

地理元胞自动机及空间动态转换规则的获取*

杨青生

(广东商学院旅游与环境学院, 广东 广州 510230 // 中山大学地理科学与规划学院, 广东 广州 510275)

摘要: 以东莞市黄江等镇城市发展为例, 利用逐步逻辑回归方法, 获取分区的元胞自动机动态规则, 用分区的动态规则 CA 模拟了研究区 1988-2004 年的城市扩张, 采用逐点对比法和 Moran I 指数对模拟结果进行了评价。结果表明, 分区的空间动态转换规则比统一的静态转换规则能获得更高的模拟精度, 动态转换规则克服了传统静态转换规则无法反映区域内部城市发展差异的缺陷。

关键词: 元胞自动机; 动态转换规则; 逐步逻辑回归; 城市模拟

中图分类号: P209; TU984 **文献标识码:** A **文章编号:** 0529-6579 (2008) 04-0122-06

元胞自动机 (Cellular Automata, 简称 CA) 具有强大的空间运算能力, 可以有效地模拟复杂的动态系统。近年来, CA 已被越来越多地运用在城市模拟中, 取得了许多有意义的研究成果^[1-3]。如 Wu 等^[4]模拟了广州市的城市扩展; 黎夏和叶嘉安^[5]模拟了东莞市的城市扩张。这些研究表明, CA 能模拟出与实际城市非常接近的特征, 可以由此预测未来城市的发展及土地利用变化, 为城市和土地利用规划提供决策依据。

CA 的特点是通过一些简单的局部转换规则, 模拟出全局的、复杂的空间模式^[6]。转换规则中的变量对应着很多参数, 这些参数值反映了不同变量对模型的“贡献”程度。如何在转换规则中合理地确定符合研究区的参数值是 CA 模拟中的难点。目前, CA 主要通过转换矩阵^[2]、多准则判断 (MCE)^[7]、层次分析法 (AHP)^[8]和灰度^[9]、主成分分析^[10]、自适应模型^[11]、人工神经网络模型^[5]、决策树^[12]等方法确定模型的参数值。这些 CA 模型的参数值在很大的地域范围内是静态的, 不随区域空间位置的差异而变化。当研究区较大, 区域内部自然、社会经济状况差异比较大时, 这种静态的转换规则难以体现研究区域内部的差异。运用这种静态的转换规则模拟整个研究区时, 不是非常合理。所以, 寻找随区域空间位置的差异, 模型参数值相应变化的动态转换规则是非常重要的。

本文以东莞市的黄江、樟木头、塘厦、清溪、凤岗等镇为例, 采用随空间而变化的动态转换规则

进行了城市扩张的模拟。根据该研究区内部自然条件、社会经济状况的差异, 将其分为若干子区。通过逐步逻辑回归法提取各个子区的 CA 转换规则, 得到研究区的动态转换规则。利用该模型模拟了研究区在 1988-2004 年的城市扩张, 并分析了各镇城市扩张的模式和特征。

1 分区的动态转换规则 CA

1.1 基于逻辑回归的 CA

利用 CA 模拟城市空间结构的演变过程是由转换规则来控制的。在定义 CA 的转换规则时, 需要确定被模拟对象的一系列影响因素。城市空间结构演变受诸多因素影响, 包括区域的社会经济条件、人口状况等。离市中心的距离、离道路的距离等空间变量也对城市的发展有重要的影响。另外, 邻近范围的影响是 CA 模型的主要考虑因素。

因此, 在城市模拟中, 元胞的城市发展概率与这些全局变量、区域变量、局部变量有关。这种关系可以用逻辑回归模型表示, 逻辑回归模型的变量参数值可通过历史数据校正。

基于逻辑回归模型校正 CA 转换规则的公式为:

$$p_{ij} = \frac{\exp(z_{ij})}{1 + \exp(z_{ij})} = \frac{1}{1 + \exp(-z_{ij})} \quad (1)$$

式中 p_{ij} 表示 i, j 元胞在道路、商业中心、居住中心等区域空间变量的作用下, 元胞的城市发展概率。其中, z_{ij} 的计算如下:

