

Urban simulation Using Neural Networks and Cellular Automata for Land Use Planning

Hamid Kiavarz Moghaddam, Farhad Samadzadegan

(Hamid Kiavarz Moghaddam ,Master Graduated in GIS, Department of Geomatics Engineering , Faculty of Engineering, University of Tehran, Karegar north St.,Tehran,Iran,hkiavarz@gmail.com)

(Associated Professor, Department of Geomatics Engineering, Faculty of Engineering, University of Tehran, Karegar north St., Tehran, Iran, samadz@ut.ac.ir)

1 ABSTRACT

Cellular automata models consist of a simulation environment represented by a gridded space (raster), in which a set of transition rules determine the attribute of each given cell taking into account the attributes of cells in its vicinities. These models have been very successful in view of their operability, simplicity and ability to embody logics- as well as mathematics-based transition rules in both theoretical and practical examples. Even in the simplest CA, complex global patterns can emerge directly from the application of local rules, and it is precisely this property of emergent complexity that makes CA so fascinating and their use so appealing.

The calibration of CA¹ models is very difficult when there is a large set of parameters. In the proposed model, most of the parameter values for CA simulation are automatically determined by the training of artificial neural network. In this paper ANN² based CA urban growth simulation and prediction of Esfahan over the last four decades succeeds to simulate specified tested growth years at a high precision level. Some real data layer have been used in the ANN-CA simulation training phase such as 1990 while others used for testing the prediction results such as 2001. Next step includes running the developed ANN- CA simulation over classified raster data for forty years in a developed .An ArcGIS extension has been developed to defined ANN-CA algorithm on real urban growth pattern. Uncertainty analysis is performed to evaluate the precision of the simulated results as compared to the historical real data. Evaluation shows promising results represented by the high average accuracies achieved.. The average precision for the predicted growth images 1975 and 2001 is over 90 %. Modifying ANN-CA model. This modification is based on the urban growth relationship for Esfahan over time as can be seen in the historical raster data. The study shows that the model has better accuracy than traditional CA models in the simulation of nonlinear complex urban systems.

2 INTRODUCTION

Cellular automata are powerful spatial dynamic modeling techniques that have been widely applied to model many complex dynamic systems. Recently, a variety of urban CA models have been developed to simulate either artificial or realistic cities (Batty and Xie, 1994; Clarke et al, 1997; Li and Yen, 2000; White and Engelen, 1993; Wu, 1998; Wu and Webster, 1998). Cities, like most geographical phenomena phenomena, are complex nonlinear systems involving spatial and sectoral interactions which cannot easily be modeled with the functionalities of current GIS software (Batty et al, 1999).CA-based approaches are useful in the study of urban and regional spatial structure and evolution..A critical issue in CA simulation is the provision of proper parameter values or weights so that realistic results can be generated. Real cities are complex dynamic systems that require the use of many spatial variables in CA simulation. Each spatial variable makes a contribution to the simulation and its influence is determined by its associated parameter or weight in the simulation. A variable associated with a larger parameter value usually indicates that it is more important than other variables with small parameter values. There are usually many parameter values to be defined in a CA model and the results of CA simulation are very sensitive to the parameter values (Wu, 2000).

Empirical data can be used to calibrate CA models to find suitable parameter values. Calibration is important to the generation of the best fit to actual urban development. Calibration procedures have been discussed in nonlinear dynamic spatial Interaction models (Lombardo and Rabino, 1986).

