[AM-07-069] Cellular Automata

Abstract

Cellular automata (CA) are simple models that can simulate complex processes in both space and time. A CA consists of six defining components: a framework, cells, a neighborhood, rules, initial conditions, and an update sequence. CA models are simple, nominally deterministic yet capable of showing phase changes and emergence, map easily onto the data structures used in geographic information systems, and are easy to implement and understand. This has contributed to their popularity for applications such as measuring land use changes and monitoring disease spread, among many others.

Keywords: cellular automata, complex systems, dynamic systems, Game of Life, land use change, modeling

Author & citation

Clarke, K. (2017). Cellular Automata. The Geographic Information Science & Technology Body of Knowledge (3rd Quarter 2017 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2017.3.9

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). CA model structure. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Definitions
2. Introduction
3. Origins and Development
4. CA Principles and Theory
5. CA in Geography
6. Advantages and Limitations of CA

1. Definitions

cellular automata: A regular framework of cells, each in one of a finite number of states. The states of all cells in the framework are updated simultaneously in discrete time steps during which the state of each cell is changed according to a set of rules that depend on the state of the cell and those of its neighbors at the previous time step.

2. Introduction

Cellular automata (CA) are computational abstractions that allow simulations of spatially

distributed phenomena and their dynamics over time. CA are among the simplest ways in which complex systems behavior can be demonstrated, and consequently have been popular in modeling geographic systems that show complexity, including land use change, urban growth, human and vehicle movement, the spread of disease, sand dune formation, and vegetation change. Advantages of CA models are that they are simple, nominally deterministic yet capable of showing phase changes and emergence, map easily onto the data structures used in geographic information systems, and are easy to implement and understand.

A CA consists of six defining components: a framework, cells, a neighborhood, rules, initial conditions and an update sequence. The framework sets the constraints and limits of the system, for example constraining a CA to one or two dimensions, and deciding what happens at a region's edges. Geographical examples usually choose a study area or region, and mask the extent so that the CA cannot apply beyond them. Cells are identical spatial units, often squares or pixels, that cover the geographic extent at a chosen resolution. Geographical data for CA are often chosen to coincide with satellite image pixels or geographic grids, that are already partitioned into tessellations. The neighborhood is the region of interaction among the cells. In classical CA, this is the immediately adjacent two cells in one dimension and the Moore (8) or Von Neumann (4) neighbors in two dimensions. Changes within a cell are made by applying the rules within the neighborhood, and changing the cell if the situation warrants it. Cells are also defined as having mutually exclusive states, a set of classes to which the cell can belong. For example, the set of states for a land use cell could be {Urban, Water, Agriculture, Rangeland, Forest, Wetland, Bare, Tundra}. All cells in the CA must be in one of these states at every time period. The rules are the determinants of the conditions under which the active cell will change states. Simple CA models have few rules, for example in Conway's classic 1970 CA game of Life (Adamatzky, 2010), cells can only have the states alive (1) or dead (0) on a square grid with an 8 cell neighborhood. At each step in time, two rules are applied based on how many of the neighbors are alive: a live cell will die if it has fewer that two or more than three live neighbors; a dead cell will come to life if it has exactly three live neighbors; otherwise it remains unchanged (Gardner, 1970; 1972). Figure 1 shows an initial distribution of live cells, and the configuration after sequences of time steps. In the design of geographical CA models, the rules set and its construction, or derivation from actual change data, is often a major part of the CA model design task.

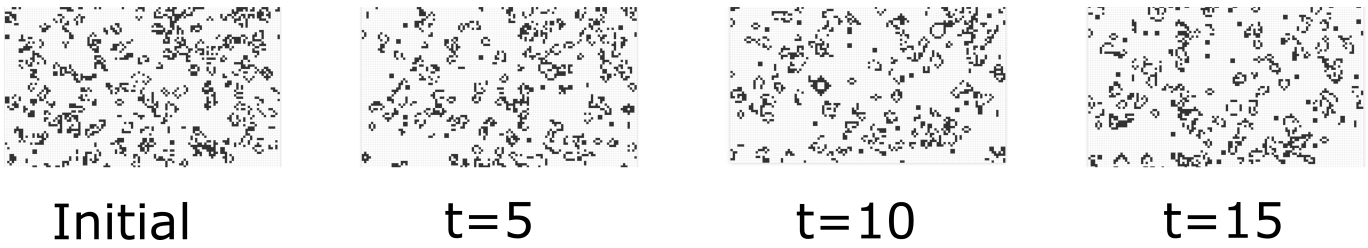


Figure 1. The initial conditions and further time steps for the Game of Life. White cells are "dead", and black "alive."

