

統計的にセル・オートマトンをモデルする： その目的と実践

川原田 茜

京都教育大学教育学部数学科

Akane Kawaharada

Department of Mathematics, Kyoto University of Education

宮路 智行

明治大学先端数理科学インスティテュート

Tomoyuki Miyaji

Meiji Institute for Advanced Study of Mathematical Sciences,

Meiji University

中野 直人

JST さきがけ / 北海道大学大学院理学研究院数学部門

Naoto Nakano

JST, PRESTO /

Department of Mathematics, Hokkaido University

概要

時空間データからセル・オートマトン (CA) モデルを統計的に構成する方法 (empirical CA construction) を紹介する。本稿は、これまでの先行研究および著者らが発表した一編の論文と二編の国際会議講演要旨のレビューである。

1 はじめに

流体力学、非線形光学や生物学など幅広い諸科学分野でパターン形成現象に関わる問題が現れる。これらの問題に対して数学・数理科学を活かした研究を行うには、数理モデルの構築 (モデリング) が不可欠である。流体力学では Navier-Stokes 方程式のような物理の第一原理に立脚したモデリングがしばしば行われる。対照的に、生物の体表模様の形成やバクテ

リアのコロニー形成といった現象に対しては、第一原理に基づくモデリングは困難であり、反応拡散系を用いた現象論的なモデリングが行われてきた。後者のモデリングには数学や物理の素養だけではないセンスがしばしば要求され、ある種の art (技術) のようなところがあり、決定的な指導原理は未だ確立されていないのが現状である。他方、計測技術や計算能力の飛躍的進歩は観測データの大量化・複雑化—ビッグデータ—をもたらし、データを積極的に活用する解析手法の研究やその応用が気象学をはじめとする諸分野で盛んに行われてきている。高度化・複雑化し様々な課題をもつ現代社会は数学者に数学・数理科学の力を十分活かした新たな方法論の提案・発展を期待しており、モデリングの重要性は今後ますます高まっていくだろう [12].

現象を記述しうる数理的なメカニズムを抽出する—あるいは、創出する—ことは容易ではない。我々は実験データに着目して、現象に対する簡易モデルの構築を試みる。本稿では、川原田・飯間 [5] によりその端緒がひらかれた、データからセル・オートマトン (CA) モデルを統計的に構成するという手法を考える。データはまさにその現象を表現するものである。川原田・飯間の手法は観測データを統計処理して、その時系列のもつ“傾向”を抽出した CA の局所規則を構成するという新しい手法である。

CA は、各セルの時間発展がその近傍のセルの状態に依存した局所規則で支配される力学系である。各セルは高々有限（それどころかごく少数）の状態しかとらず、シンプルな規則で時間発展するにもかかわらず、非常に豊富な時空間パターンが生成される。例えば Elementary Cellular Automaton (ECA) Rule 110 は Turing 完全であることが知られている [1]。Rule 90 では Sierpinski のギャスケットのような時空間パターンが生成される (図 1(a)) [11]。Belousov-Zhabotinsky 反応に現れるような螺旋パターンを生成する CA も存在する [2]。このように、CA はパターン形成のモデルとしての記述力を備えている。

偏微分方程式 (PDE) から CA を構成する手法として、超離散化手法 (ultradiscretization) が知られている [3]。これを CA モデル構成に適用するには、対象とする PDE モデルが既知であることと、可積分性のような良い構造を持っていることが要求されるという制約がある。統計的 CA 構成法はデータを元に構成するため、元となるような PDE モデルは必要ない。また、十分な量の時空間データさえあれば、原理的に任意の系に適用可能であるという利点をもつ。従って、対象とする時空間データは数値実験データでもよく、PDE の数値解をもとにその PDE を模倣する CA を構成することもできる [5]。これらの利点は手法が潜在的に広範な応用を持ちうることを示唆するとともに、既知の PDE をもとにして手法の妥当性を理論的に検証しうる可能性を示唆している。

統計的 CA 構成法は 2013 年に [5] でそのアイデアが提案されたばかりで、手法として確立したと言える状況ではない。いくつかのケーススタディ [4, 6] や実際の生物体流への応用 [10] を経て、筆者らによってその理論の深化 [7, 9] と方法の精緻化 [8] が試みられ始めたというのが現状である。実用的なモデリング手法として確立するには、理論と方法の両面にお