3 NEURAL NETWORK CA CALIBRATION AND SIMULATION

This section shows the transition rules of CA simulation are represented by the neural network calibrated by the empirical data. The essential part of the transition rules. In general CA models are to estimate the

¹ .Cellular Automata

² Artificial Neural Network

conversion probability between states. For example, a simple urban CA model can be expressed by the following neighbourhood-based transition rules (Batty, 1999):

if any cell $\{x \pm 1, y \pm 1\}$ is already developed,
then $P_d \{x, y\} = \frac{\sum_{i,j \in \mathcal{V}} P_d \{i, j\}}{8}$ and
if $P_d \{x, y\} >$ some threshold value,
then cell $\{x, y\}$ is developed with some other probability $p\{x, y\}$

Where $P_d \{x, y\}$ is the urban development probability for cell $\{x, y\}$ and cell $\{i, j\}$ is the set of all the cells which are from the Moore neighbourhood \mathcal{V} including the cell $\{x, y\}$ itself. The probability of a cell being developed is decided by the number of already developed cells in the neighbourhood. Usually, there is a higher chance of a cell being developed if it is surrounded by more developed cells.

Simulation based on a single factor of developed cells in the neighbourhood cannot address complicated urban systems. More factors have been incorporated in CA models to improve simulation performance for either hypothetical or realistic applications. Distance, direction, density thresholds, and transition or mutation probabilities have been included in the transition rules of various CA models (Batty and Xie, 1994; 1999). Various types of constraints based on site features can also be used to regulate development patterns for land-use planning (Li and Yeh, 2000; Yeh and Li, 2001a). An increase in the number of variables will result in an increase of the number of parameters in CA models. There are few concerns about the exact values of parameters if CA models are used only for hypothetical studies. However, when CA models are applied to the simulation of real cities, suitable parameter values have to be determined through some calibration procedures. The ANN-CA model consists of two separate parts—using the neural network to obtain the parameter values automatically based on training data, and using the neural network to Urban growth is dependent on these variables which may include various types of proximities, amount of development, and site conditions. A regression model or MCE method may not be the best way to reveal relationships because urban systems involve complex nonlinear processes. Neural networks can be designed to estimate development probability each iteration of the CA simulation. The neural network may have three layers (Fig.1): one input layer, one hidden layer, and one output layer. The input layer has n neurons corresponding to the n variables. The hidden layer may also have neurons. The output layer has only one neuron which indicates development probability.