* 收稿日期: 2007-12-14

基金项目: 国家杰出青年基金资助项目 (40525002); 国家自然科学基金资助项目 (40471105); 国家高技术研究发展计划资助项目 (2006AA12Z206)

作者简介: 杨青生 (1974 年生), 男, 博士, 讲师; E-mail: qsyang2002@163.com

$$z_{ij} = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n \quad (2)$$

式中, $x_1, x_2, \cdots, x_n$ 为区域空间变量, 如离公路的最短距离, 离铁路的最短距离、离商业中心、居住中心的最短距离等, $a_0, a_1, \cdots, a_n$ 为相应各变量的回归系数。

考虑到邻近范围城市化元胞对中心元胞的影响, 定义邻域作用值:

$$\Omega'_{ij} = \frac{\sum_{3 \times 3} \text{con}(s_{i,j} = \text{urban})}{3 \times 3 - 1} \quad (3)$$

式中 Ω'_{ij} 表示 t 时刻 i, j 元胞的 3×3 邻域作用值。其中, $\text{con}()$ 为条件函数, 如果元胞为城市元胞, 则值为 1, 否则为 0。

道路、河流、陡峭的山地、优质的农田等发展为城市用地的概率较小, 需加一约束条件。这样, 综合考虑局部邻近范围、元胞约束条件的作用及区域道路、商业中心等变量的作用下, 某元胞在 $t+1$ 时刻发展为城市用地的概率 $p_{c,ij}^{t+1}$ 为:

$$p_{c,ij}^{t+1} = p_{ij} \times \text{con}(s_{ij}^t = \text{suitable}) \times \Omega'_{ij} \quad (4)$$

式中, $\text{con}()$ 为条件函数, 取 $[0, 1]$ 间的值。如果元胞的约束性很大, 无法转变为城市用地, 该值为 0。如果完全没有约束性, 该值为 1。

由于城市系统的不确定性, 有必要引入一随机变量^[6]。该随机项可以表达为:

$$RA = 1 + (-\ln\gamma)^\alpha \quad (5)$$

式中, γ 为 $[0, 1]$ 间的随机数, α 为控制随机变量大小的参数。

综合考虑上述因素, 某元胞 $t+1$ 时刻发展为城市用地的概率 $p_{d,ij}^{t+1}$ 为:

$$p_{d,ij}^{t+1} = RA \times p_{c,ij}^{t+1} = (1 + (-\ln\gamma)^\alpha) \times \frac{1}{1 + \exp(-z_{ij})} \times \text{con}(s_{ij}^t) \times \Omega'_{ij} \quad (6)$$

式中, Ω'_{ij} 随着时间 t 的变化而动态计算。在每次循环中, 将该发展概率与预先给定的阈值进行比较, 确定该元胞是否发生状态的转变。

1.2 分区的动态转换规则 CA 模型

不同的地区, 区位条件、自然条件和社会经济条件不尽相同。这种差异导致了不同区域相同的空间变量对城市发展作用的程度不同。传统的逻辑回归模型校正 CA 转换规则时, 在不同的子区运用相同的空间变量参数值, 影响了模型模拟的精度。获取分区的动态规则具有重要的意义。

为获取分区的动态规则, 首先将研究区按区位、自然条件和社会经济条件的差异分成若干子区, 然后在每个子区内用逐步逻辑回归模型获取各个空间变量的参数值, 形成整个研究区的动态转换

规则。即:

$$p_{r,ij} = \frac{\exp(z_{r,ij})}{1 + \exp(z_{r,ij})} = \frac{1}{1 + \exp(-z_{r,ij})} \quad (7)$$

式中 $p_{r,ij}$ 表示 r 子区 i, j 元胞在道路、商业中心、居住中心等区域空间变量的作用下, 元胞的城市发展概率。 $p_{r,ij}$ 随子区的变化而动态变化, r 为不同的子区。其中:

$$z_{r,ij} = a_{r,0} + a_{r,1}x_1 + a_{r,2}x_2 + \cdots + a_{r,n}x_n \quad (8)$$

式中, $x_1, x_2, \cdots, x_n$ 为区域空间变量, 如离公路的最短距离, 离铁路的最短距离、离商业中心、居住中心的最短距离等。 $a_{r,0}, a_{r,1}, \cdots, a_{r,n}$ 为对应变量的回归系数。其中, $a_{r,0}, a_{r,1}, \cdots, a_{r,n}$ 等回归系数随子区 r 的变化而相应变化, 在整个研究区形成空间动态的转换规则。在每个子区校正模型参数时, 运用逐步逻辑回归的方法, 选取主要的空间变量, 剔除影响很小的变量, 达到变量约简的目的。

