

文章编号: 1007 4619 (2006) 06 0836 11

# 基于支持向量机的元胞自动机 及土地利用变化模拟

杨青生, 黎 夏

(中山大学 地理科学与规划学院, 广东 广州 510275)

**摘 要:** 提出了利用遥感数据, 并采用支持向量机来确定元胞自动机非线性转换规则的新方法。元胞自动机在模拟复杂地理现象时, 需要采用非线性转换规则。目前元胞自动机主要采用线性方法来获取转换规则, 在反映复杂的非线性地理现象时有一定的局限性。以城市扩张的模拟为例, 将模拟城市系统的主要特征变量映射到 Hilbert 空间后, 通过 SVM 建立最优分割超平面, 分割超平面的分类决策函数由径向基核 (Radial Basis Kernel) 构造。利用历史遥感数据校正超平面的决策函数, 确定城市元胞自动机的非线性转换规则, 计算出城市发展概率。利用所提出的方法, 对深圳市 1988—2010 年的城市发展进行了模拟, 取得了较理想的模拟效果。研究表明, 基于 SVM CA 模型的模拟精度比传统 MCE 方法模拟精度高, Moran I 指数与实际更为接近。

**关键词:** 元胞自动机; 支持向量机; 遥感; 土地利用变化

**中图分类号:** TP79      **文献标识码:** A

## Cellular Automata for Simulating Land Use Changes Based on Support Vector Machine

YANG Qing sheng LI Xia

(School of Geography and Planning, Zhongshan University, Guangdong Guangzhou 510275, China)

**Abstract:** This paper presents a new method to simulate complex land use systems by integrating support vector machine (SVM), cellular automata, and GIS. Recently, cellular automata (CA) have been increasingly used to dynamically simulate urban growth and land use. There are many issues that should be solved in the simulation of this type of complex systems. One major problem is how to define transition rules using training data. Linear boundaries are often used to retrieve transition rules which define the probability of state conversion. However, many geographical phenomena are very complex and transition rules should be defined using nonlinear boundaries.

In this study, a CA model based on the support vector machine (SVM) is developed using Visual Basic and ArcObjects of GIS. The GIS provides both data and spatial analysis functions for constructing SVM CA model. Training data is conveniently retrieved from remote sensing and GIS database for calibrating and testing the model. The SVM method is used to transform the data from nonlinear boundaries in the original space to linear boundaries in the Hilbert space. The nonlinear transition rules can then be defined by using the functions of SVM. The SVM CA model can be applied to the simulation of urban development. Complex global patterns can be generated from the local interactions with the SVM CA model.

This paper demonstrates that the proposed model can overcome some of the shortcomings of the existing CA

收稿日期: 2005 09 13; 修订日期: 2005 12 13

基金项目: 国家杰出青年基金资助项目 (编号: 40525002)、国家自然科学基金资助项目 (编号: 40471105)、教育部博士点基金资助项目 (编号: 20040558023)、“985 工程”GIS 与遥感的地学应用科技创新平台资助项目 (编号: 105203200400006)。

作者简介: 杨青生 (1974—), 男, 博士研究生。研究方向: 遥感和地理信息系统。E-mail: qsyang2002@163.com, qsyang2002@gmail.com, lixia@mail.sysu.edu.cn

models in simulating complex urban systems by using the nonlinear transition rules. The model has been successfully applied to the simulation of urban development in Shenzhen city of the Pearl River Delta.

Key words: cellular automata; support vector machine; remote sensing; land use changes

## 1 引言

元胞自动机 (Cellular Automata, CA) 1948 年被 Ulam 等学者提出后,已运用到许多非线性、复杂动态系统的模拟。近年来,CA 被越来越多地用来模拟城市系统,取得了许多有意义的研究成果<sup>[1-8]</sup>。如 Clarke 等模拟了美国旧金山地区的城市发展<sup>[9]</sup>,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<sup>[10]</sup>,黎夏和叶嘉安模拟了东莞市的城市发展<sup>[11]</sup>。

CA 的特点是通过简单的局部规则可模拟出复杂的空间格局及其演变过程。用 CA 模拟城市系统时,局部规则体现了邻近元胞状态对中心元胞状态的转变的影响。为了模拟出与实际更为接近的城市形态,需要引入与城市发展相关的区域变量和全局变量。这些区域变量主要有与区域商业中心的最短距离、与居住中心的最短距离、与各类道路和河流等地理要素的最短距离。全局变量主要是社会经济发展要素等变量。

CA 模拟城市的关键是如何用全局变量、区域变量和局部变量确定 CA 的转换规则。目前元胞自动机主要采用线性方法来获取转换规则,在反映如城市系统这样复杂的非线性地理现象时有一定的局限性。如 Wu 用多准则判断 (MCE) 及逻辑回归模型确定 CA 的转换规则<sup>[12]</sup>。黎夏等用主成分分析和理想点结合的方法<sup>[13]</sup>、神经网络方法<sup>[14]</sup>、决策树方法<sup>[15]</sup>、粗糙集<sup>[16]</sup>等确定 CA 的转换规则<sup>[17-20]</sup>。在这些转换规则中,逻辑回归模型方法本身要求特征变量间线性无关。而模拟城市发展的特征变量间往往存在相关性,如公路和铁路的布局在空间上接近时,离公路的距离和离铁路的距离相关性很大。特征变量间的这种相关性限制了逻辑回归模型精度的进一步提高。神经网络方法能很好地解决非线性分类,但方法本身是暗箱操作,模型内部物理机制不易理解和解释。

支持向量机通过核函数,将非线性可分的向量空间映射到线性可分的 Hilbert 空间,在 Hilbert 空间中,通过核函数的内积和原特征向量构造分类决策函数,具有良好的非线性分类能力。本文运用非线性支持向量分类机,将模拟城市系统的主要特征变量映射到线性可分的 Hilbert 空间中。Hilbert 空间中特征变量的分类决策函数通过径向基核函数的内

积由原特征空间的向量表示。在这个空间中,构造最优分割超平面。依据构造最优超平面的决策函数作为 CA 的转换规则模拟城市系统。根据历史遥感数据,采用基于支持向量机的 CA 模型模拟了深圳市 1988—2004 年城市用地的增长,并根据发展趋势,模拟了深圳市 2010 年的城市发展状况。