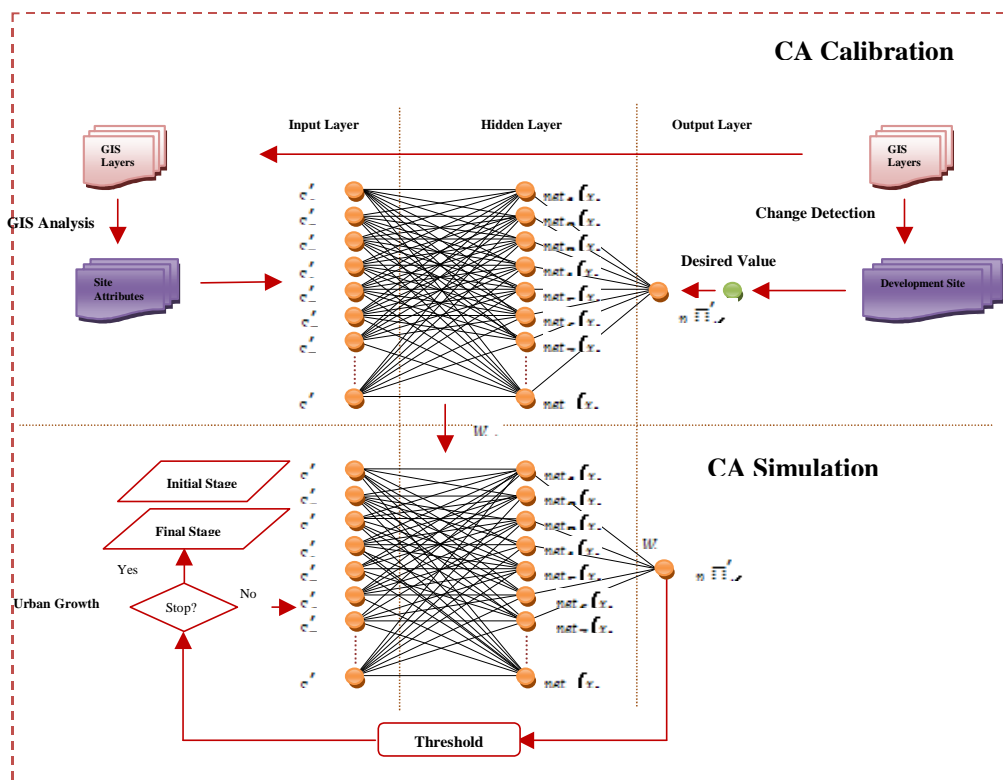


Fig. 1: A cellular automaton model based on neural networks

At each iteration, the site attributes of a cell will be input into the first layer and the neural network will determine its development probability at the output layer. Experiments indicate that it will be more appropriate for all original data to be converted into a range between 0 to 1 before they are input into neural networks (Gong, 1996). This is similar to data normalization in that it uses maximum and minimum values in scaling the original data set. Scaling variables treats them as equally important inputs to neural networks and makes them compatible with the sigmoid activation function that produces a value between 0 and 1. The following linear transformation is used (Equation 1):

$$S'_i = \frac{S_i - \text{minimum}}{\text{maximum} - \text{minimum}} \quad (1)$$

The algorithm for the CA model is devised using a neural network. In the neural network, the signal received by neuron j of the hidden layer from neuron i of the first input layer for cell x is calculated from Equation 2:

$$\text{net}_j(x, t) = \sum_i W_{i,j} S'_i(x, t) \quad (2)$$

Where x is a cell, $\text{net}_j(x, t)$ is the signal received for neuron j of cell x at time t , and $S'_i(x, t)$ is the site attributes for variable (neuron) i . The activation of the hidden layer for the signal is (Equation 3):

$$\frac{1}{1 + e^{-\text{net}_j(x,t)}} \quad (3)$$

The development probability (P_d) for cell x is then calculated from Equation 4:

$$P_d(x, t) = \sum_i W_j \frac{1}{1 + C} \quad (4)$$

A stochastic disturbance term can be added to represent unknown errors during the simulation. This can generate patterns that are closer to reality. The error term (RA) is defined as Equation 5 (White and Engelen, 1997):

$$RA = 1 + (-Lny)^\alpha \quad (5)$$

Where Y is a uniform random variable within the range 0 to 1, and α is the parameter controlling the size of the stochastic perturbation; a can be used as a dispersion factor in the simulation. The development probability is revised as Equation 6:

$$P'_d(x, t) = RA \sum_i W_j \frac{1}{1 + e^{-\text{net}_j(x,t)}} = 1 + (-Lny)^\alpha \sum_i W_j \frac{1}{1 + e^{-\text{net}_j(x,t)}} \quad (6)$$

At the end, a predefined threshold value is used to decide whether a cell is to be developed or not according to the probability $P'_d(x, t)$ at each iteration. If a cell has a probability greater than the threshold value, it will be converted for development. The number of already developed cells in the neighbourhood is recalculated and the site attributes are updated at the end of each iteration. The number of iterations is determined by the total land consumption in a given time period. The simulation will stop when all the required number of cells has been converted into urban development.

4 IMPLEMENTATION

Implementation section describes the design and implementation of ANN-CA algorithm to calibrate and simulate the urban growth of real city namely, Esfahan-Iran. The simulation process of a real city will go

through many processes starting from analyzing the area of study, data processing, an algorithm design and implementation and finally evaluation of the simulation results.

4.1 Data and Image Preparation

The data that has been used for calibration and simulation included three historical satellite images covering a period of Four years. These raw images include one 60 m resolution MSS image (1975) and four 30 m resolution TM images (1990) and 2001). The images are geometrically rectified to the same projection of Universal Transverse Mercator (UTM) zone 39N. Projected images are registered to spatially fit over each other using a second order polynomial transformation function and 15 well defined control points. Registration errors are very small represented by values far less than one pixel. After the images are geometrically rectified and registered spatially to each other, the next step to prepare the images as inputs to the CA simulation algorithm is image classification. Six classes are defined based on maximum likelihood classification system (1976): water, road, commercial, Green area where land has been covered by forest area, residential areas and nonurban areas. Commercial and residential classes will be combined after the simulation as one class called urban. Ground reference data sources including orthophotographs classification maps are used for identifying the land cover classes and for training and testing data collection. Maximum likelihood classification method is used for supervised classification. Classification precision report is prepared for each classified image using the testing data to check the quality of classification. Results indicate high precision level of classification above 75%.

