
Valentine Judge \(^1,^2\), Jean-Philippe Antoni \(^1\)

\(^1\)ThéMA (UMR 6049 CNRS - Franche-Comte University)
\&
\(^2\)LISER

Conférence ThéoQuant 21-05-2015
Supported by the National Research Fund, Luxembourg (Project INTER/CNRS/12/02)
Sommaire

Introduction

Methodology

Simulation and validation

Perspectives
Land use simulation and cellular automata

Cellular Automata have been widely used for modeling land use change in geography (i.e. Tobler, 1970; Couclelis, 1985; Phipps, 1989; White, 1997; Torrens, 2002, etc.) Main attractive features: simplicity, interaction between neighboring land units, emergence.


Automata : "Processing mechanism with characteristics that changed over time based on its internal characteristics, rules and external inputs." (Beneson & Torrens, 2004)

\[ A \sim (S, T, R) \]

- \( A \) : Automata
- \( S \) : State
- \( T \) : Transition rules
- \( R \) : Neighborhood

Tool: LucSim (developed from scratch by: Vuidel & Antoni)
Urban development study using cellular automata: A transition rules calibration approach

Hypothesis:
- The transition of a cell from one land use type to another is driven by its initial land use and by the land use of cells from its neighborhood.

Questions:
- How can transition be defined?
- How can transition be written as rules?
Study area

Extend of the study area:
67 km² (6700 Ha).

Extend of the study area:
6700 Cells.
Cell dimension:
1 Ha (100x100m cell)
Cellular approach

Land use cover (1990)

Urban Changes (1990-2006)

Re-classified CorineLandCover

<table>
<thead>
<tr>
<th></th>
<th>1990</th>
<th>2006</th>
<th>Change 1990-2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>810</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>IndusCom</td>
<td>12</td>
<td>309</td>
<td>-13</td>
</tr>
<tr>
<td>Agri</td>
<td>18</td>
<td>59</td>
<td>-100</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>48</td>
<td>-48</td>
</tr>
<tr>
<td>Gain</td>
<td>30</td>
<td>50</td>
<td>-25</td>
</tr>
</tbody>
</table>
Sommaire

Introduction

Methodology
  Methodological process
  Rules definition proposal
  Principal component analysis and hierarchical classification
  Decision tree

Simulation and validation

Perspectives
Methodological scheme

1. Land use classification
   - Neighborhood data extraction
2. Principal Component Analysis
   - Hierarchical classification
   - Rules set creation
   - Postdictive simulation
   - Validation
3. Decision Tree
   - Rules set creation
   - CA simulation & Analysis
   - Postdictive simulation
   - Validation
   - Comparison and Analysis of complementarity
Rules definition proposal

Assumption CA based: Land use of cells settled in the neighborhood can explain on a potential urban transition of a cell.

\[ A \sim (S, T, R) \]

Objective: Create a set or transition rules based on the assumption.

\[ S_{i(t)} \to S_{f(t+1)} : R_{Sa}\{N_{Sa}(D_a;n_a), N_{Sb}(D_b;n_b), N_{Sc}(D_c;n_c), \ldots \} \]

Questions:
- Which states?
- Which distances?
- How many cells?

Proposition: Create transition rules based on statistical analysis of the neighborhood of cells that carried out specific transition.
Datasets: Ernolsheim-Bruche area

- Tables {Individual, initial state, final state, a 5 pixels neighborhood}.
- Data: 1990-2006.
- Variable shows the number of cells of a land use type at a specific distance.
Objectives:

- Target variables summarizing the most of information contained in the dataset being explored.
- Create cluster of individuals (cells) considering variables defining them.
- Analyze the neighborhood of individuals from the same cluster in order to get land use influence threshold (number of cells within a specific distance).

Dataset:

- Focused on cells that carried out a transition from a $\lambda$ state to an urban state (30 individuals).
Method:

- Variable that summarize the most of the dataset information are used as a base for the rules set (i.e. IndusCom4px, Urban4px, Agri4px, Forest5px).
- A minimum and maximum threshold is set from the analysis of individuals or clusters from the hierarchical classification (i.e. 4 < IndusCom4px < 7, 0 < Urban4px < 6, etc.).
**Objective:**
- Sequential partition of the dataset using a top-down approach. The aim is to target the variable that best splits the different cluster of data left at each step.

**Dataset:**
- Every cell regardless of their transition.
- Adding a binary categorical variable to characterize an urban transition or not.
Decision tree

Method:
- Rules can be extracted directly from the tree analysis. Each "leave" corresponding to the transition into urban land use is considered and transcribe as rules.
Rules set analysis

Why do we analyse rules sets produced?

