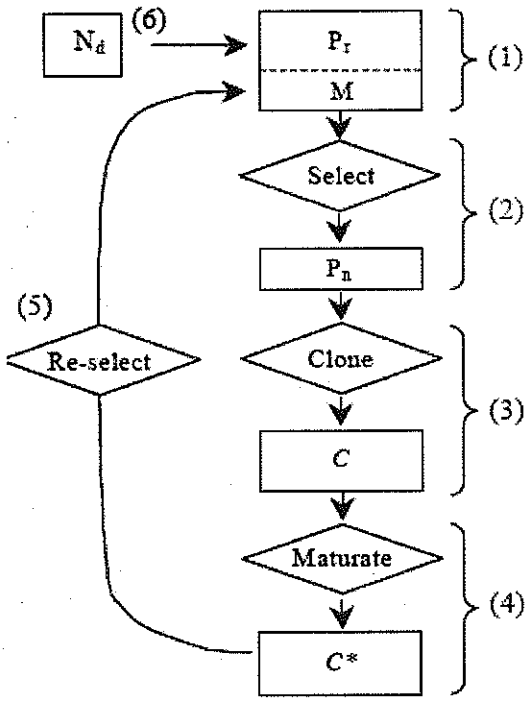


Figure 4 : Proposed Algorithm

The algorithm works as in Figure 3 (after each six steps we have one cell generation):

(1) Generate a set (P) of candidate solutions, composed of the subset of memory cells (M) added to the remaining population ($P = P_r + M$)

(2) Determine (Select) the n best individuals of the population (P_n), based on an affinity measure

(3) Reproduce (Clone) these n best individuals of the population, giving rise to a temporary population of clones (C). The clone size is an increasing function of the affinity with the antigen

(4) Submit the population of clones to a hyper mutation scheme, where the hyper mutation is proportional to the affinity of the antibody with the antigen. A maturated antibody population is generated (C*)

(5) Re-select the improved individuals from C* to compose the memory set M. Some members of P can be replaced by other improved members of C*

(6) Replace d antibodies by novel ones (diversity introduction). The lower affinity cells have higher probabilities of being replaced.

4. Set of Promoter Region Structures

The set of Promoter Regions used in this study was compiled from two different sources: from a set of representative Promoter Regions,described by Eyrich et al.,33 that contains primarily small Promoter Regions and Promoter Region domains, 34 a nonredundant set of Promoter Region chains, with less than 25% sequence identity from the Promoter Region Data Bank (PDB).35 For each Promoter Region, 2000 decoys were created with Rosetta.36 The final set contains 79 Promoter Regions with at least one decoy within 6.0 A°___ root-mean-square deviation (RMSD) from the native structure, and with a sufficiently deep enough multiple sequence alignment (see below): 1a32, 1a68, 1aa3, 1aboA, 1aca, 1acp, 1adr, 1afi, 1aho, 1ap0, 1ark, 1b3aA, 1b67A, 1bdo, 1bkrA, 1bor, 1bq9, 1c5a, 1c9oA, 1cc5, 1cc8A, 1ccwA, 1coo, 1cseI, 1csp, 1ctf,1ctj, 1cyo, 1dol, 1e6iA, 1edmB, 1ejgA, 1elkA, 1elwA,1eyvA, 1f7lA, 1fipA, 1fjlA, 1fjsL, 1fm0D, 1fna, 1fqtA,1g6xA, 1h4xA, 1h75A, 1h97A, 1hyp, 1icfI, 1jbeA, 1kjs,1lkkA, 1mzm, 1opd, 1psrA, 1ptq, 1qyp, 1r69, 1rb9, 1scjB,1sgpI, 1sro, 1stu, 1svy, 1tif, 1tuc, 1ubi, 1vig, 2af8, 2cdx,2fdn, 2fow, 2gdm, 2pdd, 2trxA, 2u1a, 3ebx, 4ubpA, 5icb,and 5pti (PDB code with specific chain in italics).

