

基于案例推理的元胞自动机及大区域城市演变模拟

黎 夏, 刘小平

(中山大学地理科学与规划学院, 广州 510275)

摘要: 元胞自动机 (CA) 被越来越多地用于复杂系统的模拟中。许多地理现象的演变与其影响要素之间存在着复杂的关系, 并往往具有时空动态性。在研究区域较大和模拟时间较长时, 定义具体的规则来反映这种复杂关系有较大的困难。为了解决 CA 转换规则获取的瓶颈问题, 提出了基于案例推理 (CBR) 的 CA 模型, 并对 CBR 的 k 近邻算法进行了改进, 使其能反映转换规则的时空动态性。将该模型应用于大区域的珠江三角洲城市演变中。实验结果显示, 其模拟的空间格局与实际情况吻合较好。与常规的基于 Logistic 的 CA 模型进行了对比, 所获得的模拟结果有更高的精度和更接近实际的空间格局, 特别在模拟较为复杂的区域时有更好的模拟效果。

关键词: 元胞自动机; 案例推理; k 近邻算法; 动态转换规则

1 引言

对地理现象的空间格局及演变过程进行模拟和预测是地理学的重要研究内容之一。对这些地理现象进行模拟, 在理论和应用中都有较大的意义, 可有助于检验关于城市发展和土地利用演变等方面的假设和理论。在许多全球资源、环境、大气模拟和预测模型中, 都涉及到土地利用空间格局演变信息的输入。地理现象的空间格局及演变过程属于复杂的动态系统, 常规的模型或一般的数学公式都很难对此进行模拟和预测。

元胞自动机 (CA) 是模拟和预测复杂系统行为的强有力工具。CA 始于 20 世纪 40 年代 Ulam 和 Von Neumann (现代计算机之父) 的研究。他们发现对简单的运算法则反复计算, 可以代替复杂的模型, 解释自然界中的许多现象。CA 的特点是通过一些十分简单的局部规则, 可以有效地模拟出复杂系统的空间格局形成过程。Wolfram 在 20 世纪 80 年代进行的 CA 研究, 为后来该领域的发展起到很大的促进作用^[1]。目前 CA 已经被广泛应用于物理、化学、生物、地学等领域中的自然或人工复杂系统的模拟中^[2, 3]。

CA 这种“自下而上”的研究思路充分体现了复杂系统局部个体行为产生全局、有秩序模式的理念, 非常适用于复杂地理过程的模拟和预测。近年来, 有越来越多的学者将 CA 应用于地理现象的模拟中。Couceleis 在 20 世纪 80 年代末利用 CA 对一些地理现象进行了模拟, 取得了许多有意义的研究成果^[4-6]。此后, 许多学者不断扩展 CA 在地理学中的应用领域, 利用 CA 模拟了城市扩张^[7]、土地利用动态演变^[8]、山火蔓延^[9]、动物群种的动态变化^[5]、荒漠演化^[10]等。

传统 CA 的转换规则是静态的, 它们不随时间和空间而变化。在模拟复杂的地理现象时, 特别是研究的区域和时间尺度跨度较大时, 采用静态的转换规则是不合适的。区

收稿日期: 2006-08-03; 修订日期: 2007-08-06

基金项目: 国家自然科学基金项目 (40471105); 国家杰出青年基金项目 (40525002); 国家高技术研究发展计划 (2006AA12Z206) [Foundation: National Natural Science Foundation of China, No.40471105; National Outstanding Youth Foundation of NSF of China, No.40525002; Hi-Tech Research and Development Program of China, No.2006AA12Z206]

作者简介: 黎夏 (1962-), 男, 教授, 从事 GIS 和遥感信息模型研究, 在国内外刊物上发表约 160 多篇学术论文。

E-mail: lixia@mail.sysu.edu.cn; lixia@graduate.hku.hk

域内部由于所处的发展过程或阶段不一样，也需要使用动态的转换规则。CA 的转换规则在时间上也不是固定不变的。随着资源环境约束条件变化和有关土地政策等的调整，使用静态的转换规则所模拟出来的结果往往与事实不吻合。在不同经济发展阶段，区域发展空间格局的演变规律是不一样的。定义具体的转换规则来反映上述地理现象的复杂关系往往是十分困难的。常规的方法在获取 CA 的转换规则时已经碰到了一定的问题，在确定模型的参数时有很大的不确定性。而且这些方法几乎不可能获取动态转换规则。

为了有效地解决 CA 模型中转换规则获取的瓶颈问题，提出了基于案例 (case-based) 的 CA 模型来代替基于规则 (rule-based) 的 CA 模型。并利用案例推理 (CBR) 方法来反映 CA 的时空动态转换规则。CBR 最大的特点是无需定义具体明确的规则，而是利用案例来隐含表达知识，从而大大减少建立模型所需要的时间，能有效地解决知识获取所碰到的模糊性和不确定性问题。它特别适合于那些专业知识难以被概括、抽象和表达的领域。最近，有些学者开始将 CBR 应用在环境、城市规划、遥感分类等地理领域中^[11-14]，并取得了一定的成果。本文尝试把 CBR 引进地理元胞自动机，并对 CBR 的 k 最近邻算法进行改进，使其能反映随空间变化的动态转换规则。此外，并根据多个年份的遥感资料对 CA 模型进行校正，把新的历史资料加入到案例库中。由此能够随着周围环境变化和经验的增加来动态更新转换规则，很好的体现了复杂系统的自适应特点。

2 地理元胞自动机

2.1 基于规则 (rule-based) 的地理元胞自动机

传统地理元胞自动机 (CA) 模型是基于规则的，即通过转换规则来确定元胞下一时刻状态的转变，并由此来模拟复杂地理现象的演变过程。目前有多种方法来确定 CA 的转换规则，包括多准则判断 (MCE)^[15]、神经网络^[16]和数据挖掘^[17]等。最常用定义转换规则的方法是利用多准则判断来决定状态转变的概率。例如，Wu 和 Webster 利用多准则判断 (MCE) 来建立城市扩张的 CA 模型^[15]：

$$P(i) = \sum_{l=1}^n w_l a_l(i) \tag{1}$$

式中： $P(i)$ 为在位置 i 处转变为城市用地的概率， $a_l(i)$ 为位置 i 第 l 个属性 (变量)， w_l 为该属性的权重。

该 MCE-CA 模型中权重的确定是通过专家知识来确定，有一定的不确定性。可以进一步利用 Logistic 回归方法来解决该方法权重确定的问题^[18]：

$$P(i) = \frac{\exp(z(i))}{1 + \exp(z(i))} = \frac{1}{1 + \exp(-z(i))} \tag{2}$$

这里 $z(i) = w_0 + w_1 a_1(i) + w_2 a_2(i) + \cdots + w_n a_n(i)$

2.2 基于案例 (case-based) 的地理元胞自动机

影响地理现象演变的因素有很多，它们之间的关系也比较复杂，往往无法用经验公式或规则来表达。采用在时空上固定的转换规则来反映这些复杂关系也是很困难的。但通过离散的地理案例，可以避免获取具体知识 (规则) 的困难，从而有效地解决自然界一些复杂推理问题。CBR 是专家系统的一种类型，它是参考过去解决问题的经验 (主要是通过案例) 来解决新问题。

