Physica 10D (1984) 45–51
North-Holland, Amsterdam

SOME APPLICATIONS OF INFORMATION THEORY TO CELLULAR AUTOMATA

Michael S. WATERMAN
Department of Mathematics and of Biological Sciences, University of Southern California, Los Angeles, California 90089-1113, USA

In this paper general deterministic one-dimensional cellular automata are identified with mappings of the unit interval into itself. This allows the machinery of dynamical systems analysis to be employed. The emphasis of the paper, however, is on applications of existing concepts and techniques of information theory to these automata. A basic paper by W.M. Conner is utilized to obtain equality of the capacity and Hausdorff dimension of each line of the automata, and existence of limiting values of these quantities is established. Assuming a probability measure on the initial line that is stationary and ergodic for the shift, a consistent ergodic theory is derived for any finite or infinite collection of lines of the automata. A body of related work by Russian authors on probabilistic automata is briefly examined. Important questions about the existence and properties of limiting distributions remain unresolved.

1. Introduction

Recently, there has been renewed interest in cellular automata, due perhaps to increased computational power and to developments in dynamical systems theory. A great variety of examples are mentioned in connection with cellular automata. Cellular automata are a focus of attention in the theory of computation, and a large array of examples from biology and physics is frequently brought into discussions of possible utility. Even the origin of life has been mentioned in these discussions of "self-organizing" systems. (See [14] for many references.)

In particular, Stephen Wolfram [14] has recently performed extensive calculations for a class of one-dimensional cellular automata. For his situation, a cellular automaton is initiated by a string of sites with each site occupied by 0 or 1. Various rules are given which allow the string of sites to evolve in discrete time steps. Wolfram requires his site values to evolve as a deterministic function of the values of the sites and their nearest neighbors.

While Wolfram's automata would appear to be of a very simple nature, many intriguing and complex patterns arise when they are enumerated, line by line, as they evolve. His paper presents many figures and their numerical properties. In a study of algebraic properties of these one-dimensional automata, Martin et al. [7] analyze global properties of some of these automata. Their results are confined to a sub-class of "additive" cellular automata of the Wolfram type.

The purpose of this paper is to bring into focus a connection of cellular automata with information theory. There will be no restrictions on alphabet size or the number of neighboring sites used in the evolution rule. The capacity and Hausdorff dimension of each line (or string) are shown to exist and to be equal, as well as the existence of limiting capacity and dimension established. A general probability measure, invariant and ergodic under the shift, is put on the initial string of sites, and the consistent probability measure on any finite or infinite collection of lines is derived. Entropy and ambiguity can then be calculated, and a theorem proved to show the almost everywhere value of the transducer ambiguity limit equal to the limiting dimension of the ambiguity set. Some obvious generalizations are shown to be easily included in this framework and important related work by Russian information theorists is mentioned.

0167-2789/84/$03.00 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

By placing these dynamical systems into a probabilistic framework, certain theorems can be proved. The work of William M. Conner [2] on the capacity and ambiguity of a noise-less transducer turns out to be an important study of the first two lines of a one-dimensional cellular automaton and is the basis of this paper. Hopefully this will begin non-empirical studies of ergodic theory of one-dimensional cellular automata. Many numerical phenomena need explanation.

2. Capacity

We first define one-dimensional cellular automata. Take an alphabet $\Lambda = \{0, 1, \ldots, b-1\}$, where $2 \leqslant b$ and b is an integer. For m a positive integer assume ϕ^* is a mapping of $\Lambda^m = \Lambda \times \Lambda \times \cdots \times \Lambda$ into Λ. Of course there are b^{b^m} such mappings. The integer m is referred to as the ϕ^*-memory.