いて今後さらなる発展が必要である。本稿では、統計的 CA 構成法に関するこれまでの研究成果を概観したい。

2 統計的 CA 構成法

本節では川原田・飯間 [5] によって提案された統計的 CA 構成法を紹介する。まず、二種類の CA の定義を与える。一列に並んだ無限個のセルの各々が 0 から $k-1$ までのいずれかの状態を持っており、ある規則に従って刻一刻とセルの状態が変化していくことを考える。以下、 $A = \{0, 1, \dots, k-1\}$ を濃度 $k = |A|$ の有限集合とする。整数全体の集合を \mathbb{Z} と表し、 A の元を要素とする両側無限列を $A^{\mathbb{Z}}$ で記す。 $A^{\mathbb{Z}}$ の元 $x = (x_i)_{i \in \mathbb{Z}}$ をコンフィギュレーション、 $A^{\mathbb{Z}}$ をコンフィギュレーション空間と呼ぶ。 x_i は i 番目のセルの状態を表している。集合 A は各セルがとりうる可能な状態の集合であり、一次元格子 \mathbb{Z} に分布したセルの状態を集めたものがコンフィギュレーションである。

定義 1. f を A^3 から A への写像とし、 $A^{\mathbb{Z}}$ からそれ自身への写像 T を次で定める。すなわち、全ての $i \in \mathbb{Z}$ に対して

$$(Tx)_i = f(x_{i-1}, x_i, x_{i+1})$$

とする。写像 T の反復を発展作用素とする $A^{\mathbb{Z}}$ 上の離散時間力学系をセル・オートマトン (CA) という。 A^3 の元 (a, b, c) から $d \in A$ への写像 f を CA の局所規則という。

定義 2. $P = (p_{\alpha, l})$ を $k^3 \times k$ の確率行列とする。すなわち、全ての $\alpha = 0, 1, \dots, k^3 - 1$ に対して $\sum_{l=0}^{k-1} p_{\alpha, l} = 1$ かつ $0 \leq p_{\alpha, l} \leq 1$ を満たすものとする (添え字は 0 からはじめている)。確率的 CA とは、次のような $A^{\mathbb{Z}}$ における確率過程 $x^0, x^1, \dots, x^n, \dots$ である: 各 $n \in \mathbb{N}, i \in \mathbb{Z}$ および $l \in A$ に対して、 $x_i^{n+1} = l$ となる確率が $p_{\alpha, l}$ である。ここで $x^n = (x_i^n)_{i \in \mathbb{Z}}$ かつ $\alpha = k^2 x_{i-1}^n + k x_i^n + x_{i+1}^n$ である。

定義 1 の CA と確率的 CA を特に区別したいときには、前者を決定論的 CA と呼ぶ。

次に、以上の定義に基づき、川原田・飯間 ([5]) の統計的 CA 構成法の手続きを紹介する。この方法では多数の実験 (あるいは数値実験) を繰り返して時空間データを収集するので、これを実験的手法と呼ぶことにする。記述の簡単化のため、各時刻・各位置におけるデータは $[0, 1)$ 上の実数値をとるとする。実数 $u \in [0, 1)$ を k 状態 CA の状態に離散化するには

$$a = \text{floor}(ku)$$

で $a \in A = \{0, 1, \dots, k-1\}$ を定めればよい。ここで、 $\text{floor}(x)$ は x を越えない最大の整数 (床関数) を表す。 u の値のとりうる範囲は任意に変更できるし、必要ならばこのような等分でなく非一様に離散化することもできる (例えば [5] では、 $u \in [0, 1)$ の 2 進法での桁数に

基づいて状態を離散化している)。また、ここでは3近傍 CA、つまり f を3変数関数と仮定しているが、局所規則が依存する近傍は任意に設定できる。

以下のような手続きでデータから決定論的 CA を構成する。

- (1) L 回の実験で空間 $i = 0, 1, \dots, I-1$ および時間 $n = 0, 1, \dots, N$ のデータセット

$$\{u_{l,i,n} \mid 1 \leq l \leq L, 0 \leq i < I, 0 \leq n \leq N\}$$

が得られたとする。 $u_{l,i,n}$ は l 回目の実験の時刻 n における第 i 番目のセルの状態を表す実数値である。

- (2) 各 $u_{l,i,n}$ から

$$a_{l,i,n} = \text{floor}(ku_{l,i,n})$$

によって離散値のデータセットを得る。

- (3) 各 $a, b, c, d \in A$ に対して、離散値のデータセット内で

$$(a_{l,i-1,n}, a_{l,i,n}, a_{l,i+1,n}, a_{l,i,n+1}) = (a, b, c, d)$$

となる要素の組み合わせの個数 $X(a, b, c, d)$ を集計する*1。

- (4) 各 $(a, b, c) \in A^3$ に対して、 $X(a, b, c, d)$ を最大化する d を見つけ、 $f(a, b, c) = d$ を CA の局所規則とする。

