Self Organizing Maps for Urban Modelling

Victor Lobo[1,2], Pedro Cabral, Fernando Bação[1]

[1]ISEGI-New University of Lisbon, Campus de Campolide, 1070-312 Lisboa, Portugal
Telephone: (+351 213870413)
Fax: (+351 213872140)
Email: {vlobo, pcabral, bacao}@isegi.unl.pt

[2]Portuguese Naval Academy, Alfeite, 2810-001 Almada, Portugal

1. Introduction

Land Use and Cover Change (LUCC) models are one potentially important way of trying to understand the urban growth phenomena. These models can be very useful for researchers that want to understand urban growth or for the general public, politicians and urban planners as an educational tool to visualize different scenarios of urban change (Wu, 1998, Wu and Webster, 2000).

Several authors have modelled the urban phenomena using different perspectives: general specification of models (White and Engelen, 1993, Cechini, 1996, Veldkamp and Fresco, 1996, Sanders et al., 1997, White et al., 1997b, Wu, 1998b, Batty et al., 1999, Pijanowski et al., 2002), location analysis (Benati, 1997, Batty et al., 2003), urban sustainable development (Li and Yeh, 2000) and regional urban growth (White and Engelen, 1997, Clarke and Gaydos, 1998, Silva, 2002, Cheng, 2003, Herold et al., 2003).

In this paper we approach the urban growth problem from a pattern recognition perspective, using unsupervised classification, specifically a Self-Organizing Map, to predict new areas of growth. We aim to establish if the Self-Organizing Map can be a competitive approach to this problem. Thus we use CA_Markov (IDRISI, 2004) method as a benchmark, and compare the results of the two approaches in a dataset representing a part of Lisbon Metropolitan Area.

2. The methods

2.1 The self-organizing map

Although the term “Self-Organizing Map” could be applied to a number of different approaches, we shall use it as a synonym of Kohonen’s Self Organizing Map (Kohonen 1982; Kohonen 2001), or SOM for short, also known as Kohonen Neural Networks. The basic idea of a SOM is to map the data patterns onto an n-dimensional grid of neurons or units. That grid forms what is known as the output space, as opposed to the input space where the data patterns are. This mapping tries to preserve topological relations, *i.e.*, patterns that are close in the input space will be mapped to units that are close in the output space, and vice-versa. So as to allow an easy visualization, the output space is usually 1 or 2 dimensional. The basic SOM training algorithm can be described as follows:

```
Let   X    be the set of n training patterns x1, x2,..xn
      W    be a p×q grid of units wij where i and j are their coordinates on
           that grid
      α    be the learning rate, assuming values in ]0,1[, initialized to a
           given initial learning rate
      r    be the radius of the neighbourhood function h(wij,wmn,r), initialized
           to a given initial radius
Repeat
   For k=1 to n
        For all wij∈W, calculate dij = ||xk-wij||
        Select the unit that minimizes dij as the winner wwinner
        Update each unit wij∈W: wij = wij + αh(wwinner,wij,r)||xk−wij||
   Decrease the value of α and r
Until α reaches 0
```

2.2 CA_Markov

CA_Markov (IDRISI, 2004) can model the change of several classes of cells using a Markov transition matrix, a suitability map and a neighbourhood filter. The changing of a cell category is defined by a rank that determines which cells are going to change class during the modelling process.

3. The study area and data

The study area is the municipalities of Sintra and Cascais, Portugal (Figure 1).