For $\boldsymbol{x}_0 \in (0, 1]$, let

$$\boldsymbol{x}_0 = \sum_{i=1}^{\infty} x_{0,i} b^{-i}$$

be the non-terminating base b expansion of x. Then define

$$\boldsymbol{x}_1 = \phi(\boldsymbol{x}_0) = \sum_{i=1}^{\infty} x_{1,i} b^{-i},$$

where

$$x_{1,i} = \phi^*(x_{0,i}, x_{0,i+1}, \ldots, x_{0,i+m-1}).$$

Clearly $\phi: (0, 1] \rightarrow (0, 1]$ and $\boldsymbol{x}_\nu$ is recursively defined by $\boldsymbol{x}_\nu = \phi(\boldsymbol{x}_{\nu-1}) = \phi^\nu(\boldsymbol{x}_0)$. The triple $((0, 1], \phi^*, \phi)$ is called a one-dimensional cellular automaton.

Note that two changes have been made to the set up in [14] where $m = 3$. The lines have been made infinite strings of sites, thereby eliminating the need for periodic boundary conditions. Also the second (third, etc.) line has been shifted one to the left, making identification into the unit interval more natural. The following elementary proposition allows easy identification of certain of Conner's results with cellular automata.

Proposition. If $((0, 1], \phi^*, \phi)$ is a one-dimensional cellular automaton with memory m, then $((0, 1], \phi^{*\nu}, \phi^\nu)$ is a one-dimensional cellular automaton with memory $m + \nu - 1$, where

$$\phi^{*\nu}(x_{0,i}, x_{0,i+1}, \ldots, x_{0,i+m+\nu-2}) = x_{\nu,i},$$

Following Conner, define

$$N_\nu(n) = \mathrm{card}\{(x_{\nu,i+1}, x_{\nu,i+2}, \ldots, x_{\nu,i+n}): x \in (0, 1]\}$$

to be the number of distinct sequences of length n on line ν. The channel capacity of line ν is defined by

$$C_\nu = \lim_{n \rightarrow \infty} \frac{\log N_\nu(n)}{n},$$

if it exists. (All logarithms are to the base b.) To prove existence, a standard lemma is necessary [2].

Lemma. Let f be non-negative, defined on the positive integers, where $f(n+k) \leqslant f(n) + f(k)$ holds for all n and k. Then $\lim_{n\to\infty} f(n)/n$ exists.

The following theorem and its proof are taken from Conner [2]:

Theorem 1. C_v exists and $0 \leqslant C_v \leqslant 1$.

Proof. $N_v(n+k)$ is the number of v-line sequences of length $n+k$. The sequence can begin in $N_v(n)$ ways and end in at most $N_v(k)$ ways so that $N_v(n+k) \leqslant N_v(n) N_v(k)$. Clearly, $\log N_v(n)$ satisfies the hypothesis of the lemma. The second statement follows from $1 \leqslant N_v(n) \leqslant b^n$.

The next theorem relates the Hausdorff dimension, dim(.), of $\phi^v(0, 1]$ with the capacity C_v. While Conner [2] employs measure theoretic techniques to establish the theorem, no measure theoretic qualifiers are needed in the theorem statement.

Theorem 2. $C_v = \dim(\phi^v(0, 1])$.

As we are interested in limits on v as well, the next theorem is a positive answer to a natural question about the existence of a limiting channel capacity C.

Theorem 3. $0 \leqslant C = \lim_{v\to\infty} C_v = \lim_{v\to\infty} \dim(\phi^v(0, 1])$.

Proof. Clearly, $N_{v+1}(n) \leqslant N_v(n+m)$, so that $0 \leqslant C_{v+1} \leqslant C_v$ and $0 \leqslant C = \lim_{v\to\infty} C_v$ exists.

Clearly, C is some index of how "thick" the set $\phi^v(0, 1]$ can become. Of course the existence of

$$\lim_{v\to\infty} \dim(\phi^v(0, 1])$$

does not assert the existence of $\lim_v \phi^v(0, 1]$. The "stability" of these sets remains an open question.

