Information Sciences 330 (2016) 358–369

Contents lists available at ScienceDirect

Information Sciences

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

Estimating spatial logistic model: A deterministic approach or a heuristic approach?

Xinxin Zhang [a], Bo Huang [b,*], Richard Tay [c]

[a] School of Computer and Information Engineering, Xiamen University of Technology, 600 Ligong Rd, Jimei Dist, Xiamen 361024, China
[b] Department of Geography and Resource Management and Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, NT, Hong Kong
[c] School of Business IT and Logistics, RMIT University, GPO Box 2476, Melbourne, VIC 3001, Australia

A R T I C L E   I N F O

Article history:
Received 11 December 2006
Revised 30 April 2015
Accepted 2 October 2015
Available online 19 October 2015

Keywords:
Logistic regression model
spatial sampling and filtering
Newton-Raphson method
genetic algorithm
land use

A B S T R A C T

This paper evaluates the performance of a deterministic method (Newton–Raphson, NR) and a heuristic method (Genetic Algorithm, GA) for solving the maximum likelihood estimation in spatial logistic analysis. A spatial logistic regression model equipped with a filtering process is formulated to examine the relationship between land use change and various spatial determinants such as population density, distance to road, distance to commercial center and neighborhood characteristics. Geographic Information System (GIS) is used to acquire spatial samples and perform spatial analysis. The NR method and GA are utilized, respectively, to estimate the coefficients that maximize the likelihood of the spatial logistic regression model. Both methods are compared in terms of the maximum likelihood and computing time. The experimental results show that the NR method can achieve a better likelihood and is also much faster than the GA method. Therefore, the NR method is recommended for estimating a spatial logistic regression model although GA can also be employed.

1. Introduction

Logistic regression has been widely used in land use – transportation planning [2,14,16]. More recently, a number of studies have focused on exploring the major determinants of rural–urban conversion and on the estimation of land transformation models [1,5,9]. In general, land transformation models are used to explore the various channels of land use and the social, economic, and spatial variables influencing that usage. It should be noted that despite some environmental concerns, urban development contributes significantly to a nation's economy and growth, and it is thus important for governmental agencies to understand the trends and determinants in land use change to facilitate proper planning and the management of scarce resources.

Most of these studies, however, employed only one specific method to estimate the coefficients for the logistic model and none of them discussed the viability of alternate methodologies for estimation of these land use and development models. The purpose of this paper is to develop a spatial logistic regression model and estimate it using a mathematical inference method (Newton–Raphson, NR) and an evolutionary computation method (Genetic Algorithm, GA). A spatial logistic regression model equipped with a filtering process will be formulated to explore the urbanization patterns in New Castle County in Delaware. The filtering process that has not been attempted in land use change modeling can minimize spatial autocorrelation existing in the

* Corresponding author. Tel.: +852 39436536; fax: +852 26035006.
E-mail addresses: zhangxinxin@xmut.edu.cn (X. Zhang), bohuang@cuhk.edu.hk (B. Huang), richard.tay@rmit.edu.au (R. Tay).

http://dx.doi.org/10.1016/j.ins.2015.10.015
0020-0255/

data, thus improving the modeling performance. Both the NR method and GA are utilized to estimate the coefficients for the regression model.

In the next section, the rural–urban land conversion model will be developed while the following two sections describe the implementation of NR method and GA for estimating the parameters of the regression model. Then the estimation results, and their comparison, are presented, followed by a discussion of the results and the performance of the two methods. The final section provides some concluding remarks.

2. Rural–urban land conversion modeling

Several disciplines such as geography, economics, real estate, urban planning and civil engineering have applied their own criteria and framework to model land usage. In general, land use change is influenced by a number of factors. Earlier studies reveal that no single set of factors can explain the changes in different places because each context is different. More often than not, land use studies reported different driving forces behind changes in land use patterns at different places.

In their case study on China, Cheng and Masser [4] focused on land conversion from rural to urban use and their literature review identified many important factors such as investment demand, industry structure, housing commercialization, land leasing and decentralization of decision-making. However, their model included only the industry structure, transportation networks and existing developed areas. The model also included the constraints imposed by water availability and places unfit for urban development. In another study, Landis and Zhang [12] investigated land use change near a railway station and included four classes of information: transportation network, urban structure (residential, commercial, public and industrial buildings), locations that are unconstrained with respect to change and locations where no change could occur because of constraints. It also demonstrated that explanatory factors do not need to be numerous, provided they are relevant.

Typically, land use changes from rural to urban seem to be influenced by a few recurrent parameters that cannot be overlooked. It is obvious that a city will grow if its population increases [17]. Consequently, new residential areas will emerge in close proximity to transportation facilities (roads, railways, bus lines) and commercial centers also develop concurrently. In the meantime, industrial buildings will develop in the vicinity of those existing previously. On the whole, urban expansion will transform vacant or low rent areas into developed areas. The agglomeration of developed areas and the availability of exploitable sites will thus significantly influence urban growth patterns. Therefore, this study will focus on developmental influences such as population migration, the proximity of residential, commercial and industrial areas as well as transportation network, and the proportion of urban cells and rural cells in neighborhood. The land available for change includes agricultural and other undeveloped land, whereas forests, wetlands, or parks are considered inappropriate for development. The latter assumption accounts for policies favoring the conservation of natural resources.

2.1. Spatial logistic regression model

Logistic regression is widely used to model the outcomes of a categorical dependent variable while the independent variables could be a mixture of continuous and categorical variables. Hence, it is a suitable approach to estimate the coefficients of explanatory factors from the observation of land use conversion because urbanization does not usually follow normal assumption and its determinants are usually a mixture of continuous and categorical variables.

The general form of logistic regression is given by:

$$u = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K \tag{1}$$

$$u = \log\left(P/(1-P)\right) = \log it(P) \tag{2}$$

$$P = e^u/(1+e^u) \tag{3}$$

where P refers to the probability of occurrence of a new unit, u is a linear-in-parameters utility function, $x_1, x_2, \ldots, x_K$ are K explanatory variables and $\beta_0, \beta_2, \ldots, \beta_K$ are K+1 parameters to be estimated.

While employing logistic analysis to model rural–urban land conversion, the heterogeneity of spatial data should be considered to remove spatial auto-correlation [15]. Otherwise, unreliable parameter estimation or inefficient estimates and false conclusions regarding hypothesis test will result. There are two fundamental approaches to consider spatial dependence: building a more complex model to incorporate an autogressive structure [18] and designing a spatial sampling scheme to expand the distance interval between sampled sites. Spatial sampling leads to a smaller sample size that loses certain information and conflicts with the large-sample requirement of asymptotic normality of maximum likelihood method, upon which logistic regression is based. Nevertheless, it is a more sensible approach to remove spatial auto-correlation and a reasonable design of spatial sampling scheme will make a perfect balance between the two sides.

Theoretically, spatial autocorrelation should be subject to distance decay. In this study, the effects of different sampling schemes on eliminating spatial autocorrelation are carefully evaluated and a novel spatial filtering process based on local Getis Statistic [13] is proposed. Specifically, an optimal sampling scheme is employed to verify the spatial autocorrelation of observations proximate in space, and a special check of *joins* is determined each time by comparing land use change types of ‘adjacent’ cells. A join is defined by sequential occurrences of similar land use changes in adjacent cells, which are those central cells of the current sampling window and the sampling windows to the east, west, north, and south of this window. In our case, we attempt