本文将以城市模拟为例，尝试把 CBR 引进地理元胞自动机 (CA) 中，自动从案例库中获取知识来反映 CA 的动态转换规则。与基于规则的 CA 模型不同，本模型是由案例来决定元胞的状态转变，包括 4 个主要部分，即案例库建立、检索相似案例、获取问题的解决方案、更新案例库。图 1 是基于案例推理的元胞自动机的具体流程。

2.2.1 建立案例库 运用 CBR 的第一步是建立 CA 模型的案例库。这些案例可以反映某元胞的状态转变(土地利用变化)与空间变量等因素的复杂关系。例如,某元胞土地利用变化的概率往往取决于一系列的距离空间变量、邻近现有城市用地量和元胞的自然属性等^[7, 15]。在建立案例库时,先只考虑各空间变量,把其作为案例的特征属性。把邻近现有城市用地量和元胞的自然属性作为外部的约束条件来考虑。

在案例库中,每个案例由两部分组成:问题的描述(案例属性)和问题的解决(决定元胞的状态转变)。在这里,问题的描述为一系列空间距离变量,问题的解决为该元胞是否转变为城市用地。一个案例具体表达如下:

$$I = (a_1(i), a_2(i), \cdots, a_N(i); s) \tag{3}$$

式中: $a_1(i), a_2(i), \cdots, a_N(i)$ 为案例 i 所对应的一系列空间距离变量,即特征向量。 s 为一布尔变量,转变为城市用地为 1,不转变为 0。

2.2.2 利用改善的 k 最近邻算法 (k-NN) 反映 CA 动态转换规则 基于 CBR 的元胞自动机的特点是用案例来隐含表达 CA 模型的转换规则。具体是在案例库中搜索最接近的案例来决定元胞的状态转变。案例搜索主要是基于 k 最近邻算法 (k-NN) 来进行的。可以利用下面欧式距离公式来计算待查询案例 (i) 与案例库中某一已知案例 (j) 的相似度:

$$d(i, j) = \sqrt{\sum_{l=1}^n (a_l(i) - a_l(j))^2} \tag{4}$$

式中: $a_l(i)$ 为某案例的第 l 个特征 (属性)。

欧式距离越小,表示两个案例之间的相似度越高。可以进一步把权重引进相似度的计算中,以反映不同的特征对相似度计算所起的贡献。公式修改为:

$$d(i, j) = \sqrt{\sum_{l=1}^n w_l^2 (a_l(i) - a_l(j))^2} \tag{5}$$

式中: w_l 为第 l 个特征 (属性) 所对应的权重。

有许多方法来确定这些权重。采用熵的方法来确定各属性的权重。在确定各属性权重前,需要对这些变量进行归一化,使它们的数值落在 $[0, 1]$ 之间。熵的计算公式如下^[19]:

$$H_n = - \sum_{i=1}^n p_i \log(1/p_i) / \log(n) \tag{6}$$

式中: n 为总样本数。

可以用熵值来判断案例某个特征的离散程度。对于某个特征,若各个案例的值没有太大区别,则该特征离散程度较小,在综合分析中所起的作用不大,对案例相似度计算的“贡献”较小;反之,若对某个特征而言,各个案例的值有很大的波动,即该特征的离散程度很大,则这个特征对综合分析有很重要的“贡献”。因此,第 1 个特征 (属性)

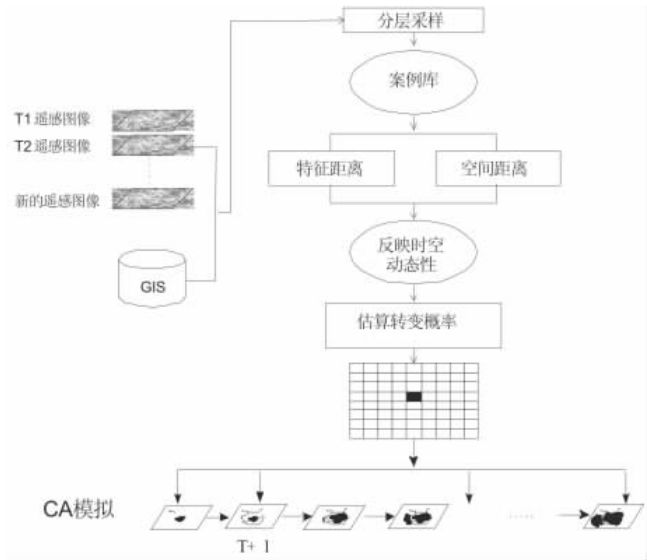


图 1 基于案例推理的 CA 模型

Fig. 1 A case-based approach for cellular automata

所对应的熵权 (entropy weigh) 可以由下式表示:

$$\phi_l = (1 - H_{nl}) / (n - \sum_{i=1}^n H_{nl}) \tag{7}$$

可以根据初始权重和熵权来获得最终的调整后的权重^[20]:

$$w_l = w_l^0 \cdot \phi_l / \sum_{i=1}^N w_l^0 \cdot \phi_i \tag{8}$$

式中: N 为总特征数目。

案例推理的实质是通过相似度的计算来寻找与待查询案例 (i) 最接近的已知案例 (j), 从而把已知案例的目标函数 $f(j)$ (问题的解, 即是否发生状态转变) 赋给待查询案例。目标函数值可以是离散的, 也可以是连续的。在本 CA 模型中, 目标函数值即是元胞的状态, 故是离散的。在实际应用中, 是通过寻找与待查询案例最接近的 k 个最近邻来实现的, 即采用 k 最近邻算法^[21]。该方法可以有效消除噪音。因此, 对离散的目标函数值, 其 k 最近邻算法公式表达如下:

$$\hat{f}(i) \leftarrow \operatorname{argmax}_{s \in S} \sum_{j=1}^k \delta(s, f(j)) \begin{cases} \delta(s, f(j)) = 1, & \text{if } s = f(j) \\ \delta(s, f(j)) = 0, & \text{if } s \neq f(j) \end{cases} \tag{9}$$