3. Ergodic theory

Whenever the ideas of time and space averages arise, the machinery of ergodic theory comes to mind. In the current setting we impose ergodic theory on (0, 1], elements of which represent the first line of our automata, and then construct the measures it induces on all succeeding lines. The measurable space will be $((0, 1], \mathscr{B})$ where $\mathscr{B}$ is the collection of Borel sets in (0, 1]. Define the shift on (0, 1] by

$$S(x) = S\left(\sum_{i=1}^{\infty} x_i b^{-i}\right) = \sum_{i=1}^{\infty} x_{i+1} b^{-i}.$$

Let $M = \{\mu : ((0, 1], \mathscr{B}, \mu)$ is a probability space, $\mu S^{-1} = \mu$, and S is ergodic for $\mu\}$.
In addition let

$$G_v = \{(x, \phi(x), \ldots, \phi^v(x)) : x \in (0, 1]\} \subset (0, 1]^{v+1}.$$

To extend to a probability measure on $(0, 1]^{\nu+1}$ consistent with $\mu \in M$, simply let

$$P_{0,1,\ldots,\nu}(B) = \mu(\mathrm{Proj}_0\,(B \cap G_\nu))\,,$$

where B is a Borel set in $(0, 1]^{\nu+1}(B \in \mathscr{B}_{\nu+1})$ and Proj_0 denotes projection onto the first or "zeroth" coordinate. The Kolmogorov consistency theorem [6] implies the existence of a measure $P_{0,1,\ldots,\infty}$ on $(0, 1]^\infty$ with finite-dimensional distributions consistent with those defined here.

Let M_ν be the set of probability measures in $(0, 1]^{\nu+1}$ which are ergodic and invariant under T_ν defined by

$$T_\nu(x_0, x_1, \ldots, x_\nu) = (S(x_0), S(x_1), \ldots, S(x_\nu))\,.$$

Theorem. $P_{0,1,\ldots,\nu} \in M_\nu$.

Proof. For $A_0 \times A_1 \times \cdots \times A_\nu \in \mathscr{B}_{\nu+1}$, $A_i \in \mathscr{B}$,

$$T_\nu^{-1}(A_0 \times A_1 \times \cdots \times A_\nu) = S^{-1}(A_0) \times S^{-1}(A_1) \times \cdots \times S^{-1}(A_\nu)\,.$$

Since $S(\phi(x)) = \phi(S(x))$, $\{x: \phi(x) \in S^{-1}A\} = \{x: \phi(S(x)) \in A\} = S^{-1}\{x: \phi(x) \in A\}$. This observation allows a proof that T_ν is a measure-preserving transformation (mpt) for $P_{0,\ldots,\nu}$

$$\begin{aligned} &P_{0,\ldots,\nu}(T_\nu^{-1}(A_0 \times A_1 \times \cdots \times A_\nu)) \\ &\quad = \mu\{x: x \in S^{-1}(A_0), \phi(x) \in S^{-1}(A_1), \ldots, \phi^\nu(x) \in S^{-1}(A_\nu)\} \\ &\quad = \mu S^{-1}\{x: x \in A_0, \phi(x) \in A_1, \ldots, \phi^\nu(x) \in A_\nu\} \\ &\quad = \mu\{x: x \in A_0, \phi(x) \in A_1, \ldots, \phi^\nu(x) \in A_\nu\} \\ &\quad = P_{0,\ldots,\nu}(A_0 \times A_1 \times \cdots \times A_\nu)\,. \end{aligned}$$

Since T_ν is a mpt on a monotone class [6] it is a mpt on $\mathscr{B}_\nu$.

In fact, these considerations also show T_ν ergodic for $P_{0,\ldots,\nu}$. Assume $T_\nu^{-1}B = B$. Then

$$\begin{aligned} T_\nu^{-1}(B \cap G_\nu) &= S^{-1}\{x: (x, \phi(x), \ldots, \phi^\nu(x)) \in B\} \\ &= \{x: (x, \phi(x), \ldots, \phi^\nu(x)) \in T_\nu^{-1}(B)\} \\ &= \{x: (x, \phi(x), \ldots, \phi^\nu(x)) \in B\} = B \cap G_\nu\,, \end{aligned}$$

and ergodicity of T_ν follows from ergodicity of S.