## 2 支持向量机及 SVM CA 模型

### 2.1 支持向量机

支持向量分类机 (Support Vector Machine, SVM) 通过度量待分向量与训练数据中的支持向量间的相似程度,对待分向量进行分类。这种相似程度的度量是通过向量的内积实现的<sup>[21]</sup>。如两个向量分别为  $x = ([x]_1, [x]_2, \dots, [x]_n)^T$  和  $x' = ([x']_1, [x']_2, \dots, [x']_n)^T$ , 它们的内积 ( $x \cdot x'$ ) 为:

$$(x \cdot x') = \sum_{i=1}^n [x]_i [x']_i \quad (1)$$

当判断待分向量属于哪个类别时,用距离进行的度量可转换为用向量内积度量。如已知  $x_+$  属于正类,  $x_-$  属于负类,判断  $x$  的归属时,运用距离进行度量的方法如下:若  $\|x - x_+\| < \|x - x_-\|$ , 则  $x \in x_+$ , 否则,  $x \in x_-$ 。若令  $\omega = (x_+ - x_-)$ ,  $m = (x_+ + x_-)/2$ , 则  $\|x - x_+\| < \|x - x_-\|$  等价于  $\omega$  和  $x - m$  呈钝角。据此,可得到决策函数:

$$y = \text{sgn} [(x - m) \cdot \omega] \quad (2)$$

式 (2) 中,

$$\begin{aligned} [(x - m) \cdot \omega] &= [(x - x_+ / 2 - x_- / 2) \cdot (x_+ - x_-)] \\ &= (x \cdot x_+) - \frac{1}{2} (x_+ \cdot x_+) - \\ &\quad \frac{1}{2} (x_- \cdot x_-) - (x \cdot x_-) + \\ &\quad \frac{1}{2} (x_+ \cdot x_-) + \frac{1}{2} (x_- \cdot x_-) \end{aligned} \quad (3)$$

式 (3) 表明,决策函数 (2) 仅仅依赖于 (3) 式中的  $x_+$ ,  $x_-$ ,  $x$  之间的内积,即这些内积决定了  $x$  的类别归属问题。

支持向量机以结构风险最小化准则和最大分类间隔为基本原则,依据“核”函数的内积,对向量进行分类<sup>[22-27]</sup>。具体分类方法如下:

(1) 设已知训练集

$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (x \times y)^l$ ,

其中:

$$x_i \in x = \mathbb{R}^n, y_i \in y = \{1, -1\}, i = 1, \dots, l$$

(2) 当训练数据线性可分及近似线性可分时 (图 1 (a), 适合 MCE 方法), 选择适当的惩罚参数 (松弛变量)  $c > 0$ , 构造并求解最优化问题:

$$\min_a \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j (x_i \cdot x_j) - \sum_{j=1}^l a_j$$

$$\text{s t } \sum_{i=1}^l y_i a_i = 0 \quad (4)$$

$0 \leq a_i \leq c, i = 1, \dots, l, a_i$  为拉格朗日系数  
得最优解  $a = (a_1, \dots, a_l)^T$ ;

(3) 计算  $\omega = \sum_{i=1}^l y_i a_i x_i$ , 选择  $a_i$  的一个小于  $c$  的正分量  $a_j$ , 并据此计算

$$b = y_j - \sum_{i=1}^l y_i a_i (x_i \cdot x_j);$$

(4) 构造分化超平面  $(\omega \cdot x) + b = 0$ , 由此

求得决策函数

$$f(x) = \text{sgn} [ (\omega \cdot x) + b ] \quad (5)$$

式 (4) 的最优化问题可由标准拉格朗日乘子法求解, 决策函数 (5) 可由原向量及内积表示, 即:

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l y_i a_i (x_i \cdot x) + b \right] \quad (6)$$

式 (4) 中, 与每个非零  $a_i$  对应的向量均为支持向量。

当训练数据线性不可分时, 可通过映射  $(x)$  将线性不可分的特征空间  $x$  转换到更高维的线性可分的 Hilbert 特征空间  $X$  中 (图 1 (b)、(c))。在 Hilbert 特征空间  $X$  中, 选择适当的惩罚参数  $c > 0$ , 构造并求解最优化问题:

$$\min_a \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j (x_i \cdot x_j) - \sum_{j=1}^l a_j$$

$$\text{s t } \sum_{i=1}^l y_i a_i = 0 \quad (7)$$

$0 \leq a_i \leq c, i = 1, \dots, l, a_i$  为拉格朗日系数

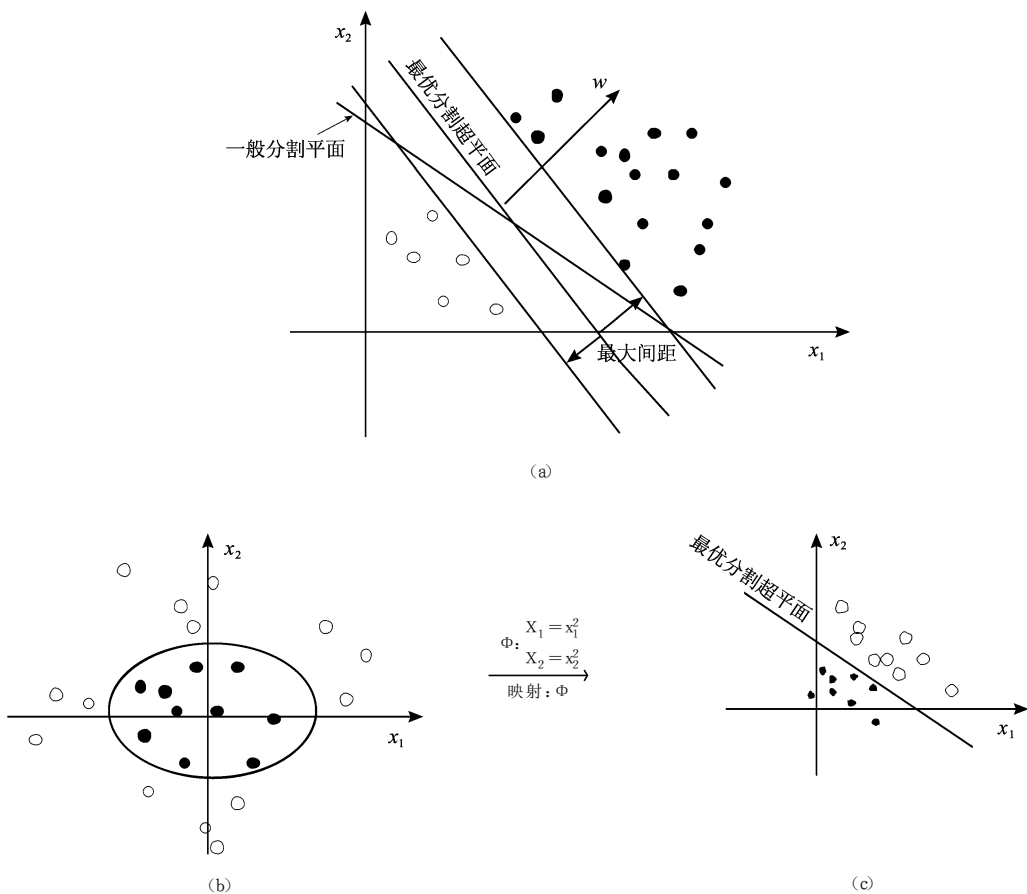


图 1 二维线性可分、非线性可分及映射后的线性可分示意图

(a) 二维空间线性可分 (适合于 MCE 方法); (b) 二维空间中用椭圆分割 (非线性可分); (c) 映射后的 Hilbert 空间中线性可分

Fig 1 Linear classified and nonlinear classified spaces and mapped into linear classified space

