Applied Geography 88 (2017) 73–83

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

A semi-automatic neighborhood rule discovery approach

Tong Wang*, Qi Han, Bauke de Vries

Built Environment, Eindhoven University of Technology, Vertigo, De Groene Loper 6, Postbus 513, 5600 MB, Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 30 March 2017
Received in revised form
15 July 2017
Accepted 28 August 2017
Available online 20 September 2017

Keywords:
Planning support
Cellular automata
Feature shape
Generic computer-based tool
Neighborhood rule detection

ABSTRACT

Cellular automata (CA) models are used a lot in urban planning for land use change simulation. Neighborhood rules in CA models are normally derived by analyzing raster maps, while in reality, urban planning is based on parcels. Moreover, new trends of land redevelopment require land use transition impacts to be addressed and incorporated in land use simulation. This can be accomplished by comparing neighborhood compositions of a particular site, before and after its transition. This paper presents a generic approach to semi-automatically discover neighborhood rules by analyzing vector maps, with a special focus on transition impact analysis. The approach contains one script for manipulating vector data and another script for visualizing neighborhood compositions. A step by step instruction of this approach is presented. The North Brabant region of the Netherlands is used as a case study area. Industrial site transition is used as an illustration for land use transition process. Three types of statistical comparison algorithms are used to compare the land use model which uses neighborhood rules discovered from the proposed approach with a benchmark model which only models land use self-influence. The robustness of the approach is studied by separating the region into urban and non-urban areas. Results show the applicability of this approach in improving transparency and applicability of CA models.

1. Introduction

1.1. Cellular automata model in land use simulation

Existing land use patterns influence future land use in three ways: the inertia of land uses in a location which reflects the land uses persist over time; the ease of land use conversions which represents the hierarchy of land use attractions for users; and the attraction or repulsion effects from surrounding land uses whose influence decreases with increasing distance till zero according to Tobler's first law of geography (Tobler, 1970; van Vliet et al., 2013).

To model these three types of influences, Cellular Automata (CA) models have been used in many applications for their simplicity and flexibility (Barredo, Kasanko, McCormick, & Lavalle, 2003; Batty, Xie, & Sun, 1999; Engelen et al., 1995; Li & Yeh, 2000; Stevens, Dragicevic, & Rothley, 2007; White & Engelen, 1997; Wu & Webster, 1998). CA is a bottom-up approach to model complex systems dynamically. Space is represented as a grid of cells. Each cell is influenced by its neighbors (Lai & Dragićević, 2011). These spatial dependencies, normally called "neighborhood rules", are crucial for CA models since they are the keys to modelling cell interactions. These local rules can generate complex patterns at global scales so that the whole is more than the sum of its parts (Barreira-González, Gómez-Delgado, & Aguilera-Benavente, 2015; Torrens, 2000; White & Engelen, 1993). Researchers have summarized eight basic types of neighborhood rules which describe the interaction between a pair of land uses (van Vliet et al., 2013). In these eight types, different land uses have various influences on other land uses based on neighborhood distances. For example, heavy industry normally has a negative influence on housing in a short distance because of the possible pollution produced by industrial activities but might attract housing at a larger distance for workers' commuting convenience.

There are several requirements for a land use change model to be useful for decision support. The CA model needs to be realistic in the sense that the predicted land use situation should resemble the actual situation to certain extent (Karimi, Mesgari, Sharifi, & Pilehforooshha, 2017). Resemble to the actual situation does not mean 100% replication, but the simulated results can be representative to what actually has happened. To what extent the model is representative enough is normally determined by the users. To

* Corresponding author. Vertigo, De Groene Loper 6, VRT 9.116, Postbus 513, 5600 MB, Eindhoven, the Netherlands.
E-mail addresses: tong.wang.0821@gmail.com (T. Wang), q.han@tue.nl (Q. Han), b.d.vries@tue.nl (B. de Vries).

http://dx.doi.org/10.1016/j.apgeog.2017.08.014
0143-6228/

explain the model to end-users, modelers can document how land use changes are modeled and what kind of rules have been applied (Sugumaran & DeGroote, 2011, p. 445). As Hagoort, Geertman, and Ottens (2008) claimed, the neighborhood rules need a better empirical foundation before they can be applied to support spatial policy. To deduce the land use change rules and compare historical data with actual data, Hansen (2012) argues, detailed data needs to be collected from the same reliable source and manipulated in the same way. Manipulation of the data includes but is not limited to rasterization, map creation, buffer setting, neighborhood determination and percentage calculation, which is a time-consuming process. The effort in acquiring and manipulating data for decision support should also be considered in the design of the land use change models.

1.2. Neighborhood rules

In the center of CA modelling, neighborhood rules are of great importance since they help to define the local interactions for modelling global behavior (Shi & Pang, 2000). However, in reality, one of the main limitations of determining neighborhood rules is the lack of generic way to set these rules systematically for each study area. Neighborhood rules are often defined by trial and error (Hagoort et al., 2008). This process is normally performed by experts on land use modelling (Karimi, Sharifi, & Mesgari, 2012). They first look at the historical land use maps and compare them to find the changes. Then they set initial neighborhood rules. Following they compare the simulated results with actual maps to adjust their neighborhood rules, which is a long and iterative process. There are a few limitations in this approach. Only experts with a lot of experience can do this job. This trial and error process requires a lot of time and is still difficult for the experts to explain why they come up with exactly these neighborhood rules instead of other rules.