The initial conditions are the start set or seed generation for the model, usually an actual spatial configuration or map of the phenomenon at the start period for the modeling, such as a land use map of a city in 2010, when the goal is to simulate changes to 2030. The choice of initial conditions is important. For example in Conway's game of Life described above, setting all cells to dead will produce no change, while setting all cells to alive will

make all cells die in just one iteration, a phase change. Lastly, the cells must be updated. Usually during one model time period all rules are applied to each cell in order for all rows and columns of cells in the framework, with the new states going into a replica buffer grid, then the replica replaces the original. Many models use discrete time steps, such as 10 minutes, one month, or a year for each time step. Updates continue until the target time is reached, or a given amount of change has taken place.

3. Origins and Development

CA's origins lie in the 1940s at Los Alamos National Laboratory, where Stanislaw Ulam was analyzing crystal growth on a lattice network while John von Neumann was working on the problem of self-replicating systems. Ulam suggested that a self-replicating system could be simulated as a mathematical abstraction, a two-dimensional cellular automata with a large number of states. Von Neumann proved mathematically that a CA pattern could indeed make endless copies of itself, proving that a self-replicating machine was possible, an important advance in solving the halting problem (von Neumann, 1951: von Neumann & Burks, 1966). However, CA remained at the margins of mathematics until 25 years later. During the 1970s, the Game of Life was devised by John Conway and popularized by Martin Gardner in his column in Scientific American. Analysis of Conway's game showed that entire systems of cellular automata could grow, extinguish themselves, or jump between chaos and organization. Beginning in 1983, Stephen Wolfram began an exhaustive systematic investigation of the behavior of one-dimensional CA, formalizing the types of patterns that could be derived (Wolfram, 1983).

That complex systems behavior can be simulated from the simplest of CAs led Wolfram to publish A New Kind of Science (Wolfram 2002). Wolfram noted that cellular automata have significance for all scientific disciplines, including biology and geography. Using only a one dimensional CA, Wolfram showed that any mathematical function can be simulated, implying that any phenomenon and any dynamic could be modeled by CA. He proposed four possible classes of CA. In Class 1, patterns quickly evolve into a stable pattern and randomness disappears. In Class 2, patterns quickly evolve into an oscillating structure, with some randomness remaining. In Class 3, patterns evolve into pseudo-random or chaotic structures in which regular structures are quickly eliminated by dissipating randomness. In Class 4, any initial pattern evolves into complex structures with unpredictable interactions. The fact that CA has been proven to replicate system behavior from extinction to stability and from chaos to complexity makes them especially suitable for geographical problems and their dynamics.

4. CA Principles and Theory

Most of the principles of CA can be learned by simple experiments with Conway's game of life. Some online software programs with which to experiment are: Golly, an open-source simulation system for Life and other cellular automata that includes extremely fast generation, and Python scripts for both editing and simulation; Mirek's Cellebration, a free 1-D and 2-D cellular automata viewer, explorer and editor for Windows; and Xlife, the standard UNIX X11 Life simulation application now ported to Windows, with up to eight possible states per cell. Experiments will reveal that regardless of initial conditions, over

time one of three outcomes will result: (1) all cells will die; (2) certain stable and unchanging geometrical configurations will emerge; or (3) stable but changing patterns that oscillate, move, or even replicate will leave the CA under constant change. Named emergent patterns that do not change include the Block, Beehive, Loaf, Boat and Tub. Those that survive by oscillation over different length time periods include the Blinker, Toad, Beacon, Pulsar and Pentadecathlon. Patterns that survive by movements are the Glider and Lightweight Spaceship. Meta structures include glider guns, rakes and puffer breeders, which sometimes take very long time periods to build and then stabilize as generators of gliders and other patterns. Examples of each are given on the Wiki entry for “Conway's Game of Life.” It is these more complex structures, Wolfram’s class 4, that are of greatest interest for models of complex spatial phenomena, since stability is rarely involved in urban growth, land use change, or traffic movement.