Since the shift is measure preserving and ergodic for all marginals, entropies for any line or collections of lines are defined and the pointwise ergodic theorem holds [1]. Instead of concentrating on generalizations of channel capacity, the remainder of this section treats an information theory concept called ambiguity.

First define

$$R_\nu(x) = \{t: t \in (0, 1] \quad \text{and} \quad \phi^\nu(t) = \phi^\nu(x)\}\,,$$

the set of points with the same νth line image as x. In addition $R_\nu(x, n)$ is the number of sequences of length $n + \nu(m-1)$ which have the same first n digits in the νth line image as the first n digits of $\phi^\nu(x)$. $R_\nu(x, n)$ is referred to as the ambiguity of line ν at the point x. The following theorem generalizes the case $\nu = 1$ which is due to Conner [2] and Fischer [3]:

Theorem. $D_\nu = \lim_{n\to\infty}(\log R_\nu(x, n))/n$ exists for almost all $x[\mu]$ and $D_\nu = \dim R_\nu(x)$ for almost all $x[\mu]$.

Corollary. $\lim_{\nu\to\infty} D_\nu = \lim_{\nu\to\infty} R_\nu(x) = D$ exists for almost all $x[\mu]$.

Proof. If $\phi^\nu(t) = \phi^\nu(x)$, then $\phi^{\nu+1}(t) = \phi^{\nu+1}(x)$ and $R_\nu(t) \subset R_{\nu+1}(x)$, so that $0 \leqslant D_\nu \leqslant D_{\nu+1} \leqslant 1$.

Once again, a limiting property of the rows has been derived.

4. Generalizations

In this section certain natural generalizations of simple automata are shown to be included in the above framework.

First of all, assume that the memory of ϕ is still m but that there is an additional temporal memory of k as well. That is

$$x_{1,i} = \phi^*\begin{pmatrix} x_{-k,i}, & x_{-k,i+1}, \ldots, x_{-k,i+m-1} \\ \vdots & \vdots \\ x_{-1,i}, & x_{-1,i+1}, \ldots, x_{-1,i+m-1} \\ x_{0,i}, & x_{0,i+1}\ , \ldots, x_{0,i+m-1} \end{pmatrix}$$

and

$$\boldsymbol{x}_1 = \phi(\boldsymbol{x}_{-k}, \boldsymbol{x}_{-k+1}, \ldots, \boldsymbol{x}_0) = \sum_{i=1}^{\infty} x_{1,i} b^{-i}.$$

Of course, $\phi: (0,1]^{k+1} \to (0,1]$ and $\boldsymbol{x}_{\nu+1} = \phi(\boldsymbol{x}_{\nu-k}, \ldots, \boldsymbol{x}_{\nu-1}, \boldsymbol{x}_\nu)$. If each column of height $k+1$ is mapped to a new integer,

$$d\begin{pmatrix} x_{-k,i} \\ \cdots \\ x_{-1,i} \\ x_{0,i} \end{pmatrix} = x_{-k,i} + x_{-k+1,i}(b) + \cdots + x_{-1,i}(b^{k-1}) + x_{0,i}(b^k).$$

Then each column can be considered a unique base B integer,

$$B = \sum_{i=0}^{k} b^i = \frac{b^{k+1}-1}{b-1}.$$

Therefore the temporal memory of ϕ can be accomodated by considering ϕ to have memory m with base B. This requires redefining ϕ as

$$\phi(\boldsymbol{x}_{-k}, \boldsymbol{x}_{-k+1}, \ldots, \boldsymbol{x}_0) = (\boldsymbol{x}_{-k+1}, \ldots, \boldsymbol{x}_0, \boldsymbol{x}_1).$$

A similar device shows that a one-dimensional cellular automaton can frequently be restricted to memory 1. First of all let

$$f(\boldsymbol{x}) = f(x_1 x_2 \ldots) = ((x_1 x_2 \ldots x_m)(x_2 x_3 \ldots x_{m+1}) \ldots)$$

and

$$\phi(f(\boldsymbol{x})) = (\phi^*(x_1 \ldots x_m)\phi^*(x_2 x_3 \ldots x_{m+1}) \ldots).$$