(a) Linear classification in 2 dimension space (b) Nonlinear classification in 2 dimension space (classified with ellipse boundary);

(c) Linear classification after projection in Hilbert space

计算  $\omega = \sum_{i=1}^l y_i a_i (x_i)$ , 选择  $a_i$  的一个小于  $c$  的正分量  $a_j$ , 并据此计算

$$b = y_j - \sum_{i=1}^l y_i a_i [(x_i) \cdot (x_j)] \quad (8)$$

构造分化超平面  $(\omega \cdot (x)) + b = 0$ , 由此求得决策函数

$$f(x) = \text{sgn} \{ [\omega \cdot (x)] + b \} \quad (9)$$

在实际运用中, 映射函数  $(x)$  可能非常复杂, 较难实现。但最优化的目标函数 (式 (7)) 和决策函数 (式 (9)) 只涉及映射后向量的点积运算, 即  $[(x_i) \cdot (x_j)]$  的形式。如果存在“核”函数  $k(\cdot)$ , 使

$$k(x_i, x_j) = [(x_i) \cdot (x_j)] \quad (10)$$

那么, 就能用原空间中的特征变量来实现 Hilbert 空间中的点积运算, 从而绕开映射  $(x)$  的具体形式。

根据泛函分析中的有关理论, 只要核函数  $k(\cdot)$  满足 Mercer 条件, 它就对应于某一变换空间中的点积, 也就是说, 存在映射  $(x)$ , 使得 (10) 式成立<sup>[21]</sup>。常见的满足 Mercer 条件的核函数有多项式核函数:  $k(x, y) = (x \cdot y + 1)^p$ , 径向基函数:  $k(x, y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$  和 Sigmoid 函数:  $k(x, y) = \tanh(kx \cdot y - \delta)$ 。此时, 支持向量分类机构造如下:

选择核函数  $k(\cdot)$  和惩罚参数  $c$  构造并求解最优化问题:

$$\begin{aligned} \min_a & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j k(x_i, x_j) - \sum_{j=1}^l a_j \\ \text{s.t.} & \sum_{i=1}^l y_i a_i = 0 \end{aligned} \quad (11)$$

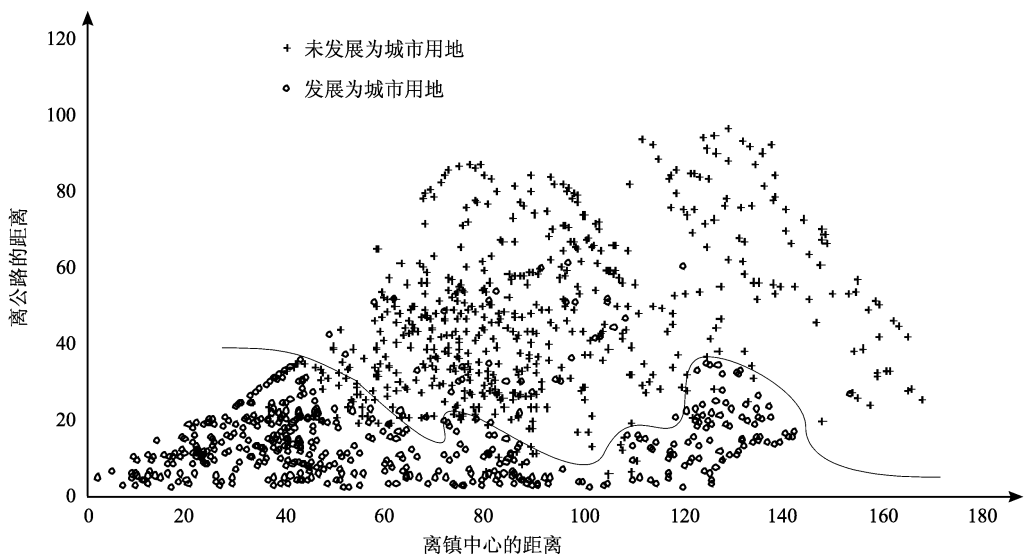


图 2 城市模拟中常碰到的非线性边界

Fig 2 Nonlinear classified boundary in simulating urban system

$0 \leq a_i \leq c, i = 1, \dots, l$ ,  $l$   $a_i$  为拉格朗日系数  
选择  $a_i$  的一个小于  $c$  的正分量  $a_j$ , 并据此计算

$$b = y_j - \sum_{i=1}^l y_i a_i k(x_i, x_j); \quad (12)$$

求得决策函数,

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l y_i a_i k(x_i, x) + b \right] \quad (13)$$

其中,

$$f(x) = \begin{cases} 1, & \sum_{i=1}^l y_i a_i k(x_i, x) + b > 0 \\ -1, & \sum_{i=1}^l y_i a_i k(x_i, x) + b \leq 0 \end{cases} \quad (14)$$

## 2.2 SVM CA 模型

城市是复杂的非线性系统。在城市演变模拟时, 采用非线性的转换规则更能反映城市系统复杂的特征。在许多情况下, 城市演变中的转变和不转变的边界是十分复杂的, 无法用简单的线性边界来区分。图 2 显示了从遥感数据获得的深圳城市扩张的转变和不转变的边界所具有的非线性特点。因此, 如果采用常用的多准则判断 (MCE) 及逻辑回归模型是有较大的弊端的。

本文提出了基于支持向量机的 CA 模型来解决非线性边界问题。运用支持向量机确定城市 CA 的非线性转换规则时, 通过支持向量机计算区域空间变量 (离市中心的距离、离商业中心的距离、离各类道路的距离等) 对每个元胞的城市发展概率的贡献。如果直接运用支持向量机的硬分类结果, 无法在 CA 模型中动态计算邻近范围的城市元胞对中心元胞城

市发展概率的影响,按照 Baty<sup>[4]</sup>, Wu<sup>[17]</sup>等提出的 CA 模型的城市发展概率计算方法,借鉴支持向量机分类的概率输出方法<sup>[28]</sup>,运用决策函数  $f(x) = \sum_{i=1}^l y_i a_i k(x_p, x) + b$  直接计算每个元胞到最优超平面的距离,对硬分类超平面软化,计算每个元胞的城市发展概率。每个元胞转变为城市用地的概率如下:

$$p_r = \frac{1}{1 + \exp\left[-\left(\sum_{i=1}^l y_i a_i k(x_p, x) + b\right)\right]} \quad (15)$$