考虑到邻近范围、约束条件、不确定性因素的影响, 某元胞 $t+1$ 时刻发展为城市用地的概率 $p_{d,ij}^{t+1}$ 为:

$$p_{d,ij}^{t+1} = RA \times p_{r,ij} \times \text{con}(s_{ij}^t) \times \Omega'_{ij} = (1 + (-\ln\gamma)^\alpha) \times \frac{1}{1 + \exp(-z_{r,ij})} \times \text{con}(s_{ij}^t) \times \Omega'_{ij} \quad (9)$$

式中, Ω'_{ij} 随着时间 t 的变化而动态计算, $p_{r,ij}^{t+1}$ 随 $p_{r,ij}$ 的变化在空间上动态变化。在每次循环中, 将该发展概率与预先给定的阈值进行比较, 确定该元胞是否发生状态的转变。基于逐步逻辑回归模型分区的动态转换规则 CA 模型如图 1。

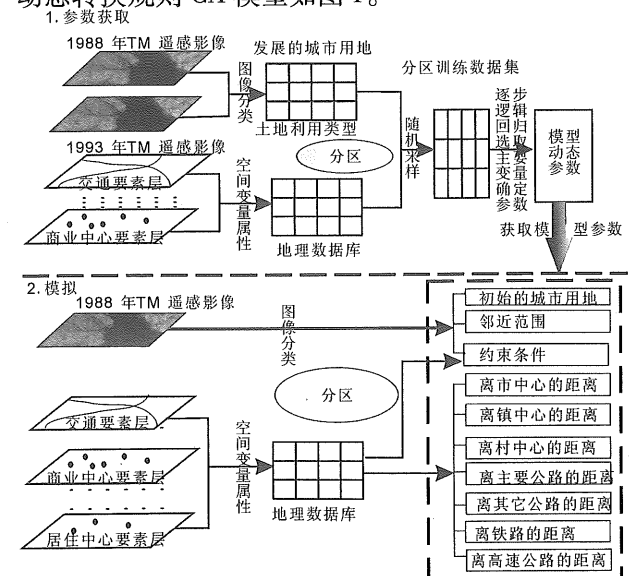


图1 基于逐步逻辑回归模型分区的动态转换规则 CA 模型

Fig. 1 Dynamical transition rule CA model based on stepwise logistical regression

2 应用

2.1 研究区概况

本文以东莞市的黄江、樟木头、塘厦、清溪、凤岗等镇为研究区,模拟 1988-2004 年的城市扩展。东莞市位于广东省中南部,珠江口东岸,北接广州,南连深圳。1988-1992 年间,黄江等镇人口不断增加,社会经济发展水平不断提高,工业企业数量迅速增长,工业总产值增长速度很快。在国内生产总值中,工业生产总值所占的比重越来越大。但各镇的人口增长速度、工业企业增长速度、工业总产值增长速度差异较大。黄江人口数从 1988 年的 15 468 万人增长到 1992 年的 16 404 万人,增长速度较快,其次是塘厦和樟木头,凤岗增长速度最慢。黄江和塘厦工业总产值的增长速度很快,黄江从 1990 年的 3 236 万元增长到 1992 年的 22 312 万元,其次是樟木头和凤岗。自然环境方面,樟木头多丘陵,其它镇地势较为平坦。

黄江等镇不断增长的人口和工业企业,对城市用地的需求越来越多。在 1988-1993 年间,由于各镇间自然环境、社会经济条件变化差异较大,城市用地的需求量、城市用地的增长速度差异较大。为反映各镇的城区用地扩展,将各个镇作为子区,模拟城市用地的扩张。基于分区的动态转换规则 CA 模型,研究各镇城市扩张模式和特征。

