

基于遗传算法自动获取 CA 模型的参数

——以东莞市城市发展模拟为例

杨青生, 黎 夏*

(中山大学地理科学与规划学院, 510275)

摘要: 本文提出了基于遗传算法来寻找 CA 模型最佳参数的方法。CA 被越来越多地应用于城市和土地利用等复杂系统的动态模拟。CA 模型中变量的参数值对模拟结果有非常重要的影响。如何获取理想的参数值是模型的关键。传统的逻辑回归模型运算简单, 常常用来获取模型的参数值, 要求解释变量间线性无关, 所以获取的城市 CA 模型参数具有一定的局限性。遗传算法在参数优化组合、快速搜索参数值方面有很大的优势。本文利用遗传算法来自动获取优化的 CA 模型参数值, 并获得了纠正后的 CA 模型。将该模型应用于东莞 1988~2004 年的城市发展的模拟中, 得到了较好的效果。研究表明, 遗传算法可以有效地自动获取 CA 模型的参数, 其模拟的结果要比传统的逻辑回归校正的 CA 模型模拟精度高。

关键词: 元胞自动机 (CA); 遗传算法 (GA); 城市模拟

文章编号: 1000-0585(2007)01-0229-09

1 引言

元胞自动机 (Cellular Automata, 简称 CA), 是一个由具有离散、有限状态的元胞组成的元胞空间, 并按照一定的局部规则, 在离散的时间维上演化的动力学系统^[1]。近年来, CA 已被越来越多地运用在地学模拟中, 取得了许多有意义的研究成果^[2,3]。如 White 等运用约束性元胞自动机模拟了辛辛那提土地利用的变化^[4]; Wu 模拟了广州市城市的扩展^[5]; 黎夏和叶嘉安模拟了东莞市土地利用状况的扩展及城市扩张^[6]。这些研究表明, CA 能模拟出与实际城市非常接近的特征, 其模拟结果与实际非常吻合。

CA 在离散时间维上的演化, 主要是通过一些简单的局部转换规则, 模拟出全局的复杂空间模式。这些局部转换规则是根据元胞当前状态及其邻近范围元胞状态的函数来确定的。为了模拟城市的实际扩张过程, 除了运用 CA 的局部转换规则外, 还要在转换规则中引入影响城市扩展的区域变量和全局变量^[5]。区域变量主要是一些空间变量, 如到交通路网的最短距离, 到商业中心的最短距离, 到居住中心的最短距离等。全局变量主要是研究范围内的社会经济变量, 如总人口数、国内生产总值等。转换规则中的这些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。研究表明, 这些参数值对模

收稿日期: 2006-03-18; 修订日期: 2006-10-09

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); “985 工程” GIS 与遥感的地学应用科技创新平台项目 (105203200400006)

作者简介: 杨青生 (1974), 男, 青海乐都人, 博士研究生。研究方向: 遥感和地理信息模型。

E-mail: qsyang2002@163.com

* 通讯作者: 黎夏, 教授。E-mail: lixia@mail.sysu.edu.cn

型模拟的结果影响很大。

定义转换规则和寻找适合研究区的参数值是 CA 模拟的关键。目前主要是通过转换矩阵^[4]、多准则判断 (MCE)^[7]、层次分析法 (AHP)^[8]、灰度^[9]、主成分分析^[10]、自适应模型^[11]、人工神经网络模型^[12]、决策树^[13]等方法确定模型转换规则及参数值。如 Wu 选取了离市中心的距离、离居住中心的距离、元胞的数字高程模型 (DEM)、坡度 (slope)、土地利用类型等变量, 利用逻辑回归模型获取了模型的参数值, 模拟了广州市 1973~1993 年的城市扩展^[5]。利用逻辑回归模型校正 CA 模型参数的方法运算方便、易于理解, 被很多学者用来定义 CA 的转换规则及计算参数值。但这种方法本身要求解释变量间线性无关, 在上述变量中, 由于行政中心设置和道路布局的相似性, 变量之间存在较大的相关性, 如离市中心的距离和离镇中心的距离、离铁路的距离和离公路的距离, 利用这种具有线性相关的解释变量用逻辑回归模型校正参数值时, 模型的精度会受到影响。