式中: s 为元胞的状态, 为城市用地 ($s = 1$) 或否 ($s = 0$)。

上述公式假定这个最近邻具有同样的贡献。对 k 最近邻算法的一个改进是根据特征空间的反距离来确定它们的贡献。即在特征空间中距离越近, 所起的作用越大, 故把较大的权重赋给较近的近邻, 则式 (9) 可修改为:

$$\hat{f}(i) \leftarrow \operatorname{argmax}_{s \in S} \sum_{j=1}^k w_{fj} \cdot \delta(s, f(j)) \tag{10}$$

其中, 特征距离权重 w_{fj} 计算公式为:

$$w_{fj} = 1 / d(i, j)^2 \tag{11}$$

这里 $\hat{f}(i) = f(j)$ if $d(i, j) = 0$

目前 k 最近邻算法是在特征空间中来查询的, 无法反映案例随空间变化的特征, 在应用中有一定的缺陷。为了获取 CA 模型转换规则随空间变化的特征, 有必要将空间距离也引进相似度的计算中。由此, 需要把案例的空间位置也作为案例属性的一部分, 即把案例的空间坐标也放在公式 (3) 中。其空间距离权重 w_{sj} 可以表达为:

$$w_{sj} = \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \tag{12}$$

式中: x, y 分别为案例 i 的横坐标和纵坐标,

最后, 公式 (10) 修改为:

$$\hat{f}(i) \leftarrow \operatorname{argmax}_{s \in S} \sum_{j=1}^k w_{fj} w_{sj} \cdot \delta(s, f(j)) \tag{13}$$

或:

$$\hat{f}(i) \leftarrow \operatorname{argmax}_{s \in S} \sum_{j=1}^k W_j \cdot \delta(s, f(j)) \tag{14}$$

式中: $W_j = w_{fj} w_{sj}$

2.2.3 基于 CBR 的 CA 模型 式 (14) 是通过布尔规则来确定查询案例所属类别, 从而确定元胞状态的转变。但地理复杂现象具有一定的不确定性, 利用布尔规则来计算效果并不理想。在实际的 CA 应用中往往利用概率的形式来确定元胞状态的转变^[18]。因此, 根据公式 (14), 元胞 i 转化为城市用地的概率可由下式表示:

$$P_{proximity}(i) = \frac{\sum_{j=1}^k W_j \cdot \delta(1, f(j))}{\sum_{j=1}^k W_j \cdot \delta(1, f(j)) + \sum_{j=1}^k W_j \cdot \delta(0, f(j))} \tag{15}$$

式中： $P_{proximity}(i)$ 是元胞 i 由距离变量所引起的转化为城市用地概率, k 为参数。

除了距离变量影响元胞状态的转变，邻近元胞的状态也是十分重要的。当一个元胞的周围有较多的元胞转变为城市用地时，会使得该元胞转变为城市用地的概率提高。由邻域影响所引起的转化为城市用地概率表达为：

$$P_{neigh}(i) = K_2 \sum_{\Omega} N(i) \tag{16}$$

式中：城市用地 $N(i) = 1$, 非城市用地 $N(i) = 0$, Ω 为邻域窗口大小。

最后，转化为城市用地的概率由 $P_{proximity}(i)$ 和 $P_{neigh}(i)$ 的联合概率构成，并乘以一些约束因子。这些约束因子包括地形和土地利用规划等，它们对转化为城市用地的概率也起到很大的约束作用。例如，在河流、陡峭的山地、生态用地和农田保护区的地方转化为城市用地的概率会大大减少。可以定义约束函数 $\delta_{\gamma}(i)$ 来反映它们的影响，其最大值为 1 时反映约束最大，禁止转化为城市用地。而最小值为 0 则反映约束条件不起作用。因此，转化为城市用地的联合概率可以由下式来表达：

$$\begin{aligned} P(i) &= P_{proximity}(i) \times P_{neigh}(i) P_{proximity}(i) \times (1 - \sum_{\gamma} \delta_{\gamma}(i)) \\ &= K \frac{\sum_{j=1}^k W_j \cdot \delta(1, f(j))}{\sum_{j=1}^k W_j \cdot \delta(1, f(j)) + \sum_{j=1}^k W_j \cdot \delta(0, f(j))} \times \sum_{\Omega} N(i) \times (1 - \sum_{\gamma} \delta_{\gamma}(i)) \end{aligned} \tag{17}$$

式中： K 为参数。 $\delta_{\gamma}(i)$ 约束条件值，其值为 0~1。

复杂系统的演变往往受到一些不确定因素的影响，可用 Monte Carlo 方法来反映这种不确定性，以使得模拟更加合理。Monte Carlo 方法在城市元胞自动机中是一种较为常用的方法，国内外许多学者在城市 CA 模型中都是运用该方法来确定单元状态的变化^[15, 17]，公式表达如下：

$$S_{t+1}(i) = \begin{cases} \text{转变为城市用地, 当 } P(i) > \text{Rand}() \\ \text{不转变, 当 } P(i) \leq \text{Rand}() \end{cases} \tag{18}$$

式中： $S_{t+1}(i)$ 为元胞在 $t + 1$ 时刻的状态。 $\text{Rand}()$ 为 0~1 之间的随机变量。

本模型是利用案例推理来决定元胞状态的转变，例如从非城市用地转变为城市用地的过程。由于对 k 最近邻算法进行了改进，将空间距离也引进相似度的计算中，使得案例推理能隐含地反映随空间而变化的动态转换规则。而且，由于在模拟过程中也将新的遥感数据加入到案例库中，所得案例库也是动态更新的，因此可以反映随时间而变化的动态转换规则。

3 模型应用及结果

选择大区域的珠江三角洲作为试验区 (图 2)，模拟了其 1988~2002 年的城市扩张过程。该研究区包含了不同等级的城市，其土地利用变化存在着明显的内部差异^[22]。使用传统单一不变的转换规则对该区域的城市扩张进行模拟存在一定的困难。

采用基于案例推理的 CA 模型来反映城市系统演变的复杂关系。首先对 1988 年、

1993 年的 TM 遥感图像进行分类, 获取土地利用变化 (状态变化) 的训练数据。通过 GIS 获取影响土地利用变化的地理要素, 包括一系列空间变量。利用随机采样获得模型所需要的原始案例库, 共获取了 4000 个案例。这些案例能隐含反映该地区土地利用变化与地理要素之间的复杂关系, 从而代替具体的转换规则。并从新的 1995 年和 1997 年的遥感影像数据获取新的案例, 使得案例库具有动态的特点。