4.2 The calibration and simulation of platform

After appropriate parameter values have been obtained, they are imported into the CA model for urban simulation. The simulation is still based on neural networks. The actual simulation model is implemented in a GIS platform by the integration of neural network, CA, and GIS. The model was programmed in ArcGIS using the Arc Object. Studies show that some distance-based variables are closely related to urban development in the region and can be used for CA simulation (Li and Yeh, 2000; Wu and Webster, 1998; Yeh and Li, 2001a). Satellite images can also indicate that development sites are usually located along major transportation lines or around an existing urban area. In this study, seven spatial variables were defined to represent the site attributes of each cell for the simulation of urban development:

- (1) Distance to the major (city proper) urban areas S_1 ;
- (2) Distances to suburban (town) areas S_2 ;
- (3) Distance to the closest road S_3 ;
- (4) Distance to the closest expressway S_4 ;
- (5) Distance to the closest water S_5 ;
- (6) Amount of development in the neighbourhood S_6 ;
- (7) Agricultural suitability (Green Area) S_7 .

4.3 Neural Network based CA Calibration

Classification Images in 1975 - 90 that show in fig. 2 was overlaid with the seven layers of site attributes. The overlay provides the training data that can reveal the relationship between site attributes and development probability. A sampling procedure was carried out by using a stratified sampling method which ensures that sampling effort can be distributed in a rational pattern so that a specific number of observations are assigned to each category to be evaluated. The sampling points which were generated by ArcGIS. There were 1000 random sampling points, which were used to train the neural network.

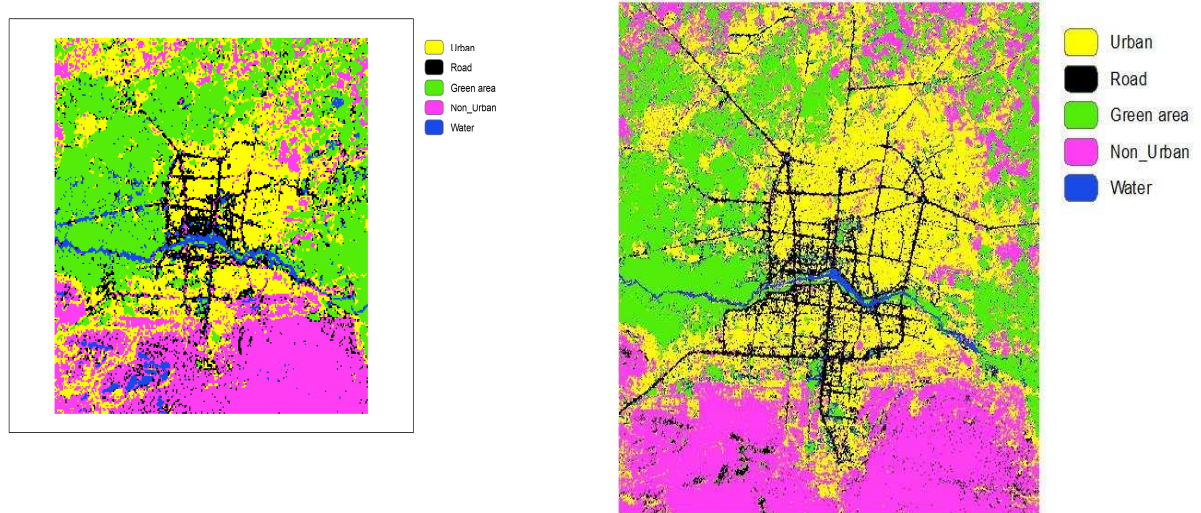


Fig 2: Ground truth of classification Image for 1975-90

Table 1 shows examples of the training data set and the calculated development probability from the ArcGIS environment.