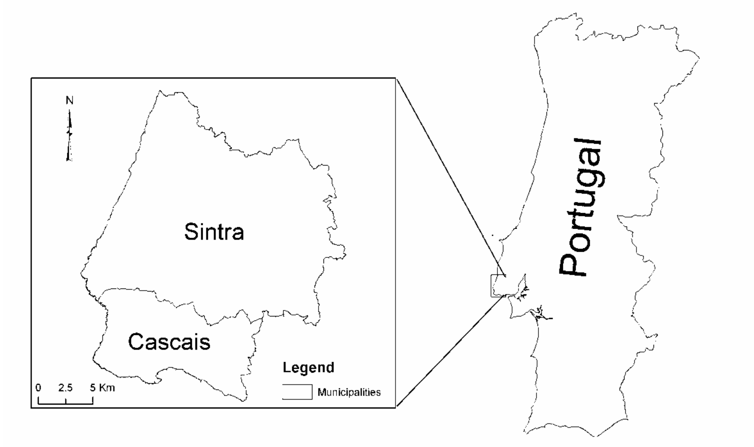

Table 1 The study area

The input data for modelling was obtained using a combined classification procedure of image segmentation and texture analysis over two Landsat TM and one ETM+ images (Cabral et al., 2005). The 1989 (14-03-1989) and 2001 (8-04-2001) images were downloaded from Global Land Cover Facility of the University of Maryland (USA). The 1994 TM (8-02-1994) image was specifically purchased for the purpose of this research.

In this research, cell a_{ijmn} is a reclassified pixel of a row i and column j of a satellite image for years 1989, 1994 and 2001. It can have two states s=”urban” and r=”rural” in time t over a space of results L={ s, r }.

The Markov matrix used in this research considers t_0 = 1989, t_1 = 1994 and t_2 = 2001. Also the suitability map in CA_Markov is based on the data from 1989 and 1994 to estimate year 2001.

4. Results

Data from 1989 were used to train the SOM and data from 1994 were used to label the resulting matrix. This way we were able to pinpoint neurons in which urban cells were clustered. To compute the probability of a given SOM unit (type of area) being urban, we calculated the number of urban areas mapped to that unit, and divided that value by the total number of areas mapped to the unit.

In order to evaluate the results we calculated the confusion matrix for each of the two methods. Based on the confusion matrix we calculated the accuracy of the classifications, obtaining 82% for the CA_Markov and 80% for the SOM, which indicates a similar level of accuracy (Tables 2 and 3). The major difference in performance relates to the “false true”. In other words, the pixels that were classified by the methods as “changing” and didn’t. In this case the SOM was less conservative and while having an almost similar value of correct predictions, the SOM made more incorrect predictions of change.

		CA Pred.		
		C	NC	Total
Real	C	11566	44186	55752
	NC	20431	289858	310289
	Total	31997	334044	366041

		CA Pred.		
		C	NC	Total
Real	C	20.7	79.3	100
	NC	6.6	93.4	100
	Total	-	-	-

		CA Pred.		
		C	NC	Total
Real	C	36.1	13.2	-
	NC	63.9	86.8	-
	Total	100	100	-

Table 2 Confusion matrix for predicted CA_Markov for 2001 against Real 2001 (C:change, NC: no change). On the left the absolute numbers are presented, in the centre the percentage of actual changes predicted, and on the right the percentage of predicted changes that actually occurred.

		SOM Pred.		
		C	NC	Total
Real	C	11285	44467	55752
	NC	28324	281965	310289
	Total	39609	326432	366041

		SOM Pred.		
		C	NC	Total
Real	C	20.2	79.8	100
	NC	9.1	90.9	100
	Total	-	-	-

		SOM Pred.		
		C	NC	Total
Real	C	28.5	13.6	-
	NC	71.5	86.4	-
	Total	100	100	-

Table 3 Confusion matrix for predicted SOM for 2001 against Real 2001 (C:change, NC: no change). On the left the absolute numbers are presented, in the centre the percentage of actual changes predicted, and on the right the percentage of predicted changes that actually occurred.

5. Conclusions

While not improving the predictive accuracy of a widely used methodology for LUCC, we showed that SOMs are a competitive alternative. Besides being one more alternative for predicting urban growth, the SOMs have the advantage of being easily parallelized for implementation on parallel machines, thus enabling effective processing of very large datasets which are common in this type of problems.

It must be pointed out that in these preliminary tests, we used exactly the same data inputs that where required by CA_Markov. There is a lot of work to be done in pre-processing and feature extraction in order to optimize the way the problem is cast to the SOM. This will certainly yield better results, and warrants future work.

6. References

Batty, M., Desyllas, J. and Duxbury, E., 2003. The Discrete Dynamics of Small-Scale Spatial Events: Agent-Based Models of Mobility in Carnivals and Street Parades. International Journal of Geographical Information Science. 17, 673-697.

Batty, M., Shie, Y. and Sun, Z., 1999. Modeling urban dynamics through GIS-based cellular automata. Computers, environment and urban systems. 23, 205-233.

