

셀룰라 오토마타와 Markov Chain 모형을 이용한 토지이용 변화 예측

Land Use Change Prediction using Cellular Automata and Markov Chain

김성준*, 정인균**

Kim, Seong Joon and Jung, In Kyun

1. 서론

GIS는 공간정보를 다루는 여러 분야에서 널리 활용되고 있다. 수자원 분야에서는 토지이용도, 토양도, 지질도, 유역도, 하천도 등의 자료가 GIS자료화 되어 구축되어 있으며, 강우, 유출량, 증발량 등 시간의 흐름에 따라 변화하는 점 자료들이 주로 이용되고 있다. 여기서 강우·증발량 등과 같은 점 자료는 GIS의 보간 기법을 이용하면 시간-공간적인 데이터의 구축이 가능하고, 토지이용도(혹은 토지피복도)와 같이 시간개념과 밀접한 공간자료는 일정 시간계열에 대하여 위성사진이나 항공사진을 이용하여 구축함으로써 시간-공간적 개념을 부여 할 수 있다. 그러나 수자원 분야에서의 지리정보시스템을 이용한 공간분석은 중첩·인접성·버퍼링·보간 등의 단순한 분석기법을 주로 활용하고 있다.

단순한 공간분석에서 벗어나 시간적 개념을 고려한 공간분석 기법을 적용한다면 어떤 사건에 대한 확률이나 향후 변화양상을 예측할 수 있을 것이다. 이러한 공간적인 사물의 동적형태를 모델링하는 방법 중 대표적인 것으로 셀룰라

오토마타(CA: Cellular Automata) 이론과 Markov Chain모형을 적용한 기법이 있다. 이미 국내외에서 셀룰라 오토마타와 Markov Chain을 이용한 도시·인구·농업·주택가격·토지피복 등의 예측에 관한 다양한 연구 성과가 발표되고 있다.

수자원분야에서 위와 같은 공간분석방법을 이용한다면, 어떤 지역의 향후 도시 확장, 산림감소, 경지면적 감소 등에 대한 변화를 예측함으로써 유출량 변화를 미리 파악할 수 있을 것이다.

본 소강좌에서는 셀룰라 오토마타와 Markov Chain 모형을 살펴보고 이 두 가지 개념이 적용되어 있는 Idrisi Kilimanjaro의 CA_MARKOV 모듈에 의한 토지이용변화 예측 분석과정을 간략히 소개하고자 한다.

2. 셀룰라 오토마타

셀룰라 오토마타의 태동은 1930년대 Turing 으로부터 기원한다고 할 수 있다. 그는 '계산한다는 것은 무엇인가'에 대하여 연구하며 일종의 수학적인 모델인 튜링머신이라는 가상의 기계를 제시하였다. 이는 순차적 접근

* 건국대학교 사회환경시스템공학과 교수(kimsj@konkuk.ac.kr)

** 건국대학교 사회환경시스템공학과(nemoik@konkuk.ac.kr)

방식을 이용한 계산기의 모태가 되었으며, 본 노이만(Von Neumann)은 자신을 복제하는 생명체와 같이 행위(연산)를 반복하는 계산절차에 관한 연구를 진행하였는데, 이것이 본 노이만의 셀룰라 오토마타 이론이다. 그의 초기 오토마타 이론은 운동학적 오토마타인데 운동학적인 오토마타는 이전의 성질을 이후에 가지고 있지 못해 반복적으로 적용될 수 없었다(Xie, 1994). Ulam은 이를 개선할 수 있는 방법을 제안하였고, 비로소 셀룰라 오토마타가 탄생하게 되었다.

셀룰라 오토마타는 복잡한 자연계 현상을 모델링하기 위한 수학적 방법론 중 하나로(Wolfram, 1984), 이상적인 시공(discrete time and space)에 기초한 비선형 동적모델로 정의될 수 있다(Park, 1996). 셀룰라 오토마타는 임의의 시점에서 임의의 개체(cell)가 셀 공간(Cell-space)에서 주변의 상태 그리고 정해진 변이규칙(transition rule)에 따라 다음시점의 상태로 변화하는 연산을 반복적으로 수행하는 원리를 통해 현상을 모델링한다. 통상 변화는 임의의 개체 주변 국지적 차원에서 일어나고, 이러한 과정이 수없이 반복되면 전체적 차원에서의 모습을 알 수 있게 된다. 셀룰라 오토마타의 구성요소는 아래와 같이 정의할 수 있다.