Creation of Multiple Sequence Alignments

For each query Promoter Region, a set of homologous sequences was collected by an iterative PSI-BLAST search.37 Based on the output, a multiple sequence alignment was created that includes sequences with less than 90% sequence identity to any other sequence and that span more than 80% of the query sequence. Three different stringency levels were used for the PSI-BLAST runs: (1) level10: 10 rounds of PSI-BLAST with an acceptance threshold of 10E_10; (2) level7: 5 rounds with an acceptance threshold of 10E_7; and (3) level5: 5 rounds with an acceptance threshold of 10E_5. The first level that resulted in a deep-enough multiple sequence alignment (defined as including more than 24 sequences) was retained for further analysis.

5. AIS FMACA Based Tree-Structured Classifier

Like decision tree classifiers, FMACA based tree structured classifier uses the distinct k-means algorithm recursively partitions the training set to get nodes (output of proposed algorithm 3.3) belonging to a single class. Each node (attractor basin) of the tree is either a leaf indicating a class; or a decision (intermediate) node which specifies a test on a single AIS FMACA.

Suppose, we want to design a AIS FMACA based pattern classifier to classify a training set S = {S1, S2,..., SK} into K classes. First, a AIS FMACA with k-attractor basins is generated. The training set S is then distributed into k attractor basins (nodes). Let S be the set of elements in an attractor basin. If S belongs to only one class, then label that attractor basin for that class. Otherwise, this process is repeated recursively for each attractor basin (node) until all the examples in each attractor basin belong to one class. Tree construction is reported in [7]. The above discussions have been formalized in the following algorithm. We are using genetic algorithm classify the training set.

Algorithm 1: AIS FMACA Tree Building (using proposed algorithm 3.3)

Input : Training set S = {S1, S2, · · ·, SK} with antigen

Output : FMACA Tree.

Partition(S, K)

Step 1 : Generate an AIS FMACA with k number of attractor basins.

Step 2: Distribute S into k attractor basins (nodes).

Step 3: Evaluate the distribution of examples in each attractor basin (node).

Step 4: If all the examples (S') of an attractor basin (node) belong to only one class, then label the attractor basin (leaf node) for that class.

Step 5: If examples (S') of an attractor basin belong to K' number of classes, then Partition (S', K').

Step 6: Stop.

6. Identification of Promoter Region Coding Region in DNA Sequence

In this section we concentrate on application of AIS FMACA to Promoter Region coding region identification. The idea of new method is to use the existing work of AIS FMACA based tree structure classifier. Lot of research has been done for finding Promoter Region statistically. By using the standard codon frequencies, [13] we can identify whether the sequence contain Promoter Region coding regions or not.

Example 3:

Consider the sequence AGGACC,

Since Codons will be in the form of triplets we split the input into three base sequences

So P(S) = F (AGG) ·F (ACC) = 0.22 * 0.38= 0.0836 using tables from, [11], [12].

In general, Let F0(c) be the frequency of codon c in a non-coding sequence.

P0 (C) =F0 (c1) F0 (c2)...F0 (cm)

Assuming the random model of non-coding DNA, F0(c) = 1/64= 0.0156 for all codons, P0 (S) = 0.0156 · 0.0156 = 0.000244. The log-likelihood (LP) ratio for S is LP(S) = log (0.000836/0.000244) = log (3.43) = 0.53.If LP(S) > 0, S is coding.

Like wise we can use Bayesian classifier to calculate the probability of finding the Promoter Region coding regions with accuracy up to 49. With our approach the average accuracy achieved is 75%.

6.1 Data and Method

The data used for this study are the human DNA data collected by Fickett and Tung. All the sequences are taken from GenBank in May 1992. Fickett and Tung have provided the 21 different coding measures that they surveyed and compared.

The benchmark human data include three different datasets. For the first dataset, non-overlapping human DNA sequences of length 54 have been extracted from all human sequences, with shorter pieces at the ends discarded.

Every sequence is labeled according to whether it is entirely coding, entirely non-coding, or mixed, and the mixed sequences (i.e., overlapping the exon-intron boundaries) are also included.

The dataset also includes the reverse complement of every sequence. This means that one-half of the data is guaranteed to be from the non-sense strand of the DNA.

In the next section we will give the experimental results for finding this coding region for all sequence lengths. It was compared with AIS FMACA and the accuracy reported was 2.2% more than that of standard ways of finding Promoter Region coding region.