为了更有效地模拟城市和土地利用复杂系统的特征, 主成分分析、人工神经网络、决策树等非线性方法被用来构造模型及获取参数值。人工神经网络方法本身属于暗箱操作, 对物理机制不容易理解, 且计算过程繁琐。用决策树提取规则时, 提取的规则数量众多, 从中选取规则也比较费时。所以, 有必要寻找更简单有效的方法校正 CA 模型优化的参数。

本文运用遗传算法, 经过染色体编码、建立适应度函数、选择遗传算子和确定遗传算法的进化策略, 对 CA 模型的参数值进行优化组合, 用自适应的方法获取了优化的 CA 模型参数。运用遗传算法优化的 CA 模型, 模拟了东莞市 1988~2004 年的城市扩张。

2 研究方法

2.1 遗传算法 (GA)

遗传算法是一种基于自然选择和遗传变异等生物进化机制的全局性概率搜索算法。与基于导数的解析方法和其他启发式搜索方法 (如爬山方法、模拟退火算法, Monte Carlo 方法) 一样, 遗传算法在形式上是一种迭代方法。从选定的初始解出发, 通过不断迭代逐步改进当前解, 直到最后搜索到最优解或满意解。在遗传算法中, 迭代计算过程采用了模拟生物体的进化机制, 从一组解 (群体) 出发, 采用类似于自然选择和有性繁殖的方式, 在继承原有优良基因的基础上, 生成具有更好性能指标的下一代解的群体^[14~16]。

遗传算法以编码空间代替问题的参数空间, 以适应度函数为评价依据, 以编码群体为进化基础, 以对群体中个体位串的遗传操作实现选择和遗传机制, 建立起一个迭代过程。在这一过程中, 通过随机重组编码位串中重要的基因, 使新一代的位串集合优于老一代的位串集合, 群体的个体不断进化, 逐渐达到最优解, 最终达到求解问题的目的。其中, 参数编码、初始群体的设定、适应度函数的设计、遗传操作的设计和控制参数的设定是遗传算法的五大要素。遗传算法优化问题求解的流程见图 1。

运用遗传算法优化问题求解的步骤如下:

- 1) 选择编码策略, 把参数集合 X 和域转换为位串结构空间 S ;
- 2) 定义适应度函数 $f(x)$;
- 3) 确定遗传策略, 包括选择群体大小 n , 选择、交叉、变异方法, 以及确定交叉概率 p_c 、变异概率 p_m 等遗传参数;
- 4) 随机初始化形成群体 P ;

5) 计算群体中个体位串解码后的适应值 $f(x)$;

6) 按照遗传策略, 运用选择、交叉和变异算子作用于群体, 形成下一代群体;

7) 判断群体性能是否满足某一指标, 或者已经完成预定迭代次数, 不满足则返回步骤 6, 或者修改遗传策略再返回步骤 6.

2.2 基于遗传算法的 CA 模型

2.2.1 城市扩展的 CA 模型

城市系统是动态的复杂系统, 具有开放性、动态性、自组织性和非平衡性的特点, 城市的发展变化受到自然、社会、经济、

文化、政治等多种要素的影响, 其行为过程具有高度的复杂性^[17, 18].

在城市发展模拟中, 与城市社会经济模型等城市发展模型相比, CA 模型因具有自组织性、“自下而上”的研究思路、高分辨率的空间尺度和城市空间的动态变化对城市发展的反馈作用等特点, 被认为是模拟复杂城市系统非常有效的模型之一^[19, 20].