가. 셀(Cell)과 셀공간(Cell-space)

셀룰라 오토마타에서의 셀은 1, 2, 3차원으로 정의할 수 있으며 일반적으로 공간현상을 다루기 용이한 2차원 셀이 가장 널리 이용되고 있다. 우리가 수문분석에 사용하고 있는 자료의 대부분이 2차원 공간을 나타내고 있기도 하다. 2차원 셀공간에서 셀의 형태는 4각형, 3각형, 6각형, Voronoi region과 Delauney 삼각형과 같은 불규칙한 형태의 셀이 적용되기도 하나 컴퓨터를 이용한 연산과정에서 효율성이 가장 우수한 정방형의 4각형 셀을 이용한다.

나. 상태(State)

상태(State)는 셀의 속성을 나타내는 것으로 가장 단순한 형태로는 0 또는 1의 값을 가진 이진형태를 들 수 있다. 그러나 입력 자료에 따라 여러 가지 값을 가지는 것이 보통이다.

다. 이웃(Neighborhood)

셀룰라 오토마타에서 이웃은 주변 임의 개수의 셀로 정의될 수 있다. 가장 널리 쓰이는 주변의 형태로는 셀들을 자신의 동, 서, 남, 북에

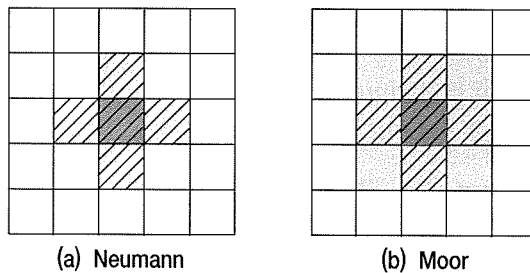


그림 1. 이웃 (Neighborhood) 셀

있는 셀들로 정의하는 Von Neumann neighborhood와 동, 서, 남, 북, 남동, 남서, 북동, 북서에 있는 셀들로 정의하는 Moore neighborhood가 있다(그림 1.참조).

라. 변화규칙(Transition rule)

변화규칙은 셀룰라 오토마타에서 가장 중요한 역할을 하는 요소로서 실제 셀의 상태를 변화시키는 제어기 역할을 한다. 셀룰라 오토마타에서 각 셀은 해당 셀의 상태와 주변의 상태에 따라 변화된다.

마. 시간(Temporal space)

셀룰라 오토마타에서 시간은 연속변수가 아닌 이산적인 변수이다. 즉 시간 t에서 셀의 상태, 이웃의 상태, 변화규칙을 적용하여 시간 t+1에서의 셀의 상태를 만들어 낸다.

3. Markov Chain

Markov Chain은 다양한 경영시스템을 모형화하는 수학적 기법으로 과거에 있었던 변화를 토대로 시스템 내에서 어떤 상태(state)에서 다른 상태로 연속적으로 전개되는 동적성격을 파악하여 미래에 있을 변화를 연속적으로 예측하는 확률과정(Stochastic process)모형으로서 이용한다.

확률과정 $[x(t)]$ 가 있을 때,
 시점 t_1, t_2, \dots, t_n 에 관해서
 $P = [x(t_1)=x_1 | x(t_2)=x_2, x(t_3)=x_3, \dots, x(t_n)=x_n]$
 $= [x(t_1)=x_1 | x(t_1)=x_1]$

가 성립하면 이 과정을 Markov process 라고 한다. 이때 시간모수 t와 상태공간이 이산적인 것이 Markov chain이다. 이 경우 현재의 상태는