S_1	S_2	S_3	S_4	S_5	S_6	S_7	Desired Value	ANN Output Value
305	20	2	256	21	22	0.2	1	0.874
271	13	2	140	18	16	0.4	0	0.076
275	5	0	18	250	17	0.6	1	0.651
131	8	1	56	33	26	0.4	1	0.865
93	25	1	110	34	9	0.4	0	0.059
95	38	3	23	159	18	0.2	0	0.045
287	35	2	10	270	19	0.6	1	0.893
310	22	0	300	16	21	0.6	1	0.921
137	16	3	129	18	14	0.4	0	0.087
152	35	2	19	143	17	0.2	1	0.595

Table.1: Examples of parameters, desired values (actual), and calculated development probability from ANN

Development in the region is influenced by some unexplained variables. It is also confirmed by other studies that the region is characterized by a chaotic development pattern because of severe land speculation (Yeh and Li, 2001a). The second step was to predict possible future urban development in 1990 - 2001 for planning purposes (Fig.3).

Due to the good results achieved by ANN Ca based simulation with $\alpha = 1$, we decided to simulate the growth from 1990 to the year 2001. Table. 2 summarizes the urban growth prediction evaluation results based on ANN based CA for year 2001.

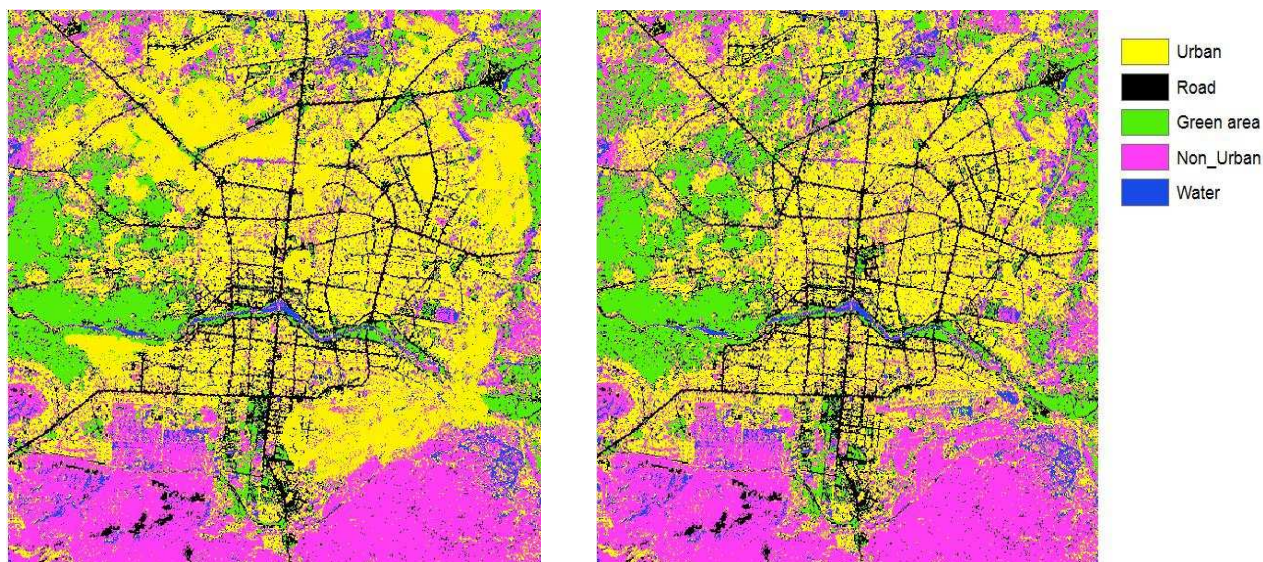


Fig. 3: The ground truth and ANN based CA simulated images for 2001

Region	2001 Ground Truth(Urban)	2001 Simulated(Urban)	Error	Precision (%)
Esfahan	103070	110960	7890	92.34

Table 2. ANN based CA simulation evaluation results of year 2001

The achieved precision for the predicted year 2001 of 92.34% is encouraging for such long term prediction of 11 years interval. Based on the ANN based CA simulation strategy we used to predict 2001. Such prediction will help municipalities identify the future growth trend and design the sustainable infrastructure plans to accommodate such trend.