5. CA in Geography

In geography, early considerations of cellular automata were by Tobler (1979) and later Couclelis (1985). The emergence of GIS and remote sensing as sources of gridded geographic data suitable for cellular modeling, and improvements in structured programming languages to manipulate grids raised the level of interest in CA during the 1990s. Much research was centered at the Santa Fe Institute, and carried into the geographic information science community after the 1996 Santa Fe GIS and Environmental Modeling conference. A host of CA models have been used to simulate urban growth and its consequences, reviewed by Sante et al. (2010). CA models have proven highly popular in modeling land use change, and are also used throughout the spatial sciences.

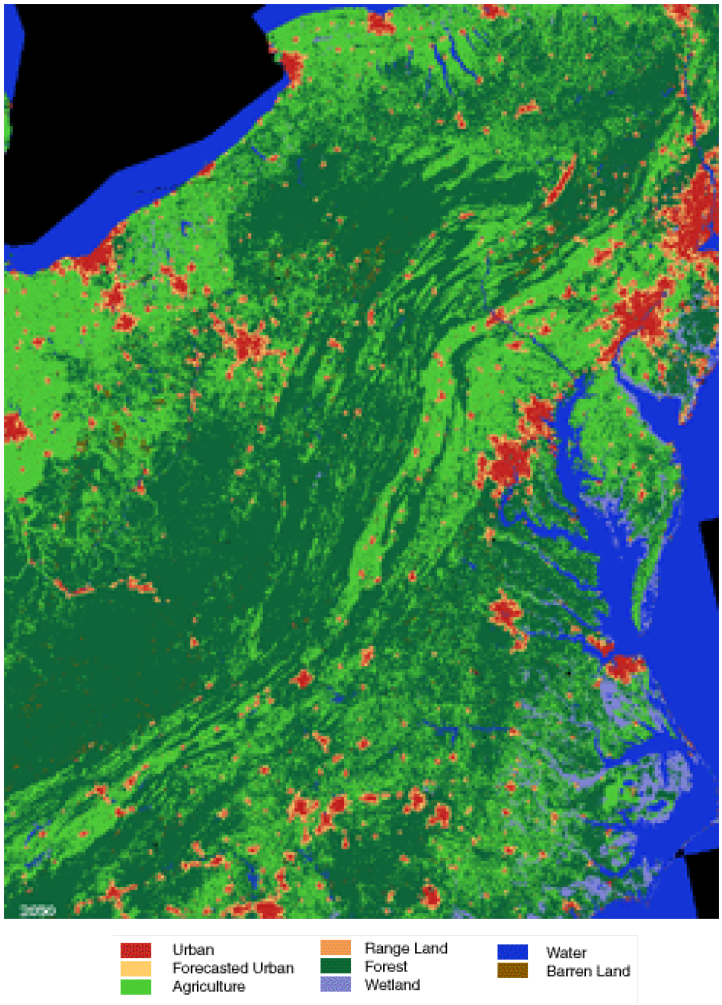


Figure 2. Cellular Automaton Model (SLEUTH) Forecast for 2050 of Land Use in the Eastern United States for the Mid-Atlantic Integrated Assessment by the U.S. Environmental Protection Agency. Source: Candau, J., Rasmussen, S. and Clarke, K. C. (2000) A coupled cellular automaton model for land use/land cover dynamics. 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs. Banff, Alberta, Canada, September2 - 8, 2000. [Public Domain image EPA/USGS].

