Computers, Environment and Urban Systems 59 (2016) 38–49

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Comparing three transition potential models: A case study of built-up transitions in North-East India

Chitrini Mozumder [a,b,*], N.K. Tripathi [a], Chudech Losiri [a]

[a] Remote Sensing and GIS, School of Engineering and Technology, Asian Institute of Technology, P.O. Box 4, Khlong Luang, Pathumthani 12120, Thailand
[b] Remote Sensing and GIS Center, Sultan Qaboos University, P.O. Box 17, Al Khoudh, 123 Muscat, Oman

A R T I C L E   I N F O

Article history:
Received 22 January 2015
Received in revised form 31 March 2016
Accepted 27 April 2016
Available online 21 May 2016

Keywords:
Land Use Land Cover
Transition potential modelling
Multi-Layer Perceptron Neural Network
Logistic Regression
Similarity Weighted Instance-based Learning

A B S T R A C T

Change allocation is an important step in the Land Use Land Cover (LULC) change modelling. Many established LULC models use transition potential maps for the allocation of the estimated land demand. This study compares three commonly used techniques for transition potential modelling: (1) Multi-Layer Perceptron Neural Network (MLP), (2) Logistic Regression (LogReg), and (3) Similarity Weighted Instance-based Learning (SimWeight); and evaluates their applicability for built-up transitions. A case study has been taken from Guwahati city, in North-East India which experiences heterogeneous built-up growth in a limited area within the large topographic variations. With the same set of input and tested driving factors, all three models were simulated for the period 1989–2001 to produce the transition potential maps for 2011 and same amount of land demands, as in 2011 were allocated on the potential maps. The validation was done by (1) a multi-resolution validation method and (2) a region based method using the wards of the city. For this particular study, with the specific landscape environment and scale, MLP produced the most accurate change and predicted areas. The LogReg simulated the no change areas the most accurately, while the SimWeight could generate the edge extensions satisfactorily. We presented a detailed comparison of the change potentials and simulated maps and discuss the importance of evaluating the ability of the transition potential model used for LULC model. The results from this study can assist the LULC modelers to validate their transition potential models for generating accurate prediction maps. It can be also useful for planners and decision makers of Guwahati city and similar landscape, environment, scale in producing accurate transition potential zones for precise built-up growth modelling.

1. Introduction

With the ever increasing popularity of Land Use Land Cover (LULC) change models, the numbers of tools and techniques to efficiently mimic the dynamic processes of change have also increased (Agarwal, Green, Grove, Evans, & Schweik, 2002; Matthews, Gilbert, Roach, Polhill, & Gotts, 2007; Soares-filho, Rodrigues, & Follador, 2013; Verburg & Overmars, 2009). Although these models have different underlying principles, most of these can be compared in terms of the (1) land demand/quantity estimation and (2) change allocation (Pontius, Huffaker, & Denman, 2004), and the accuracy of these two building blocks determines the projection of particular LULC types. The land demand/quantity estimation often involves the interventions from policies and must be flexible depending on the scenarios chosen for the particular study (Batista, Koomen, Diogo, & Lavalle, 2014; Geneletti, 2013; Mozumder & Tripathi, 2014b).

While there are different approaches for change allocation, a large panel of the models uses transition potential or suitability maps as the intermediate softened images to allocate the estimated quantity of change. These images depict range order indices and do not entirely represent the future predicted LULC classes (Camacho Olmedo, Paegelow, & Mas, 2013). The preparation and evaluation of these intermediate maps is therefore critical, since the final land demands are allocated based on them (Eastman, Van Fossen, & Solorzano, 2005). Till date, two different approaches have been adopted for presenting these intermediate softened images: (1) a suitability map which explains the suitability of a particular LULC class for a specific purpose, usually simulated from a later date (T1) of a calibration period (T0-T1) and (2) a transition potential map demonstrating the relative likelihood of transition of a particular pixel of particular LULC class, simulated based on the transition of LULC in the calibration period (Camacho Olmedo et al., 2013; Mas et al., 2014). Both the models, however, are based on the relationships between the driving variables and LULC or the transition type (Chow & Sadler, 2010; Paegelow & Camacho Olmedo, 2008; Veldkamp & Lambin, 2001).

Comparison of the suitability and transition potential maps has been studied previously by Camacho Olmedo et al. (2013). In this study, we focus on comparing three commonly used techniques, Multi-Layer

* Corresponding author at: Remote Sensing and GIS Center, Sultan Qaboos University, P.O. Box 17, Al Khoudh, 123 Muscat, Oman.
E-mail address: chitrini@gmail.com (C. Mozumder).

http://dx.doi.org/10.1016/j.compenvurbsys.2016.04.009

Perceptron (MLP), Logistic Regression (LogReg), and Similarity Weighted Instance-based Learning (SimWeight), to produce the transition potential maps. The MLP Neural Network uses a machine learning approach and is capable of efficiently handling the complex non-linear relationship between the multiple driving variables and the transitions (Bhatti, Tripathi, Nitivattananon, Ahmad Rana, & Mozumder, 2015; Mozumder & Tripathi, 2014b; Oñate-valdivieso & Bosque, 2010; Pijanowski, Brown, Shellito, & Manik, 2002; Thapa & Murayama, 2012). Moreover, it is distribution free and do not consider any underlying model for the multivariate distribution of class specific data in the feature space. However, it is computationally extensive and do not allow much user intervention during the simulation process (Mas, Puig, & Palacio, 2004; Mas et al., 2014; Pérez-vega, Mas, & Ligmann-zielinska, 2012; Sangermano, Labs, & Eastman, 2010). On the other hand, the regression models are the most popular techniques among the LULC modelling researchers. However, multi-collinearity is one of the focal points of criticism for regression models, which is often observed in the individual factors used in the regression analysis (Kolb, Mas, & Galicia, 2013). The SimWeight is also a machine learning process, which is computationally much simpler than the MLP (Sangermano et al., 2010). But, since it works on distance principle, its suitability for heterogeneous transitions modelling is limited.