7. Experimental Results

7.1 AIS MACA Rule Space In Successive Generations

The motion of CA rule space in successive generations is characterized by evaluating the entropy and mutual information of CA rule vectors of a population. The rule vectors for study are sampled out at a gap of 5 generations. The top most fit rule vectors of the selected population is subjected to closer scrutiny.

The entropy and mutual information of the CA in successive generations of GA are reported in Fig 5,6 ,7,8 for four different CA size (n= 10, 15, 20, 30). For each of the cases, the values of entropy and mutual information reach their steady state once the AIS FMACA for a given pattern set gets evolved. For understanding the motion, the initial population (IP) is randomly generated. All these figures points to the fact that as the CA evolve towards the desired goal of maximum pattern recognizing capability, the entropy values fluctuate in the intermediate generations, but saturate to a particular value (close to the critical value 0·84 [245]) when fit rule is obtained. Simultaneously, the values of mutual information fluctuate at the intermediate points prior to reaching maximum value that remains stable in subsequent generations. All these figures indicate that the CA move from chaotic region to the edge of chaos to perform complex computation associated with pattern recognition.

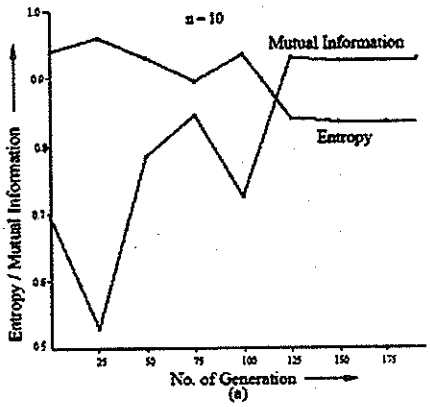


Figure 5 : Entropy & Mutation Information For n=10

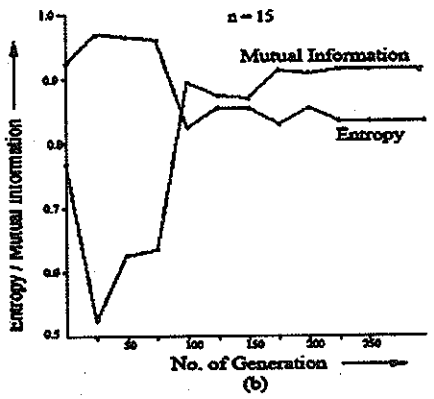


Figure 6 : Entropy & Mutation Information For n=15

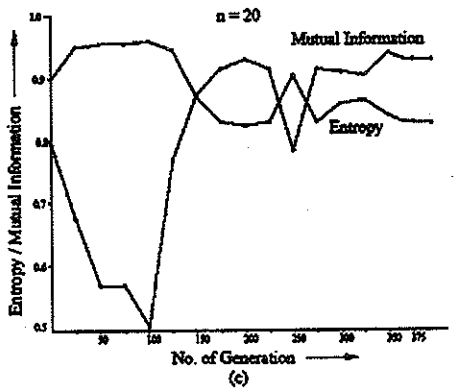


Figure 7 : Entropy & Mutation Information For n=20

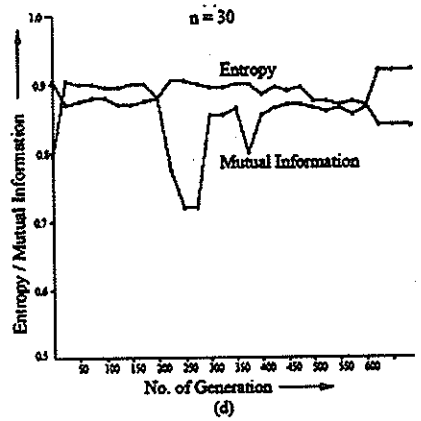


Figure 8 : Entropy & Mutation Information For n=30

7.2 Accuracy Calculations

The below tables shows the predictive accuracy of different algorithms on both coding and non-coding DNA sequences.

In this section we present the results on using AIS FMACA for Fickett and Tung's dataset. Values are given for the percentage accuracy on test set coding sequences and the percentage accuracy on test set non coding sequences.