式中,  $p_r$  为区域变量作用下元胞的城市发展概率,  $y_p, a_i, x_p, x, b$  的含义同式 (13)。核函数  $k(\cdot)$  选用高斯径向基核,即  $k(x, x_i) = e^{-\|x-x_i\|^2/2\sigma^2}$ 。

考虑到元胞邻近范围的影响、不确定性因素的影响及约束条件的影响,元胞  $k$  在  $t+1$  时刻的城市发展概率为:

$$p_{t+1, k} = [1 + (-\ln r)^a] \times \frac{1}{1 + \exp\left[-\left(\sum_{i=1}^l y_i a_i e^{-\|x_k-x_i\|^2/2\sigma^2} + b\right)\right]} \times \Omega_{3 \times 3, k}^t \times \prod_{i=1}^m c_{i, k} \quad (16)$$

式 (16) 中,  $r$  为 0—1 间的随机数,  $a$  为控制离散变量大小的参数,一般取 5;  $\Omega_{3 \times 3, k}^t$  为  $t$  时刻第  $K$  个元胞  $3 \times 3$  邻近范围已城市化的数目,在每次模拟过程中迭代计算;  $\prod_{i=1}^m c_{i, k}$  表示第  $K$  个元胞的约束条件,  $c_i$  为约束要素。

$$\|x_k - x_i\|^2 = (x_k - x_i)^T \cdot (x_k - x_i),$$

$$x_k = (x_{k1}, x_{k2}, \dots, x_{km})^T \quad (17)$$

每次迭代运算中,计算出每个元胞发展为城市的概率后,将其与预先设定的阈值  $p_{th\_reshold}$  进行比较,如果元胞的城市发展概率大于或等于该阈值,则该元胞转换为城市用地;否则,该元胞不转换为城市用地,用公式 (18) 表示如下:

$$\begin{cases} p_t \geq p_{th\_reshold}, & \text{转变为城市用地} \\ p_t < p_{th\_reshold}, & \text{不转变为城市用地} \end{cases} \quad (18)$$

运用支持向量分类机确定元胞自动机的转换规则,并模拟城市系统的流程如图 3 所示。

### 3 模型应用

#### 3.1 数据及获取方法

本文的 SVM CA 模型是利用 VB 结合

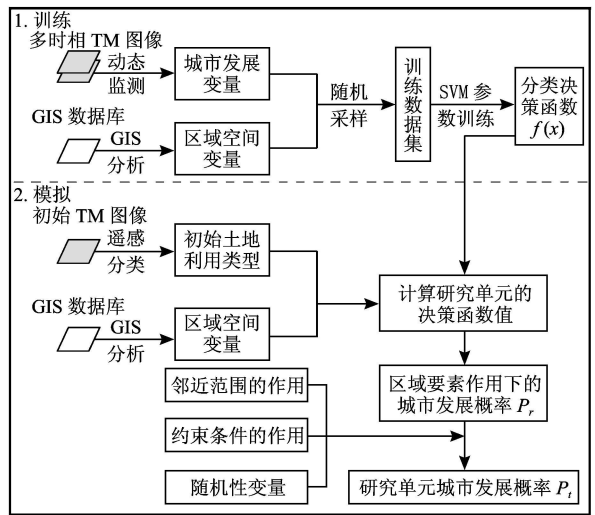


图 3 基于 SVM CA 模型模拟城市系统流程图  
Fig 3 The SVM CA model for the simulation of urban systems

ArObjects 来构成的。ArObjects 管理 GIS 数据库,并直接应用 GIS 的功能动态计算距离变量和邻近范围已城市化的元胞数。模型中决策函数参数的训练独立进行。运用专门的支持向量机软件 OSU SVM 训练,获取决策函数的参数值。VB 读取决策函数的参数值后结合 ArObjects 进行真实的城市模拟。

本文以珠江三角洲的深圳市为例进行了模拟研究。运用 1988—1993 年的 TM 遥感图像获取城市发展的历史资料,将该时间段内转变为城市用地的元胞编码为 1,其他元胞编码为 -1。模拟时,主要选取了以下变量:

因变量 ( $y$ ): 城市发展变量 (转化为城市用地,  $y=1$ , 否则,  $y=-1$ )

区域空间变量:

- (1) 离市中心的距离 ( $x_1$ );
- (2) 离镇中心的距离 ( $x_2$ );
- (3) 离国道、省道的距离 ( $x_3$ );
- (4) 离铁路的距离 ( $x_4$ );
- (5) 离高速公路的距离 ( $x_5$ );

局部变量:

- (1) 邻近范围 ( $3 \times 3$  邻域) 已城市化的元胞数 ( $x_6$ );
- (2) 元胞的土地利用类型 (约束条件) ( $x_7$ )

上述变量中,因变量通过遥感图像分类的方法获取,重采样后的像元分辨率为  $50m \times 50m$ 。距离变量通过 ArcGIS 的空间分析功能中的 Euclidean 函数获取。邻近范围已城市化的元胞数通过

ArGIS 的空间分析功能中的 Neighbor 函数动态获取。

利用随机采样方法获取 1988—1993 年深圳市城市发展及空间变量。首先,运用分层采样的方法产生随机点的空间坐标,并利用 ArGIS 的 Sample 功能读取这些空间坐标点对应的城市发展变量及空间变量。获取训练数据集后,运用支持向量机软件 OSU SVM 训练获取非线性决策函数的参数值。决策函数的惩罚参数  $c$  和径向基函数的参数  $\sigma$  通过常

用的分组交叉试验获取。 $c$  和  $\sigma$  分别取 0.1, 0.5, 1, 5, 10 形成 25 组参数组合。用每组参数对训练数据集训练,获取不同参数组合的支持向量和对应的拉格朗日系数  $a_i$ , 并用检验数据集检验各组模拟的精度,选取对检验数据集模拟精度最高 (86.4%) 的参数组合 ( $c=1, \sigma=1$ ) 和对应的支持向量及其参数  $a_i$ 。从 7996 个向量中获取了 7980 个支持向量及对应的拉格朗日系数  $a_i$ 。获取的部分支持向量及对应的拉格朗日系数  $a_i, y_i$  和  $y_i a_i$  见表 1。