若用常规的 k 最近邻算法, 无法反映转换规则随空间而变化的特点。当研究区域很大时, 该方法有较大的弊端。当采用所提出的 k 最近邻改善算法, 利用几何空间的权重可以十分有效地反映转换规则的空间动态性。公式 (14) 由于把特征空间近邻和原始空间近邻同时考虑进来, 可以更有效地反映 CA 转换规则随空间而变化的特征。图 3 的案例 Q1 和 Q2 具有相同的特征空间近邻。若用常规的 k 最近邻算法, 其状态转变的概率是不变化的。案例 Q1 和 Q2 的 10 个相同近邻按照特征距离从小到大顺序为: N1、N2、N3、N4、N5、N6、N7、N8、N9、N10; 但在原始的几何空间中, 案例 Q1 的近邻根据其空间距离从小到大排序为: N1、N7、N3、N8、N9、N2、N10、N4、N5、N6。而案例 Q2 的近邻根据其空间距离从小到大排序则为: N2、N9、N4、N10、N8、N3、N7、N1、N5、N6。如果把原始空间近邻也同时考虑进来, 案例 Q1 和 Q2 的状态转变的概率是不同的, 从而可以有效地反映 CA 的动态转换规则。

另外, 由于利用新的遥感



图 2 珠江三角洲城市扩张模拟试验区

Fig. 2 Simulation of urban expansion in a large region-the Pearl River Delta

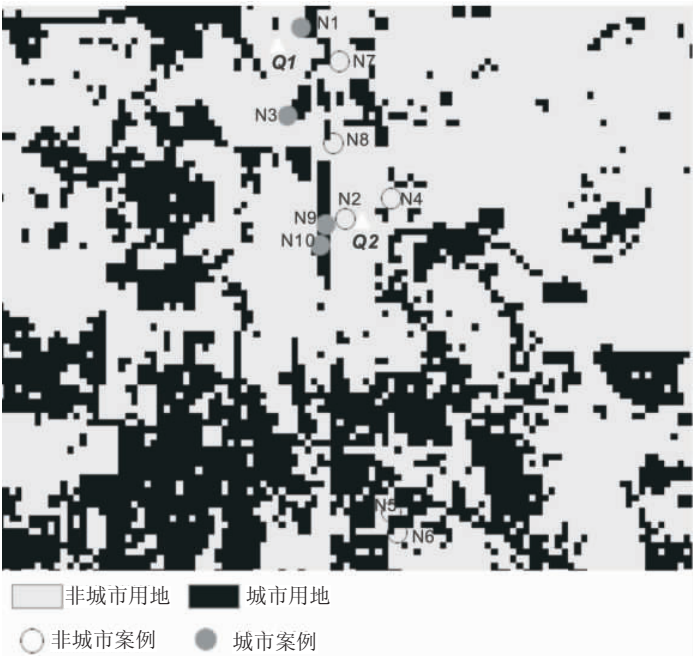


图 3 利用 CBR 改善的 k 最近邻算法反映转换规则的空间动态性

Fig. 3 Representing spatio-temporal dynamic transition rules using improved k -NN algorithm of CBR

图像, 案例库在不断更新, 元胞的近邻会随着时间发生动态变化, 使得模型能反映转换规则的时间动态性。这种方法对模拟复杂的资源环境系统有较大的优越性, 很适合研究区域较大、模拟时间跨度较长的情况使用。

在案例学习中, 近邻数目 k 的取值对案例推理过程有一定的影响。图 4 是对训练数据分析的结果, 显示了近邻数目 k 取值与预测准确率的关系曲线。 k 取值为 1 时, 预测准确率为 69.1%; k 取值为 10 时, 预测准确率为 73.2%; k 取值为 20 时, 预测准确率为 73.4%。可以发现, 当 k 取值超过 10 时, 精度的改善并不明显, 因此, 在运用 CBR 提取 CA 动态转换规则时, 采用 k 取值为 10 来对案例进行推理, 以提高模型运算的速度。

在进行 CA 模型校正时, 使用了多年的遥感资料。模型初始案例主要来源 1988 年和 1993 年遥感分类图像与 GIS 数据。随着模拟时间的推进, 通过分层随机采样的方法从新的遥感图像获取新的案例, 对案例库进行更新, 以反映随时间变化的动态转换规则。并从遥感图像上获取不同时段转化为城市用地的总量 ΔQ_0 。CA 模型需要循环迭代运算多次才能体现局部的相互作用。一般来讲, CA 进行迭代运算 100-200 次是较为正常的^[15]。假设模型运行 N 次, 每次转化的数量控制为 $\Delta Q_0/N$ ^[23], 引入控制变量 $\text{Rand}()$, 其为 $(1-\Delta Q)$ 的随机整数。于是 CA 的转换规则可表达如下:

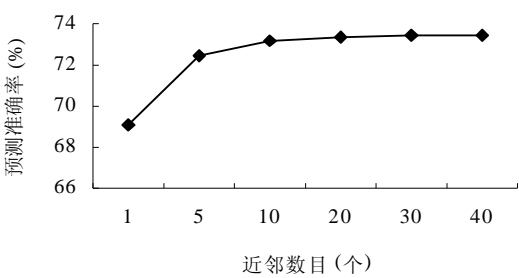


图 4 近邻数目 k 和预测误差的关系

Fig. 4 The relationship between the number of k and prediction accuracy

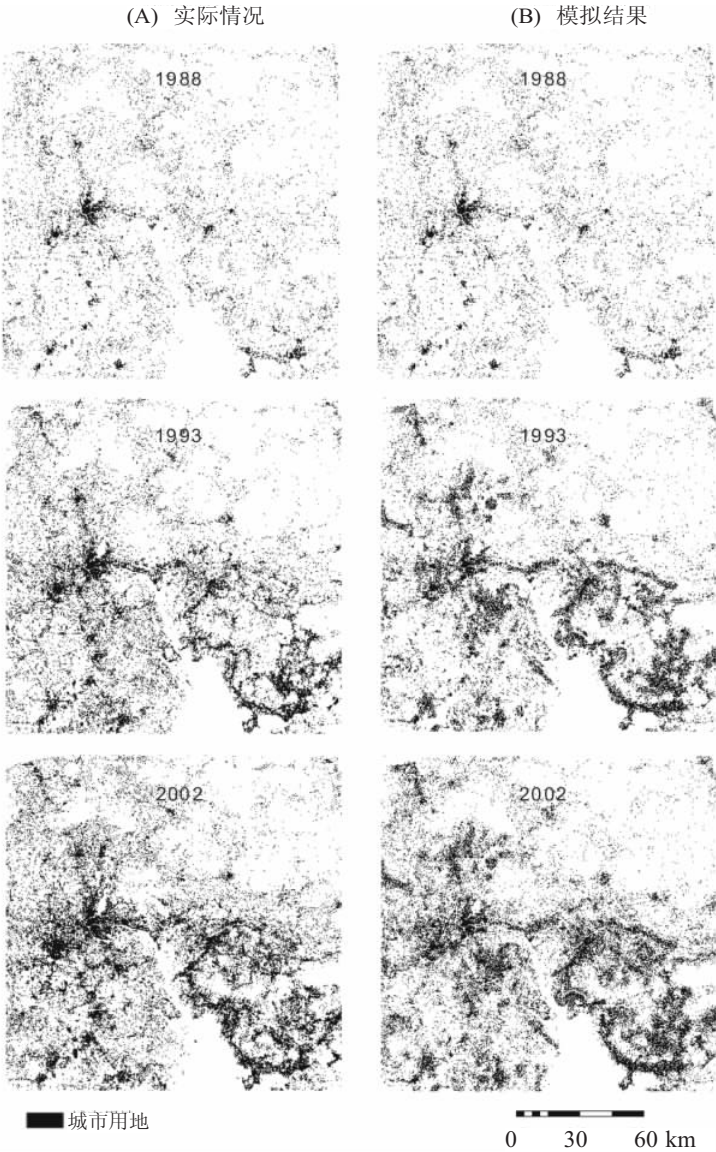


图 5 1988 年、1993 年和 2002 年珠江三角洲实际与模拟的城市演变对比

Fig. 5 The actual and simulated patterns of urban development in the Pearl River Delta in 1988, 1993 and 2002