研究表明, 用 CA 模型模拟城市发展时, 元胞的城市发展概率主要与邻近范围元胞的状态, 离商业中心的最短距离、离居住中心的最短距离、离道路的最短距离等变量相关^[1, 5, 9]. 模拟真实城市的发展时, 这些变量的参数值需要通过历史数据校正, 逻辑回归模型常用来校正 CA 模型的参数. 基于逻辑回归模型, 某元胞 $t+1$ 时刻发展为城市用地的概率 $p_{d,ij}^{t+1}$ 为:

$$p_{d,ij}^{t+1} = RA \times \frac{1}{1 + \exp(-z_{ij})} \times \Omega_{ij}^t \times \text{con}(s_{ij}^t) \quad (1)$$

式中, RA 为反映城市系统不确定性的一随机项, 按下式计算:

$$RA = 1 + (-\ln \gamma)^\alpha \quad (2)$$

式中, γ 为 $[0, 1]$ 间的随机数; α 为控制随机变量大小的参数, 一般取 0.5; z_{ij} 是描述元胞 ij 城市发展程度的向量, 按下式计算:

$$z_{ij} = a^0 + a^1 x_1 + a^2 x_2 + \dots + a_m x_m \quad (3)$$

式中, a^0, a^1, \dots, a_m 为空间变量的权重, x_1, x_2, \dots, x_n 为空间变量, 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.

Ω_{ij}^t 表示 t 时刻 ij 元胞的 3×3 邻近范围元胞影响值, 按下式计算:

$$\Omega_{ij}^t = \frac{\sum_{3 \times 3} \text{con}(s_{ij}^t = \text{urban})}{3 \times 3 - 1} \quad (4)$$

$\text{con}()$ 为约束性条件函数, 值域范围 $0 \sim 1$; s_{ij}^t 为 t 时刻 ij 元胞的状态, 如果该状态越为有利于城市发展的状态 (如元胞为开发用地、耕地、低坡度的果园等), 则 $\text{con}(s_{ij}^t)$ 取值越接近于 1, 若该状态为不利于城市发展的状态 (如元胞为水体、建设用地、道路、林地等), 则 $\text{con}(s_{ij}^t)$ 取值越接近于 0.

$\Omega_{ij}^t, \text{con}(s_{ij}^t)$ 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 $P_{\text{threshold}}$ 进行比较, 确定该元胞是否发生状态的转变.

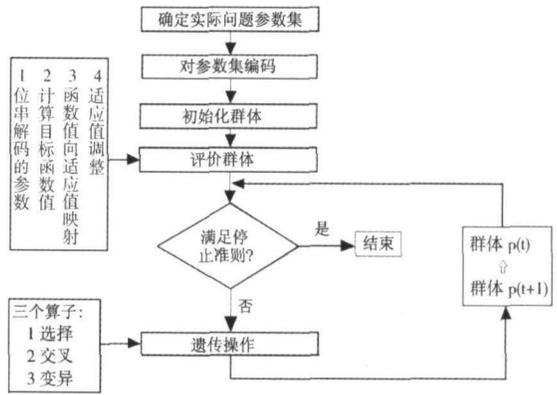


图 1 遗传算法优化问题求解流程图

Fig 1 Work flow of problem optimization with GA

用逻辑回归模型校正式中的参数时, 由于逻辑回归模型本身的求解过程, 得到的参数值 a_1, \dots, a_m 可能有正有负, 当空间变量为离公路的最短距离, 离铁路的最短距离、离商业中心、居住中心的最短距离时, 元胞的城市发展概率与这些空间变量间应为负相关^[5], 正的参数值不利于模型的解释, 有必要通过遗传算法优化这些变量的权重。