Table 2 : Predictive Accuracy For Length 54 Human DNA Sequence

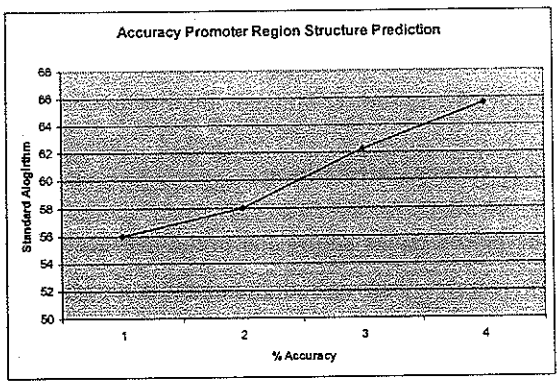
Algorithm	Coding	Non Coding
Dicodon Usage	61%	57%
Bayesian	51%	46%
CA	78%	72%
Sup FMACA	79%	72.5%
AIS FMACA	81%	73.5%

Table 3 : Predictive Accuracy For Length 108 Human DNA Sequence

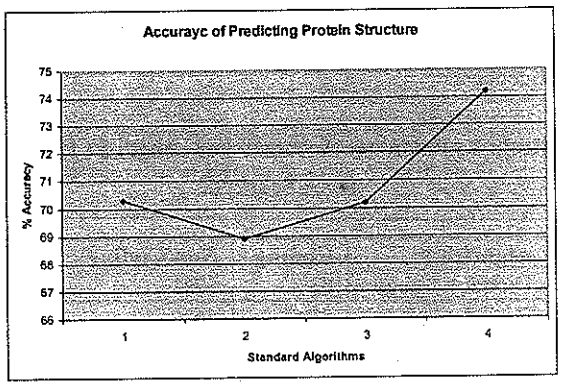
Algorithm	Coding	Non Coding
Dicodon Usage	58%	50%
Bayesian	45%	36%
CA	74%	69%
Sup FMACA	75%	69%
AIS FMACA	77%	74.5%

Table 4 : Predictive Accuracy For Length 252 Human DNA Sequence

Algorithm	Coding	Non Coding
Dicodon Usage	65%	54%
Bayesian	50%	44%
CA	71%	70%
Sup FMACA	71%	71%
AIS FMACA	72%	71.5%



Graph 1 : Promoter Region Structure Prediction (1: Zib, 2:Baker, 3: Rychlewski, 4: AIFMACA)



Graph 2 : Accuracy of Promoter Region Structure Prediction(1: Edvard Luie, 2: Paul Stalings, 3: Murty, 4: AIS FMACA)

The graph 1 compares the accuracy of finding the Promoter Region coding regions with existing standard algorithms like Bayesian, Data Base Search, Splicing Algorithm, Sup FMACA, and AIS FMACA. We can observe the accuracy increased considerably. The graph 1 compares the best Promoter Region coding region identification algorithm NPCRIT [19] with AIS FMACA. Both the graphs shows accuracy of AIS FMACA is comparable with any standard algorithm. AIS FMACA can be used to identify Promoter Region coding regions among all DNA sequence lengths.

AIS FMACA overcomes all the disadvantages of previous standard algorithms like fixing the position of the gene and static order of the DNA sequence. The average accuracy reported is 76.6%. It also finds the Promoter Region coding regions in mixed and non overlapping exon-inton boundary DNA sequences with average accuracy 74.5.Graph 2 shows the AIS FMACA can also predict the structure of Promoter Region with greater accuracy.

8. Conclusion

This paper presents the application of AIS MAFCA based supervised pattern classifier to address the problems of Promoter Region coding region identification in DNA sequences and finding the Structure of Promoter Regions. It also finds the Promoter Region coding regions in mixed and non overlapping exon-inton boundary DNA sequences with considerable accuracy. Aside from developing a good classifier for this particular problem, the proposed model may be very much useful to solve many other bioinformatics problems like RNA structure prediction, promoter region identification, etc.

References

[1] P.Kiran Sree, I .Ramesh Babu ,*"Identification of Promoter Region Coding Regions in Genomic DNA Using Unsupervised FMACA Based Pattern Classifier"*, in International Journal of Computer Science & Network Security with ISSN: 1738-7906, Vol.8, No.1,2008.