The selection of the intermediate technique in LULC modelling is user dependent and may also vary with the LULC type. Hence, it is of utmost importance to evaluate which model can explain the interactions between the past changes of the particular LULC type and its driving factors in a more robust and precise way which will be distinct in that particular area of study. However, relatively fewer studies have presented and validated the results of transition potential models exclusively (Camacho Olmedo, Pontius, Paegelow, & Mas, 2015; Camacho Olmedo et al., 2013; Conway & Wellen, 2011; González, Aguilera-Benavente, & Gomez-Delgado, 2015).

Therefore, this study aims at comparing the intermediate softened images (transition potential maps) and the simulated results produced by the MLP, LogReg and SimWeight. All these methods require the inputs of the desired transitions to be modelled and the appropriate explanatory factors driving those transitions. The resultant in either case is the transition potential map for each transition demonstrating an expression of time specific potential for changes. If there is more than one transition with same driving variables, those can be grouped to form a sub-model for producing the transition potential maps. We used same land demand/quantity and driving factors for the simulation to make sure that the only difference in the whole experiment is the use of the method. In addition, we validated transition potential maps as well as the predicted results using a well-established multi-resolution validation technique (Pontius et al., 2004) and a stratified validation technique.

This study particularly considers the built-up growth as the transition class for modelling. A case study was chosen from the major city Guwahati, in North East (NE) India. The city lies between the large river Brahmaputra and several smaller to larger hills leaving very limited space to expand. However, being the centre of the NE India, this city is experiencing a rapid growth in the last two decades. The economic scenario of the city has changed due to rapid industrial development and establishment of commercial companies, especially from the beginning of 21st century. The construction of the expressway from Guwahati University (Jalukbari) to Khanapara (Fig. 1) has catalysed built-up growth in a random fashion away from the city centres (Borthakur & Nath, 2012). Most of these new industrial and educational buildings have been constructed along the roads which were earlier the wetlands or protected areas. This heterogeneous built-up growth makes it difficult to establish its relationship with the factors driving the changes. Therefore, in order to have a better planning for the city and the outskirts, it is essential to analyse the built-up growth and produce accurate transition potential maps for accurate predictions of built-up in the future. This study can help the urban planners in the city and areas of similar geographical and topographical conditions to select a correct method for generating accurate transition potentials and hence develop a precise LULC model.

2. Methods

2.1. Transition potential models: MLP, LogReg, and SimWeight

We compared and evaluated three commonly used transition potential models used in LULC modelling for their similarities, differences, and capabilities to produce accurate transition potential maps. These three methods differ in terms of the algorithm used (statistical versus machine learning approach) and complexity to produce the potential maps.

MLP is one of the most common form of Artificial Neural Network (ANN) models for the feed forward natural network architecture. In general, ANN models are based on networks of biological neurons and are consisted of multiple layers of simple computing nodes that operate as nonlinear summing devices. The MLP constructs a network of neurons between the driving factors and the classes of "change" and "persistence" and a web of connections between the neurons that are applied as a set of (initially random) weights. For instance, if there are three transitions, MLP constitutes six examples of classes: three transition classes and three persistent classes which are fed with the appropriate driving factors.

The second method undertakes binomial Logistic Regression and prediction using the maximum likelihood technique. Unlike the other two methods used in this study, it is not a machine learning process and assumes that if the probability of a pixel changing from one class to another (for example, built-up) follows a logistic curve as described by a Logistic Regression (Eq. (1)); then the probability of the transition of a pixel into built-up can be estimated by the LogReg model given by Eq. (2). In the equation, $\mathrm{P}(\mathrm{Y}=1|\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_k)$ is the probability of the dependent variable Y being 1 given $(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_k)$, i.e. the probability of a cell being urbanized; X_i is an independent variable representing a driving force of urbanization, which can be of interval, ordinal or categorical nature; and β_i is the coefficient for variable X_i (Hu & Lo, 2007).

$$f(z)=\frac{1}{1+e^{-z}} \tag{1}$$

$$\mathrm{P}(\mathrm{Y}=1|\mathrm{X}_1,\mathrm{X}_2,\ldots\mathrm{X}_k)=\frac{1}{1+e^{-\left(\alpha+\sum(\beta_i\mathrm{X}_i)\right)}}. \tag{2}$$

SimWeight is based on a modified K-nearest neighbour machine learning algorithm which calculates the weighted distances in variable space to known events for the classes (Sangermano et al., 2010). For the generation of suitable maps in land change modelling, SimWeight identifies "change" and "persistence" for each transition. For each pixel to be evaluated, the procedure then extracts the K-nearest neighbours (either change or persistence) and computes the distance in the variable space from each unidentified location to the events of change that occurs in the range of K. This distance is used in an exponential weighting function in order to calculate a continuous surface of class membership (Eq. (3)). A higher membership in the change class implies that a pixel has similar environmental conditions of a changed pixel, and therefore it can be considered to be highly potential for the change. In the Eq. (3), 'K' is the number of closest pixels (change + persistence) of a pixel, 'c' is the number of change pixels within the K-nearest neighbours and d is the distance to a change event "*i*". It is recommended to work with a K about 1/10th of the sample size.

$$Membership_{change}=\frac{\sum_{i=1}^{n}\left(1.0-\frac{1}{1+e^{\frac{1}{d_i}}}\right)}{\mathrm{K}}(\mathrm{c}\leq\mathrm{K}) \tag{3}$$