바로 이전의 상태에 의존하며, 과거의 경로에는 무관하다는 것을 나타낸다. 여기서 상태(state)란 집단의 성원을 특징지을 수 있는 변수를 일정한 구간으로 잡아 그 교체에 의해 형성되는 유한개의 범주를 정의한다. Markov chain 모형에서는 상태들의 집합인 상태공간이 정의되며, 주어진 시점에서 이들 상태집합 가운데 하나의 상태에서부터 집합의 다른 상태로 계속적 이동을 정의한다. 이 경우 각 이동을 단계(step)라고 한다. 예를 들면 농가에서 농민들이 동절기 작물을 경작한 경우 n년의 재배패턴을 보면 마늘을 경작하는 경우 $X_n = 1$, 양파를 경작하는 경우 $X_n = 2$, 보리를 경작하는 경우는 $X_n = 3$, 시금치나 시설재배 작물을 경작하는 경우는 $X_n = 4$, 유희지로 남겨두는 경우 $X_n = 5$ 라 한다면 매년의 경작패턴은 하나의 Markov process를 이룬다. 이 경우 상태공간 $S = \{1, 2, 3, 4, 5\}$ 이다. 이상의 설명을 Vector나 행렬로 나타내면 기간 m에서 I state가 기간 n에 j state로 이동될 확률은 다음과 같이 나타낼 수 있다.

$$P_{(m,n)} = \begin{bmatrix} P_{11(m,n)}, \dots, P_{1n(m,n)} \\ P_{21(m,n)}, \dots, P_{2n(m,n)} \\ P_{n1(m,n)}, \dots, P_{nn(m,n)} \end{bmatrix}$$

이때 $P_{(m,n)}$ 을 변이확률행렬(transition probability matrix)이라 하고 이것은 간단하게 다음과 같이 표시된다.

$$P_{(m,n)} = (P_{ij(m,n)})$$

한편 변이확률은 어떤 시점에서 변하는 경우를 Non-stationary Markov chain이라 하며, 시간의 경과에도 불구하고 항상 동일한 경우를 Stationary Markov chain이라 한다. 변이확률이란 어떤 상황이 주어졌을 때 다음의 어떤 상황이 일어날 비율을 나타내는 조건확률이므로

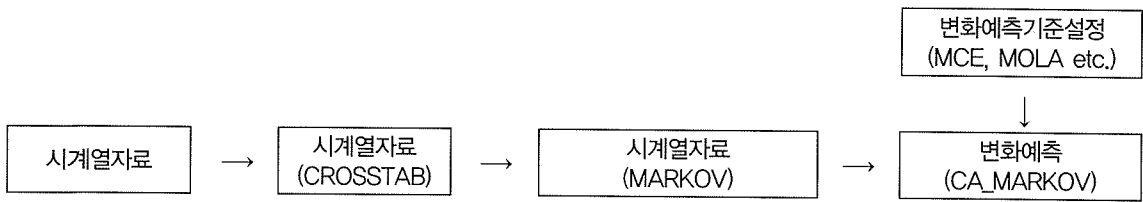


그림 2. Idrisi를 이용한 토지이용변화 예측과정

이러한 조건확률 법칙에 따라서 초기상태가 주어졌을 때 시간 t에 있어서 각상태의 변이확률 행렬의 추정이 가능하다.

4. 토지이용 변화 예측

본 소강좌에서는 셀룰라 오토마타와 Markov chain을 이용한 토지이용변화예측이 가능한 Idrisi Kilimanjaro Edition의 CA_MARKOV 모듈에 대하여 간략히 소개하고자 한다.

Idrisi는 Clark University Clark 연구실에서 개발한 GIS도구로 벡터와 래스터처리가 가능하여 지리정보 및 위성영상분석 등 다양하게 이용되고 있는 툴이다.

Idrisi는 1099년 북아프리카 해안의 Geuta(현재의 Morocco)라는 스페인령 식민지에서 태어나 스페인의 코르도바 대학에서 교육을 받았으며, 유럽 북아프리카, 중동, 중앙아시아 등지를 광범위하게 여행했다. 중세 시대에 있어서 중요한 위치를 차지했던 지도학자의 조사를 부탁받았던 Idrisi는 Palermo에 있는 Rogerian Court에서 학자들 및 기술자들의 공동연구를 이끌었다. 고문헌 조사와 실제 야외조사에 기초한 공동연구결과로부터 얻은 지도와 서적들은 500년 이상 동안이나 중요한 참고자료로 활용되었다.

이와 같은 지리적 공동연구의 정신을 기념하기 위하여 Idrisi 소프트웨어가 제작되었다.

Idrisi는 저렴하고 GIS/원격탐사 통합분석도구로 기능이 다양할 뿐만 아니라 응용프로그램 개발이 쉬워 상업적인 목적보다는 대학에서 주로 많이 이용되고 있다.