確率的 CA をつくるには、上記 (4) を次で置き換える：

- (4') 各 $(a, b, c) \in A^3$ に対して

$$p_{\alpha,d} = \frac{X(a, b, c, d)}{\sum_{d \in A} X(a, b, c, d)}$$

とする。ここで α とは $\alpha = ak^2 + bk + c$ である。

3 数値実験と理論的考察

統計的に構成された CA によって、元のデータの背後に潜む（と仮定している）力学系を本当に模倣できるかどうかは非自明であり、検討の必要がある。例えば数値解析学においてある種の方程式の近似解法を提案するとき、その解法の収束性や安定性を論じるように、何らかの方法で統計的 CA 構成法の妥当性を示したい。このためには、性質が既知である偏微分方程式や CA から生成された時空間データに対して統計的に CA を構成して、それが元の系を模倣できるかどうか検証することが考えられる。本節では、手法の妥当性に関するこれまでの研究についてまとめる。

*1 もし周期境界条件を課すならば i は I を法とする。

3.1 ノイズを付加した ECA からノイズを除去する

[5] と [6] では、ECA から生成した時空間データから統計的に CA を構成する数値実験について報告されている。 $f: \{0, 1\}^3 \rightarrow \{0, 1\}$ を ECA の局所規則とする。確率 $0 \leq p \leq 1$ でセルの状態を反転させてデータセットを作る。すなわち、時刻 n での第 i 番目のセルの状態を x_i^n とするとき、時刻 $n+1$ での状態を次で定める：

$$x_i^{n+1} = \begin{cases} f(x_{i-1}^n, x_i^n, x_{i+1}^n) & \text{with probability } 1-p, \\ 1-f(x_{i-1}^n, x_i^n, x_{i+1}^n) & \text{with probability } p, \end{cases} \quad (1)$$

ただし、減算は体 $\mathbb{F}_2 = \{0, 1\}$ の元として行う。

[5] では Rule 90 を確率 $p = 0$ および 0.1 で反転させたデータにそれぞれ実験的手法を適用し、いずれの場合も統計的に構成された決定論的 CA として Rule 90 を得ている。[6] では Rule 150 に対して同様の実験を行い、やはり Rule 150 を得ている。(周期境界条件で一セルだけ状態 1、他のセルはすべて状態 0 という初期値を用いている。) これらの報告をまとめると以下の通りである：

1. $p = 0$ すなわち決定論的な ECA に対しては、そのまま元の ECA が復元される。
2. $p = 0.1$ でノイズを付加した場合、決定論的 CA の構成を通じてノイズが除去される。

[5] で示されている通り、データセットの作り方から明らかに $p < 0.5$ ならば全ての ECA に対してノイズの除去が可能である。

我々は 256 個の各 ECA に対して、確率 $p = 0.6$ でノイズを付加して生成したデータセットから決定論的 CA を構成する数値実験を行った。予想される通り、元の ECA の状態を反転したルールが得られた。

図 1 は $p = 0$ および $p = 0.1$ に対して計算した Rule 90 の時空間プロットである。 $p = 0$ の場合は Sierpinski のギヤスケットが見られる ([11])。一方で、[5] で指摘されているように、 $p = 0.1$ の場合はそれとは似ても似つかないにもかかわらず、統計的 CA 構成法を適用すると決定論的 ECA として Rule 90 が得られる。このように、大域的な時空間パターンは大きく異なるのに、局所的な決定論的規則として共通のものが見出されることは驚くべきことである [5, 6]。

これと同様のノイズ除去の有効性については、ECA だけでなく、3 以上の状態をもつ k 状態 CA に対しても直ちに拡張できる。このときは、(1) における状態反転の代わりに、

$$x_i^{n+1} = f(x_{i-1}^n, x_i^n, x_{i+1}^n) + i \quad \text{with probability } p_i \quad (i = 0, 1, \dots, k-1)$$

と確率 p_i で状態を i だけずらすことを考えればよい。ただし、 $\sum_{i=0}^{k-1} p_i = 1$ であり、 f は k 状態 CA の局所規則とし、加算は体 $\mathbb{F}_k = \{0, 1, \dots, k-1\}$ の元として行う。この場合においても、(1) の例と同様に、 $p_0 \geq 0.5$ ならばノイズが除去され、元の CA が復元される。

これにより、観測誤差などのノイズを含むデータセットであっても、統計的 CA 構成法によって得られる局所規則はノイズに対して頑健であることが期待できる。これは、統計的に数理モデルを構築する際に、未知の空間局所的な支配法則の抽出可能性を担保する上で極めて重要な性質である。