2.2 空间变量及数据获取

元胞的城市发展概率跟一系列空间距离变量、邻近范围的城市化元胞数、元胞本身的属性等关系密切。结合研究区的实际情况,本文选取了以下空间变量。这些变量及获取情况见表 1。

表 1 空间变量及获取方法

Tab. 1 Spatial variables and obtaining methods

变量	获取方法
因变量 (是否转变为城市元胞)	遥感分类, reclass
空间 距离 变量	离市中心的距离 (x_1)
	离镇中心的距离 (x_2)
	离村中心的距离 (x_3)
	离国道、省道的距离 (x_4)
	离其它公路的距离 (x_5)
	离铁路的距离 (x_6)
局 部 变 量	离高速公路的距离 (x_7)
	3×3 邻域已城市化元胞数 (x_8)
	元胞的土地利用类型 (x_9)

为了在各个子区获取模型所需的参数,需要用历史数据来校正。本文选取黄江等镇 1988 年和 1993 年的 TM 遥感图像,通过遥感分类,获取元胞的土地利用类型。TM 的分辨率是 30 m×30 m,重采样成 50 m×50 m 分辨率。

选取历史数据时,本文运用随机分层取样的方法,从转换为城市用地的元胞和可以转换为城市元胞而尚未转换的元胞中分别选择 20% 的样点,获取这些样点的空间坐标。运用 ARC/INFO 的 Sample 功能分层读取对应的城市发展和空间变量数据,利用逐步逻辑回归校正的方法获取各个子区空间变量的参数。

2.3 分区的动态规则

为获取空间动态转换规则,运用 Visual Basic 6.0 和 Arcobjects 建立 CA 模型。用 Arcobjects 组件开发的方法,既能充分利用现有 GIS 软件的强大功能,在计算邻域元胞数和最短距离时,快速准确,而且不用考虑数据格式的转换,同时能用高级语言功能强大、灵活的特点,编制适合 CA 空间动态转换规则的工具。

由于各个子区城市发展的差异性,在不同子区选择空间变量,确定参数值时,采用了逐步逻辑回归的方法。各个子区选取的空间变量及其参数值见表 2。

表 2 的参数值表明,离市中心的距离和离高速公路的距离对黄江等镇城市发展概率的“贡献”很小,可以忽略不计。这是因为这些镇离市中心和高速公路距离都很远。道路对这些镇的城市发展概率“贡献”很大,尤其是国道和省道影响非常大。铁路对城市发展的概率影响较小。镇中心对城市发展的概率影响较大,小商业中心对城市发展概率也有一定的影响。

就各镇之间的发展模式而言,黄江、塘厦和凤岗三镇镇中心和村中心的权重较大,达到 -0.02,然后是道路的权重较大,达到 -0.2 及 -0.08。所以,黄江、塘厦和凤岗表现为沿镇中心和主要道路(国道、省道)的点-轴同步发展模式。清溪的权重分配比较均匀,镇中心、村中心的权重较大,城市扩张倾向于沿镇中心的中心发展模式。樟木头镇道路的权重很高,达到 -0.297 和 -0.203,其它变量的权重比较均匀,樟木头镇更多地表现为沿省道发展的轴线(道路)发展模式。

表2 研究区各子区空间变量及参数值

Tab. 2 Parameters of spatial variables in different subdistrict

项目	con	x_1	x_2	x_3	x_4	x_5	x_6	x_7
黄江	0.525	—	-0.026	0.087	-0.220	—	—	—
樟木头	1.171	—	0.020	-0.009	-0.297	-0.203	-0.02	—
清溪	1.939	—	-0.039	-0.004	-0.144	-0.067	—	—
塘厦	0.537	—	-0.016	-0.027	-0.040	-0.079	0.033	—
凤岗	0.164	—	0.022	-0.025	-0.082	0.076	-0.017	—
整个研究区	0.919	—	-0.005	-0.011	-0.067	-0.050	-0.006	—

