

# HYBRID CLONAL ALGORITHM USING HYBRID DEEP LEARNING AUGMENTED WITH CA

P KIRAN SREE<sup>\*</sup>, SSSN USHA DEVI N<sup>\*\*</sup>

## ABSTRACT

This article uses the basic framework of Hybrid Deep Learning Mechanism (DLM) to build a potential, versatile and robust classifier which can be applied to the diverse fields like bioinformatics, networks, image processing, compression and information retrieval. Detailed studies of various types of DLM and some classes of DLM referred to as MADLM (Multiple Attractor DLM), GMADLM (General Multiple Attractor DLM) and FMADLM (Fuzzy Multiple Attractor DLM) were presented with their possible applications. This article concentrates in projecting H-DL-MACA as a potential classification tool for specific applications in the domain of bioinformatics and network security.

## INTRODUCTION

An  $n$ -bit H-DL-MACA with  $k$ -attractor basins can be viewed as a natural classifier. It classifies a given set of patterns into  $k$  number of distinct classes, each class containing the set of states in the attractor basin. To enhance the

classification accuracy of the machine, most of the works have employed H-DL-MACA to classify patterns into two classes (say I and II). The following example illustrates an H-DL-MACA based two class pattern classifier.

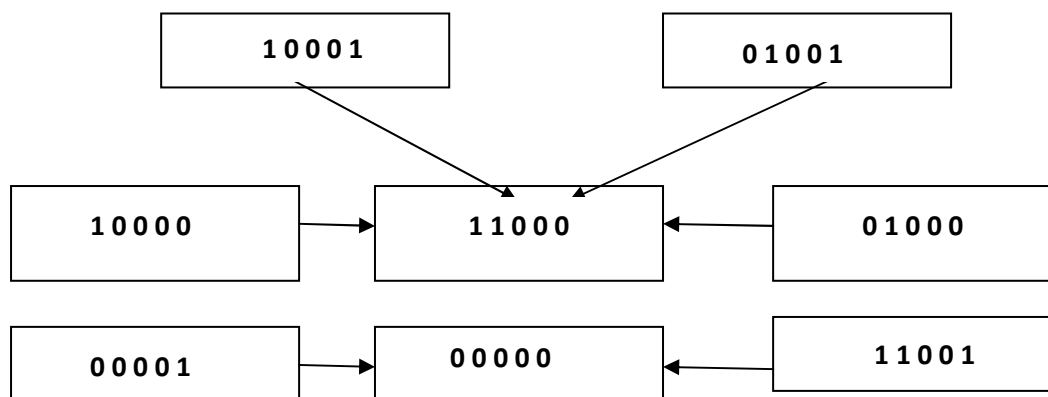


Figure 1. Example of MADLM with basin 0000

<sup>\*</sup> Professor, Dept of CSE, SVECW, Bhimavaram.

<sup>\*\*</sup> Assistant Professor, Dept of CSE, UCEK, JNTUK.

Correspondence E-mail Id: editor@eurekajournals.com

## MADLM TREE BUILDING

Input : Training set  $S = \{S_1, S_2, \dots, S_K\}$

Output : MADLM Tree.

Partition( $S, K$ )

Step 1 : Generate a MADLM 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.

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.

## GENETIC ALGORITHM FOR EVOLUTION OF TWO STAGE CLASSIFIER

In order to evolve the two stage classifier realizing the above design objective, we have developed a special type of GA formulation. It utilizes the concept of Dependency Vector. The fitness function for GA evolution to arrive at the desired MADLM based pattern classifier is reported next.

### FITNESS FUNCTION

In pattern classification, the fitness  $F$  of a particular chromosome in a population is determined by two factors.

1. The capability of the evolved DS (Classifier #1) for classifying the given input pattern set  $S_1$  and  $S_2$  into separate set of attractor basins; this is referred to as the factor  $F_1$ .

2. The capability of the evolved DV (Classifier #2) for classifying the pseudo-exhaustive field (PEF) set  $k_1$  and  $k_2$  into different attractor basins (one in zero basin, another in non-zero basin); this is referred to as the factor  $F_2$ .