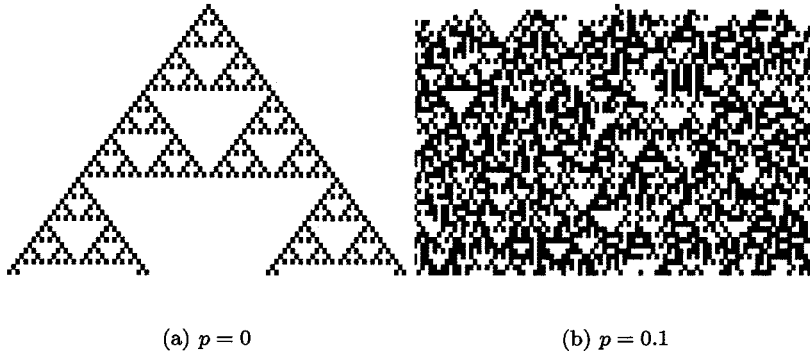


図 1 Rule 90 の時空間プロット ([5] の Fig.2 と同様の設定で計算し、描画した)。

3.2 偏微分方程式の数値解から CA モデルを構成する

より実践的なテスト問題として、時間発展偏微分方程式 (PDE) の数値解をデータセットとして統計的に CA を構成することが考えられる。[5] および [6] では熱方程式と粘性 Burgers 方程式, [7] および [9] ではこれらに加えて移流方程式に対する統計的 CA 構成法を適用した数値実験を行っている。

偏微分方程式から統計的に CA を構成するには、次のようにする。すなわち、初期条件をランダムに設定し、偏微分方程式の初期境界値問題に対する数値計算を多数行うことによってデータセットを作り、2 節の川原田・飯間の実験的手法を適用する。例えば、熱方程式

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad (2)$$

に対して、時間前進空間中心差分 (Forward-Time Central-Space, FTCS) スキーム

$$u_i^{n+1} = u_i^n + K(u_{i-1}^n - 2u_i^n + u_{i+1}^n) \quad (3)$$

によって時空間データを生成するというのである。

著者らは偏微分方程式の数値解から CA を構成するもう一つの方法を提案した ([7])。最小限の数値実験を行うことから、この方法をミニマル法と呼ぶ。以下、一般に (3) のような三点差分に基づく数値計算スキームを関数 $H: \mathbb{R}^3 \rightarrow \mathbb{R}$ で表す。すなわち、(3) に対応す

る H は

$$H(u, v, w) = v + K(u - 2v + w)$$

である。提案手法の手続きは以下のようなものである：

1. 各 $(a, b, c) \in A^3$ に対して, $X(a, b, c, d) = 0$ として次の手順を L 回繰り返す：

(a) 集合

$$\left[\frac{a}{k}, \frac{a+1}{k} \right) \times \left[\frac{b}{k}, \frac{b+1}{k} \right) \times \left[\frac{c}{k}, \frac{c+1}{k} \right)$$

上の一様乱数 (x, y, z) をとる.

(b) $d = \text{floor}(kH(x, y, z))$ を計算する

(c) $X(a, b, c, d)$ に 1 を加える.

2. 各 $(a, b, c) \in A^3$ に対して, $X(a, b, c, d)$ が最大となる d を見つけ, $f(a, b, c) = d$ を局所規則とする.

確率的 CA の確率行列は, 上記ステップ (2) を実験的手法のステップ (4') と置き換えることで得られる.

ミニマル法は実験的手法を簡約化したものと見なすことができ, 実験的手法よりも数学的な解析が遥かに容易である. PDE の数値解に対して実験的手法を適用した結果得られる CA を先験的に決定するのは非常に困難である. なぜなら, これを行うには数値解の分布が各計算ステップでどのように変化していくか知っていなければならないが, 分布は非常に複雑になりうるからである. ミニマル法は, 空間全体の代わりに 3 近傍にフォーカスして, 1 ステップだけ時間発展を調べてデータセットを集めており, この困難をかなり軽減できる. ミニマル法で構成される CA の局所規則は実験的手法のそれと必ずしも一致しないが, [7] の数値実験によれば, 生成される時空間パターンは定性的に似たものが得られる.

[9] はミニマル法に対する数学的な解析を行った. その結果を述べるために, まず準備を行う. (a, b, c) を A^3 の任意の元とし, D を

$$D = \left[\frac{a}{k}, \frac{a+1}{k} \right) \times \left[\frac{b}{k}, \frac{b+1}{k} \right) \times \left[\frac{c}{k}, \frac{c+1}{k} \right)$$

とする. すなわち D は, その離散化された状態が (a, b, c) に対応するような実数の組 $(x, y, z) \in [0, 1]^3$ の集合である. 各 $(x, y, z) \in D$ に対して, $H(x, y, z)$ は

$$Y = H\left(\left[\frac{a}{k}, \frac{a+1}{k}\right), \left[\frac{b}{k}, \frac{b+1}{k}\right), \left[\frac{c}{k}, \frac{c+1}{k}\right)\right)$$

に含まれる. もし (x, y, z) が D 上の一様分布に従う確率変数ならば, $w = H(x, y, z)$ も Y にサポートをもつ確率分布に従う確率変数である. 十分大きい実験回数 L に対してミニマル法を適用するという理想的な状況を考える. このとき, ミニマル法は確率

$$P(w \in Z_d), \quad Z_d := Y \cap \left[\frac{d}{k}, \frac{d+1}{k} \right)$$

を最大化する $d \in A$ をもって局所規則 $f(a, b, c) = d$ を与える. もし集合 Z_d が空となるような $d \in A$ に対しては, $f(a, b, c) = d$ は起こりえない.