$$S_{t+1}(i) = \begin{cases} \text{转变为城市用地, 当 } P(i) > \text{Rand}() \text{ 及 } \text{Rand}() < \Delta Q_0/N \\ \text{不转变, 当其它情况} \end{cases} \quad (19)$$

初始的城市用地从 1988 年 TM 图像分类获得。模型进行 225 次迭代运算后，获得模拟的 1993 年珠江三角洲城市用地分布；运行 600 次后得到模拟的 2002 年珠江三角洲城市用地分布。图 5 显示了珠江三角洲城市格局演变的模拟过程和实际的城市用地变化的对比。实际的城市用地变化从遥感图像分类获取。可以发现，模拟结果的整体空间布局与实际情况相当接近。

尽管本研究区域很大，而且包括不同类型的城市，但所提出的方法可以很好地同时模拟出该区域不同城市的空间格局演变情况。图 6、7、8、9 分别是深圳、东莞、广州市

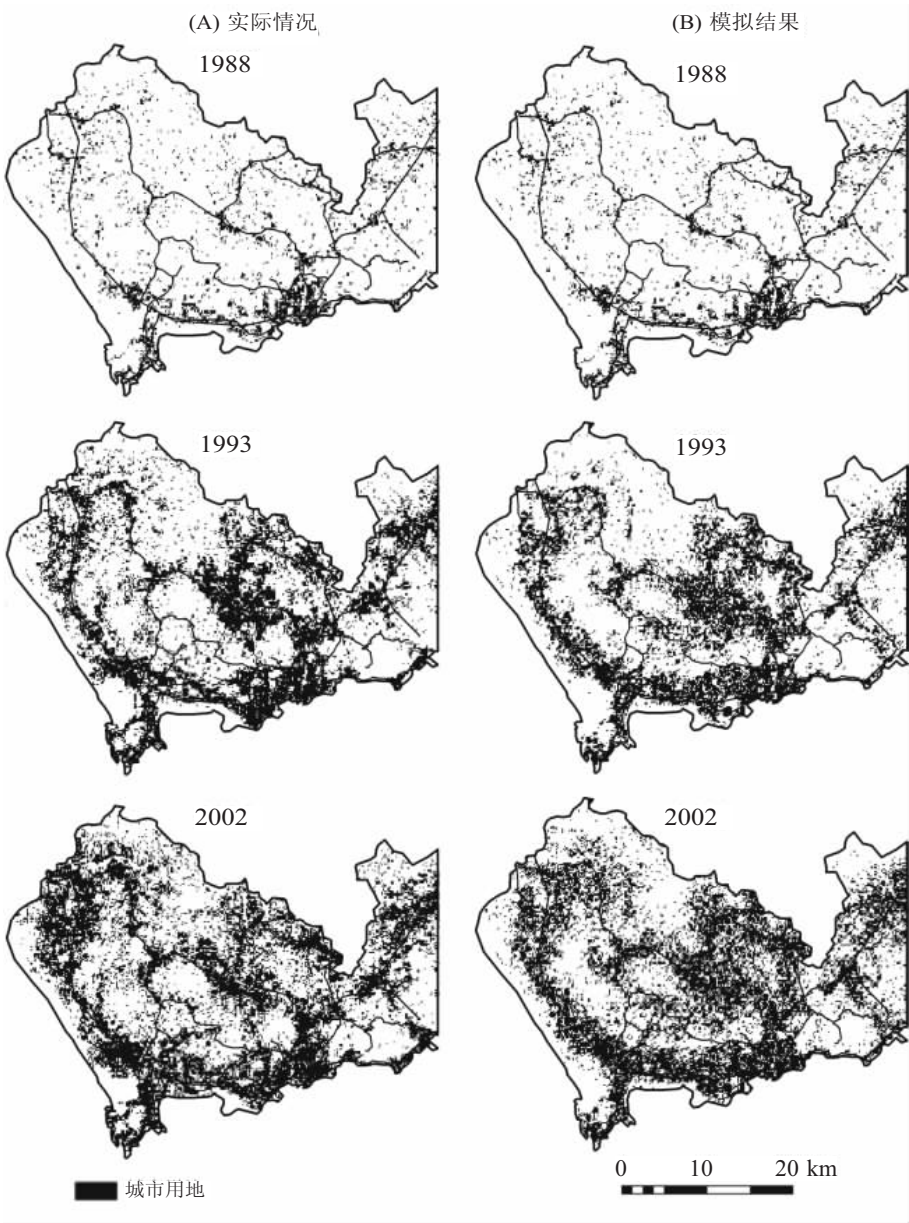


图 6 1988 年、1993 年和 2002 年深圳实际与模拟的城市演变对比
Fig. 6 The actual and simulated patterns of urban development in Shenzhen in 1988, 1993 and 2002