2.4 模拟结果与精度评价

获取参数后, 利用分区空间动态规则 CA 模型对城市发展进行模拟。模拟时, 依据 1988 年的土地利用类型, 将城市元胞、林地元胞、河流元胞在约束条件中设置为不可能转换为城市用地。同时, 根据各镇的特点, 依据道路的通达性, 河流、铁路、高速公路等的阻碍程度, 分别设置约束条件。模拟的初始状态利用 1988 年的 TM 图像获取, 经过 200 次迭代运算, 模拟出了 1993 年的城市用地 (图 2)。由于城市的发展受到许多不确定性因素的影响, 完全模拟出城市发展是非常困难的。

型模拟的混淆矩阵, 总精度为 81.2%, Kappa 系数为 0.55, 该精度在可接受的范围之内。

对采用分区的动态规则和统一的静态规则的模拟结果 (图 3D) 进行了比较, 表 4 是统一的静态规则模拟的混淆矩阵, 总精度为 77.4%, Kappa 系数为 0.46。这表明利用分区的动态规则 CA 模型模拟的精度要高于统一的静态规则 CA 模型模拟的精度, 因为分区的动态规则更符合各个子区城市发展模式, 能在各个子区内达到最优模拟效果, 从而达到研究区局部-整体最优模拟结果。

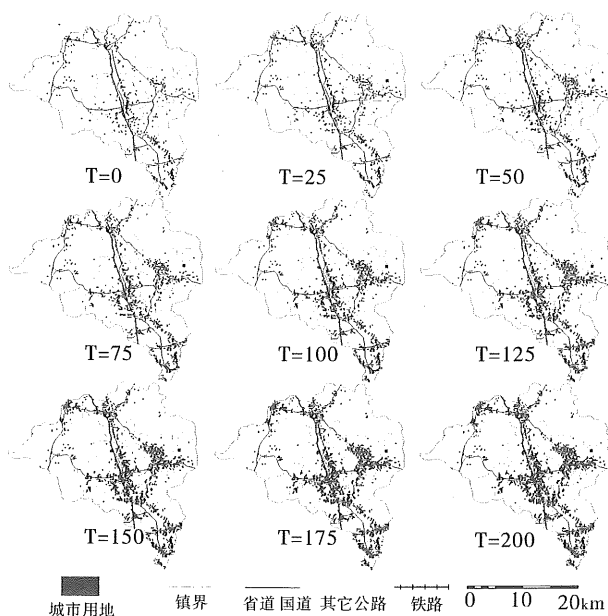


图2 1988-1993年研究区城市发展过程模拟

Fig. 2 Simulation procedure of urban development from 1988 to 1993

图 3A 是 1988 年从遥感图像分类中获取的城市用地, 作为模拟的初始用地。我们将 1993 年从遥感图像分类中获取的实际城市用地 (图 3B) 与分区的动态规则 CA 模型模拟的结果 (图 3C) 进行了比较, 发现分区的动态规则 CA 模型模拟的结果与实际非常接近。表 3 是分区的动态规则 CA 模

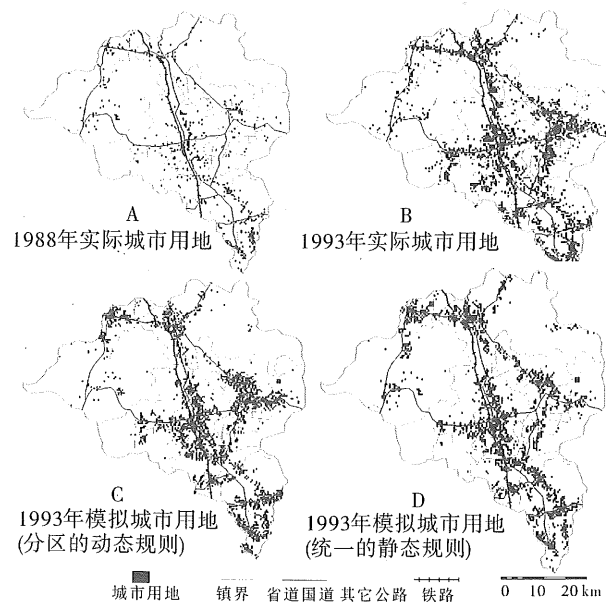


图3 研究区利用不同转换规则模拟对比

Fig. 3 Real urban distribution and simulation result with dynamical CA model and static CA model