表 1 部分支持向量及对应的系数  $a_i, y_i, y_i a_i$

Table 1 Support vectors and  $a_i, y_i, y_i a_i$

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$a_i$	$y_i$	$y_i a_i$
519.64	40.311	10.63	30.017	558.68	0.94142	1	0.94142
369.49	35.805	2.8284	21	287.71	0.94184	1	0.94184
490.29	116.91	64.008	83.199	417	0.99738	1	0.99738
433.24	68.964	30.067	65.054	160.38	1	1	1
395.38	89.051	69.029	74.169	222.61	0.97771	1	0.97771
526.82	110.86	64	17	539.45	0.77101	1	0.77101
509.2	31.145	96.607	121.06	324.08	0.9417	1	0.9417
531.61	108.04	70.214	48.662	488.93	0.96395	1	0.96395
398.18	78	19.698	30.414	442.12	0.25	1	0.25
330.46	18.682	1	36.056	384.69	0.9785	1	0.9785
395.38	89.051	69.029	74.169	222.61	0.2131	1	0.2131
488.33	80.262	73.763	56.462	338.02	0.25	1	0.25
612.67	54.562	19.416	57.723	356.48	0.80858	-1	-0.80858
504.6	57.245	5.3852	31.064	248.21	0.5	-1	-0.5
217.59	106.02	14.56	55.027	93.984	0.76227	-1	-0.76227
43.382	73.98	5	34	96.519	0.52561	-1	-0.52561
198.35	82.28	15	67.417	54.644	0.80813	-1	-0.80813
323.56	74.431	18	7.2801	142.3	0.75	-1	-0.75
176.82	94.668	50.448	11	169.85	0.78667	-1	-0.78667
267.42	97.098	13.038	1	19.925	0.81278	-1	-0.81278

在区域空间变量作用下,基于 SVM 的元胞城市发展概率如下:

$$P_r = \frac{1}{1 + \exp \left[ - \left( \sum_{i=1}^l y_i a_i e^{-0.5 \times \|x_k - x_i\|^2} - 0.19154 \right) \right]} \quad (19)$$

式 (19) 中,  $x_i$  为 7980 个支持向量,  $y_i$  为  $x_i$  对应的  $y$  值,  $a_i$  为  $x_i$  对应的拉格朗日系数,  $x_k$  为第  $k$  个元胞对应的向量。

考虑到元胞邻近范围、随机变量和约束条件的影响,深圳市某元胞  $k$  在  $t+1$  时刻城市发展的概率如下:

$$p_{i,k} = [1 + (-\ln D)^5] \times \frac{1}{1 + \exp\left[-\left(\sum_{i=1}^l y_i a_i e^{-0.5 \times \|x_k - x_i\|^2} - 0.19154\right)\right]} \times \Omega_{3 \times 3, k}^t \times \prod_{i=1}^m c_{i,k} \quad (20)$$

模拟时,首先用 SVM 计算每个元胞在区域空间变量作用下的城市发展概率,在此基础上计算在初始状态、约束条件和随机变量共同作用下的元胞城市发展概率。初始状态从 1988 年遥感图像分类中获取。模拟过程中,邻近范围已城市化的元胞数依迭代次数动态计算。

约束条件的设置从初始的土地利用类型和政府规划资料中获取。如:林地的城市发展概率非常小,约束值设置较小;河流、湖泊等水体和城市绿地的城市发展概率极小,约束值可设为 0;政府确定的开发区的城市发展概率很大,约束值可设为 1。

模拟过程中,需要确定城市发展的阈值  $P_{\text{threshold}}$ 。如果确定的阈值  $P_{\text{threshold}}$  过大,模拟得到的城市形态过于集中;如果过小,模拟得到的城市形态过于分散。太大或太小的阈值都不能获得合理的城市形态。按照 Batty<sup>[4]</sup>, Clarke<sup>[9]</sup>, Wu<sup>[17]</sup>, Li Yeh<sup>[18]</sup> 等的方案,经过试验,模拟过程中确定的阈值  $P_{\text{threshold}} = 0.65$ ,并在每次迭代过程中以 0.00008 的速率动态递减。

CA 需要循环迭代运算多次才能获得最终的模拟结果。但对循环迭代运算次数的多少,目前并没有统一的意见。在每次循环迭代运算中,局部的相

互作用是模拟的关键。循环迭代运算次数太少,就很难产生较真实的空间分布细节。以国内外学者常采用的迭代次数为参照<sup>[4, 15]</sup>,CA 每年进行 100 次左右的迭代运算是比较合适的。本文在实验的基础上,确定了 CA 模型的迭代次数。1988—1993 年为 600 次,1993—2004 年为 1400 次,2004—2010 年为 600 次。

### 3.2 模拟结果

利用 1988—1993 年的训练数据,获取转换规则的参数,以此转换规则模拟了深圳市 2004 年和 2010 年的城市发展状况。模拟时,以 1988 年从遥感图像分类得到的城市用地作为初始城市用地(图 4(a)),以 1993 年、2004 年从遥感图像分类得到的城市用地(图 4(b)、图 4(d))总量为参照,经过 600 次迭代运算,得到了深圳市 1993 年的模拟城市用地(图 4(c)),经过 2000 次迭代运算,得到了 2004 年的模拟城市用地(图 4(e))。表 2 为典型数据的最终分类结果。图 5 是 1988—1993 年的城市用地扩张模拟过程图。同时,根据 1988—2004 年的城市发展趋势,经过 2600 次迭代运算,获取了 2010 年模拟的城市用地(图 6)。

表 2 基于 SVM CA 的典型数据分类结果

Table 2 The simulation result based on SVM CA model

x	y	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	实际值	模拟概率
818166.7	2499786	36.23534	36.68787	9.433981	32	12	5	1	0.85
818830.4	2500516	56.72742	50.08992	7.071068	17.11724	27	4	1	0.79
822149	2504034	152.8987	59.77457	15.13275	41.48494	85.16455	1	1	0.38
822348.1	2506689	197.4968	70.3278	32.01562	34.20527	66.00758	0	-1	0.23
838675.6	2514720	542.603	70.5762	41.67733	326.9557	51.97115	0	-1	0.017
833764.1	2523348	601.0208	104.6566	75.21304	297.3113	134.4061	0	-1	0.15
793741.8	2520029	635.3849	72.20111	33	503.746	24.18677	0	1	0.64
791020.6	2522352	603.4318	16.27882	10.29563	543.5679	13.45362	0	1	0.59
800910	2519432	530.0236	63.78872	57	363.3208	153.16	0	-1	0.44
805224.2	2520096	494.0162	86.26703	142.2146	276.4218	115.6936	0	-1	0.31
810932.2	2518237	415.0301	48.05206	50.21952	165.9789	30.8707	0	-1	0.03
802502.9	2509410	364.6107	90.47099	34.36568	363.4556	45.88028	0	-1	0.12
794870.2	2500317	444.7134	38.94868	24.83949	497.0362	5	0	-1	0.25
805489.7	2508812	310.7748	73.97973	38.60052	302.3673	68.26419	0	-1	0.38
825998.6	2500317	184.1765	99.82484	11	124.7878	79.40403	2	1	0.71
819826	2509808	235.7478	88.60023	54.45181	45.60702	17.08801	0	-1	0.17
820290.6	2514189	322.8188	15.13275	3.605551	12.80625	66.70832	2	1	0.49
801905.6	2501246	307.3321	113.2166	23.76973	356	23	4	1	0.51
813786.2	2512131	282.3491	107.5407	40.79216	142.2849	11.18034	0	-1	0.07
829516.3	2502043	260.44	59.94164	14.86607	192.7511	14.21267	0	-1	0.34