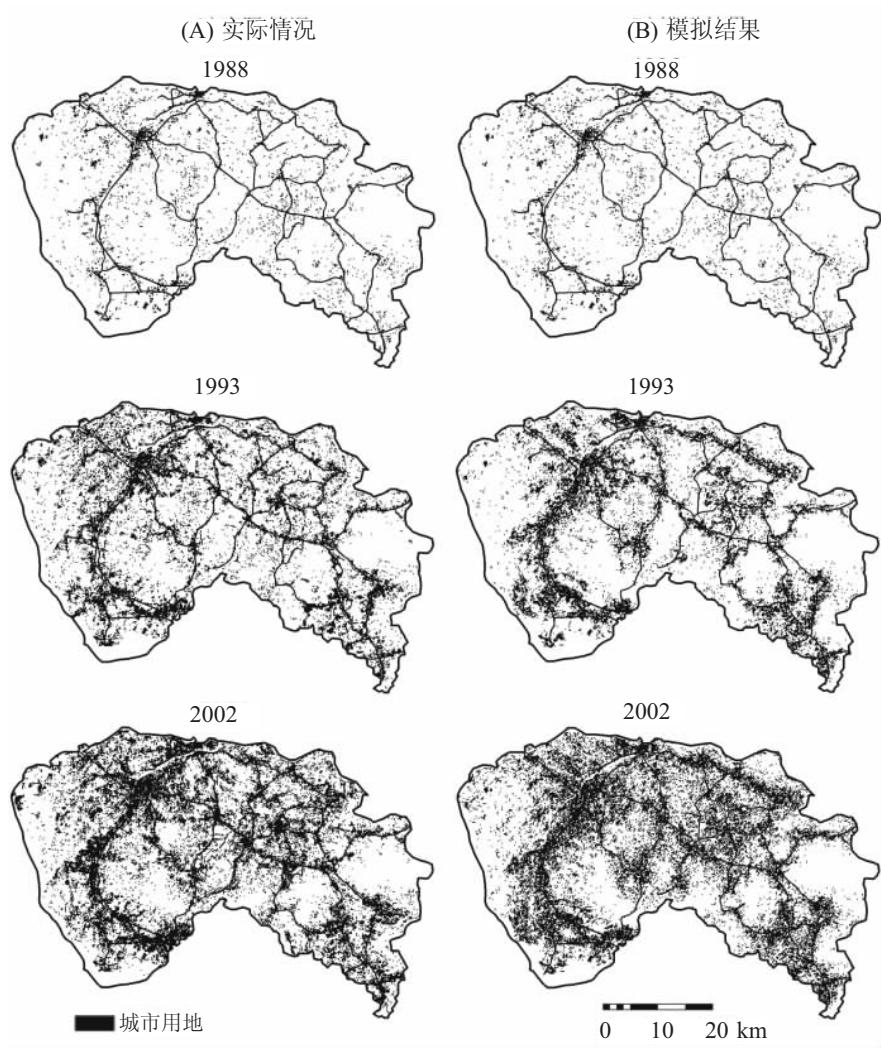


图 7 1988 年、1993 年和 2002 年东莞实际与模拟的城市演变对比
Fig. 7 The actual and simulated patterns of urban development in Dongguan in 1988, 1993 and 2002

区和增城不同类型城市演变的局部放大细节。这四个区域的城市演变过程各具差异，深圳自 1988 年起城市发展非常迅速，主要是沿着南山、福田、罗湖、龙华以及高速公路、国道发展，发展方式为典型的点轴式发展；东莞的城市发展也非常快，但主要是沿着道路进行发展，是典型的轴线式发展方式。广州市区主要是以市中心向外围辐射的方式发展；而增城的城市发展不明显，属于多中心城市发展方式。对这几个不同类型的城市演变进行局部放大处理 (图 6-9)，发现在局部情况下所获得的模拟结果与实际情况也是十分吻合的。由此，本模型能有效地反映大区域下城市演变的复杂关系。

可以进一步采用定量的方法来检验模拟与实际情况的吻合程度，包括逐点对比和格局对比这两种方法。前一种方法是将模拟的结果和实际情况迭合起来，然后逐点对比来计算其精度；后一种方法是检验所模拟出来的整个空间格局与实际空间格局相符合的程度^[15]。首先将 1993 年、2002 年珠江三角洲城市用地的模拟结果与实际情况 (遥感分类) 进行逐点对比，计算模拟精度。表 1 列出了它们的精度，1993 年和 2002 年模拟总精度分别为 0.86 和 0.82，其 kappa 系数分别是 0.53 和 0.51。同时，我们计算了四个局部放大城

市的模拟总精度 (表 2)。

也计算了模拟结果和实际情况的形态指数, 以检验模型的空间格局是否与实际情况相符。采用 Moran I 指数进行对比。Moran I 指数一般用来描述空间的自相关性, 但该指数也反映了空间集中和分散的程度^[13, 15]。1993 年模拟结果的 Moran I 指数值为 0.382, 实际情况的 Moran I 指数值为 0.376; 2002 年模拟结果的 Moran I 指数值为 0.393, 实际情况的 Moran I 指数值为 0.385。模拟结果和实际情况的 Moran I 指数值非常接近, 这证明模拟结果的空间格局和实际情况较为相近。

所提出的方法能有效地反映转换规则随时空而变化的特征, 比常规的采用静态转换规则的 CA 模型应该有更好的模拟精度。为了对比, 也利用常规的基于 Logistic 回归的 CA 模型模拟了本研究区城市的扩张过程, 并计算了模拟结果的逐点对比精度和 Moran I 指数值。表 3 是 Logistic 回归模型模拟的逐点对比精度。1993 年和 2002 年模拟总精度分别为 0.81 和 0.75, 但其 Kappa 系数分别只有 0.34 和 0.28。由此, 基于案例推理的 CA 模型比基于 Logistic 的 CA 模型有更好的模拟精度, 特别是 Kappa 系数方面的对比。

在形态的一致性对比方面, 基于案例推理的 CA 模型也有更好的效果。Logistic 的 CA 模型 1993 年模拟结果的 Moran I 指数值为 0.489, 实际情况的 Moran I 指数值为 0.376; 2002 年模拟结果的 Moran I 指数值为 0.526, 实际情况的 Moran I 指数值为 0.385。其差异要比本模型的大得多。

4 结 论

对复杂的资源环境系统进行模拟有较大的理论和应用意义。基于方程式的方法无法反映复杂系统的特点。而基于“自下而上”的 CA 模型能有效地模拟复杂系统的演变过程。CA 模型被越来越多地应用在地理现象演变的模拟中, 取得了许多有意义的研究成果。CA 模型的关键是如何定义转换规则, 但目前常规的 CA 模型是采用静态的转换规则, 在研究区面积较大和模拟间隔较长时有一定的弊端。而且, 获取具体的转换规则往往是十分困难的。许多地理现象的演变受众多的因素影响, 它们之间的关系也较复杂, 难以用具体的转换规则来表达。