Benati, S., 1997. A cellular automaton for the simulation of competitive location. Environment and Planning B: Planning and design. 24, 205-218.

Cabral, P., Gilg, J.-P. and Painho, M., 2005. Monitoring urban growth using remote sensing, GIS and spatial metrics. In: Gao, W. (Ed.). Proceedings of the SPIE Optics & Photonics: Remote sensing and modeling of ecosystems for sustainability, San Diego, USA, pp.

Cechini, A., 1996. Urban Modelling by Means of Cellular Automata: Generalised Urban Automata with the Help On-Line (AUGH) model. Environment and Planning B. 23, 721-732.

Chen J., Gong P., He C., Luo W., Tamura M. & Shi P. (2002) Assessment of the urban development plan of Beijing by using a CA-Based urban growth model. Photogrammetric Engineering & Remote Sensing 68: 1063-1071

Cheng, J., 2003. Modeling Spatial & Temporal Urban Growth. In: Faculty of Geographical Sciences), pp. 203. Utrecht University, Utrecht.

Clarke, K. and Gaydos, L., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. International Journal of Remote Sensing. 12, 699-714.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20, 37-46.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment. 37, 35-46.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map. Photogrammetric Engineering & Remote Sensing. 47, 343-351.

Hagen, A., 2002a. Comparison of map containing nominal data). Institute for Knolwedge Systems.

Hagen, A., 2002b. Multi-method assessment of map similarity. Proceedings of the 5th Agile Conference, Mallorca, Spain, pp.

Herold, M., Goldstein, N. and Clarke, K., 2003. The spatio-temporal form of urban growth: measurement, analysis and modeling. Remote Sensing of Environment. 85, 95-105.

IDRISI, 2004. Idrisi Kilimanjaro. Clark Labs, Worcester, MA.

INE, 2001. Censos 2001 - Resultados preliminares). Instituto Nacional de Estatística, Lisboa.

Li, X. and Yeh, A., 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. International Journal of Geographical Information Science. 14, 131-152.

Li X., Yeh A. (2000) Modelling sustainable urban development by the integration of constrained cellular automata and GIS. International Journal of Geographical Information Science 14: 131-152

Li X., Yeh A. (2002) Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science 16: 323-343

Pijanowski, B., Brown, D., Shellito, B. and Manik, G., 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. Computers, environment and urban systems. 26, 553-575.

Pontius R., Malanson J. (2005) Comparison of the structure and accuracy of two land change models. Int. J. Geogrphical Information Science 19: 243-265

Pontius, R., 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. Photogrammetric Engineering & Remote Sensing. 68, 1041-1049.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F. and Bura, S., 1997. A multiagent system for the study of urbanism. Environment and Planning B: Planning and design. 24, 287-305.

Silva, E., 2002. Beyond modeling in environmental and urban planning. Planning support systems and the case study of Lisbon and Porto Metropolitan Areas, Portugal. In: PhD Thesis. Regional Planning), pp. 293. University of Massachussets, Massachussets.

UNESCO, 2005. Cultural landscape of Sintra), vol. 2005. no. 13 août 2005. UNESCO.

Veldkamp, A. and Fresco, L., 1996. CLUE: a conceptual model to study the conversion of land use and its effects. Ecological Modelling. 85, 253-270.

White, R. and Engelen, G., 1993. Cellular-automata and fractal urban form - a cellular modeling approach to the evolution of urban land-use patterns. Environment and Planning A. 25, 1175-1199.

White, R. and Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environment and Planning B: Planning and design. 24, 235-246.

White, R., Engelen, G. and Uljee, I., 1997b. The use of constrained cellular automata for high-resolution modelling of urban land use dynamics. Environment and Planning B. 24, 323-343.

Wu, F., 1998. An emprirical model of intra-metropolitan land-use changes in a Chinese city. Environment and Planning B: Planning and design. 25, 245-263.

Wu, F., 1998b. An experiment on the generic polycentricity of urban growth in a cellular automatic city. Environment and Planning B: Planning and design. 25, 731-752.

Wu, F. and Webster, C., 2000. Simulating artificial cities in a GIS environment: urban growth under alternative regulation regimes. International Journal of Geographical Information Science. 14, 625-648.