2.2.2 利用遗传算法优化 CA 模型的参数 用遗传算法优化 CA 模型参数时, 首先要把实际问题参数集合和位串空间相对应, 并对位串空间的染色体进行编码。本例中, 我们将需要求解的 CA 模型的参数权重 a_0, a_1, \dots, a_m (m 为自变量个数) 定义为位串空间的染色体。染色体的编码采用实数编码。染色体 (CM) 可表达为:

$$CM = [a_0, a_1, \dots, a_m] \quad (5)$$

式中: $0 < a_0 < 1, -0.1 < a_1, \dots, a_m < -0.0001$

其次, 定义适应度函数。本例中, 适应度函数的定义如下:

$$f(x) = \sum_{i=1}^n (f'_i - f_i)^2 \quad (6)$$

$$f'_i(x_1, x_2, \dots, x_m) = \frac{1}{1 + \exp(-(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m))} \quad (7)$$

式中, f_i 为第 i 个样本对应的因变量值 ($f_i = 1$, 发展为城市用地; $f_i = 0$, 未发展为城市用地)。此时, $f(x)$ 表示模型的误差, $f(x)$ 越小表明模型的精度越高, 获取的参数值越好。

第三, 设置遗传算法的参数值。本例中, 初始化种群数量设置为 50, 随机生成; 染色体的初始参数设置如下: 将 a_0 的初始值设为 0.5, 其他染色体的初始值设为 0.01; 操作算子的参数设置如下: 运用选择、交叉、突变等遗传算子模拟进化, 其中交叉率为 0.9, 突变率为 0.01, 运用精英选择策略 (Elitist selection); 模型终止条件的设置: 当最佳适应值在 50 代内不发生变化时, 进化终止。运用遗传算法寻找 CA 模型的最佳参数的流程见图 2。

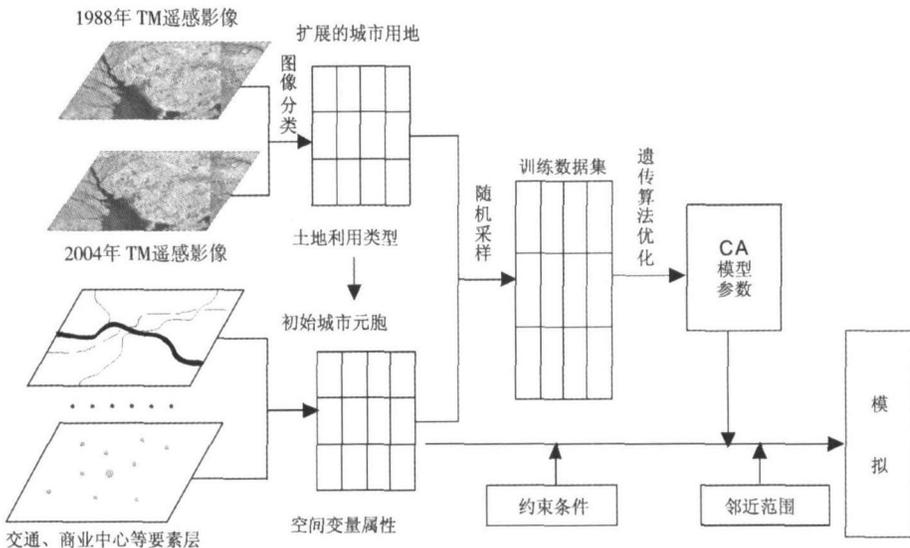


图 2 遗传算法寻找 CA 模型最佳参数流程图

Fig 2 A CA model based on optimization parameters with GA

3 模型应用

3.1 用遗传算法来寻找 CA 模型的最佳参数

本文以珠江三角洲的东莞市 1988~2004 年城市用地的扩张作为例, 研究遗传算法优化 CA 模型参数的方法。东莞市位于广东省中南部, 珠江口东岸, 北接广州南邻深圳, 经济发展区位优势明显。随着广州和深圳市产业结构的不断调整, 东莞市面临着良好的发展机遇, 过去 20 年中东莞的经济得到了飞速发展, 与之相应的城市建设用地规模快速增长, 已经带来了一系列资源环境问题^[21]。为此, 寻找合理的城市发展模拟方法, 可以为东莞市城市发展的政策制定提供科学的依据。本研究中, 为模型运算方便, 将建成区和建设用地作为城市用地, 由于城市绿地等用地与建设用地的特点和发展特点不同, 所以, 城市绿地等其他城市用地在本研究中尚未考虑。