현재 Idrisi에서 셀룰라 오토마타와 Markov chain 기능은 Kilimanjaro Edition 이상에서만 가능하다.

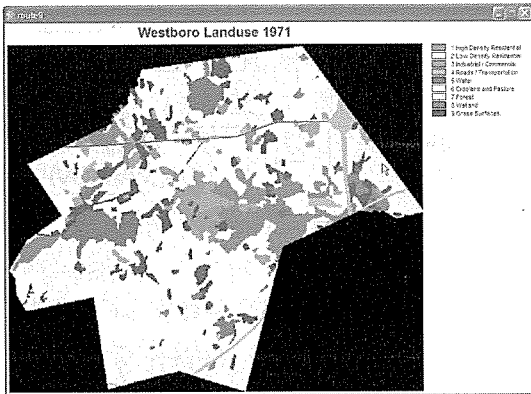
Idrisi에서 토지이용변화 예측을 위한 과정은 그림 2와 같고, 주로 사용되는 기능은 CROSSTAB, MARKOV, CA_MARKOV이며, 토지이용변화 예측을 위한 기준설정을 위하여 Decision 지원기능을 이용할 수 있다.

가. 시계열 자료의 준비

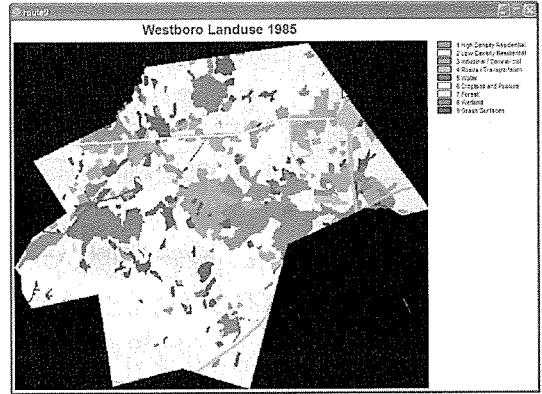
항공사진 또는 위성사진으로부터 분류하여 구축된 토지이용도를 이용하여 분석할 수 있으며, 5년 또는 10년과 같이 특정주기의 전후시기 토지이용도가 필요하다. 소강좌에서는 소프트웨어의 예제로 제공되는 미국 매사추세츠주 Westborough 지역의 1971년, 1985년의 14년 간격의 토지이용도를 이용하였다.

나. CROSSTAB 분석

토지이용변화를 예측하기 이전에 전후시기 영상을 CROSSTAB 기능으로 교차비교하면 토지이용변동 사항을 교차행렬표를 통하여 손쉽게



(a) 1971년 토지이용 현황



(b) 1985년 토지이용 현황

그림 3. Westborough의 토지 이용도

파악할 수 있다. 예제의 분석결과에 의하면 1985년 토지이용은 1971년에 비하여 농경지, 목초지, 산림, 습지, 초지가 주거, 산업, 상업, 도로화가 진행되었음을 간단히 파악할 수 있다.

다. MARKOV에 의한 변이확률행렬 추정

MARKOV 모듈은 1971년과 1985년의 14년간의 토지이용변화에 대한 확률을 파악하고 같은 주기인 14년 후 즉, 1999년의 토지이용변화를 예측할 수 있도록 변이확률행렬을 산정해주는 역할을 한다. 주요입력 사항을 간략히 설명하면 다음과 같다.

- 1) First (earlier) land cover image,
Second (later) land cover image

비교할 전/후 단계의 토지피복 이미지를 입력한다. 토지피복도의 분류코드는 반드시 1부터 시작해야 하고, 분류코드 사이에 공백이 없어야 한다. 예를 들어 6개 분류항목이 있는 토지피복도는 반드시 1, 2, 3, 4, 5, 6의 분류코드로 존재해야 하며 1, 2, 3, 4, 5, 7 과 같이 입력되어 있으면 안 된다. 두 토지피복도에서 각 분류 항목은 정확히 일치하고 있어야 한다.

- 2) Prefix for output conditional probability image

확률조건을 출력하는 이미지의 prefix를 입력하는 란으로 각 분류 항목별 조건확률 이미지 이름 앞에 prefix가 붙게 된다. 만약 legend caption이 있으면 prefix 다음에 legend caption이 붙고, 없다면 prefix 다음에 "class_#" 가 붙는다. 추가적으로 각 확률이미지가 prefix이름으로 그룹화 되어 래스터 그룹파일(.rgf)이 생성된다.