精度检验时, 为了比较城市总体的形态差异, 计算了 Moran I 指数和 Geary c 指数。Moran I 指数和 Geary c 指数通常是描述空间自相关的变量, 也可以表示集中和分散的程度。Moran I 指数取值 0-1 之间, 越接近于 1, 表明集中程度越高。Geary c 指数与 Moran I 指数存在负相关关系。分区的动态规则模拟结果的 Moran I 指数和 Geary c 指数分

别为 0.485 1 和 0.509 1, 实际的指数分别为 0.457 4 和 0.5355, 而统一的静态规则模拟结果的指数分别为 0.480 6 和 0.514 6, 表明利用分区的动态规则模拟的结果比统一的静态规则模拟结果效果好。

表 3 分区的动态规则模拟的精度

Tab. 3 Overall accuracy of urban landuse with dynamical CA model

项目	模拟			
	不转变	转变	正确比/%	
实际	不转变	10632	2118	83.4
	转变	1182	3614	75.4
	总精度	81.2		
Kappa 系数				0.55

表 4 统一的静态规则模拟的精度

Tab. 4 Overall accuracy of urban landuse with static CA model

项目	模拟			
	不转变	转变	正确比/%	
实际	不转变	10889	2568	80.9
	转变	1558	3238	67.5
	总精度			77.4
Kappa 系数			0.46	

表 3、4 的结果表明, 利用分区的动态规则模拟得到的逐点对比精度要高于利用统一的静态转换规则模拟的结果, 但在总体形态上要比实际更为集中。利用统一的静态规则模拟的总体形态更好, 但逐点对比精度相对低于分区的动态规则模拟的精度。形成这种结果的原因是, 利用分区获取动态转换规则时, 由于区域较小, 随机样点的选择集中在镇中心地带, 运用这些样点训练得到的参数值中“离中心的距离”变量对模型“贡献”更大, 模拟时在镇中心地区城市转换量稍多。而区域扩大时, 随机样点的选择沿道路分布更多, 运用这些样点训练得到的参数值中“离道路的距离”变量对模型“贡献”更大, 模拟时沿道路转换的城市用地比实际更多 (图 3 C, D), 模拟的城市总体形态更为完整。

利用 1988 - 1993 年训练数据中获取的分区的动态规则, 以 2004 年从遥感图像分类中得到的城市用地总量为参照, 经过 1000 次迭代运算, 模拟了研究区 2004 年的城市用地状况 (图 4)。

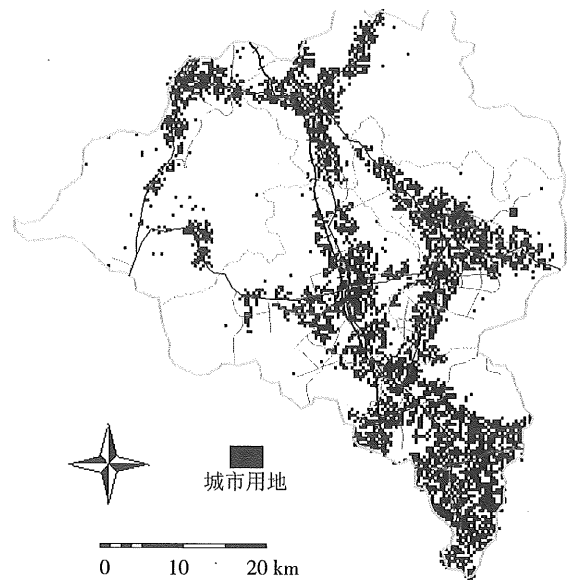


图 4 利用分区的动态规则模拟的 2004 年城市用地

Fig. 4 Urban development simulation of 2004 with dynamical CA model

3 结 论

城市 CA 模拟中, 如何合理地确定符合研究区的转换规则至今没有非常好的解决方法。尽管不同的学者研究不同区域时, 得到了不同的转换规则, 但在较大的研究区或较长时间中这种静态转换规则有局限性。获取空间动态变化的转换规则具有重要的意义。