首先利用遗传算法来寻找 CA 模型的最佳参数。研究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^[1,5,9]。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。

为了获取模型所需的参数, 需要用历史数据来校正。本文选取东莞市 1988 年和 2004 年的 TM 遥感图像, 通过遥感分类, 获取元胞的土地利用类型。遥感分类通过训练模板建立专家分类器在 Erdas 软件中进行, 并经过野外验证和修改, 1988 年的分类总精度为 86.3%, 其中建设用地的分类精度为 92.1%; 2004 年的分类总精度为 87.1%, 其中建设用地的分类精度为 92.7%。TM 的分辨率是 30m×30m, 为 CA 模型方便运算, 将分类结果重采样成 50m×50m 分辨率的格网作为元胞空间。

模拟时, 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 模型的工具。

遗传算法自动获取 CA 模型最佳的参数, 需要用历史资料作为训练数据。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, 利用遗传算法训练优化 CA 模型空间变量的参数。所获得的最佳的空间变量参数值见表 2。图 3 是遗传算法的适应值改善的过程图, 反映了该算法能很快地收敛。

表 1 空间变量及获取方法

Tab 1 Spatial variables in CA model

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)	利用 ArcGIS
离镇中心的距离 (x_2)	的 Euclidean 获取
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)	
离高速公路的距离 (x_7)	
局部变量	
3×3 邻域已城市化元胞数 (x_8)	利用 ArcGIS 的 focal 函数
元胞的土地利用类型 (x_9)	遥感分类

表 2 遗传算法及逻辑回归 CA 模型的参数表

Tab 2 Parameters of CA model obtained with GA and logistic model

	常数	离市中 心距离	离镇中 心距离	离村中 心距离	离省道、国 道距离	离公路 的距离	离铁路 的距离	离高速公 路的距离
GA	0.8937	-0.0001	-0.00284	-0.00206	-0.036	-0.088	-0.0001	-0.00323
MCE	1.906	0.003	-0.009	-0.017	-0.048	-0.112	-0.014	-0.027

由表 2 可以发现, 1988~2004 年间, 对东莞市城市用地发展概率贡献最大的是公路 (其中省道和国道的权重值为 0.036, 其他公路的权重值为 0.088), 其次是高速公路 (权重值为 0.00323), 然后是镇中心 and 村中心的影响 (权重值分别为 0.00284 和 -0.00206); 市中心对城市用地的发展概率影响较小, 达到模型的临界值 (权重值为 -0.0001); 铁路对东莞市 1988~2004 年城市用地的发展概率影响非常小, 也达到了参数值的极限值 (权重值为 -0.0001)。

3.2 与基于逻辑回归模型的 CA 的比较

对遗传算法获得的 CA 模型与逻辑回归获得的 CA 模型进行了比较。由于逻辑回归模型要求变量间线性无关, 该模型中离市中心的距离与城市发展概率呈正相关, 这与实际情况不太符合。而经过遗传算法优化的模型, 距离因子全部表现为负相关。对相关性较强的变量, 通过优化变量间的参数组合, 如离市中心的距离和离镇中心的距离相关性较大, 优化组合时, 弱化了市中心的影响 (权重值为 -0.0001), 而强化了镇中心的影响 (权重值为 -0.00284), 得到了易于理解和基于地理意义解释的模型。

获取模型的参数后, 模拟了东莞市的城市发展。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将条件函数设置为接近于 0 的约束条件, 即这些单元的城市发展概率非常小。