次の定理は状態数無限大の極限でミニマル法により構成された CA の局所規則が元の差分スキーム H に収束することを示す.

定理 1 ([9]). $H : \mathbb{R}^3 \rightarrow \mathbb{R}$ を連続関数とする. 任意の $(u, v, w) \in [0, 1]^3$ に対して $(a, b, c \in A^3)$ を

$$a = \text{floor}(ku), \quad b = \text{floor}(kv), \quad c = \text{floor}(kw) \quad (4)$$

によって定義する. $z = H(u, v, w)$ とし, H に対してミニマル法によって構成された CA の局所規則を f とする. このとき $k \rightarrow \infty$ のとき $f(a, b, c)/k \rightarrow z$ となる.

従って, たとえば熱方程式, 移流方程式, 粘性バーガース方程式に対する三点差分に基づく FTCS スキームからミニマル法で構成された CA の局所規則は, 状態数無限大の極限で, 元の FTCS スキームに収束する ([9, Corollary 1]). この意味において, 統計的 CA 構成法は差分方程式と関連を持っており, 数値解析的な観点から手法の妥当性に対する部分的な回答を与えている. ただし, 「状態数 k が小さいと間違っただ CA が得られる」と結論づけることはできないし, 「大きい k がよい」とも言えない. CA の利点の一つは少ない状態数で豊富な時空間パターンを生成することにあるので, 実用上はなるべく k を小さくとりたい. あとに紹介する統計的 CA 構成法 [8] は, k が大きくなくても, ある程度定量的に適合する CA の構成が可能であることを示唆している.

さて, もし $P(w \in Z_d)$ を最大化する d が一意でない場合, $L \rightarrow \infty$ の極限で決定論的 CA の局所規則を定められない. この d が一意であることを, 局所規則を一意的に決定可能であるといおう. 実用上は, 有限な L しか利用できないので, ミニマル法のステップ 1(a) で利用する疑似乱数に敏感に依存して, 試行ごとに異なる局所規則が得られてしまう. この依存性を回避するためには, 例えば差分熱方程式のパラメータ K の選び方を気をつけなければならない.

定理 2 ([9]). H を差分熱方程式の FTCS スキームとする. 局所規則が一意的に決定可能でないならば, K が奇数 m と偶数 n の既約分数として表せる ($K = m/n$).

これは必要条件なので, 奇数/偶数 だからといって局所規則が定まらないとは限らない. この定理の対偶が一意的な具体的な判定条件を与える.

系 1 ([9]). 熱方程式の FTCS スキームにおいて, K が以下のいずれかを満たすならば, ミニマル法で得られる局所規則は一意的に決定可能である.

1. K は無理数である,
2. K は二つの奇数の既約分数である,

3. K は偶数 m と奇数 n によって既約分数 $K = m/n$ として表せる。

一意性に関するこれらの結果は差分移流方程式に対しても同様に成り立つ [9].

上記の結果は、 $L \rightarrow \infty$ の極限におけるものである。[7] は有限の L への依存性について数値実験結果を報告している。 $K = 4/5$ の差分熱方程式に対して状態数 $k = 8$ の CA をミニマル法で構成する場合、この場合は局所規則が一意的に決定可能であり、およそ $L = 150$ 程度で極限の局所規則に到達する。一方、 $K = 1/2$ の場合は一意的に決定可能でない場合である。この場合に $L = 10^8$ で得られる CA と一致する局所規則の個数は常におよそ 100 個から 150 個の間を揺らいでおり、試行のたびに構成される CA が異なってしまう。これに対して、確率的 CA の確率行列は概ね $1/\sqrt{L}$ に比例して収束していく。ミニマル法は、ある意味で、モンテカルロ法であり、期待通りの収束オーダーを示している。

3.3 超離散化との比較

偏微分方程式と CA を理論的に結び付ける既存手法として、超離散化・逆超離散化が知られている [3]. [5] と [6] は、熱方程式および粘性 Burgers 方程式に対して超離散化と実験的手法で構成された CA とを比較した。