Again the alphabet has effectively been enlarged to reduce the memory. This device is employed by Fischer

[3]. Of course

$$Sf(x) = f(S(x)) \quad \text{and} \quad \phi(x) = \phi(f(x)).$$

5. Probabilistic automata

The automata considered this far are all deterministic. Some important work by Russian information theorists considers probabilistic automata. See for example [5, 8, 9, 12, 13]. In Conner's framework, they study the noisy transducer.

As a sample of their reults, Vasershtein and Leontovich [13] take a string of zeros and ones with transition rules

$$Q(\phi^*(11) = 1) = 1,$$
$$Q(\phi^*(00) = 1) = Q(\phi^*(01) = 1) = Q(\phi^*(10) = 1) = \theta.$$

M is the space of Borel probability measures on (0, 1]. A measure $\mu \in M$ includes (as above) measures on (0, 1] for time step 1 (P_1), time step 2 (P_2), ...

A question of importance is that, for $\mu S^{-1} = \mu$, when does $\lim_{\nu\to\infty} P_\nu$ exist? When is the limit unique?

Obviously if the measure μ is defined by

$$\mu\{.111\ldots\} = 1,$$

then $\mu S^{-1} = \mu$ and for the Q given above, $P_\nu \equiv \mu$ for all ν. When θ is large, no other invariant measure exists. But when θ is small ($\theta < 1/14$) there is at least one other invariant measure. Other deep results have been found.

In addition, this group considers the geometries obtained with the evolution of the automata [4, 10, 11]. Many of these results should carry over to the automata considered in this paper.

6. Conclusion

The object of this paper was to give a mathematical framework for one-dimensional cellular automata and to prove some simple results. The most important work remaining relates to the Markov process $P_{0,1,\ldots,\nu}$. For example, when does

$$\lim_{\nu\to\infty} P_\nu$$

exist and when is the limit unique? What is the set of support of P_ν? When is P_ν absolutely continuous with respect to Lebesque measure? Singular?

To obtain information about self-similarity the limits of

$P_{\nu-1,\nu,\nu+1}$,

$P_{\nu-1,\nu,\nu+1}$,

In addition, a number of questions arise for specific ϕ^*. For example, explicit computation of C_ν, C, D_ν, and D would be of real interest but is likely to be difficult.

Acknowledgement

This work was partially supported by System Development Foundation. The author is grateful to John R. Kinney for useful discussions.

References

[1] P. Billingsley, Ergodic Theory and Information (Wiley, New York, 1965).
[2] W.M. Conner, Ann. Math. Stat. 41 (1970) 2093.
[3] R. Fischer, J. of Math. Anal. and Apl. 75 (1980) 549.
[4] G.A. Galperin, Lecture Notes in Mathematics 653 (1978) 56.
[5] G.L. Kurdyumov, Lecture Notes in Mathematics 653 (1978) 37.
[6] M. Loeve, Probability Theory, 3rd ed. (Van Nostrand, New York, 1963).
[7] O. Martin, A.M. Odlyzko and S. Wolfram, Algebraic Properties of Cellular Automata, preprint.
[8] M.G. Shriyman, Lecture Notes in Mathematics 653 (1978) 31.
[9] A.L. Toom, Problemy Peredachi Informatsii 10 (1973) 70.
[10] A.L. Toom, Problemy Peredachi Informatsii 12 (1976) 48.
[11] A.L. Toom, Lecture Notes in Mathematics 653 (1978) 1.
[12] L.N. Vasershtein, Problemy Peredachi Informatsii 5 (1969) 64.
[13] L.N. Vasershtein and A.M. Leontovich, Problemy Peredachi Informatsii 6 (1970) 71.
[14] S. Wolfram, Rev. Mod. Phys. 55 (1983) 601.