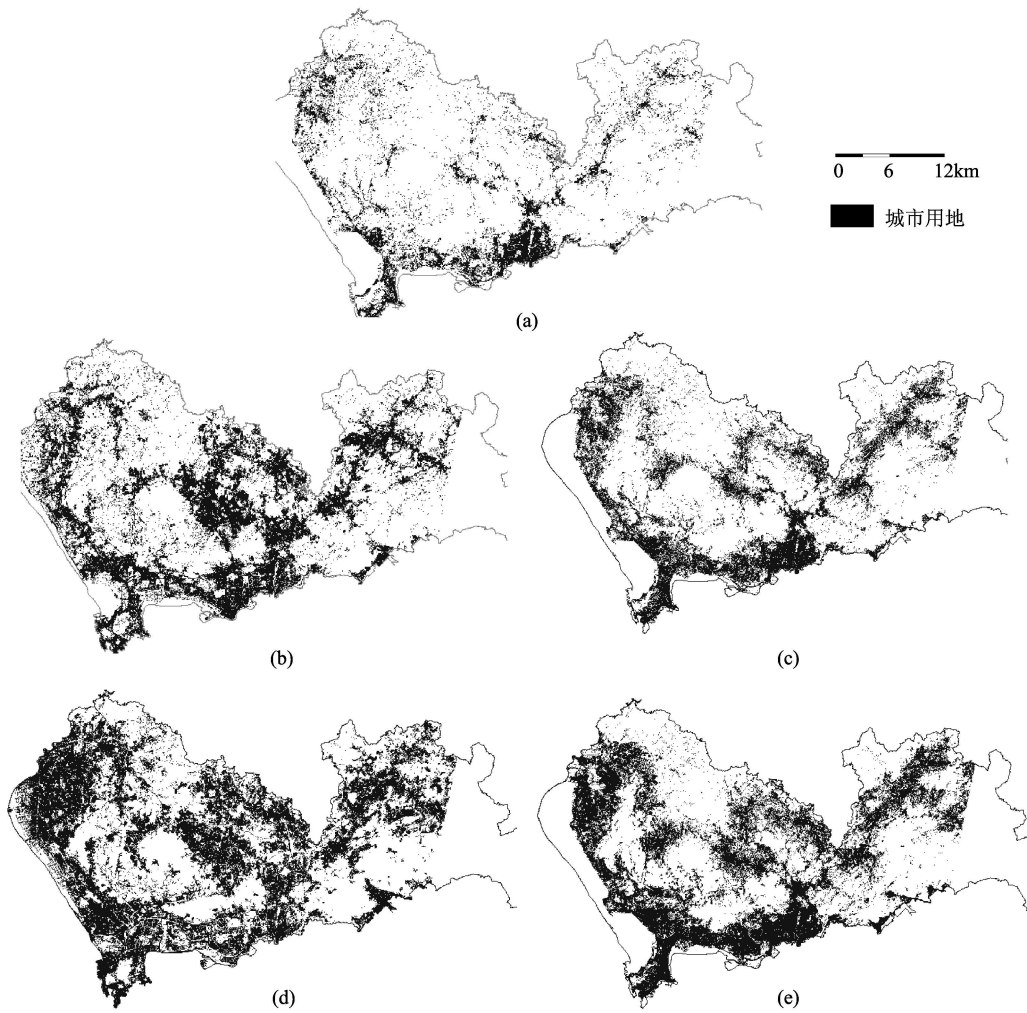


图 4 深圳市 1988—2004 年城市发展模拟结果与实际对比

(a) 1988 年城市用地 (初始); (b) 1993 年实际城市用地; (c) 1993 年模拟城市用地;  
(d) 2004 年实际城市用地; (e) 2004 年模拟城市用地

Fig 4 The simulation of urban system from 1988 to 2004 based on SVM CA model

(a) Urban land use of 1988 (initiation); (b) Actual urban land use of 1993; (c) Simulation urban land use of 1993;  
(d) Actual urban land use of 2004; (e) Simulation urban land use of 2004

可以发现,深圳市在 20 世纪 90 年代初期处于城市高速发展阶段,到 2000 年前后,城市发展速度有所减慢。这种结果的形成除了受到城市发展的一般因素影响外,还与将深圳作为经济特区、大力发展深圳经济的政策关系非常密切。从城市发展形态看,90 年代初期,城市增长主要沿道路进行,同时,龙岗区和龙华镇周围地区增长的城市用地更多。由于这个阶段主要沿道路和新中心增长,城市用地分布相对较为分散;90 年代末期开始,城市主要沿道路和已存在的城市中心增长,城市用地分布更为密集。

### 3.3 精度检验与对比分析

对模型模拟的结果用逐点对比和总体形态评价

的方法进行了检验。1993 年模拟的逐点对比总体精度达到 87.25% (表 3), 2004 年模拟的逐点对比总体精度达到 84.90% (表 4)。需要说明的是,模拟精度除了受模型本身影响外,还受遥感分类精度的影响,包括利用多时相遥感数据提取土地利用变化的驱动因子会存在误差。有关遥感分类精度对模拟的影响,还待进一步的研究。

同时,计算了评价总体形态的 Moran I 指数 (表 5)。将此方法与逻辑回归模型对深圳市模拟的精度进行了对比,表 5 和表 6 是逻辑回归模型的 Moran I 指数和混淆矩阵。表 3—表 6 表明,基于支持向量机的 CA 模型比 MCE 的逻辑回归模型模拟精度高, Moran I 指数与实际值更为接近。本文也