模拟时, 初始状态利用 1988 年的 TM 图像获取 (图 4A)。由 1988~2004 年的实际用地总量作为控制, 经过迭代运算, 模拟出了 2004 年的城市用地 (图 4B)。由于城市的发展受到许多不确定性因素的影响, 模拟出完全正确的城市发展是不可能的。为了比较遗传算法优化的 CA 模型对模拟结果的改进状况, 还需要对模拟结果进行评价。参考 Wu、黎评价 CA 模型模拟结果的方法^[5, 12], 本文运用逐点对比评价和总体形态评价的方法对模拟结果进行了评价。逐点对比评价将从遥感图像获取的实际城市用地与对应时段的模拟城市用地进行叠加分析, 得到模拟精度评价的混淆矩阵。我们将 2004 年从遥感图像分类得到的实际城市用地 (图 4A) 与遗传算法优化的模型模拟的结果 (图 4B) 进行了比较, 发现模拟结果与实际非常接近。表 3 是遗传算法优化模型模拟的混淆矩阵, 总精度为 83.98%。但需要说明的是, 真正的模拟精度还需要考虑遥感分类的精度, 这里只是采用了一种比较简单的常用评价方法。

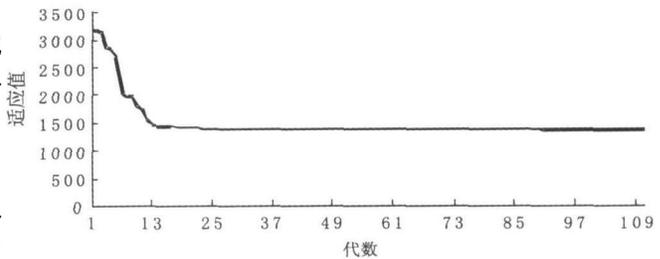


图 3 遗传算法的适应值改善过程

Fig. 3 Training time and prediction error

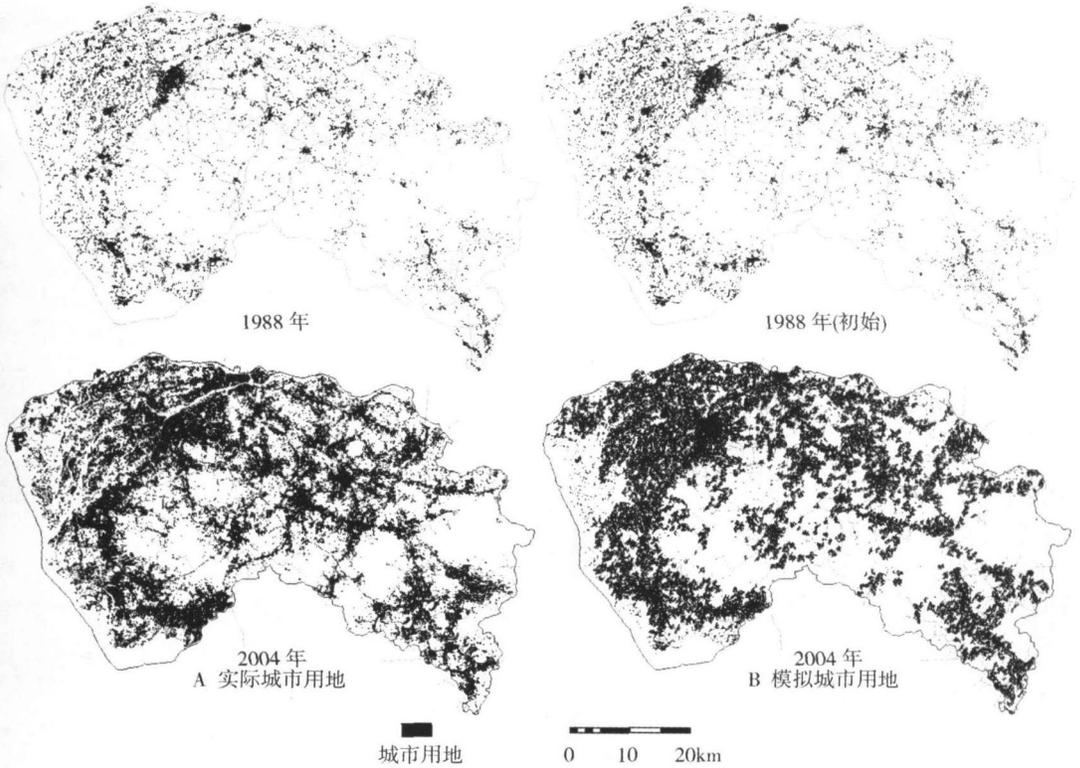


图4 东莞市1988~2004年城市模拟对比图

Fig. 4 Actual urban land (A) and simulated land (B) of Dongguan city from 1988 to 2004

表3 遗传算法优化模型模拟的混淆矩阵

Tab 3 Confusion matrix of simulated with GA-CA model from 1988 to 2004

		模拟		
		不转变	转变	正确比
实际	不转变	54631	8532	86.49%
	转变	8784	36181	80.46%
总精度		83.98%		

表4 逻辑回归模型模拟的混淆矩阵

Tab 4 Confusion matrix of simulated with logistic CA model from 1988 to 2004

		模拟		
		不转变	转变	正确比
实际	不转变	49296	13867	79.76%
	转变	14119	30846	68.60%
总精度		74.10%		

