GeoInformatics · Practical 1

Dave Murray-Rust, Corentin M. Fontaine

CECS, School *of* GeoSciences

12th January 2009

This first week's practical encourages you to understand the principles of a cellular automata (CA). It applies a theoretical equation that describes urban growth into an agent-based framework, Repast Symphony (Java coded). You might wish first to read what is required by way of assessment, at the end of this document.

The following exercise is based on a working paper (Batty et al., 1999) that you may want to read for more details (available on WebCT).

You are given a CA framework programmed in Java with Repast Symphony. First, you put the urban growth rules described below in the model and run it. Play around with the parameters and compare the outputs. Second, you add a rule to make permanent cells become available again for redevelopment. Third, you alter the model to make the rule stochastic.

You will need to read the document (also available on WebCT) about using Repast and checking out course content from SVN. After reading the theory, you should start by checking out the Geoinformatics1 project.

1 Theoretical background

Cellular automata (CA) concept appears with two mathematicians of the 40s, S. Ulam and J. von Neumann. Their application starts with the rise of the computers in the 70s and they become very popular after J. Conway created his well-known 'Game of Life'. They are very powerful for spatial and dynamic modelling. The main objective for using a CA is to seek to represent a complex and dynamic (spatial) system at the aggregate level with as few as possible rules to describe local behaviours.

CA can be described as a group of cells which have simple and local interactions leading to a global more complex behaviour. CA are discrete: space, time and cell states are integers. At each time t, a cell c is at the state k. Rules of transition define how cells change state at each time step $(t + 1)$ and usually consider the local situation at time t: the states of the target cell and of its neighbours. The neighbourhood is explicitly defined in the rule(s).

CA are defined by a minimum of 4 elements:

1. A group of cells which have any dimension and an infinite size. In practice, dimension are limited to 1 (a line), 2 (a plane) or 3 (a volume). In programming, the infinite size is approached with a finite shape. An infinite line is represented by a circle (dim 1). An infinite plane is represented by a torus, a finite rectangle whose opposing edges are considered connected (dim 2). The group of cells has an initial configuration, i.e. a defined state for each of the cell at time $t = 0$.

2. A states space that defines the different k states a cell can have. In theory, it can be a n-dimensional space. von Neumann has mathematically studied a 29 states CA. In practice, it is limited to a few, e.g. Conway's 'Game of Life' has two (dead or alive).

3. A neighbourhood function that defines which cells have an influence over the target cell. In theory, the neighbourhood is defined by a circle of radius r, measured in number of cells. In practice, this is usually limited to the closest cells, i.e. the 4 contiguous cells (von Neumann neighbourhood, $r = 1$) or the 8 surrounding cells (Moore neighbourhood $r = 1.44$).

4. A transition function that defines the dynamic of the CA, i.e. how the neighbourhood influence at each time step the target cell according to the state configuration at the previous time step. The rules apply to all cells simultaneously. In other words, CA work cells in parallel. The number of rules can rapidly be unmanageable because there is r^k different possible neighbourhood configurations. In practice, the transition function is based on thresholds and/or probabilities to reduce the number of rules, e.g. "if x cells within r are of state k, then ..., otherwise ...", "there is a $x\%$ chance for the cell to change into state k if r ...",etc. The fundamental difference between these two example is the stability of the outcome. For a given neighbourhood configuration, the deterministic approach (rules with thresholds only) alway produces the same result at each time step. The stochastic approach (rules with probability) may lead to very different configurations after some time. There is as many probabilitic rules as there are statistical functions.

The 5 key notions to remember when working with CA:

1. **Discrete**: space, time and cell states are integer measures.
2. **Homogeneous**: same rules apply to all cells.
3. **Synchronous**: rules apply simultaneously to all cells.
4. **Spatial**: the neighbourhood configuration drives the rule.
5. **Deterministic** or **Stochastic**

2 CA in practice: the dynamics of urban sprawl

You are going to explore the process of suburbanization (urban sprawl). It is considered as the cumulative growth of the city through additions to its periphery (after Batty et al., 1999, p.6). Urban development is divided at any time t into three constituent parts:

- $P(t)$: development which is established, i.e. surrounded by other development
- $N(t)$: new development which has just made the transition from undeveloped land
- $A(t)$: available land which drives the process of development in the first place