The most common way to calibrate a land use change simulation model and to check the rule settings is to compare the simulated land use maps with the actual maps from the same year. Several statistics are developed to calibrate the model such as Kappa (Monserud & Leemans, 1992), Kappa simulation (van Vliet, Bregt, & Hagen-Zanker, 2011), and Fuzzy Kappa simulation (Hagen, 2003; van Vliet et al., 2013). Kappa is a commonly used statistical measure that expresses the agreement between two categorical maps, corrected for expected agreement. Kappa simulation references to the initial land use map by comparing the amount of each land use change. Values above 0 indicate predictive power. 1 represents the perfect fit. Fuzzy Kappa Simulation combines both initial map and geographical fuzziness for distinguishing between small and large disagreement in position and in land use classes in the comparison.

1.3. Limitations

Despite CA's significant contributions to land use change modelling, limitations have also been widely criticized. It is difficult to set neighborhood rules and to calibrate the model (de Almeida et al., 2003; Engelen & White, 2008; Kamusoko & Gamba, 2015; Verburg, de Nijs, Ritsema van Eck, Visser, & de Jong, 2004; Wu, 2002). As explained in 1.2, neighborhood rules are normally set by experts using trial and error on an ad hoc basis, and adjusted based on the calibration results (Hagoort et al., 2008). Each change in the neighborhood rule setting requires a new calibration process which is tedious and time-consuming.

Additionally, in CA modelling, the space is presented by regular homogeneous cells. Raster-based CA have been used extensively for simulating land use changes because of the simplicity of computations and their conformity with pixel-based data (Abolhasani, Taleai, Karimi, & Rezaee Node, 2016). However, simulating through irregular CA can better represent the actual land use change process since urban unit is organized based on parcels in reality (Barreira-González et al., 2015; Lu, Cao, & Zhang, 2015). Applying irregular parcel structures can help to match reality and modelling practice. Moreover, using raster-based data costs information and precision loss since many grids contain not only just one land use type. Specific land use types such as industrial sites are relatively large and irregular which require more precise space structure to be represented in the CA modelling process.

Furthermore, the recent land use redevelopment trend requires more attention to model the after transition impacts. Land use transition impact analysis should also be incorporated in the land use change analysis process. Technically speaking, many platforms or software packages for land use simulation require specific data formats which need extensive data collection period and processing time in advance (Hansen, 2007; Koomen, Hilferink et al., 2011,; Koomen, Koekoek et al. 2011; van Delden, Escudero, Uljee, & Engelen, 2005).

To overcome these shortcomings of traditional CA models, attempts have been made. A thorough review of applications is presented by van Vliet et al. (2016). Artificial intelligence techniques such as neural networks, decision trees, support vector machines and random forests are used to generate the neighborhood rules automatically (Basse, Charif, & Bódis, 2016; Kamusoko & Gamba, 2015; Li & Yeh, 2001, 2002; Liu, Li, Shi, Wu, & Liu, 2008; Liu, Feng, & Pontius, 2014; Pijanowski, Pithadia, Shellito, & Alexandridis, 2005; Yang, Li, & Shi, 2008). Another way to overcome the difficulty of generating neighborhood rules is to use statistical analysis approaches such as regression models and Bayesian networks (Celio, Koellner, & Grêt-Regamey, 2014; Ku, 2016; Liao et al., 2016; Verstegen, Karssenberg, van der Hilst, & Faaij, 2014). Fuzzy set theory is also applied for urban growth CA models (Al-Ahmadi, Heppenstall, Hogg, & See, 2009). However, the complexity and the limited interpretability of such models make it hard to implement these techniques and explain to the users. Overfitting is also common in this type of approaches. Land use change modelling not only requires acceptable correctness of simulated results. More importantly, it seeks for a better understanding of the underlying mechanism of land use change and an easier communication process to the users.

To analyze neighborhood characteristics of land use patterns more transparently, Verburg et al. (2004) have studied the enrichment of the neighborhood by specific land use types for every location in a rectangular grid. However, big irregular shape features' impacts might be lost if they are rasterized first for land use change analysis. Urban space transformations do not follow the pattern of a regular raster structure but fit into the pre-existing land structure with irregular parcels, such as industrial site transitions. Efforts have been made to incorporate vector shapes into CA models but further calibration and validation are still underway (Barreira-González et al., 2015).

To conclude, finding a generic way to set neighborhood rules using vector data which also incorporates land use transition impact analysis is necessary. It should also be possible to explain the method to the users transparently. Land use modelling should have empirical foundation. By means of empirical study, rule settings ought to be tailored to individual regions.

1.4. Objectives

Computer-based tools can provide analysis possibilities for land use change analysis for different regions. In this paper, a generic semi-automatic approach for neighborhood rule discovery is proposed. It uses vector maps to minimize the information loss.

To make the process easier to explain, industrial land transition