

Cellular Automata and Urban Development Simulation : A Transition Rules Creation Process Based on Statistical Analysis.

Valentine Judge^{1,2}, Jean-Philippe Antoni¹

¹Laboratoire ThéMA

UMR 6049 CNRS - Université de Franche-Comté

²CEPS/INSTEAD

Département développement urbain et mobilité

valentine.judge@univ-fcomte.fr

Key words - Cellular Automata, Spatial Analysis, Principal Component Analysis, Hierarchical Clustering, Decision Tree.

1 Context

Nowadays land use evolution study has become a major stake in urban planning. The main focus is to understand the way in which land use evolves across time and to understand processes that take place. This understanding would allow to plan urban developments based on a knowledge as complete as possible covering as many fields as possible (i.e. urban planning, politics, sociology, etc.). Simulation tools can be used to merge and display different points of view and stakes from different stakeholders (Parrott & Meyer, 2012). Indeed simulation tools have proven to be an efficient way to visualize different kind of stakes that take place on a territory and to come up with solutions that help to achieve general urban planning or organization over a territory (Parrott et Al., 2012). Different studies highlight the usefulness and needs of simulation tools in research focus on urban development and planning (White & Engelen, 1994, Batty et Al., 1999, Parrott et Al., 2012). Several types of simulation tools have been used so far considering urban planning like Cellular Automata (White & Engelen, 1994, Batty et Al., 1999, Jantz C.A., et Al., 2010, Van Vliet J., et Al., 2012) or Agent Based Model (Parrott et Al., 2012).

Cellular Automata (CA) are known for

their geometric properties and their potential to show the emergence of trends that could appear on landuse evolution over a spatial area. They present the advantage to link the state evolution of a cell (a parcel) to its current state (its land use) and to states that characterize cells in its neighborhood. The potential evolution of cells is represented by mean of transition rules (Santé I., et Al., 2010).

However issues related to calibration and validation of transition rules in urban CA remain. One of the issues using CA is related to the creation of the transition rules settlement that will drive the simulation. Indeed, “in the context of urban systems, we often have no idea what the “right” rules are” (Torrens, 2011). This paper will focus on CA transition rules creation and proposes two methods to achieve this task. It addresses the hypothesis that landuse composition of the neighborhood has an influence on the landuse change of a cell.

2 Methodology

2.1 Data set

In order to fulfill the objective, different datasets showing the composition of the neighborhood from 100m cells are analyzed on the Strasbourg-Kehl area between

1990, 2000 and 2006. Each cell is characterized by variables that represent the neighborhood at a defined distance (eg. : number of agricultural cells within a 500m radius). The method is developed to be transferable over several areas. This is the reason why only open source data available at the European scale are used : Corine Land Cover landuse classification. The choice has been made considering two specific requirements : (i) generic data available over all of the European area, and (ii) data available from different time periods (making a diachronic analysis possible).

The main hypothesis will be tested on cells that make a transition from a non-urban to an urban type of landuse in order to address the urban development evolution issue. The objective is to understand what type of neighborhood leads to this particular transition and to facilitate the creation of the set of transition rules in CA. In our case study, we propose to base the automation process of the rules on statistical analysis of the neighborhood. Two different methods will be used : a Principal Component Analysis associate to a Hierarchical Classification and a Decision Tree.

2.2 Principal Component Analysis (PCA)

Several descriptive methods for multivariate analysis exist, amongst them Principal Component Analysis. One of the main purpose of this method is to study the resemblance and the difference between individuals (in our case cells defined by a specific landuse) based on the different variables that characterize each one of them : “The core idea to all Principal Component methods is to describe a data set using a small number of uncorrelated variables while retaining as much information as possible. The reduction is achieved by transforming the data into a new set of continuous variables called principal component” (Husson et Al., 2010).

Basically, each component/factor can be described by the variables originally

present in the data set. Factors characterize the cells as new variables that summarize the previous variables information. According to the Principal Component Analysis the first factor contains the most significant information ; the second factor the second most significant information and so on. The projection of the individuals according to the two first factors helps to perform a Hierarchical Classification based on their similarities. Several clusters will be obtained from this classification. Each cluster is described and this description leads to the construction of transition rules allowing calibrating CA on the base of PCA results.