Recent research has investigated constrained CA models, where CA processes are limited by weighted potentials, geographic distributions or economic constraints. CA methods have proven effective in modeling traffic flow, human movement, natural systems such as erosion and streams, wildfire, and many other areas. CA have proven useful in assisting the classification process for remotely sensed images. Many applications use geocomputation and high performance computing with CA models, since the simple operations are easily parallelized. Since about 2010, there has been a new interest in CA model rule selection and model calibration using machine learning methods, such as support vector machines, genetic algorithms and artificial neural networks. These techniques attempt to automate the rules for a CA by using actual change information, by analysis of the classes of the surrounding pixels. For example, a 2000 and a 2010 land use map can be differenced, and the cells that changed states isolated and used to extract the change conditions as a function of contributory factors, such as distance or topographic slope.

6. The Advantages and Limitations of CA

There is a clear match between geographical regions and CA. GIS and remote sensing allow data to be rapidly formatted, resampled and selected to provide layers of geographic information for CA modeling. Indeed, some GIS packages include CA tools within the software. They also set the initial conditions, i.e. the land configuration prior to the time period used by the CA, and can be used to display and evaluate future conditions by map comparison. When multiple time slices of data exist, the model can be calibrated using hind-casting, i.e. the model map distribution at any time can be evaluated against the actual data to test model accuracy. In calibration, the rules of the CA and often the weights given to those input layers that influence the CA can be adjusted until the model provides the best simulation of the “present”, or at least the time period with the most recent data. The calibrated model can then be run into the future to produce forecasts. It is the ease of data assembly, along with the simple nature of CA models, that are the method’s chief advantages.

An important component of CA models is that they are “bottom-up” models. All change takes place at specific locations, and aggregate behavior emerges from the coalescence of the local. Thus all interactions in CA are independent and local, with no influence-at-a-distance effects. At high spatial resolutions, these influence distances can be very small, just meters on the ground. Of course many have argued that this makes CA unsuitable for modeling systems with more complicated and longer range interactions, such as social, demographic and economic models. Others have pointed out that CA models also absorb immense amounts of computer time, and can take months or years of work to yield accurate forecasts. Increased computing power, lower computing costs and high performance computing have countered this claim. Lastly, in CA models an equal time step is assumed (say one year) that may differ from the appropriate time resolution of the actual city or data, where a census may be available only decadally.

Overall, CA have enjoyed an extended period of being among the most frequently applied spatial models, especially for spatially distributed dynamical systems. They are flexible enough to integrate well with other forms of modeling, such as Agent-Based models and Multi-Criterion Evaluation, and can be evaluated and calibrated with many geostatistical and machine learning methods. Refinements and improvements in CA methods continue unabated, and the spread of the models to other disciplines and applications will ensure this status for the immediate future.

References

Adamatzky, A. (ed.). (2010). Game of Life Cellular Automata. Springer.

Couclelis, H. (1985). Cellular Worlds: A Framework for Modeling Micro—Macro Dynamics. Environment and Planning A: Economy and Space. 17:585-596.

Gardner, M. (1970). Mathematical Games: The fantastic combinations of John Conway's new solitaire game "life". Scientific American, 223, 120-123.

Gardner, M. (1972). Mathematical Games. Scientific American, 226, 114-118. Reprint of the 1970 article.

Sante, I., Garcia, A. M., Miranda, D., & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landscape and Urban Planning, 96, 108-122.

Tobler, W. R. (1979). Cellular Geography. In: Gale, S., Olsson, G. (eds) Philosophy in Geography. Theory and Decision Library, vol 20. Springer, Dordrecht.

von Neumann, J. (1951). "The general and logical theory of automata" in L.A. Jeffress, (ed.) Cerebral Mechanisms in Behavior – The Hixon Symposium, John Wiley & Sons, New York, 1-31.

von Neumann, J., & Burks, A. W. (1966). Theory of Self-Reproducing Automata. University of Illinois Press.

Wolfram, S. (1983). Statistical Mechanics of Cellular Automata. Reviews of Modern Physics, 55(3), 601-644.

Wolfram, S. (2002). A New Kind of Science. Wolfram Media.