The growth limit is set to when what can be brought onto the market is reached. The city and its hinterland are capacitated so that:

$$P(t) + N(t) + A(t) = C \quad \forall \quad t = 0, 1, 2, ..., T \tag{1}$$

Equation 1 immediately gives an important condition on the growth mechanism (note the equation is different from Batty et al. (1999) where it is incorrect):

$$\frac{dP(t)}{dt} + \frac{dN(t)}{dt} + \frac{dA(t)}{dt} = 0 \tag{2}$$

The mechanism in the model is one-directional: available land passes to new development which then passes to established development as $A(t) \rightarrow N(t) \rightarrow P(t)$. Figure 1 shows that new development is an essential filter in the growth process which is articulated as the transition between undeveloped or available land and established development. The filtering process can be expressed as:

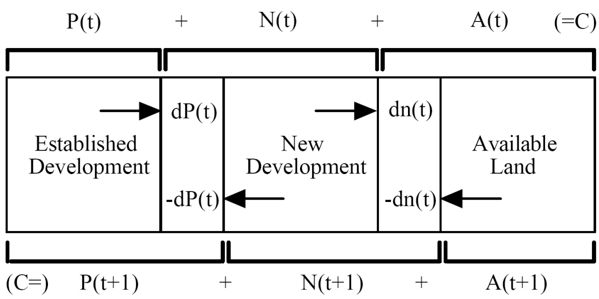
$$\frac{dN(t)}{dt} = \alpha N(t)A(t) - \gamma N(t) \tag{3}$$

And the other parts of the system can be derived from equation 2 and 3 as follow:

$$\frac{dP(t)}{dt} = \gamma N(t) \tag{4}$$

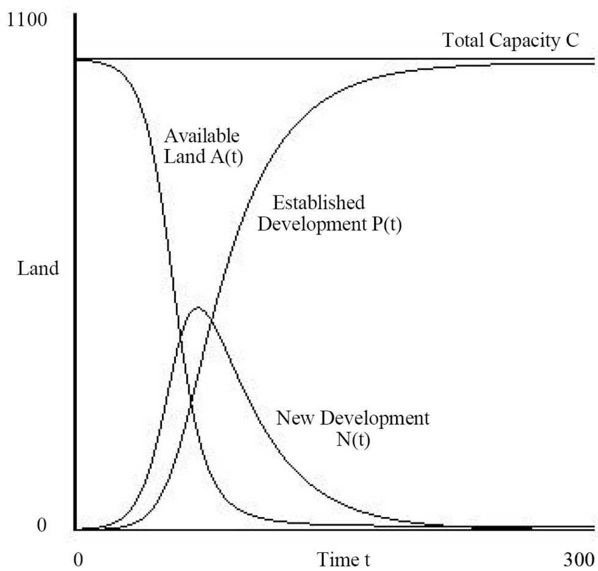
$$\frac{dA(t)}{dt} = -\frac{dN(t)}{dt} = -\alpha N(t)A(t) \tag{5}$$

Figure 1: An aggregate model of land development (after Batty et al., 1999, p.6)



It is quite easy to show how this aggregated model behaves and to compute the steady state values of the three key components for different parameter values as well as the parameter values which are able to reproduce a steady state from fixed initial conditions. Figure 2 shows graphically the behaviour of the model components. Mature development grows according to the classic S-shaped curve (a classic logistic growth) while available land declines as a mirror image of this.

Figure 2: Trajectories of development from the basic model (after Batty et al., 1999, p.14)



The translation of the continuous model to the cellular space involves several changes in interpretation. The cellular space is based on a regular tessellation of grid squares, the cells, with coordinates (x,y). The sum of the cells indicates the carrying capacity (C) of the system. In analogy to the continuous model, the three possible states for a cell are: available land, new development or established development. To indicate cell occupancy, we set the relevant activity equal to 1 with a constraint of no more than one activity per cell at any given time t:

$$A_{xyt} = 1 \quad \text{or} \quad N_{xyt} = 1 \quad \text{or} \quad P_{xyt} = 1 \qquad \forall \quad x, y, t \tag{6}$$

and

$$A_{xyt} + N_{xyt} + P_{xyt} = 1 \qquad \forall \quad x, y, t \tag{7}$$

The development process of the continuous model is the transition function of the CA. Both consists of three key transitions based on: (1) the diffusion effect which adds to the available land supply in the neighbourhood of each unit of new development, (2) the transition from available land to new development, and (3) the transition of new development to established development.