2.3 Decision tree

Decision trees are also a well-known method for extracting classification rules from data set. This sorting method arises from Data Mining field and belongs to the supervised descending hierarchical classification group. A decision tree “grow in a top-down way when we successively partition the training data into subsets having similar or the same output (class labels)” (Wang et Al., 2014). In order to build a Decision tree, the subsets represented by “branch” that end in “leaf” are analyzed. A rule that classifies the data into a class can be defined from each leaf. “Once the building process ends, the decision rules generated can be used to predict the classes of the remaining records in the data set” (Padua et Al., 2013).

Our objective is to use the Decision Tree method based on the major criteria of landuse change from a non-urban type to an urban type. As each “leaf” represents a rule, the latter must be adapted, integrated and tested into the Cellular Automata.

3 Results

Results obtained using a Principal Component Analysis and a Decision Tree help to create a set of transition rules based on statistically significant information about type

of neighborhood that lead to a specific transition, and to test them in a Cellular Automata based simulation. Two rules sets are produced by the statistical analysis mentioned above. They are not identical but reveal themselves as complementary and allow highlighting parameters or trends from different points of view.

The efficiency of each transition rules set is tested making a comparison between actual landuse cover data from Corine Land Cover database and simulation landuse data obtained running the simulation according to the transition rules created on the base of on statistics analysis of the neighborhood.

The understanding of the neighborhood environment that leads to a landuse change from a non-urban type of landuse to an urban type provides significant intelligence concerning the landuse evolution study in general, with an urban development specific scope. A further objective is to integrate this knowledge into the prospective scenario process of creation. This study will then contribute to the general stake of urban growth and artificial land sprawl studies.

References

- Batty M., Xie Y., Sun Z., 1999, « Modeling urban dynamics through GIS-based cellular automata », *Computers, Environment and Urban Systems* 23, 205-233.
- Husson, F., Josse, J. & Pagès J., 2010, «Principal component methods - hierarchical clustering - partitional clustering : why would we need to choose for visualizing data? », *Technical report*.
- Jantz C. A., Goetz S. J., Donato D., Claggett P., 2010, « Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model », *Computers, Environment and Urban Systems*, 34, 1-16.
- Kolossov V., Scott J., 2013, « Selected conceptual issues in border studies », *Belgeo* [Online], 1 | 2013, Online erschienen am : 31 Oktober 2013, nachgeschlagt am 04 Februar 2014. URL : <http://belgeo.revues.org/10532>
- Moullé F., 2013, « La frontière et son double. Un modèle à partir de l'expérience européenne », *Belgeo* [En ligne], 1 | 2013, mis en ligne le 31 octobre 2013, consulté le 21 janvier 2014. URL : <http://belgeo.revues.org/10620>
- Padua L., Schulze H., Matkovi? K., Delrieux C., 2013, « Interactive exploration of parameter space in data mining : Comprehending the predictive quality of large decision tree collections », *Computers & Graphics* 41 (2014) 99-113
- Parrott, L., Meyer, W., 2012, « Future Landscapes : Managing within complexity », *Frontiers in Ecology and the Environment*, 10(7) : 382-389 doi :10.1890/110082.
- Parrott, L., Chion C., Gonzalès R. and Latombe G., 2012. « Agents, individuals, and networks : modeling methods to inform natural resource management in regional landscapes », *Ecology and Society* 17(3) : 32. <http://dx.doi.org/10.5751/ES-04936-170332>
- Torrens, P.M., 2011, « Calibrating and validating cellular automata models of urbanization », *Urban Remote Sensing : Monitoring, Synthesis and Modeling in the Urban Environment*, Yang, Xiaojun (Ed.). Chichester : John Wiley & Sons, pp. 335-345
- Van Vliet J., Hurkens J., White R., Van Delden H., 2012, « An activity-based cellular automaton model to simulate land-use dynamics », *Environment and Planning B : Planning and Design*, 39, 198 - 212
- Wang X., Liu X., Pedrycz W., Zhang L., 2014, « Fuzzy rule based decision trees », *Pattern Recognition* 48 (2015) 50-59