5 CONCLUSION

The Implemented model can significantly reduce the requirements for explicit knowledge in identifying relevant criteria, assigning scores, and determining criteria preference. Variables used in spatial decisions are very often dependent on each other. ANN based CA urban growth simulation and prediction of Esfahan city over the last Four decades succeeds to simulate specified tested growth years at a high precision level however the interval time of satellite images of has got strong effect on predicate precision. The result of ANN based CA simulation has got more than precision (92.34%) rather than traditional CA models. Some ground truth images have been used in the CA simulation training phase such as 1990 while 2001 used for testing the prediction results. Calibrating the ANN based CA growth rules is important through comparing the simulated images with the real ground truth to obtain feedback. An important notice is that ANN based CA need also to be modified over time to adapt to the urban growth pattern. The evaluation method used on region basis has its advantage in covering the spatial distribution component of the urban growth process. These investigation shows that Esfahan has got a noticeable growth in urban area and it can be anticipate that the rate of urban growth will be more rapid rather than previous years.

6 REFERENCES

- ALMEIDA, C. M., 2003, Spatial Dynamic Modelling as a Planning Tool: Simulation of Urban Land Use Change in Bauru and Piracicaba (SP), Brazil. Sao Jose dos Campos. 321p. Published PhD thesis – Brazilian National Institute for Space Research (INPE-10567-TDI/942/B).
- Batty M, XieY, 1994, "from cells to cities" Environment and Planning B: Planning and Design 21 S31 - S48
- Batty M, XieY, Sun Z L, 1999, "Modeling urban dynamics through GIS-based cellular automata" Computers, Environment and Urban Systems 23 205 – 233.
- BISHOP, C.M., 1995, Neural Networks for Pattern Recognition (Oxford, UK: Oxford University Press).
- Clarke K C, Hoppen S, Gaydos L, 1997, "A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area" Environment and Planning B: Planning and Design 24 247 - 261
- Gong P, 1996, "Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping" Photogrammetric Engineering and Remote Sensing 62 513 - 523
- Li X, Yeh A G O, 2000, "Modeling sustainable urban development by the integration of constrained cellular automata and GIS" International Journal of Geographical Information Science 14 131 - 152

- Lombardo S T, Rabino G A, 1986, "Calibration procedures and problems of stability in nonlinear dynamic spatial interaction modeling" *Environment and Planning A* 18 341 - 350
- Soares-Filho, B.S., Cerqueira, G.C. and Pennachin, C.L., 2002, DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modeling*, 154, 217–235.
- White R, Engelen G, Uljee I, 1997, "the use of constrained cellular automata for high-resolution modeling of urban land-use dynamics" *Environment and Planning B: Planning and Design* 24 323 - 343
- Wu F, 2000, "A parameterized urban cellular model combining spontaneous and self-organizing growth", in *GIS and Geocomputation* Eds P Atkinson, D Martin (Taylor and Francis, New York) pp 73 - 85
- Wu F, Webster C J, 1998, "Simulation of land development through the integration of cellular automata and multicriteria evaluation" *Environment and Planning B: Planning and Design* 25 103 - 126
- Yeh A G O, Li X, 2001a, "A constrained CA model for the simulation and planning of sustainable urban forms by using GIS" *Environment and Planning B: Planning and Design* 28(5) forthcoming
- YEH, A.G. and LI, X., 2003, Simulation of development alternatives using neural networks, cellular automata, and GIS for urban planning. *Photogrammetric Engineering and Remote Sensing*, 69, 1043–1052.