Extraction of Spatial Rules Using a Decision Tree Method: A Case Study in Urban Growth Modeling

Jungyeop Kim[1], Yunhee Kang[2], Sungeon Hong[3], and Soohong Park[1]

[1] Dept. of Geoinformatic Engineering, Inha University, Incheon, Korea
jyfloo@hotmail.com, shpark@inha.ac.kr
[2] Dept. of Computer Science & Information Engineering, Inha University, Incheon, Korea
yjfloo@datamining.inha.ac.kr
[3] Dept. of Welfare−Land Information, Cheongju University, Cheongju, Korea
hongsu2005@cju.ac.kr

Abstract. Data mining refers to the extraction of knowledge from large amounts of data using pattern recognition, statistical methods, and artificial intelligence. The decision tree method, a well-known data mining technique, is used to extract decision rules because using it, we can understand the grounds of classification or prediction. The decision tree method thus can be used to analyze spatial data. There is an enormous amount of spatial data in GIS (Geographic Information System), which has been studied in modeling with these data. Spatial modeling as applied to a decision tree method can be carried out more effectively. In this paper, we extracted spatial rules using the decision tree method, and then applied them to urban growth modeling based on Cellular Automata (CA). An evaluation comparing the model using the decision tree method (the proposed model) with the standard UGM showed that the proposed model is more accurate.

1 Introduction

Simulation modeling is one way of predicting future changes under a plausible scenario using constraints and a relevant data set. Simulation models generally have low prediction accuracy due to the incompleteness of both (simulation) model rules and input data sets. These problems may be partly solved using data mining techniques. There are many data mining techniques available. Some of these include association rule algorithms, genetic algorithms, neural networks, decision tree methods, link analyses, clustering, and fuzzy sets. The decision tree method is one of the techniques that has been used for extracting patterns and discovering features efficiently in large database sets. Accordingly, many research communities have used decision tree methods. Most of the resultant studies have focused on non-spatial data; however, considering the features of the decision tree methods, it is possible to use one of them to model spatial data, particularly in simulation modeling. This paper applies a decision tree method to a (simulation) modeling application, more specifically Cellular Automata (CA)-based urban growth modeling, and evaluates and demonstrates the

B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part I, LNAI 4251, pp. 203 – 211, 2006.

feasibility of the decision tree method to simulation modeling by comparing it to a well known modeling example. CA-based urban growth modeling was chosen for this study because it has recently received much attention from the GIS research community as a new way of modeling urban growth. In addition, the CA-based urban growth models entail universal transition rules that predict or simulate future trends. Transition rules are basically spatial rules and are generally derived from previous research or developed by the researcher in a subjective manner. Thus, constructing transition rules for the CA-based model is one of the most important steps affecting modeling results.

Combining a CA-based urban growth model and decision tree methods also has very significant practical implications. When considering the impacts of rapid urbanization, unemployment, disease and other factors, it is vital for us to analyze how cities have been developed and to predict how cities will be expanded in the near future, in order to establish a plan. As a result, analysis and prediction of urban growth yields important data for urban policies (Kang, Park 2000). Accordingly, there have been many studies conducted on urban growth modeling (Jeong, Lee, Kim 2002; Chen, Gong, He 2002; Choi 2001; Zhang, Wang 2001). However, there is no study on modeling through spatial-rule extraction. The present study aimed both to extract spatial rules using a decision tree method and to show that an urban growth model applying the decision tree type of method is more effective than other models.

2 Cellular Automata (CA) and Decision Tree Algorithm

2.1 CA

Cellular Automata (CA) are both a body of knowledge and set of techniques for solving complex dynamic-systems problems. CA evolve in discrete time steps by updating their local state according to a universal rule that is applied to each cell synchronously at each time step. The value (local state) of each cell is determined by a geometrical configuration of neighbor cells, and is specified in the transition rule. Updated values of individual cells then become the inputs for the next iteration. A Cellular Automaton include four components: a cellular space, local states, neighborhoods and a transition rule. The cellular space is a multidimensional space commonly consisting of triangular, square, and hexagonal shapes. Local states represent the status of each cell encoded by numerical values at a given time step. A neighborhood is defined as a set of cells located around a cell of interest. Transition Rules specify how each automaton is updated according to the defined neighborhood and current states of cells (Park 1997). In Fig 1, the lighter color gray cell within the 3 x 3 window (i.e, the neighborhood) is changed to a darker gray cell in a discrete time step according to the transition rule.