超離散熱方程式は $K = 0.5$ に対する (3) を通して導かれ、

$$u_i^{n+1} = \max(u_{i-1}^n, u_{i+1}^n) \quad (5)$$

で与えられる。[5, 6] によれば、この超離散熱方程式に対応する CA と統計的に構成された決定論的 CA とは定性的によく似た振る舞いを示す。しかしどちらも差分解とは随分かけ離れている。

超離散 Burgers 方程式は超離散 Cole-Hopf 変換を通して導かれ、

$$u_i^{n+1} = u_i^n + \min(u_{i-1}^n, k - 1 - v_i^n) - \min(u_i^n, k - 1 - v_{i+1}^n) \quad (6)$$

で与えられる。この場合は超離散 Burgers 方程式から得られる CA と統計的に構成された CA とでは、波面の拡がる速度は等しいが、時空間パターンが少し異なる [5, 6]. 粘性 Burgers 方程式の波動の伝播速度は波の高さに依存するが、いずれの CA でも伝播速度が常に 1 となってしまう、複数の波の伝播を適切に表現できない。4 節で紹介する方法で統計的に CA を構成すれば、この困難を克服できる [8].

4 適切な時空間スケール選択の重要性

統計的 CA 構成法は原理的に任意の時空間データに適用することができるという利点をもつ。しかし、構成された決定論的 CA が元のデータの背後にある力学系を模倣できるとは必ずしも限らない。むしろ、これまでほとんどの場合、決定論的 CA はどんな初期状態から出

発しても膠着状態へと陥ってしまっていた。確率的 CA はこの難点を回避しているが、決定論的 CA ではたとえ状態数 k を大きくしても、やがて膠着状態に至ってしまう (図 2)。これは、ナイーブな実験的手法やミニマル法による決定論的 CA の構成では、状態が変わろうとする傾向に寄与する情報が欠落 (あるいは埋没と言えるかもしれない) してしまうことを意味している [8]。

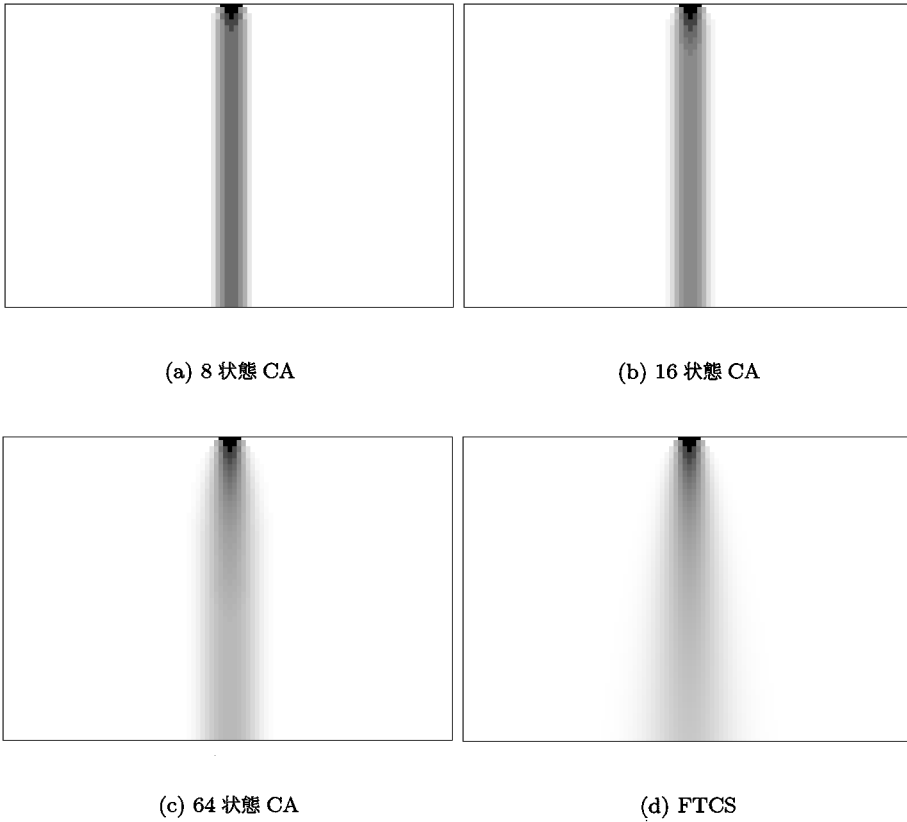


図 2 差分熱方程式 ($K = 0.4$) に対する FTCS スキーム (d) とミニマル法で統計的に構成された決定論的 CA の時空間プロット (a)–(c)。

情報の欠落が発生する主な要因は、CA のセルの時空間スケールとデータセットのそれと

の不整合にある。例えば、状態離散化の結果次のような時空間データが得られたとする：

$$\begin{array}{cccccccccccc}
 \dots & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \dots \\
 \dots & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \dots \\
 \dots & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \dots \\
 \dots & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \dots \\
 \dots & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \dots \\
 \dots & 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \dots
 \end{array} \tag{7}$$

