

Cellular Automata Evolution : Theory and Applications in Pattern Recognition and Classification

Niloy Ganguly

A report in partial fulfillment for the Degree of

Doctor of Philosophy

in

Computer Science & Technology

Under the supervision of

Prof. P. Pal Chaudhuri

Dept. of Computer Science Technology

Bengal Engineering College(D U)

Department of Computer Science & Technology
Bengal Engineering College (A Deemed University)

June, 2003

Synopsis

Researchers in artificial life conceive the universe as a computer implementing transformations of information. If the universe can be viewed as a computation, it should be possible to build computing models of physical systems of the universe [6]. Many people in Artificial life have been enamored of a mathematical formalism of computing model known as the Cellular Automata. This modeling tool can be regarded as parallel processing computer. In an abstract sense, it can be also viewed as a logical universe with its own local physics and with emergent structures as artificial molecules. The researchers have explored the possibility of building Cellular Automata model having the capacity for self-reproduction and other essential functions of biomolecules leading to the possibility of life-like behavior. As a result, *Cellular Automata (CA)* came to be just as essential to the study of physical systems as the microscope was to study microbes and the telescope to the exploration of deep space.

The cellular automata (*CA*) is sufficiently complex to develop an entire universe as sophisticated as the one in which we live. Our own universe might be thought of as one mammoth Cellular Automaton. *CA* allows a programmer to specify rules for local interaction between ‘cells’ on a lattice-like grid and to study the emergent consequences of those rules. Von Neumann [1, 4] first envisioned that proper specification of rules empower *CA* to build models simulating bacterial growth, the growth of patterns on seashells, fluid dynamics, and the voting patterns of individuals who made decisions based on their local neighbors.

One of the most important milestones in the history of development of the simple homogeneous structure of cellular automata is due to Wolfram [2]. The suggested simplification leads to a one/two dimensional structure of simple cells, each having only two states with uniform three-neighborhood dependence. This simplified structure motivated a number of researchers [3] to undertake the study of *CA* behavior with linear/additive next state function amenable for matrix algebraic analysis. This new group of researchers explored innovative applications of this simple, regular, modular, and cascable structure of *CA* machine.

However, some important theoretical constraints and design issues partially restrict the capability of this modeling tool. The construction/synthesis/search of Cellular Automata(*CA*) having the ability to simulate a given modeling task is extremely difficult. Scientists have emulated the evolution mechanism of the living organisms of the universe to solve this hard problem. Evolution is mostly directed by the

popular genetic algorithm whose underlying philosophy is *survival of the fittest*. The fitter rules slowly overpower the less fit rules to arrive at the *CA* with the desired configuration. In order to implement genetic algorithm, the rules of *CA* are encoded in chromosome format. The scheme is termed by a variety of name like *EVCA*, *CAGA* etc [5].

However, the scheme suffers from a few inherent problems, discussed next. First, in order to find the fit rules, *GA* has to traverse a huge search space. For example, in an n -cell *CA*, if each cell has a modest 256 different rule options, the search space becomes a staggering 256^n . Secondly, specific to a problem, we can analytically discard many of the rules/ rule sequences. Alternatively, we may be able to specify the particular rule sequences as candidate solutions of the problem. However, when we implement genetic algorithm, it is difficult for the designer to accommodate such analysis into consideration. The genetic algorithm has to consider all those discarded rule sequences as the candidate solution to the problem.

In order to overcome the above problems, one of the major challenges in the research of *CA* evolution is to develop schemes which can initiate evolution within a sub-class of *CA*. A subclass differ from a class of *CA* in the sense that not only its state transition behavior depends upon a subset of rules but it also depends upon the sequence in which the rules are arranged. The conventional chromosome operations of *GA* cannot be applied for such solution, since a different sequence of the bits of the same chromosome will take us out of the search domain. The thesis handles this problem in the domain of Linear/Additive Cellular Automata and shows that vector algebraic analysis embedded in genetic algorithm can drastically reduce the search space.

The last decades of twentieth century have witnessed a colossal stride in the power of computation and communication leading to Internet technology. While pushing human civilization to new heights, such a large stride has opened up new challenges to the human society of cyber age. One of the major challenges we are facing is the exponential explosion of data in every field of our daily life. The need of the hour is to have good knowledge extraction methodology to extract meaningful and perceptible information from the large volume of data that are apparently uncorrelated and random in nature.

In the above context, Pattern Recognition/Classification has become an extremely important class of problem of the internetworked society of the twenty first century. In this class of problems, a machine identifies patterns of interest from its back-

ground. Several type of machines, most notable among them is neural network, has been proposed and widely used for solving pattern recognition/classification problems. However, in order to tackle the large volume of data-set and as well as to have fast decisions, high speed, low cost hardwired implementation of pattern recognition & classification algorithm/machine is a necessity. The sparse network of cellular automata (*CA*) holds enormous importance and potential in this domain.

The thesis reports a comprehensive approach in tackling the problems of pattern recognition and classification. The versatility of *CA* modeling tool is explored in this dissertation to identify special subclasses of Linear/Additive *CA* which can perform pattern recognition/classification task. A framework for efficient evolution of desired with required configuration has been proposed.