The diffusion effect is computed around each cell of new development by making adjacent Moore cells available for development:

$$\begin{array}{rll} \text{if} & N_{xyt} = 1 & \text{where} \quad i = x \pm 1, j = y \pm 1 \\ \text{then} & A_{ij(t+1)} = 1 & \\ \text{unless} & N_{ijt} = 1 \quad \text{or} \quad P_{ijt} = 1 & \end{array} \tag{8}$$

The second transition is from available land to new development:

$$\begin{array}{rl} \text{if} & A_{xyt} = 1 \\ \text{then} & N_{xy(t+1)} = 1 \quad \text{and} \quad A_{xy(t+1)} = 0 \end{array} \tag{9}$$

The last transition is from new to established development. A cell of new development steps into the mature state when it is surrounded by new or permanently developed land:

$$\begin{array}{rl} \text{if} & \sum_i \sum_j [N_{ijt} + P_{ijt}] = 8 \\ \text{then} & P_{xy(t+1)} = 1 \quad \text{and} \quad N_{xy(t+1)} = 0 \end{array} \tag{10}$$

This is a very simple process for in this form, available land and new development are separated by one time period. From a single seed, the process produces a wave of available land advancing one time period ahead of new development while behind the wave, new development becomes established. On a grid, this can be visualized as a square band of land one cell wide advancing in front of a band of new development one cell wide which is then converted to established development, the lag separating each component being one time period. When the edge of the system is approached, growth immediately stops for the capacity limit is only recognized locally (Batty et al., 1999, p.18). Of course, it is possible to vary this process by changing the size of the neighbourhood or incorporating differential time lags or adding randomness to the transition (see next section).

TASKS:
(1) Translate the three rules (diffusion, new development, established development) into the model code.

> **HINT:** Look at the file `Cell` for where to place these changes

(2) Run the model and observe the predicted behaviour.

3 CA in practice: urban dynamics

It is much more realistic to assume that new development is generated from available land according to a certain probability, thus reflecting the fact that different developers of different cells vary in the way they finance development and react to the market. Such an extension would clearly break spatial symmetry, producing development clusters with irregular edges much more characteristic of real cities. A straightforward modification is by introducing noise into the system with a random component into equation 9:

$$\begin{array}{ll} \text{if} & A_{xyt} = 1 \quad \text{and} \quad random(\Lambda_{xy}) < \Phi \\ \text{then} & N_{xy(t+1)} = 1 \quad \text{and} \quad A_{xy(t+1)} = 0 \end{array} \tag{11}$$

TASKS:
(3) Incorporate the above alteration (release probability, building probability) into the code.

> **HINT:** Use the existing variables in `Cell` for the probabilities

(4) Run the model several time for several values of probabilities.
(5) Compare both deterministic and stochastic outputs.

4 CA in practice: advanced urban dynamics

One of the key processes of urban dynamics is its redevelopment or its regeneration process. Two additional components need to be included to consider how established development reverts back to available land and enters the development process as new development again: the age of a unit of development (B_{xyt}) and an age threshold (τ). Only both combined can indicate the time at which a cell must be redeveloped or renewed through demolishing the structure, clearing the land, and placing the land in question back into the pool of available land to be considered for new development.

$$\begin{array}{ll} \text{if} & t - B_{xyt} = \tau \\ \text{then} & A_{xy(t+1)} = 1 \quad \text{and} \quad P_{xy(t+1)} = 0 \end{array} \tag{12}$$

This rule can also be randomized on the assumption that there is variation in this decision.

$$\begin{array}{ll} \text{if} & t - B_{xyt} = \tau \quad \text{and} \quad random(\Lambda_{xy}) < \Phi \\ \text{then} & A_{xy(t+1)} = 1 \quad \text{and} \quad P_{xy(t+1)} = 0 \end{array} \tag{13}$$

TASKS:
(6) Add the ageing rule into the model.

> **HINT:** The cell's ages are already updated automatically, and reset to 0 every time they change state

(7) Run to observe the system behaviour and compare with previous ones.
(8) Randomise the ageing rule.
(9) Run the model several time for several values of probabilities.
(10) Compare both deterministic and stochastic outputs.

5 Assignment 1 (10% of final marks)

Write no more than 300 words to present and discuss the different outputs: what is the influence of redevelopment? of randomness? ... Reproduce no more than six graphs of the aggregate evolution of the system to support your analysis.

Extra credit: To which extent different seed shapes (such as a circle, a line, a star,...) influence the results?

HINT: To explore different seed shapes, look at `CityInitialiser`, which contains a selection of seeds for you to use

References

Batty, M., Xie, Y., Sun, Z., November 1999. The dynamics of urban sprawl. Working Paper Series 15, Centre for Advanced Spatial Analysis, University College of London.