SciencePG
Science Publishing Group

Non Linear Cellular Automata Enhanced with Active Learning for Pattern Classification in Highly Dense Images

P. Kiran Sree[1], Sssn Usha Devi N.[2]

[1]Dept of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India

[2]Dept of Computer Science and Engineering, University College of Engineering, Jawaharlal Nehru Technological University, Kakinada, India

Email address:
drkiransree@gmail.com (P. K. Sree), usha.jntuk@yahoo.com (Sssn U. D. N.)

To cite this article:
P. Kiran Sree, Sssn Usha Devi N. Non Linear Cellular Automata Enhanced with Active Learning for Pattern Classification in Highly Dense Images. *Machine Learning Research.* Vol. 1, No. 1, 2016, pp. 15-18. doi: 10.11648/j.mlr.20160101.12

Abstract: This paper introduces a new approach to classify several high density images based on the properties of Non Linear 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.

Keywords: Cellular Automata (CA), Active Learning (DL), Non Linear CA

1. Introduction

In the first part of the paper we have developed a classifier based on Linear DLM and Non Linear Active 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 Cellular Automata.

In the second part of the paper we explored a Heuristic based Non Uniform ActiveLearningMechanisam 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.

A protein is a mind boggling, high-atomic weight, natural intensifies that contains of amino acids joined by peptide bonds. Proteins are basic to the structure and capacity of every single living cell and infections. The proteins in a cell figure out what that cell will look like and what employments

that cell will do. The qualities likewise decide how the various cells of a body will be orchestrated. In the event that we recognize the protein coding district and we can extricate parcel of data like, how DNA controls what number of fingers you have, where your legs are put on your body, and the shade of your eyes. DNA is sorted out as introns and exons. Introns shape the significant part of the DNA strand and exons frame the minor part of the DNA strand. Be that as it may, exons just comprise of protein coding locales. Recognizing protein coding districts in the exons is a genuine test. The proposed calculations LMADLM, NPCRITDLMDLM can handle DNA successions of various lengths. Trial comes about affirm the versatility of the proposed FDLM based classifier to handle extensive volume of datasets regardless of the quantity of classes, tuples and traits. Great grouping exactness has been set up.

Bioinformatics Problems

A protein is a complex, high-molecular weight, organic compounds that contains of 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 and we can extract lot of information like, how DNA controls how many 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 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. Fickettand 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.

2. AIS Augmented with Activelearing

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.

The fundamental unit of Artificial Deep Learning Mechanisam (DLM) is a cell that has a basic structure advancing in discrete time and space. A standout amongst the most essential turning points in the historical backdrop of improvement of the straightforward homogeneous structure of DLM is because of Wolfram. Answers for complex issues requests parallel registering environment. Most parallel PCs contain more than a couple of dozen processors. DLM can accomplish parallelism on a scale bigger than hugely parallel PCs. DLM is described by nearby availability of its cells. All communications occur on a simply nearby premise. A phone can just speak with its neighboring cells. Promote, the interconnection connects for the most part convey just a little measure of data. One ramifications of this rule is that no cell has a worldwide perspective of the whole framework.

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 Setsare be 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.

3. Complexity of DLM

DLM performs computations in a distributed fashion on a spatially extended grid. It differs from the conventional approach to parallel computation in which a problem is split into independent sub-problems, each solved by a different processor; the solution of sub-problems are subsequently combined to yield the final solution

The evolution process is directed by the popular Genetic Algorithm (GA) with the underlying philosophy of survival of the fittest gene. This GA framework can be adopted to arrive at the desired CA rule structure appropriate to model a physical system. The goals of GA formulation are to enhance the understanding of the ways DLM performs computations and to learn how DLM may be evolved to perform a specific computational task and to understand how evolution creates complex global behavior in a locally interconnected system of simple cells.

The task of pattern recognition is encountered in a wide range of human activity. In a broader perspective, the term could cover any context in which some decision or forecast is made on the basis of currently available information. The problem deals with the construction of a procedure to be applied to a set of inputs; the procedure assigns each new input to one of a set of classes on the basis of observed attributes or features. The construction of such a procedure

on an input dataset is defined as pattern recognition

4. DLM in Pattern Recognition

Pattern recognition algorithm has two phases, the learning or training phase and the testing phase. In the training phase, the algorithm is trained with some patterns. Based upon the nature of training, there are two broad categorization of pattern classification

This model is built describing a predefined set of data classes. A sample set from the database, each member belonging to one of the predefined classes, is used to train the model. The training phase is termed as supervised learning of the classifier. Each member may have multiple features. The classifier is trained based on a specific metric. Subsequent to training, the model performs the task of prediction in the testing phase. Prediction of the class of an input sample is done based on some metric, typically distance metric.

5. Conclusion

This paper can be extended by formulating the memorizing capacity of non linearDLM based associative memory model. FDLM (Fuzzy 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 paper can be extended to propose a hybrid system with a combination of Non Linear DLM (NLDLM) and fuzzy sets.