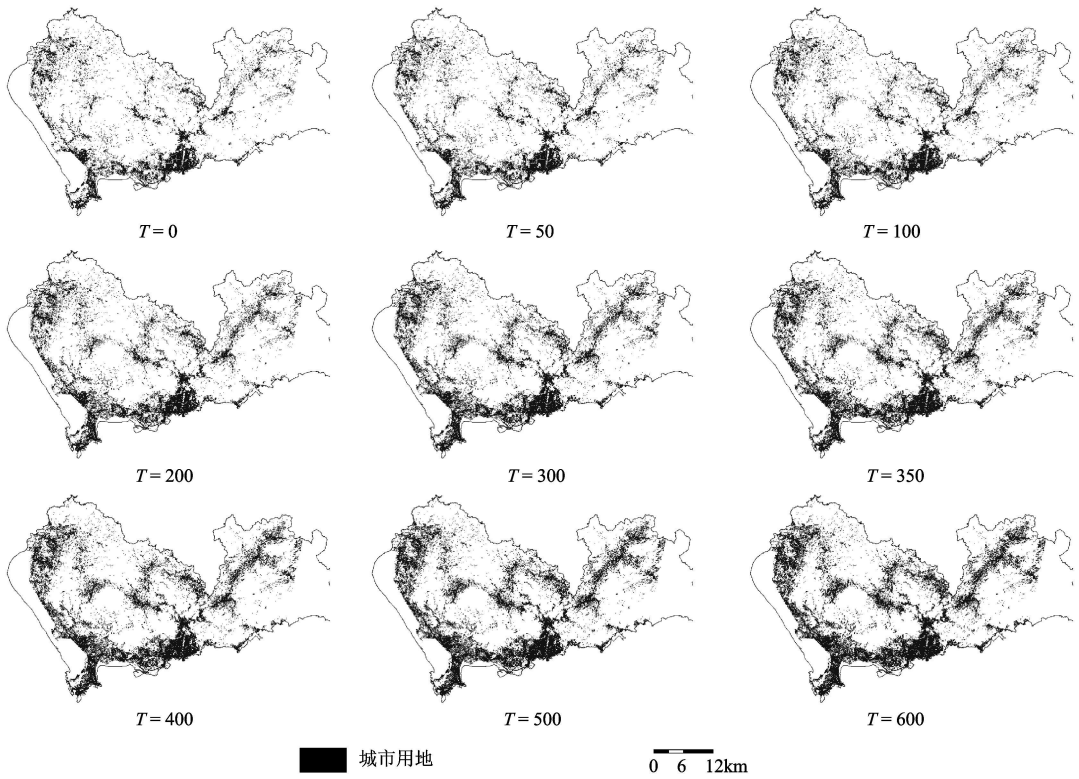


图 5 深圳市 1988—1993 年城市发展模拟过程图

Fig 5 The simulation procedure from 1988 to 1993 based on SVM CA model

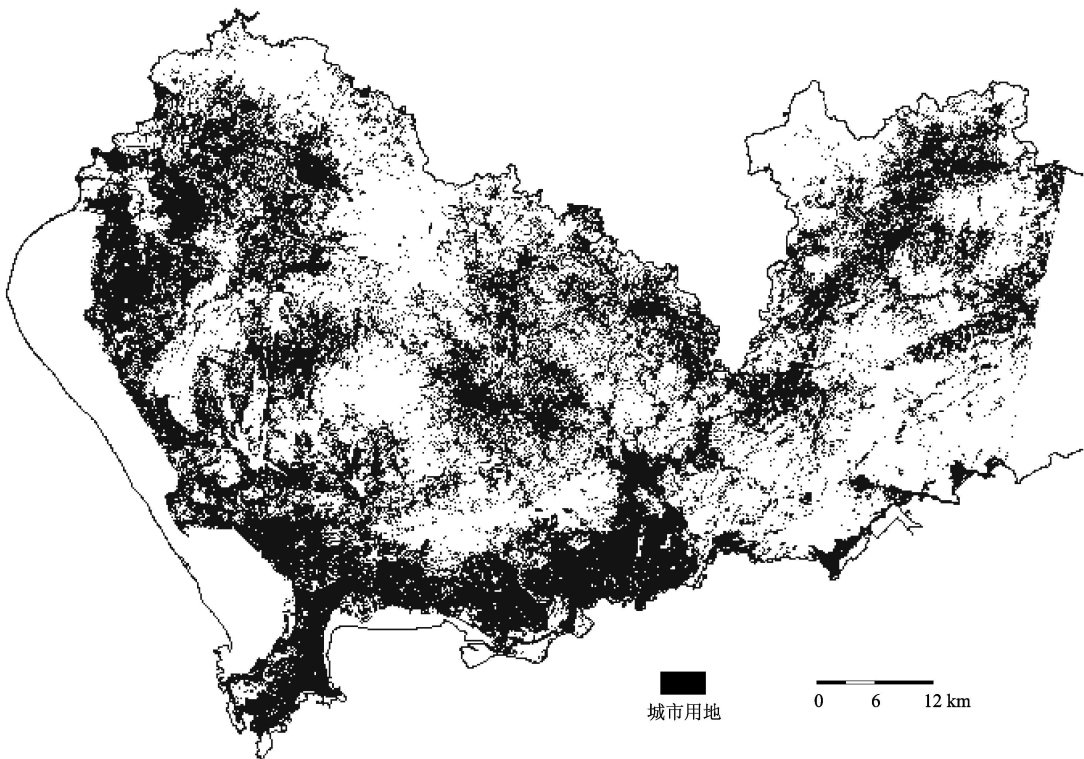


图 6 根据发展趋势模拟的深圳市 2010 年城市发展状况

Fig 6 The simulation result based on SVM CA model in 2010

表 3 SVM CA 模型精度评价混淆矩阵 (1993 年)

Table 3 The confusion matrix between the actual and simulated urban in 1993 based on SVM CA (1993)

		1993 年模拟		
		不转变	转变	正确比 /%
实际	不转变	495838	9388	98.14
	转变	91229	192804	67.9
	总精度			87.25
Kappa		0.70		

表 4 SVM CA 模型精度评价混淆矩阵 (2004 年)

Table 4 The confusion matrix between the actual and simulated urban in 2004 based on SVM CA (2004)

		2004 年模拟		
		不转变	转变	正确比 /%
实际	不转变	432064	22391	95.07
	转变	96781	238023	71.09
	总精度			84.90
Kappa		0.68		

表 5 Moran I 指数对比表

Table 5 Moran I Index

1988 年实际	1993 年实际	1993 年 SVM CA 模拟	1993 年 MCE CA 模拟	2004 年 SVM CA 模拟
0.618	0.719	0.734	0.741	0.768

表 6 基于 MCE 的逻辑回归模型精度评价混淆矩阵 (1993 年)

Table 6 The confusion matrix between the actual and simulated urban in 1993 based on MCE CA (1993)

		1993 年模拟		
		不转变	转变	正确比 /%
实际	不转变	485233	75166	86.6
	转变	99804	129056	56.4
	总精度			76.97
Kappa		0.44		

计算了 Kappa 系数来检验模拟的一致性 (表 3)。本文提出的方法的 Kappa 系数为 0.70 (1993 年) 和 0.68 (2004 年)。Kappa 系数表明, 模拟的精度在可接受的范围之内。传统的 MCE CA 模型模拟结果的 Kappa 系数为 0.44 (1993 年)。可见, 本方法对

模拟精度的改善效果较为理想。

## 4 结 论

CA 是模拟地理现象演变的非常有用的工具。如何合理地确定 CA 模型的转换规则及其参数是 CA 的关键, 也是目前 CA 研究所碰到的主要难题。目前 CA 模型主要采用线性的转换规则, 在反映非线性的复杂地理现象时, 线性 CA 转换规则具有较大的局限性。