ここで、横方向が空間であり、下方に進むにつれて時間ステップが経過するものとする。このデータでは3ステップごとに2と1の界面および1と0の界面が右にシフトしていく。しかし、素朴に統計的CA構成法を適用すると、 $f(2,2,1) = 2$ と $f(1,1,0) = 1$ を得て、CAのコンフィギュレーションは膠着する。このような状況は、状態解像度に対して時間・空間の解像度が高すぎると生じてしまう。

(7)の例は時空間スケールの適切な選択の必要性を示唆している。例えば(7)のようなデータセットに対しては、現時刻と3ステップ先の時刻とを見比べることにすれば、 $f(2,1,1) = 2$ および $f(1,0,0) = 1$ となり右にシフトするCAの局所規則が得られる。すなわち、データセットに対する適切なサブサンプリングによって、元のデータセットが持つ情報の欠落を免れる。[8]で提案した方法は、統計的CA構成法を適用する前に以下のような前処理を行うものである：

1. J 回の実験（または数値実験）によって、時空間データ

$$U^{(j)} = \{u_{i,n}^{(j)} \mid i = 0, 1, \dots, I-1, n = 1, 2, \dots, N-1\} \tag{8}$$

を $j = 1, 2, \dots, J$ に対して得る。 $u_{i,n}^{(j)}$ は j 回目の実験における時刻 i 、位置 n におけるデータ値を表す。

2. “サンプリングレート” $q \in \mathbb{Z}$ でサブサンプリングして、データセット

$$\tilde{U}^{(j)} = \{\tilde{u}_{i,n}^{(j)} \mid i = 0, 1, \dots, I-1, n = 1, 2, \dots, N_0-1\} \tag{9}$$

を得る。ここで、 $\tilde{u}_{i,n}^{(j)} = u_{i,qn}^{(j)}$ であり、 N_0 は N/q の整数部分である。

このようにして得られるデータセット $\tilde{U} = \sum_{j=1}^J \tilde{U}^{(j)}$ に対して2節の川原田・飯間の実験的手法を適用する。

[8]は粘性 Burgers 方程式の初期境界値問題の数値解として生成したデータセットに対してサブサンプリングを適用し、統計的にCAを構成した。その結果、 \tanh 型の進行波の伝播を模倣する決定論的CAを構成することができた。特筆すべきは、そのCAでは波の伝播速度がFTCSスキームのものと良い一致を示す点である。また、高さの異なる二つの界面の伝播・衝突を模倣することにも成功している。特に、衝突するということは、伝播速度の

波の高さへの依存性の抽出にも成功しているということである。統計的に構成された状態数 $k = 7$ の CA のコンフィギュレーションは以下のようになっている：

$$\begin{array}{cccccccccccc}
 \dots & 6 & 6 & 6 & 5 & 3 & 3 & 3 & 3 & 2 & 1 & 0 & 0 & 0 & \dots \\
 \dots & 6 & 6 & 6 & 6 & 5 & 3 & 3 & 3 & 3 & 1 & 0 & 0 & 0 & \dots \\
 \dots & 6 & 6 & 6 & 6 & 6 & 5 & 3 & 3 & 3 & 2 & 0 & 0 & 0 & \dots \\
 \dots & 6 & 6 & 6 & 6 & 6 & 6 & 5 & 3 & 3 & 2 & 1 & 0 & 0 & \dots \\
 \dots & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 5 & 3 & 3 & 1 & 0 & 0 & \dots \\
 \dots & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 5 & 3 & 2 & 0 & 0 & \dots
 \end{array} \tag{10}$$

大きい波面 $(6, 5, 3)$ は 1 ステップごとに右へ 1 セル進行している。一方、小さい波面は $(3, 2, 1, 0)$, $(3, 3, 1, 0)$, $(3, 2, 0, 0)$ という三つの状態を経てから右へ 1 セル進行している。様々な速度の波の形成には、このような内部遷移層の形成が重要な役割を果たす [8].