[2] P.Kiran Sree, R.Ramachandran, *"Identification of Promoter Region Coding Regions in Genomic DNA Using Supervised Fuzzy Cellular Automata"*, in International journal of Advances in Computer Science and Engineering, with ISSN: 0973-6999, Vol: 1,2008.

[3] Eric E. Snyder, Gary. D. Stormo, *" Identification of Promoter Region Coding Regions In Genomic DNA"*, ICCS Transactions 2002.

[4] E E Snyder and G D Stormo, *"Identification of coding regions in genomic DNA sequences: an application of dynamic programming and neural networks"*, Nucleic Acids Res. Vol : 21(3), PP. 608-613, 1993.

[5] P. Flocchini, F. Geurts, A. Mingarelli and N. Santoro and physica D, *"Convergence and Aperiodicity in Fuzzy Cellular Automata: Revisiting Rule 90,"* 2000

[6] P. Maji and P. P. Chaudhuri, *"FMACA: A Fuzzy Cellular Automata Based Pattern Classifier"*, Proceedings of 9th International Conference on Database Systems , Korea, PP. 494-505, 2004.

[7] C.G. Langton, *"Self-reproduction in cellular automata"*, Physica D, Vol.10, PP.135-144, 2000

[8] T. Toffoli and J.W. De Bakker and J. Van Leeuwen *"Reversible computing in Automata, Languages and Programming, ed."*, PP.632-644, 1994.

[9] G. Vichniac and physica. D, *"Simulating physics with cellular automata"*, Vol:10, PP.96-115, 1994

[10]S. Chattopadhyay, S. Adhikari, S. Sengupta and M. Pal, *"Highly regular, modular, and cascadable design of cellular automata-based pattern classifier"*, IEEE Trans. Very Large Scale Integr. Syst., Vol. 8, No.6, 2000.

[11] J. Fickett, *"Recognition of Promoter Region coding regions in dna sequences.,"* Nucleic Acids Res., Vol. 10, PP. 5303-5318, 1982.

[12] B. E. Blaisdell, *"A prevalent persistent global non randomness that distinguishes coding and non-coding eukaryotic nuclear dna sequence"*, J. Molec. Evol., Vol. 19, PP. 122-133, 1983.

[13] R. Farber, A. Lapedes and K. Sirotkin(1992), *"Determination of eukaryotic Promoter Region coding regions using neural networks and information theory"*, J. Mol. Biol., Vol. 226, PP. 471-479, 1992.

[14] E. Uberbacher and R. Mural, *"Locating Promoter Region-coding regions in human dna sequences by a multiple sensor-neural network approach"*, Proc. Natl. Acad. Sci., USA, Vol. 88, PP. 11261-11265, 1991.

[15] Aickelin, U and Cayzer. S., *"The Danger Theory and Its Application to AIS."*, Proceedings 1st International Conference on AIS, PP. 141-148, Canterbury, UK, 2002.

[16] D. Dasgupta, *"Artificial Immune Systems and Their Application"* Berlin, Germany: Springer-verlag, 1999.

[17] De Castro. L. N and Timmis. J, *"Artificial Immune Systems: A New Computational Intelligence Approach"*, Springer-Verlag, 2002.

[18] Krishna Kumar K, Kaneshige J and Satyadas. A, *"Challenging Aerospace Problems for Intelligent Systems."*, Proceedings of the von Karman Lecture series on Intelligent Systems for Aeronautics, Belgium, May 2002.

[19] P.Kiran Sree, *"NPCRIT: A Novel Promoter Region Coding Region Identifying Tool using Decision Tree Classifier with Trust-Region Method & Parallel*

Scan Algorithm," IEEE International Conference .(BIOTECHNO 2008). Proceeding published by IEEE Computer Society Press.

[20] Mitchison N. A, *"Cognitive Immunology"*, The Immunologist, Vol. 2 No.4 PP. 140-141, 1994

[21] Tauber. A. I, *"Historical and Philosophical Perspectives on Immune Cognition"*, Journal of the History of Biology, 30, PP. 419-440, 1997

[22] Jerne. N. K, *"Towards a Network Theory of the Immune System"*, Ann. Immunol. (Inst. Pasteur) 125C, PP. 373-389, 1974.