- 3) Number of time periods between the first and second land cover images & Number of time periods to project forward from the second image

첫 번째와 두 번째 토지피복사이의 시간주기와 장래 변화상태를 나타낼 이미지의 시간주기를 입력한다. 사용되는 특정시간단위(년, 10년 등)는 그리 중요치 않으나, 두 주기는 반드시 동일(5년 이면 5년, 10년이면 10년 단위)해야 한다.

- 4) Background cell option

배경영역을 어떻게 처리할 것인지 세 가지 옵션

선 중에 하나를 선택하도록 되어있다.

- Assign 0.0 : 확률이미지의 배경영역에 0.0을 할당한다.

- Assign equal probabilities : 확률이미지의 배경영역에 1/분류 항목수 값을 할당한다.

- Assign relative frequency : 두 번째 입력 이미지에서 큰 영역을 차지하는 토지피복에 대한 높은 확률을 할당한다.

5) Proportional error

입력이미지와 관련된 비례오차(Proportional error)를 입력한다. 입력되는 토지피복이 완벽한 정확도를 가지고 있다면 0.0을 입력할 수 있으나, 일반적으로 0.15를 입력한다. 0.15는 일반적으로 대부분의 토지이용도가 85%의 정확도를 가지고 있다고 보는 것을 의미한다.

라. CA_MARKOV를 이용한 토지이용 변화예측

CA_MARKOV 모듈은 MARKOV에 의해 산정된 변이확률자료를 셀룰라 오토마타를 이용하여 토지이용변화를 예측하는 기능을 수행하며 계산시간이 매우 길다.

1) Basis land cover image

기준이 되는 토지피복이미지를 입력한다. 여기서는 1985년의 토지이용도가 기준이 된다.

2) Markov transition area file

위 MARKOV모듈의 실행결과로 생성된 변이 영역파일을 입력한다. 여기서는 "7185 transition_areas.txt"파일이 된다.

3) Transition suitability image collection

변이에 적합한 영역에 대하여 분석된 여러 향

목들을 그룹화하여 입력하는 부분이다. 토지이용변화 예측결과에 가장 큰 영향을 미칠 수 있는 부분으로, MCE(Multi-criteria evaluation) 분석결과를 입력하게 된다.

4) Output land cover projection

예측결과가 저장되는 파일이름을 입력 한다.

5) Number of Cellular Automata iterations

시간주기와 관련된 사항이다. 여기서는 1971년 ~1985년의 14년간 토지이용변화 자료를 이용하여 같은 주기의 14년 후인 1999년의 토지피복변화를 예측하는 것이므로 같은 주기14년을 입력한다.

6) Cellular Automata filter type

- a) Standard 5X5 contiguity filter,
- b) User-defined filter

CA를 위한 필터를 설정하는 부분이다. 기본값으로 5X5의 연속필터를 입력한다.

CA_MARKOV 모듈에서 가장 중요한 것은 transition suitability image collection 이다. 이것은 MCE 분석과정을 통하여 얻어지는 결과 파일들이다. MCE는 GIS Analysis의 Decision Support에 포함되어 있다. MCE를 요약하면 변화에 적합한 영역을 결정하는 것으로 1) Constraints (강제사항) 설정, 2) Factors (변화의 인자) 설정, 3) constraints 와 Factors 이미지를

0	0	1	0	0
0	1	1	1	0
1	1	1	1	1
0	1	1	1	0
0	0	1	0	0

그림 4. 5×5 contiguity filter

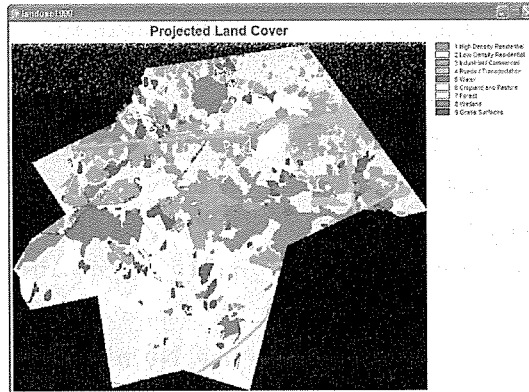


그림 5. 1999년 토지이용변화 예측

MCE 모듈에 입력하여 Transition suitability image를 생성하는 과정으로 요약된다. MCE 분석의 일련과정은 IDRISI Kilimanjaro Tutorial 2-7 ~ 2-11장에 자세히 설명되어 있다.