Aim of the dissertation

The major focus of this dissertation is to explore the applications of the powerful modeling tool of Cellular Automata in two important areas - pattern classification and pattern recognition. This demands a framework of *CA* evolution to meet the challenges of such ubiquitous applications. In order to design efficient evolutionary algorithms, an analytical foundation for characterization of Linear/Additive *CA* is an essential prerequisite.

In the above background this thesis concentrates on building the theoretical framework for complete analysis of *CA* state transition behavior in terms of its cyclic/non-cyclic vector subspaces. The solution to the reverse problem of *CA* synthesis is also necessary to realize the stated objective of this dissertation.

Once the theoretical foundation of *CA* analysis and synthesis is laid down, the framework for *CA* evolution is developed. The framework enables evolution of different types of *CA*. The evolution scheme enables us to direct search within Group *CA* - that is, the sub-class of *CA* whose state transitions always form cycles. The scheme can be further extended to direct search within a special type of Group *CA* to drastically reduce the search space. A special sub-class of Non-group *CA* which forms basins of attraction has been also evolved with the framework. The evolution scheme can be directed in such a manner so that the number of basins either dynamically changes over generations or it remains same throughout the period of evolution.

The evolution of Non-group *CA* has provided the platform to build the *CA* based model for design of a general pattern classifier. The same sub-class of Non-group *CA* is also used to model associative memory.

Non-linear *CA* evolved with the framework adds versatility and strength to the *CA* model. This model has been also employed for design of associative memory and its consequent application for pattern recognition. The comparison between the two models of associative memory realized through Linear and Non-linear *CA* has been also investigated.

Organization of the Dissertation

Prior to dealing with the proposed work on *CA* theory and its applications, we have reported a comprehensive survey of the relevant research publications in *Chapter 2*. The survey covers the following aspects:

- A general study of the methods employed to characterize Cellular Automata.
- Applications of Cellular Automata with special emphasis on modeling Pattern Recognition/Classification problem.
- A general overview of the existing schemes employed to address the Pattern Recognition/Classification problems

Chapter 3 reports the characterization of Additive *CA* in terms of its vector subspaces. The chapter establishes an analytical framework to study the state transition behavior of Linear/Additive *CA*.

If analysis is viewed as one side of a coin, its reverse is ‘Synthesis’. In *Chapter 4*, we report the schemes for synthesizing a Linear/Additive *CA* from its state transition behavior.

Once the general analysis and synthesis methodologies are in place, we start characterizing a special class of *CA* referred to as Multiple Attractor *CA* (*MACA*) to model different applications. Two such important applications, Modeling Associative Memory and Design of a General Pattern Classifier are addressed in the subsequent chapters.

Chapter 5 describes *MACA* as a special class of hash family referred to as Hamming Hash Family (*HHF*). The probability of collision of two patterns in *HHF* varies inversely with the hamming distance between them. A constrained Genetic Algorithm (*GA*) has been formulated to ensure directed search within the *MACA* family. The desired *MACA* is derived for a specific application through *GA* evolution.

In *Chapter 6* we report two major applications of *MACA* - Modeling Associative Memory and developing Pattern Classifier. The *MACA*-based Associative Memory

has been modeled using both single and multiple *MACA*. The Associative Memory Model and its applications in Pattern Recognition are the major focus of this chapter. Subsequent study and experimental results establish *MACA* as an efficient tool for modeling a Pattern Classifier.

The evolution framework, developed in *Chapter 5* and implemented in different applications in *Chapter 6*, is further generalized in *Chapter 7* to cover Group *CA*. While extending the study for Group *CA* evolution, we observed an interesting phenomena of practical relevance. Pseudo-random patterns are employed in many applications such as simulation, testing *VLSI* circuits etc. Some of these applications demand generation of pseudo-random patterns without a given set of prohibited patterns. Group *CA* have been evolved in *Chapter 7* to exclude the prohibited patterns from the pseudo-random pattern set. The solution proposed has its direct relevance for design of Pseudo-Random Pattern Generator for testing *VLSI* circuits.

The *Chapters 3 to 7* have dealt with Linear/Additive *CA* that employ *XOR/XNOR* logic function. On the other hand, Non-Linear *CA* employs Non-Linear rules involving AND/OR logic. Such *CA* are not amenable to the mathematical analysis proposed in earlier chapters. Non-Linear *CA* are evolved with Genetic Algorithm to design generalized *MACA* referred to as *GMACA*. The *Chapter 8* has established *GMACA* as an efficient model of Associative Memory. A comparative study of both the models of Associative Memory - *GMACA* and *MACA* are also reported.

The last chapter concludes the thesis and provides direction for future research.

Bibliography

- [1] J.Von Neumann *The Theory of Self Reproducing Automata*, A. W. Burks ed. Univ. of Illinois Press, Urbana and London, 1966.
- [2] S. Wolfram. *Statistical mechanics of cellular automata*. Rev. Mod. Phys., 55(3):601-644, July 1983.
- [3] P.Pal Chaudhuri et al. *Additive Cellular Automata Theory and Applications*. IEEE Computer Society Press, California, USA, 1997.
- [4] J.Thatcher. *Universality in Von Neumann cellular model*. Tech Report 0315-30-T, ORA, University of Michigan, 1964.
- [5] M.Mitchell *An Introduction to Genetic Algorithm*. MIT Press, 1999.
- [6] Trinh Xuan Thuan *Chaos and harmony*. OXFORD University Press, 2001.