References

[1] Dr P. KiranSree&DrInampudi Ramesh Babu et al, 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, Vol 5, Number 6, pp 647-662, ISSN: 1744-5493. (2009) (Inderscience Journals, UK) Listed & Recognized in US National Library of Medicine National Institutes of Health. National Center for Biotechnology Information (Government of USA)PMID: 19887338 [PubMed-indexed for MEDLINE] H Index (Citation Index): 08 (SCImago, www.scimagojr.com) (Nine Years Old Journal).

[2] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. The International Arab Journal of Information Technology (IAJIT), Volume 7, No 1, 2010, pp 75-78. ISSN: 1683-3198H Index (Citation Index): 05 (SCImago, www.scimagojr.com)(Eleven Years Old Journal)(SCI Indexed Journal).

[3] Dr P. KiranSree&DrInampudi Ramesh Babu et al, A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier for Coding Region Prediction in Human Genome, Journal of Bioinformatics and Intelligent Control, Vol. 3, 2014, pp 1-6. DOI: 10.1166/jbic. 2014. 1077 (American Scientific Publications, USA).

[4] Dr P. KiranSree&DrInampudi Ramesh Babu et al, A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier Promoter Region Prediction in Eukaryotes. Journal of Bioinformatics and Intelligent Control, Vol. 3, 1–6, 2014. DOI: 10.1166/jbic. 2014. 1077 (American Scientific Publications, USA).

[5] Dr P. KiranSree&DrInampudi Ramesh Babu et al, 5. MACA-MCC-DA: A Fast MACA with Modified Clonal Classifier Promoter Region Prediction in Drosophila and Arabidopsis. European Journal of Biotechnology and Bioscience, 1 (6), 2014, pp 22-26, Impact Factor: 1.74.

[6] Dr P. KiranSree & DrInampudi Ramesh Babu et al, Cellular Automata in Splice Site Prediction. European Journal of Biotechnology and Bioscience, 1 (6), 2014, pp 36-39, Impact Factor: 1.74.

[7] Dr P. KiranSree&DrInampudi Ramesh Babu et al, AIX-MACA-Y Multiple Attractor Cellular Automata Based Clonal Classifier for Promoter and Protein Coding Region Prediction. Journal of Bioinformatics and Intelligent Control 3, no. 1 (2014): 23-30. DOI: 10.1166/jbic. 2014. 1071, (American Scientific Publications, USA).

[8] Dr P. KiranSree&DrInampudi Ramesh Babu et al, PSMACA: An Automated Protein Structure Prediction Using MACA (Multiple Attractor Cellular Automata). Journal of Bioinformatics and Intelligent Control 2, no. 3 (2013): 211-215. DOI:10.1166/jbic. 2013. 1052 (American Scientific Publications, USA).

[9] Dr P. KiranSree&DrInampudi Ramesh Babu et al, An extensive report on Cellular Automata based Artificial Immune System for strengthening Automated Protein Prediction. Advances in Biomedical Engineering Research (ABER) Volume 1 Issue 3, September 2013, pp 45-51. Science Publications (USA).

[10] Dr P. KiranSree&DrInampudi Ramesh Babu et al, 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 (IJBB) Volume 4, 177-189 Number 2 (December 2008). (Eight Years Old Journal) Listed in Indian Science Abstracts, ISSN: 0019-6339, Volume 45, Number 22, November 2009.

[11] Dr P. KiranSree&DrInampudi Ramesh Babu et al, HMACA: Towards proposing Cellular Automata based tool for protein coding, promoter region identification and protein structure prediction. International Journal of Research in Computer Applications & Information Technology, Volume 1 Number 1, pp 26-31, 2013.

[12] Dr P. KiranSree&DrInampudi Ramesh Babu et al, PRMACA: A Promoter Region identification using Multiple Attractor Cellular Automata (MACA) in the proceedings CT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol I Advances in Intelligent Systems and Computing Volume 248, 2014, pp 393-399 (Springer-AISC series).

[13] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Towards Proposing an Artificial Immune System for strengthening PSMACA: An Automated Protein Structure Prediction using Multiple Attractor Cellular Automata proceedings ofInternational Conference on Advances in electrical, electronics, mechanical and Computer Science (ICAEEMCS)-2013, ISBN: 978-93-81693-66-04 on September 2nd 2013, Hyderabad.

[14] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Multiple Attractor Cellular Automata (MACA) for Addressing Major Problems in Bioinformatics in Review of Bioinformatics and Biometrics (RBB) Volume 2 Issue 3, September 2013, pp70-76.

[15] Dr P. KiranSree&DrInampudi Ramesh Babu et al, Protein coding region Identification, in proceedings of 2nd International Conference on Proteomics Bioinformatics, July 2-4, 2012 Embassy Suites Las Vegas, USA ",(Special Issue of Journal of Proteomics & Bioinformatics. (USA), Volume 5 Issue 6 – 123, ISSN: 0974-276X, H Index (Citation Index): 06 (SCImago, www.scimagojr.com)Impact Factor: 2.2, (Five Years Old Journal).