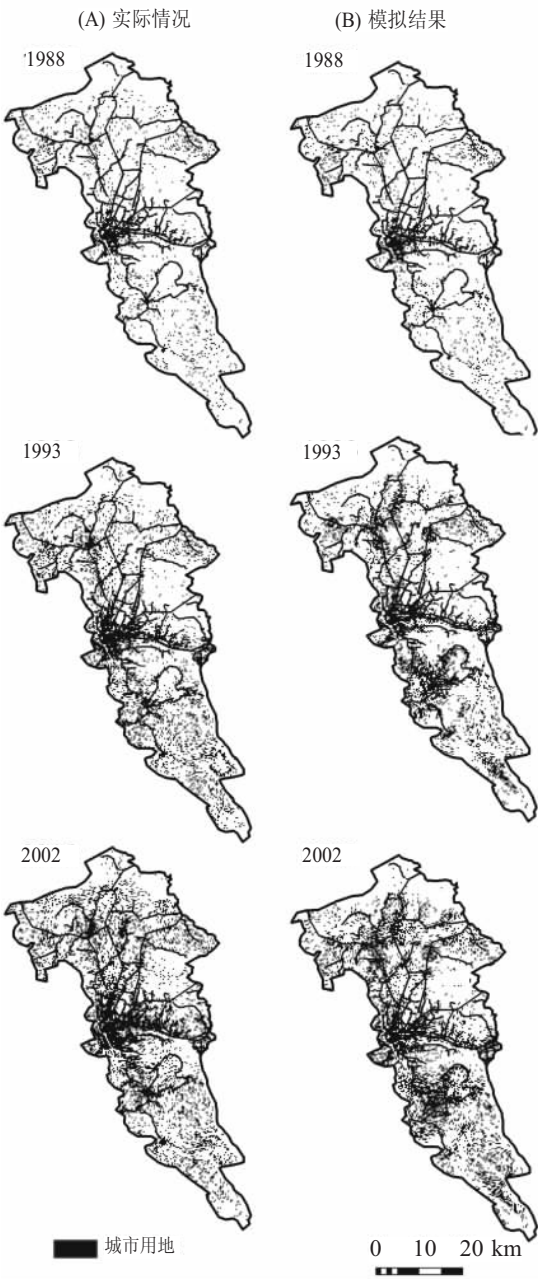


图 8 1988 年、1993 年和 2002 年广州实际与模拟的城市演变对比

Fig. 8 The actual and simulated patterns of urban development in Guangzhou in 1988, 1993 and 2002

CA 模型被越来越多地应用在地理现象演变的模拟中, 取得了许多有意义的研究成果。CA 模型的关键是如何定义转换规则, 但目前常规的 CA 模型是采用静态的转换规则, 在研究区面积较大和模拟间隔较长时有一定的弊端。而且, 获取具体的转换规则往往是十分困难的。许多地理现象的演变受众多的因素影响, 它们之间的关系也较复杂, 难以用具体的转换规则来表达。

为了解决 CA 转换规则获取的瓶颈问题，本文首次提出了基于案例推理 (CBR) 的 CA 模型。该模型无需定义具体的转换规则，而是通过案例来隐含地表达转换规则，从而大大提高了建模的效率。并通过对 CBR 中的 k 最近邻算法的改善，使得模型能反映随空间而变化的转换规则，在研究区面积较大时有一定的优势。由于案例库是动态更新的，该模型还有自适应的特点，能更好地适应快速变化的资源环境。因此 CBR-CA 模型在模拟较为复杂的地理现象时具有非常明显的优势。

将该模型应用于大区域的珠江三角洲城市扩张的模拟中。该区域包含了不同等级的城市群，用传统单一的静态转换规则对其城市演化进行模拟效果并不理想。本模型利用 GIS 和不同年份的卫星遥感图像来建立案例库，这些案例反映了城市扩张和地理要素的复杂关系。通过案例推理来决定元胞的状态转变 (城市扩张)，模拟了研究区在 1988-2002 年的城市扩张情况。根据逐点对比与 Moran I 计算，模拟结果与实际情况很接近。并与常规的基于 Logistic 的 CA 模型进行了对比。分析表明，CBR-CA 模型所获得的模拟结果有更高的精度和更接近实际的空间格局，在模拟较为复杂的区域时比基于规则的 CA 模型有更好的优势。此外，CBR-CA 模型具有自适应的特性，能较好的适应资源环境的变化，因此，更适合对未来城市发展的情景进行模拟。

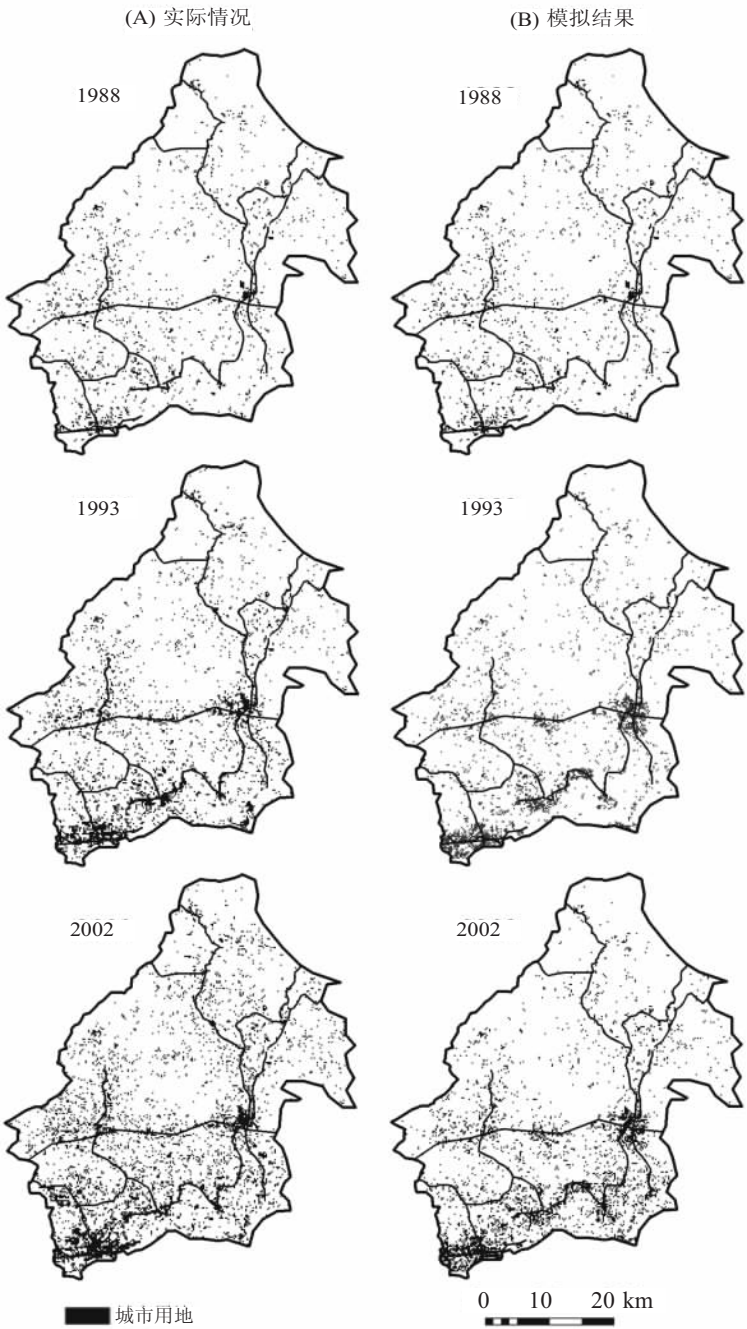


图 9 1988 年、1993 年和 2002 年增城实际与模拟的城市演变对比
Fig. 9 The actual and simulated patterns of urban development in Zengcheng in 1988, 1993 and 2002