5. 결과 및 고찰

위 MARKOV 및 CA_MARKOV 모듈을 이용하여 매사추세츠주 Westborough 지역의 1999년 토지이용변화를 예측한 결과는 그림 5와 같다.

토지이용변화 예측결과를 CROSSTAB 기능으로 1985년의 토지이용과 비교하면 1999년 가장 큰 변화를 나타내는 영역은 초지로 나타났다. 초지는 저밀도 주거지역과 산업/상업지역으로 당초 면적의 30% 정도가 변화되는 것으로 예측하였고, 경작/목초지의 경우 약 19% 정도가 산업/상업지로, 산림의 17%가 주거/산업/상업지로 변화되는 것으로 예측하였다. 또한 산업/상업지역의 7% 영역에 도로가 확장될 것으로 예측하였다.

토지이용변화를 예측하는 것은 매우 흥미로운 연구과제이며 셀룰라 오토마타와 Markov chain에 의한 토지 이용변화 예측 시 다음 두 가지 사항만큼은 중요하게 고려해야 한다고 생각된다. 첫 번째로 어떤 토지이용변화를 정확

하게 예측한다는 것은 불가능하지만 과거의 시계열적 자료를 이용하여 어떤 시점의 토지이용을 예측하려면 사용되는 자료의 정밀도가 영향을 많이 미칠 것으로 판단되었다. 일반적으로 Landsat TM, ETM+ 영상을 이용하여 분류한 토지피복도는 일반화과정을 거치지 않고 그대로 사용되는 경우가 있다. 어떤 토지이용경계에 대하여 내부에 다른 토지피복 격자가 포함되어 있는 상태 그대로 사용하는 것을 말한다. 행정구역 또는 유역개념에서 분류된 클래스의 격자수가 수적인 통계치와 비슷하다고 해서 그 토지피복 분류가 잘되었다고는 할 수 없을 것이다. 이러한 일련의 일반화과정을 거치지 않는다면 10년 전에 산림이 5년 전에는 도로가 되고 다시 현재에는 산림이 되는 현상이 반복될 수 있다. 특별한 경우가 아닌 이상 산림에서 도로로 변경되었다면 거의 대부분 도로로 토지이용이 유지될 것이다. 따라서 통계적인 토지피복 분류결과에 연연하지 말고 도상 또는 영상에서 확실한 지역들은 분류결과를 수정하고 서로 군집화하여 일반화함으로써 보다 정확한 토지이용변화 예측에 이용할 수 있을 것이다. 두 번째로는 어느 지역의 토지이용변화를 예측하고자 한다면 향후의 도시개발계획지역을(각 지자체별로

2016년 또는 2020년까지의 도시개발 계획도를 비치하고 있으며 열람 가능토록 하고 있음)포함하거나, 개발제한 사항 등 지역/지리적인 어떠한 규칙과 요소를 토지이용변화 예측에 필수적으로 고려해야 한다고 생각한다. 본 소강좌에서는 자세히 설명하지 않았으나 Idrisi를 이용하는 경우 MCE와 같은 Decision support 도구를 사용하여 일련의 규칙들을 생성하고 적용함으로써 더욱 타당성 있는 토지이용변화 예측이 가능할 것이라고 생각된다.

참고문헌

1. 박세권의, 1988, 동기작물 식부면적 예측 전산모형 개발-Markov Chain 기법을 이용하여, 한국농촌경제연구원 통합농촌정보시스템 연구보고.
3. 정재준, 한동엽, 김용일, 이재원, 2001, 11, 셀룰라 오토마타를 이용한 수도권의 도시성장 예측, 한국 GIS학회지, p(3), pp.397-412.
4. Bryan, B. A., 2000, Modeling the propagation of land clearance using cellular automata and the implications for nature conservation, 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4).
5. Eastman, J. R., 2003, Idrisi Kilimanjaro Tutorial, Clark Labs.
6. von Neumann, J., 1966, The Theory of Self-Reproducing Automata, University of Illinois Press.