- Are the same variables extracted with both methods?
- If yes: Are the threshold similar?
- If not: what is the extend of the dissimilarity?

Principal Component Analysis based rules set

\[ \text{Agri or IndusCom } \rightarrow \text{Urban :} \]

\[ 0 \leq \text{Urban4px} \leq 6 \text{ And } 4 \leq \text{IndusCom 4px} \leq 7 \text{ And } 3 \leq \text{Agri3px} \leq 11 \text{ And } 1 \leq \text{Forest5px} \leq 4 \]

Decisions tree based rules set

\[ \text{Agri or IndusCom } \rightarrow \text{Urban :} \]

\[ \text{Urban4px} > 0 \text{ And } \text{IndusCom 1px} \geq 1 \text{ And } \text{Induscom5px} < 9 \text{ And } \text{Agri5px} < 28 \text{ And } \text{Forest5px} > 0 \]
Sommaire

Introduction

Methodology

Simulation and validation
  Validation
  Rules sets simulation and analysis
  Comments

Perspectives
Validation

Objective:
- Evaluation of the results quality comparing with data from 2006.
- Compare result produced by the PCA/HC method and the decision tree method.

Validation method: GIS based
- Pixel to pixel validation
- buffer 100m, 200m, 500m
**Case:** Simulation matching the most real NewUrban cells.

<table>
<thead>
<tr>
<th>Units</th>
<th>PCA Simulation</th>
<th>DT Simulation</th>
<th>Similarity Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>NewUrban</td>
<td>604</td>
<td>746</td>
<td></td>
</tr>
<tr>
<td>Px vs Px</td>
<td>30</td>
<td>18</td>
<td>49</td>
</tr>
<tr>
<td>100m</td>
<td>19</td>
<td>13</td>
<td>72</td>
</tr>
<tr>
<td>200m</td>
<td>13</td>
<td>45</td>
<td>82</td>
</tr>
<tr>
<td>500m</td>
<td>38</td>
<td>50</td>
<td>95</td>
</tr>
</tbody>
</table>
Rules sets simulation and analysis

**Case:** Simulation matching the most real NewUrban cells.

<table>
<thead>
<tr>
<th>Unites</th>
<th>PCA Simulation Number of cell</th>
<th>DT Simulation Number of cell</th>
<th>Similarity Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>NewUrban</td>
<td>604</td>
<td>746</td>
<td>49</td>
</tr>
<tr>
<td>Px vs Px</td>
<td>30</td>
<td>18</td>
<td>72</td>
</tr>
<tr>
<td>100m</td>
<td>19</td>
<td>13</td>
<td>49</td>
</tr>
<tr>
<td>200m</td>
<td>13</td>
<td>45</td>
<td>82</td>
</tr>
<tr>
<td>500m</td>
<td>38</td>
<td>50</td>
<td>95</td>
</tr>
</tbody>
</table>
Rules sets analysis:

- Some identical variables pop up from both analysis (i.e. Urban4px, Forest5px)
- Not all variables that pop up are identical (distances can differ i.e. IndusCom5px vs IndusCom4px)

Simulation analysis:

- PCA simulation is more efficient (localize 100% of the real 2006 NewUrban cells)
- DT simulation is more compact than PCA simulation
- Both simulation seems to spatially complement on another having a 72% similarity when a 100m accuracy is taken into account.

General comment: Both set of rules are partly satisfying or interesting to explore but miss accuracy.
Sommaire

Introduction

Methodology

Simulation and validation

Perspectives
Urban development study using cellular automata: A transition rules calibration approach

First results:

- Rules can be written based on the analysis of the neighborhood.
- Several solution can be use, two have been tested:
  - Principal component analysis with hierarchical classification
  - Decision trees
- Both set of rules are partly satisfying according to the validation but miss accuracy
First results: perspectives?

- Improve rules sets according to both methodologies
- Improve the accuracy of the rules (i.e. add new variables? accessibility?)
- Apply the methodology developed here on a larger area (reduce the bias from the data)
- Explore results more deeply from a thematic point of view
Methodological scheme

1. Land use classification
2. Principal Component Analysis
   Hierarchical classification
   - Rules set creation
   - Postdictive simulation
     Validation
   - CA simulation & Analysis
   - Postdictive simulation
     Validation
   - Comparison and Analysis of complementarity
3. Decision Tree
   - Rules set creation
A methodological approach

Valentine Judge$^{1,2}$, Jean-Philippe Antoni$^1$

$^1$ThéMA (UMR 6049 CNRS - Franche-Comte University)
$^2$LISER

Conférence ThéoQuant 21-05-2015
Supported by the National Research Fund, Luxembourg (Project INTER/CNRS/12/02)