CBR-CA 模型所获得的模拟结果有更高的精度和更接近实际的空间格局，在模拟较为复杂的区域时比基于规则的 CA 模型有更好的优势。此外，CBR-CA 模型具有自适应的特性，能较好的适应资源环境的变化，因此，更适合对未来城市发展的情景进行模拟。

表 1 案例推理方法的珠江三角洲的总体模拟精度

	1988 -1993年			1993-2002年		
	模拟非城市用地	模拟城市用地	精度	模拟非城市用地	模拟城市用地	精度
实际非城市用地	476678	58315	0.89	432997	65273	0.87
实际城市用地	31953	71454	0.69	48513	91616	0.65
总精度			0.86			0.82
Kappa系数			0.53			0.51

表 2 案例推理方法的各城市的总体模拟精度

	广州市区	深圳	东莞	增城
1988-1993年	0.833	0.872	0.845	0.881
1993-2002年	0.812	0.842	0.797	0.848

表 3 Logistic 回归方法的模拟精度

	1988 -1993年			1993-2002年		
	模拟非城市用地	模拟城市用地	精度	模拟非城市用地	模拟城市用地	精度
实际非城市用地	461699	73294	0.86	416056	82214	0.84
实际城市用地	50462	52944	0.51	76651	63478	0.45
总精度			0.81			0.75
Kappa系数			0.34			0.28

参考文献 (References)

[1] Wolfram S. Cellular automata: A model of complexity. Nature, 1984, 31: 419-424.

[2] Goles E. Cellular automata, dynamics and complexity. In: Manneville P, Boccarra N, Vichniac G Y et al. (eds.). Cellular Automata and Modeling of Complex Physical Systems. Berlin: Springer-Verlag, 1989. 10-20.

[3] Binder P. Evidence of Lagrangian Tails in a Lattice Gas. In: Manneville P, Boccarra N, Vichniac G Y et al. (eds). Cellular Automata and Modeling of Complex Physical Systems. Berlin: Springer-Verlag, 1989. 155-160.

[4] Couclelis H. Cellular worlds: A framework for modeling micro-macro dynamics. Environment and Planning A, 1985, 17: 585-596.

[5] Couclelis H. Of mice and men: What rodent populations can teach us about complex spatial dynamics. Environment and Planning A, 1988, 20: 99-109.

[6] Couclelis H. Macrostructure and microbehavior in a metropolitan area. Environment and Planning B, 1989, 16: 141-154.

[7] Batty M, Xie Y. From cells to cities. Environment and Planning B: Planning and Design, 1994, 21: 531-548.

[8] 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.

[9] Clarke K C, Brass J A, Riggan P J. A cellular automata model of wildfire propagation and extinction. Photogrammetric Engineering & Remote Sensing, 1994, 60: 1355-1367.

[10] Chen Jianping, Ding Huoping, Wang Gongwen et al. Desertification evolution modeling through the integration of GIS and cellular automata. Journal of Remote Sensing, 2004, 8(3): 254-260. [陈建平, 丁火平, 王功文 等. 基于 GIS 和元胞自动机的荒漠化预测模型. 遥感学报, 2004, 8(3): 254-260.]

[11] Lekkas G P, Avouris N M, Viras L G. Case-based reasoning in environmental monitoring applications. Applied Artificial Intelligence An International Journal, 1994, 8: 359-376.

[12] Holt A, Benwell G L. Applying case-based reasoning techniques in GIS. International Journal of Geographical Information Science, 1999, 13(1): 9-25.

[13] Yeh A G O, Shi X. Applying case-based reasoning to urban planning: A new planning-support system tool. Environment and Planning B, 1999, 26(1): 101-115.

[14] Li X, Yeh A G O, Liao Q F. Case-based reasoning (CBR) for land use classification using radar images. Journal of Remote Sensing, 2004, 8(3): 246-253. [黎夏, 叶嘉安, 廖其芳. 利用案例推理 CBR 方法对雷达图像进行土地利用分类. 遥感学报, 2004, 8(3): 246-253.]

[15] Wu F, Webster C J. Simulation of land development through the integration of cellular automata and multicriteria

- evaluation. *Environment and Planning B*, 1998, 25: 103-126.
- [16] Li Xia, Yeh A G O. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 2002, 16(4): 323-343.
- [17] Li X, Yeh A G O. Data mining of cellular automata's transition rules. *International Journal of Geographical Information Science*, 2004, 18(8): 723-744.
- [18] 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.
- [19] Theil H. *Economics and Information Theory*. Amsterdam: NorthHolland, 1967. 488.
- [20] Xu X Z. A note on the subjective and objective integrated approach to determine attribute weights. *European Journal of Operation Research*, 2004, 156: 530-532.
- [21] Dasarathy B V. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. Los Alamitos, CA: IEEE Computer Society Press, 1991. 447.
- [22] Li X, Yeh A G O. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. *Landscape and Urban Planning*, 2004, 69(4): 335-354.
- [23] Li X, Yeh A G O. Mining cellular automata's transition rules. *Science in China (Series D)*, 2004, 34(9): 865-872. [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.]

Case-based Cellular Automaton for Simulating Urban Development in a Large Complex Region

LI Xia, LIU Xiaoping

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

Abstract: The essential part of geographical cellular automata (CA) is to provide appropriate transition rules so that realistic patterns can be simulated. Transition rules can be defined by a variety of methods, such as multicriteria evaluation (MCE), logistic regression, neural networks, and data mining. The solicitation of concrete knowledge (transition rules) is often difficult for many applications. There are problems in representing complex relationships by using detailed rules. This study demonstrates that the case-based approach can avoid the problems of the rule-based approach in defining CA. The proposed method is based on the case-based reasoning techniques, which don't require the procedure of soliciting explicit transition rules. The knowledge for determining the state conversion of CA is inexplicitly embedded in discrete cases. The lazy-learning technology can be used to represent complex relationships more effectively than detailed equations or explicit transition rules.

This paper presents an extended cellular automaton in which transition rules are represented by using case-based reasoning (CBR) techniques. The common k-NN algorithm of CBR has been modified to incorporate the location factor to reflect the spatial variation of transition rules. Multi-temporal remote sensing images are used to obtain the adaptation knowledge in the temporal dimension. This model has been applied to the simulation of urban development in the Pearl River Delta which has a hierarchy of cities. Comparison indicates that this model can produce more plausible results than rule-based CA in simulating large complex regions.

Key words: cellular automata; case-based; k-NN; dynamic transition rules