5 おわりに

ここまでで紹介した研究では、既知の ECA や PDE から生成した数値データに統計的 CA 構成法を適用した。しかし、こうした研究だけでは統計的 CA 構成法の重要性を示したことはない。より説得力のある研究には、有効な応用や、あるいは、それでなければ解けないような問題の解決が求められるだろう。

川原田ら [10] は、ミドリムシ (*Euglena gracilis*, *E. gracilis*) が生成する生物対流の時空間パターンに対して統計的 CA 構成法を適用した。円環状の容器に *E. gracilis* の懸濁液を入れ、底面から LED 光を照射する。*E. gracilis* は光を嫌って水面に集まるが、*E. gracilis* は水よりも重い密度の高いところでは *E. gracilis* が沈み、局所的に流れを引き起こす。その結果、*E. gracilis* が希薄な部分と密集している部分とが時間とともに変動する時空間パターンが形成される。円環状容器のある高さにおける時空間パターンを平面にプロットすると、密集部分が樹状構造をつくるのが典型的なパターンである。これは、時間経過とともに枝分かれするのではなく、まず梢ができてから別の梢から伸びる“枝”と合流するものである。生成された“枝”はそのまま長時間維持されることもあれば、他の枝と合流することもある。また、ふいに消失することもある。その発生メカニズムは未だ解明されていない。統計的に構成された CA は、元の時空間パターンを完全には模倣できていないが、梢の発生とそれが真っすぐ伸長する様子は再現できた。

実際の実験で収集するデータには、数値実験で生成されるものと大きく異なる点がある。それは、必ずしも全ての $(a, b, c) \in A^3$ に対して局所規則を定められるとは限らないことである。生物対流の例では、8 状態 CA を構成しており、512 個ある $(a, b, c) \in A^3$ の組み合わせのうち 164 個（およそ 32 パーセント）しか局所規則が決定されなかった。[10] では、そのような決定されない規則を empty rule と呼んでいる。CA の時間発展で empty

rule に対応する $(a, b, c) \in A^3$ が現れたとき, 川原田ら [10] は $f(a, b, c) = k - 1$ とすることにして計算を行った. 興味深いことに, CA の時間発展が初期の遷移過程を経た後, empty rule に遭遇する頻度が低下したようである. 対象とする現象に典型的なパターンの生成に寄与する $(a, b, c) \in A^3$ の割合は, A^3 のうちでごく一部分ということがありうる. アトラクタ上での力学系の挙動を見ていると思えば納得できるような気がするが, よくわからない. $(a, b, c) \in A^3$ の遷移には, 低次元制約があるのかもしれない. empty rule の影響を理論的に理解するにはさらなる研究が必要である.

現在のところ, 実際の実験データへの応用はまだほとんど行われていない. [10] はデモンストレーションの一つにすぎず, これをもって成功例とは言えない. 統計的 CA 構成法研究をより一層駆動するような, 興味深い応用例の蓄積が求められる.

謝辞

本研究は日本学術振興会科学研究費助成事業 (学術研究助成基金助成金) 挑戦的萌芽研究課題番号 16K13772 の助成を受けた.

参考文献

- [1] M. Cook, Universality in Elementary Cellular Automata, *Complex Systems* **15** (2004), 1–40.
- [2] M. Gerhardt, H. Schuster, J. J. Tyson, A cellular automaton model of excitable media: Iii. fitting the Belousov-Zhabotinskii reaction, *Physica D* **46** (1990), 416–426.
- [3] 広田良吾, 高橋大輔, 『差分と超離散』, 共立出版, 2003.
- [4] M. Iima, T. Yamaguchi, T. Watanabe, A. Kawaharada, Y. Tasaka, and E. Shoji, Toward understanding global flow structure, to appear in *Proceedings of International Conference on Mathematical Fluid Dynamics, Present and Future*.
- [5] A. Kawaharada and M. Iima, Constructing cellular automaton models from observation data, In 2013 First International Symposium on Computing and Networking, pages 559–562, (2013).
- [6] A. Kawaharada and M. Iima, An application of data-based construction method of cellular automata to physical phenomena. to appear in *J. Cell. Autom.*.
- [7] A. Kawaharada, T. Miyaji, and N. Nakano, An analyzable method for constructing a cellular automaton from a continuous system, In 2015 Third International Symposium on Computing and Networking, pages 418–423, (2015).

- [8] A. Kawaharada, T. Miyaji, and N. Nakano, Proper choice of spatio-temporal scale and dataset subsampling for empirical CA construction, In 2015 Third International Symposium on Computing and Networking, pages 424–429, (2015).
- [9] A. Kawaharada, T. Miyaji, and N. Nakano, Analysis of a method for constructing a cellular automaton from a continuous system, *Int. J. Networking and Computing* **6** (2016), 230–242.
- [10] A. Kawaharada, E. Shoji, H. Nishimori, A. Awazu, S. Izumi, and M. Iima, Cellular automata automatically constructed from a bioconvection pattern, *Recent Advances in Natural Computing, Mathematics for Industry*, **14** (2015), 15–25.
- [11] S. Wolfram, Statistical mechanics of cellular automata, *Rev. Mod. Phys.*, **55** (1983), 601–644.
- [12] 文部科学省, 『数学・数理科学と諸科学・産業との連携による数学イノベーションの推進』, http://www.mext.go.jp/a_menu/math/