本文运用非线性支持向量分类机, 从训练数据中获取非线性可分的支持向量及其对应的拉格朗日系数和核函数, 构造非线性的 CA 模型转换规则。基于支持向量机的 CA 模型能依靠“核”函数很好地反映了复杂城市系统演变的非线性特征。通过“核”函数将非线性可分的特征空间映射到更高维的线性可分的 Hilbert 空间中, 在线性可分的 Hilbert 空间中构造线性可分的最优分类超平面, 解决线性不可分问题。该方法的优点是 SVM CA 模型的转换规则能反映城市系统的非线性特征。

将该模型应用在珠江三角洲的深圳市, 以不同年份的卫星遥感图像为主要观测数据, 根据支持向量机方法提取能反映非线性特征的转换规则, 模拟了该地区 1988—2004 年的城市增长情况, 并对 2010 年的城市用地进行了模拟。利用逐点对比和 Moran I 指数来评价模型的精度。研究结果表明, 基于支持向量机的非线性 CA 模型能获得更为理想的模拟结果, 模拟精度比传统的 MCE 方法模拟精度高, 为复杂城市系统的演变模拟提供了一种十分有用的工具。

用支持向量机确定城市 CA 的非线性转换规则时, 支持向量机参数的确定是非常重要的, 本文用传统的分组交叉试验的方法确定, 如何在模型中优化组合支持向量机的参数, 以后需要更深入的研究。同时, 交通网的动态扩张对模型模拟的结果有非常重要的影响, 如何动态模拟交通网的扩张并将其对城市发展的动态影响反映在 CA 模型中, 也有待进一步研究。

## 参 考 文 献 (References)

- [1] Batty M, Xie Y. From Cells to Cities [J]. Environment and Planning B: Planning and Design, 1994, 21: 531—548.
- [2] Ward D P, Murray A T, Phinn S R. A Stochastically Constrained Cellular Model of Urban Growth [J]. Computers Environment and Urban Systems, 2000, 24: 539—558.
- [3] Zhou C H, Sun Z L, Xie Y C. Geocellular Automata [M].

- Beijing Science Press 1999. [周成虎, 孙战利, 谢一春. 地理元胞自动机研究 [M]. 北京: 科学出版社, 1999.]
- [4] Batty M, Xie Y. Possible Urban Automata [J]. *Environment and Planning B: Planning and Design*, 1997, 24: 175—192.
- [5] White R, Engelen G. Cellular Automata as the Basis of Integrated Dynamic Regional Analysis [J]. *Environment and Planning B: Planning and Design*, 1997, 24: 235—246.
- [6] Takeshi A, Tetsuya A. Empirical Analysis for Estimating Land Use Transition Potential Functions: A Case in the Tokyo Metropolitan Region [J]. *Computers Environment and Urban Systems* 2004, 28: 65—84.
- [7] Nijs T C, Niet R D. Crommentijn. Constructing Land use Maps of the Netherlands in 2030 [J]. *Journal of Environmental Management* 2004, 72: 35—42.
- [8] Silva E A, Clarke K C. Calibration of the SLEUTH Urban Growth Model for Lisbon and Porto, Portugal [J]. *Computers Environment and Urban Systems* 2002, 26: 525—552.
- [9] Clarke K C, Hoppen S, Gaydos L. A Self-modifying Cellular Automaton of Historical Urbanization in the San Francisco Bay Area [J]. *Environment and Planning B: Planning and Design*, 1997, 24: 247—261.
- [10] White R, Engelen 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.
- [11] Li X, Yeh A G O. Constrained Cellular Automata for Modeling Sustainable Urban Forms [J]. *Acta Geographica Sinica*, 1999, 54 (4): 289—298. [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.]
- [12] Wu F. Calibration of Stochastic Cellular Automata: the Application to Rural-urban Land Conversions [J]. *International Journal of Geographical Information Science* 2002, 16 (8): 795—818.
- [13] 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 (8): 1501—1518.
- [14] Li X, Yeh A G O. Neural network based Cellular Automata for Realistic and Idealized Urban Simulation [J]. *Acta Geographica Sinica*, 2002, 57 (2): 159—166. [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.]
- [15] Li X, Yeh A G O. Mining Cellular Automata's Transition Rules [J]. *Science in China Ser D Earth Science* 2004, 34 (9): 865—872. [黎夏, 叶嘉安. 知识发现及地理元胞自动机 [J]. *中国科学 D 辑, 地球科学*, 2004, 34 (9): 865—872.]
- [16] Yang Q S, Li X. Mining Transition Values for Geo simulation Using Rough Set [J]. *Acta Geographica Sinica*, 2006, 61 (8): 882—894. [杨青生, 黎夏. 基于粗集的知识发现与地理模拟 [J]. *地理学报*, 2006, 61 (8): 882—894.]
- [17] Wu F, Webster C J. Simulation of Land Development Through the Integration of Cellular Automata and Multicriteria Evaluation [J]. *Environment and Planning B: Planning and Design*, 1998, 5: 103—126.
- [18] Wu F L. SimLand: A Prototype to Simulate Land Conversion Through the Integrated GIS and CA with AHP derived Transition Rules [J]. *International Journal of Geographical Information Science* 1998, 12 (1): 63—82.
- [19] Li X, 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.
- [20] Li X, Yeh A G O. Spatial Decision and Urban Simulation Based on Principal Component Analysis and CA Model [J]. *Science in China Ser D Earth Science* 2001, 31 (8): 683—690. [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 [J]. *中国科学 (D 辑), 地球科学*, 2001, 31 (8): 683—690.]
- [21] Deng N Y, Tian Y J. A New Way of Data Mining Support Vector Machine [M]. Beijing Science Press 2004. [邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量机 [M]. 北京: 科学出版社, 2004.]
- [22] Vapnik V N. *Statistical Learning Theory* [M]. John Wiley & Sons Inc New York 2004. [许建华, 张学工译. 统计学习理论 [M]. 电子工业出版社, 2004.]
- [23] Tom M. Mitchell. *Machine Learning* [M]. McGraw Hill Companies Inc 2003. [曾华军, 张银奎等译. 机器学习 [M]. 北京: 机械工业出版社, 2003.]
- [24] Dong B, Cao C, Lee S E. Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region [J]. *Energy and Buildings* 2005, 37: 545—553.
- [25] Cherkassky V, Ma Y. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression [J]. *Neural Networks* 2004, 17: 113—126.
- [26] David V, A Sanchez. Advanced Support Vector Machines and Kernel Methods [J]. *Neurocomputing*, 2003, 55: 5—20.
- [27] Chang C C, Lin C J. Training  $\nu$  support Vector Classifiers: Theory and Algorithms [J]. *Neural Computation*, 2001, 13 (9): 2119—2147.
- [28] Ana M B, Dragana N, Leopold C. Probabilistic SVM Outputs for Pattern Recognition Using Analytical Geometry [J]. *Neurocomputing* 2004, 62: 293—303.