本文运用逻辑回归方法, 在分析自然、社会、经济差异的基础上, 依据行政边界划分子区, 提取了分区的动态规则, 用动态转换规则 CA 模型对研究区进行模拟, 并用逐点对比和 Moran I 指数进行了精度评价。结果表明, 运用空间上动态的转换规则模拟的效果要比静态转换规则模拟精度高。

逐步逻辑回归方法校正 CA 模型参数, 获取的空间动态转换规则选择了模型的主要空间变量, 达到了变量约简的目的, 节省了模型的运行时间。同时, 通过获取的空间动态转换规则, 能明确地体现大区域中各个子区间的城市发展模式和特征, 为分析区域内部的差异提供了研究方法和手段。

参考文献:

- [1] BATTY M, XIE Y. From cells to cities[J]. Environment and planning B, 1994, 21: 531 - 548.
- [2] WHITE R, ENGELN G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns[J]. Environment and Planning A, 1993, 25: 1175 - 1199.

- [3] LI X, YEH A G O. Principal component analysis of stacked multi-temporal images for monitoring of rapid urban expansion in the Pearl River Delta[J]. *International Journal of Remote Sensing*, 1998, 19: 1501 – 1518.
- [4] WU F. Calibration of stochastic cellular automata: the application to rural-urban land conversions[J]. *Int J geographical information science*, 2002, 16(8): 795 – 818.
- [5] 黎夏, 叶嘉安. 基于神经网络的单元自动机 CA 模拟及真实和优化城市的模拟[J]. *地理学报*, 2002, 57(2): 159 – 166.
- LI Xia, YEH Anthony Gar-On. Neural – network-based cellular automata for realistic and idealized urban simulation[J]. *Acta Geographica Sinica*, 2002, 57(2): 159 – 166.
- [6] 黎夏, 叶嘉安. 约束性单元自动演化 CA 模型及可持续发展城市形态的模拟[J]. *地理学报*, 1999, 54(4): 289 – 298.
- LI Xia, YEH A G O. Constrained cellular automata for modelling sustainable urban forms[J]. *Acta Geographica Sinica*, 1999, 54(4): 289 – 298.
- [7] WU F, WEBSTER C J. Simulation of land development through the integration of cellular automata and multicriteria evaluation[J]. *Environment and Planning B*, 1998, 25: 103 – 126.
- [8] WU F. Simland: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules[J]. *Int J geographical information science*, 1998, 12(1): 63 – 82.
- [9] LI Xia, YEH A G O. Modelling sustainable urban development by the integration of constrained cellular automata and GIS[J]. *International Journal of Geographical Information Science*, 2000, 14(2): 131 – 152.
- [10] LI X, YEH A G O. Integration of principal components analysis and cellular automata for spatial decisionmaking and urban simulation[J]. *Science in China: Series D*, 2002, 45(6): 521 – 529.
- [11] ALLEN P M. Cities and regions as self-organizing systems: Models of complexity [M]. Amsterdam: Gordon and Breach Science Publishers, 1997.
- [12] LI X, YEH A G O. Knowledge discovery for geographical cellular automata[J]. *Science in China: Series D*, 2005, 48(10): 1758 – 1767.

Dynamic Transition Rules for Geographical Cellular Automata

YANG Qing-sheng

(School of Tourism and Environment, Guangdong University of Business Studies, Guangzhou 510230 //
School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: There is rapid development of CA models for simulation of land use patterns and urban systems recently. The existing traditional models use the same transition rules in large areas or at long times. These models have limitations because they use the same factors and weights for different places and time intervals. In this paper, a method using dynamic transition rules was presented for producing more reliable urban simulation. The advantages of dynamic transition rules are apparent. Firstly, it uses different factors and weights in different place which is consistent with actual urban systems. Secondly, the main factors can be obtained from the dynamic transition rules in different places which may produce different urban development patterns. Based on the stepwise logistical regression model, dynamic transition rules were obtained for simulating the evolution of the urban systems in five towns of the Dongguan city from 1988 to 2004. Compared with the general CA model, the simulation precision is improved by using this method. From the factors and weights of dynamic transition rules, three types of development patterns were obtained: center development pattern, road development pattern and center-road development pattern.

Key words: cellular automata; dynamic transition rules; logistical regression; urban systems