We have worked on the analysis and synpaper of a type of DLM called as MADLM which works in linear fashion. This type of DLM uses XOR logic for defining the transitions between the cells. The tree based pattern classifiers have been modeled to build the proposed pattern classifier. Computation cost will be more in evaluating decision for non linear data sets. Different MADLM's can be identified by genetic algorithm (GA) simple decision logic. The versatility of the proposed classifier is demonstrated by applying this to the fields of network security and bioinformatics. Experimental results prove the scalability of MADLM to handle more number of data sets with number of attributes and classes. This result proves the fact that MADLM is a promising tool for simulating complex linear and non linear problems. All the above said functionalities and features make it an attractive and efficient pattern classifier.

## RANDOM GENERATION OF INITIAL POPULATION

To form the initial population, it must be ensured that each solution randomly generated is a combination of an  $n$ -bit DS with  $2m$  number of attractor basins (Classifier #1) and an  $m$ -bit DV (Classifier #2). The chromosomes are randomly synthesized according to the following steps:

1. Randomly partition  $n$  into  $m$  number of integers such that  $n_1 + n_2 + \dots + n_m = n$ .
2. For each  $n_i$ , randomly generate a valid Dependency Vector (DV).
3. Synthesize Dependency String (DS) through concatenation of  $m$  number of DVs for Classifier #1.

4. Randomly synthesize an  $m$ -bit Dependency Vector (DV) for Classifier #2.
5. Synthesize a chromosome through concatenation of Classifier #1 and Classifier #2.

A pattern classifier was built using proposed to model decision tree based pattern classification scheme. Evaluation of decisions for non linearly related dataset involves costly computation. Consequently, we evolve GA based evolutionary scheme to identify a set of linear MADLMs with simple decision logic. Theoretical formulations supported by extensive experimental results confirm the scalability of the proposed classifier to handle large datasets irrespective of the number of classes and attributes.

## CONCLUSION

We have successfully developed a classifier based on Modified CLONAL Algorithm using Deep Learning. This was applied on various domains of image processing and bioinformatics. After a thorough analysis we can say that this is applicable to major domains where the rules can be represented in the form of binary numbers.

## REFERENCES

- [1]. Sree PK, Babu IR, Devi NSSNU. Investigating an Artificial Immune System to Strengthen the Protein Structure Prediction and Protein Coding Region Identification using Cellular Automata Classifier. *International Journal of Bioinformatics Research and Applications* 2009; 5(6): 647-62.
- [2]. Sree PK, Babu IR, Murty JVR et al. Power-Aware Hybrid Intrusion Detection System (PHIDS) using Cellular Automata in Wireless Ad Hoc Networks. *World Scientific and Engineering Academy and Society Transactions on Computers* Dec 2008; 7(11): 1848-74.
- [3]. Sree PK, Babu IR. Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. *The International Arab Journal of Information Technology* 2010; 7(1): 75-78.
- [4]. Sree PK, Babu IR, Raju GVS et al. Improving Quality of Clustering using Cellular Automata for Information retrieval. *International Journal of Computer Science* Dec 2008; 4(2): 167-71.
- [5]. Sree PK, Babu IR, Devi NSSNU. A Novel Protein Coding Region Identifying Tool using Cellular Automata Classifier with Trust-Region Method and Parallel Scan Algorithm (NPCRITCACA). *International Journal of Biotechnology & Biochemistry* Dec 2008; 4(2): 177-89.
- [6]. Sree PK, Babu IR et al. Non Linear Cellular Automata in identification of protein coding regions. *Journal of Proteomics & Bioinformatics* 5(6): 123.
- [7]. Sree PK, Babu IR. Face Detection from still and Video Images using Unsupervised Cellular Automata with K means clustering algorithm. *ICGST International Journal on Graphics, Vision and Image Processing* 2008; 8(2): 1-7.
- [8]. Sree PK, Babu IR, Raju GVS et al. Cellular Automata with Biological Sequence Analysis Approach to Robotic Soccer. *International Journal of Computer Science and Applications* Nov/Dec 2009; 2(2): 127-34.