表5 遗传算法优化模型模拟与实际的指数对比表

Fig 5 Moran I index and Geary c index in simulated city form and actual city form

1988年		2004年模拟		2004年实际	
Moran I	Geary c	Moran I	Geary c	Moran I	Geary c
0.338	0.662	0.458	0.542	0.467	0.533

运用遗传算法优化的CA模型模拟的结果与逻辑回归CA模型模拟的结果进行了比较,表4是逻辑回归模型模拟的混淆矩阵,总精度为74.1%。这表明利用遗传算法优化的CA模型模拟的精度要高于逻辑回归模型模拟的精度,因为基于遗传算法的CA模型对模型的参数进行了优化组合。

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^[5]。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。表 5 是遗传算法优化模型模拟结果的 Moran I 指数和 Geary c 指数对比表。表 5 表明,优化模型模拟的结果在总体形态上与实际非常接近,反映了区域 2004 年相对集中的城市形态。

4 结论

CA 的核心问题是如何定义转换规则和寻找模型的最佳参数,使模拟结果更接近真实城市状况。CA 在模拟复杂城市系统时,涉及到许多区域空间变量及相应的参数值,这些参数值反映了各变量对模拟结果的“贡献”大小。用逻辑回归模型校正 CA 模型参数的传统方法在确定 CA 的转换规则和模型参数值时有一定局限性。

本文采用基于遗传算法的全局优化搜索程序来快速寻找 CA 模型的最佳参数,用优化参数的 CA 模型模拟了东莞市 1988~2004 年的城市发展状况。研究表明,运用遗传算法建立的 CA 模型比常用的由多准则判断和逻辑回归模型建立的 CA 模型能获取更高的模拟精度,模型可以方便地模拟快速城市化地区的城市发展状况。

所建立的 CA 模型能清楚地反映各空间要素对研究区的城市格局的影响。可以发现,对城市扩张贡献最大的是公路,然后是镇中心 and 村中心的影响,而市中心和铁路对城市扩张的影响较小。从这些空间变量所对应的权重值可以定量地反映它们的关系,从而可以寻找出影响城市发展的规律。

运用遗传算法优化 CA 模型参数时,在影响城市发展的参数特征选择,选取影响城市发展主要的空间特征变量组合方面需要进一步研究。

参考文献:

- [1] Batty M, Xie Y. From cells to cities. *Environment and Planning B: Planning and Design*, 1994, 21: 531~ 548
- [2] 周成虎,孙战利,等. 地理元胞自动机研究. 北京: 科学出版社, 1999. 55~ 59
- [3] 黄翀,刘高焕. 基于元胞模型的河道纵剖面演化动态模拟. *地理研究*, 2006, 25(2): 342~ 349.
- [4] White R, Engelen G. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, 1993, 25: 1175~ 1199
- [5] Wu F. Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science*, 2002, 16(8), 795~ 818
- [6] 黎夏,叶嘉安. 基于神经网络的元胞自动机及模拟复杂土地利用系统. *地理研究*, 2005, 24(1): 19~ 27
- [7] Wu F, Webster C J. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B*, 1998, 5: 103~ 126
- [8] Wu F. SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 1998, 12(1), 63~ 82
- [9] Li X, Yeh A G O. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, 2000, 14(2): 131~ 152
- [10] 黎夏,叶嘉安. 主成分分析与 CA 在空间决策与城市模拟中的应用. *中国科学(D 辑)*, 2001, 31(8): 683~ 690
- [11] Clarke K C, Hoppen S, et al. A self-modifying cellular automaton of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, 1997, 24: 247~ 261.
- [12] 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. *地理学报*, 2002, 57(2): 159~ 166.
- [13] 黎夏,叶嘉安. 知识发现及地理元胞自动机. *中国科学(D 辑)*, 2004, 34(9): 865~ 872.
- [14] 李敏强,寇纪淦,等. 遗传算法的基本理论与应用. 北京: 科学出版社, 2002. 26~ 74

- [15] Hyun M, Kim J Hybrid evolutionary programming for heavily constrained problems *Biosystems*, 1996, 38: 29~43
- [16] Booker L B, *et al.* Classifier systems and genetic algorithms *Artificial Intelligence*, 1989, 40(1-3): 235~282
- [17] 李铭, 方创琳. 基于自组织模型的酒嘉玉地区城市化动态演变. *地理研究*, 2006, 25(3): 551~559
- [18] 陈述彭 主编. 城市化与城市地理信息系统. 北京: 科学出版社, 1999 15~27
- [19] Allen P M. Cities and regions as self-organizing systems: models of complexity. *Cordon and Breach*, 1997, 33~49
- [20] 黎夏, 刘凯. GIS与空间分析—原理和方法. 北京: 科学出版社, 2006 201~233
- [21] 黎夏, 叶嘉安. 利用遥感监测和分析珠江三角洲的城市扩张过程——以东莞市为例. *地理研究*, 1997, 16(4): 56~61

Calibrating urban cellular automata using genetic algorithms

YANG Qing-sheng, LI Xia

(School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: This paper presents a new method to calibrate urban cellular automata (CA) using genetic algorithms (GA). The GA is used to find the optimal parameter values so that CA models can simulate urban expansion in a more realistic way. Traditional multi-criterion evaluation (MCE) and logistic methods have limitations for deriving the transition rules of CA models. The variables should be independent so that the parameter values (coefficients) can be properly estimated by regression analysis. This assumption is not true in most situations. The limitations can be overcome by using GA to estimate these parameter values for these correlated variables.

When calibrating urban cellular automata with GA, the parameters of CA models are set to the chromosomes in GA program. The real number encoding way is used to encode chromosomes. The fitness function is defined with mean square error between simulated and actual urban forms. The initial population is set to be 50 randomly. And crossover probability is set to be 0.9, and mutation probability is set to be 0.01. The elitist selection is used to heredity the better individual. If the fitness does not change in the past 50 generations, the genetic procedure will be finished. After properly encoding the chromosomes, the optimal parameter values are automatically found by the evolutionary approach.

This method is applied to the simulation of urban expansion in Dongguan, a fast developing city in the Pearl River Delta in South China. The model is able to simulate urban development in 1988-2004 by using the training data from remote sensing data. The analysis indicates that the proposed model can produce better simulation results than MCE-based CA models and logistic calibrated CA models. Moreover, the parameter values can be used to explain the relationships between spatial variables and urban development.

Key words: cellular automata; genetic algorithms; urban expansion; Pearl River Delta