[23] Farmer. J. D, N. H. Packard, et al, *"The Immune System, Adaptation, and Machine Learning"* Physica 22(D): 187-204, 1986

[24] Timmis J and M. Neal, *"A resource limited artificial immune system for data analysis"*, Knowledge Based Systems PP 14(3-4): 121-130, 2001

[25] Ph. Tsalides T. A. York and A. Thanailakis, *"Pseudo-random Number Generators for VLSI Systems based on Linear Cellular Automata"*, IEE Proc. E. Comput. Digit. Tech., PP. 138(4):241-249, 1991.

[26] Marco Tomassini and Mattias Venzi, *"Artificially Evolved Asynchronous Cellular Automata for the Density Task"* Proceedings of Fifth International Conference on Cellular Automata for Research and Industry, ACRI 2002, Switzerland, PP. 44-55, October 2002.

[27] N. Tolstrup, J. Toftgard, J. Engelbrecht and S. Brunak, *"Neural network model of the genetic code is strongly correlated to the ges scale of amino-acid transfer free-energies"*, J. Mol. Biol., PP 243:816-820, 1994.

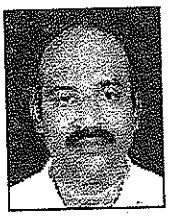
[28] S. Tan, J. Hao and J. Vandewalle, *"Determination of weights for hopfield associative memory by error back propagation"*, In Proc. IEEE Int. Symp. Circuits Systems PP, 5:2491, 1991.

[29] H. Szu, *"Fast TSP Algorithm based on Binary Neuron Output and Analog Input using Zero-diagonal Interconnect Matrix and Necessary and Sufficient Conditions of the Permutation matrix"* In IEEE International Conference on Neural Networks, PP 259-266, 1988.

[30] S. R. Sternberg *"Language and architecture for parallel image processing"*, PP. 35. North Holland, Amsterdam, 1980.

Author's Biography

P.kiran Sree received his B.Tech in Computer Science & Engineering, from J.N.T.U and M.E in Computer Science & Engineering from Anna University. He is pursuing Ph.D in Computer Science from J.N.T.U, Hyderabad. He has published many technical papers; both in international and national Journals & Conferences .His areas of interests include Parallel Algorithms, Artificial Intelligence, Compiler Design and Computer Networks. He also wrote books on Analysis of Algorithms, Theory of Computation and Artificial Intelligence. He was the reviewer for many IEEE Society Conferences in Artificial Intelligence and Networks. He was also member in many International Technical Committees.He was the Associate Editor for Asian Journal of Scientific Research (ISSN: 1992-1454), Journal of Artificial Intelligence (ISSN: 1994-5450), Information Technology Journal (ISSN: 1812-5646), Journal of Software Engineering(ISSN: 1819-4311), and Research Journal of Information Technology. He is the member of C.S.I, I.E.T.E, I.S.T.E (India),ICST (Europe) and IAENG (U.S.A). He is now associated with S.R.K Institute of Technology, Vijayawada.



Inampudi Ramesh Babu received his Ph.D in Computer Science from Acharya Nagarjuna University, M.E in Computer Engineering from Andhra University, B.E in Electronics & Communication Engg from University of Mysore. He is currently working as Professor in the department of computer science, Nagarjuna University. Also he is the senate member of the same University from 2006. He held many positions in Acharya Nagurjuna University as Head, Director - Computer Centre, Chairman- Board of studies. He was a special officer, convenor of ICET. He is also a member of Board of Studies for other universities. His areas of interest include Image Processing, Computer Graphics, Cryptography, Artificial Intelligence and Network Security. He is a member of IEEE, CSI, ISTE, IETE, IGISS, Amateur Ham Radio (VU2 IJZ). He is currently supervising 10 Ph.D students who are working in different areas of image processing & Artificial Intelligence. He has published 35 papers in international journals and conferences.

N.S.S.S.N Usha Devi was a graduate student of C.S.E from J.N.T.U. She has published two research papers in international conferences and one in an international journal. Her interests include Cellular Automata and Adhoc Networks. She was the member of IAENG (U.S.A), ICST (Europe). She was also Associate Editor of Journal of Software Engineering (ISSN: 1994-5450).