International Journal of Recent Development in Computer Technology & Software Applications
Vol. 2, Issue 2 - 2018

NON UNIFORM ADDITIVE NON UNIFORM CELLULAR AUTOMATA WITH DEEP LEARNING (ACADL) FOR PATTERN CLASSIFICATION

P KIRAN SREE[*], SSSN USHA DEVI N[**]

ABSTRACT

This article depicts a new approach to classify several problems based on the properties of Additive Non uniform cellular automata. We use a state-transition which consists of a set of disjoint trees rooted at cyclic states of unit cycle length thus forming a natural classifier. The framework proposed is strengthened with genetic algorithm to find the desired local rule of the modeling as a global state function.

INTRODUCTION

In the first part of the article, we have developed a classifier based on Linear DLM and Non Linear Deep Learning Mechanism which can address major problems in bioinformatics like protein coding region identification, protein structure prediction and promoter region identification. We have also proposed Artificial Immune System, a novel computational intelligence technique for strengthening the system with more adaptability and incorporating more parallelism to the system. We have also shown how the quality of clustering can be improved with Non uniform cellular automata.

In the second part of the article, we have explored a Heuristic based Non Uniform Deep Learning Mechanism based Intrusion detection system that monitors network for malicious activities or policy violations and produces reports to a management station. We found a pattern of abstract IDS that define the general features and patterns for behavior based IDS and signature based IDS which will be used to find the potential threats in the network.

BIOINFORMATICS PROBLEMS

A protein is a complex, high-molecular weight, organic compound that contains amino acids joined by peptide bonds. Proteins are essential to the structure and function of all living cells and viruses. The proteins in a cell determine what that cell will look like and what jobs that cell will do. The genes also determine how the many different cells of a body will be arranged. If we identify the protein coding region, we can extract lot of information like how DNA controls the number of fingers you have, where your legs are placed on your body, and the color of your eyes. DNA is organized in the form of introns and exons. Introns form the major part of the DNA strand and exons form the minor part of the DNA strand. But, exons only consist

---

[*] Professor, Dept of CSE, SVECW, Bhimavaram.
[**] Assistant Professor, Dept of CSE, UCEK, JNTUK.
Correspondence E-mail Id: editor@eurekajournals.com

ISSN: 2581-6276

of protein coding regions. Identifying protein coding regions in the exons is a real challenge. The proposed algorithms LMADLM, NPCRITDLMDLM can process DNA sequences of different lengths. Experimental results confirm the scalability of the proposed FDLM based classifier to handle large volume of datasets irrespective of the number of classes, tuples and attributes. Good classification accuracy has been established. Fickett and Tung data sets are used for measuring the efficiency of the classifier.

In genetics, a promoter is a region of DNA that initiates transcription of a particular gene. Promoters are located near the genes they transcribe, on the same strand and upstream on the DNA. An algorithm was proposed to identify the promoter regions with DLM. Eukaryotic Promoter Database new data sets are used.

Protein structure prediction is the prediction of the three dimensional structure of a protein from its amino acid sequence that is, the prediction of its secondary, tertiary, and quaternary structure from its primary structure. Structure prediction is fundamentally different from the inverse problem of protein design. Protein structure prediction is one of the most important goals pursued by bioinformatics. Data set used was taken from DLMSP.

AIS AUGMENTED WITH DEEP LEARING

An artificial immune system (ARTIS) is described which incorporates many properties of natural immune systems, including diversity, distributed computation, error tolerance, dynamic learning and adaptation, and self-monitoring. ARTIS is a general framework for a distributed adaptive system and could, in principle, be applied to many domains. This AIS-MADLM system was used to strengthen the protein coding region identification system and protein structure predicting system.

A fundamental problem for network intrusion detection systems is the ability of a skilled attacker to evade detection by exploiting ambiguities in the traffic stream as seen by the monitor. We discuss the viability of addressing this problem by introducing a new network forwarding element called a traffic MADLM normalizer. The MADLM normalizer sits directly in the path of traffic into a site and patches up the packet stream to eliminate potential ambiguities before the traffic is seen by the monitor, removing evasion opportunities. We examine a number of tradeoffs in designing a MADLM normalizer, emphasizing the important question of the degree to which normalizations undermine end-to-end protocol semantics.

We discuss the key practical issues of “cold start” and attacks on the MADLM normalizer, and develop a methodology for systematically examining the ambiguities present in a protocol based on walking the protocol’s header. We then present norm, a publicly available user-level implementation of a MADLM normalizer that can normalize a TCP traffic stream at 100,000 pkts/sec in memory-to-memory copies, suggesting that a kernel implementation using PC hardware could keep pace with a bidirectional 100 Mbps link with sufficient headroom to weather a high-speed flooding attack of small packets. DARPA Intrusion Detection Data Sets are used to evaluate the developed classifier.

We did an extensive survey on the key features of DLM which will be useful for pattern recognition. We have reported all the characteristics of DLM with their classes and applicability of the classes in various fields. After this study, we have successfully developed a linear and non linear classifier to address various problems in bioinformatics. Then the proposed algorithm is strengthened with artificial immune system with better stability and accuracy. The proposed algorithm was slightly modified to identify intrusions in the network also.

CONCLUSION

This article can be extended by formulating the memorizing capacity of non linear DLM based associative memory model. FDLM (Fuzzy Non uniform cellular automata) based model for complex functions involving datasets with attributes of real numbers can be explored. The proposed algorithm with some minor changes can be used as compression algorithm also. This article can be extended to propose a hybrid system with a combination of Non Linear DLM (NLDLM) and fuzzy sets.

REFERENCES

[1]. Sree PK, Babu IR et al. Investigating an Artificial Immune System to Strengthen the Protein Structure Prediction and Protein Coding Region Identification using Non uniform cellular automata Classifier. *International Journal of Bioinformatics Research and Applications* 2009; 5(6): 647-62.

[2]. Sree PK, Babu IR et al. Power-Aware Hybrid Intrusion Detection System (PHIDS) using Non uniform 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 et al. Identification of Promoter Region in Genomic DNA Using Non uniform cellular automata Based Text Clustering. *The International Arab Journal of Information Technology* 2010; 7(1): 75-78.

[4]. Sree PK, Babu IR et al. Improving Quality of Clustering using Non uniform cellular automata for Information retrieval. *International Journal of Computer Science* Dec 2008; 4(2): 167-71.

[5]. Sree PK, Babu IR et al. A Novel Protein Coding Region Identifying Tool using Non uniform 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 Non uniform cellular automata in identification of protein coding regions. *Journal of Proteomics & Bioinformatics* 5(6): 123.

[7]. Sree PK, Babu IR et al. Face Detection from still and Video Images using Unsupervised Non uniform 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 et al. Non uniform cellular automata with Biological Sequence Analysis Approach to Robotic Soccer. *International Journal of Computer Science and Applications* Nov/Dec 2009; 2(2): 127-34.