

基于时空 Logistic 回归模型的漳州城市扩展预测分析

杨云龙, 周小成, 吴 波

(福州大学福建省空间信息工程研究中心, 空间数据挖掘与信息共享教育部重点实验室, 福州 350002)

摘要: 本文提出一种以时空 Logistic 回归模型来预测城市扩展的新方法。其首先在传统 Logistic 回归模型中加入空间自相关结构构建空间 Logistic 回归模型, 然后, 利用漳州市区近 20 年(1989 - 2009 年)的数据, 建立不同时期城市扩展模拟的多个子空间 Logistic 回归模型 M_i , 再采用一次平滑指数法综合处理这些时间序列的 M_i , 构建出顾及空间复杂性和时间序复杂性的时空 Logistic 回归预测模型。新方法一方面克服了传统 Logistic 回归模型法受限于预测年份影响因素数据难以获取的缺点, 另一方面由于模型考虑了城市扩展的长时间序列复杂性, 即综合了城市扩展不同时期影响因素不同的情况, 使它更接近城市扩展的实际, 因而预测精度会提高。以福建省漳州市区为例, 分别运用传统 Logistic 回归模型方法, 在传统 Logistic 回归模型中单独加入空间自相关结构的时空 Logistic 回归模型法和基于时空 Logistic 回归模型的新方法这 3 种方法, 对 2009 年城市扩展进行了预测分析。结果表明, 基于时空 Logistic 模型的新方法比传统 Logistic 回归模型法和空间 Logistic 回归模型法的预测精度都要好, 总体预测精度分别为 81.02%、83.82% 和 87.00%, 预测城市用地的精度从 63.59% 提高到 67.35% 和 73.34%, ROC 曲线下的面积 AUC 从 0.826 提高到 0.883 和 0.924。

关键词: 空间 Logistic 回归模型; 扩展预测; 指数平滑法; 空间自相关; 漳州市

DOI: 10.3724/SP.J.1047.2011.00374

1 引言

城市空间扩展是一个长期而复杂的过程, 从土地利用转换的角度来看, 在这个过程中将导致大规模以自然为主的土地覆盖转变为以人工构筑物为主的土地利用结果, 这种发展变化是自然、社会等多因素共同作用的结果。而一个城市在其发展过程中的某一特定时期, 总存在着一些主要的影响因子, 主导着城市空间扩展的速度和方向^[1], 因此对城市扩展过程的描述、模拟及预测成为可能。目前, 国内外学者提出了各种土地利用变化模型来分析理解城市扩展的变化过程, 主要方法有线性回归^[2]、Logistic 回归、元胞自动机(CA)^[3-5]、系统动力模型和 CLUE, 以及 CLUE-S^[6]等, 这些方法在城市扩展变化中得到了较广泛的应用。由于 Logistic 回归模型简单且自变量没有正态假设等优点, 国内外许多学者利用 Logistic 回归模型结合 GIS 技术在土地利用分布和城市扩展预测方面做

了许多研究工作^[7-8], 其效果已得到了公认。然而由于海量数据和计算的复杂性, 目前, Logistic 回归模型在土地利用空间分布和城市预测应用中很少考虑到空间相关和时间序列的特性。本文提出一种时空 Logistic 回归模型对城市扩展进行预测的新方法。

2 城市扩展预测方法与技术路线

2.1 空间 Logistic 回归模型

传统 Logistic 回归模型是城市扩展模拟预测研究中常用的一种方法。其一般是将研究区细分为许多栅格单元, 模型的目标变量(城市用地类型)由二分类变量数据表示, 1 表示城市用地出现, 0 表示不出现, 模型的解释变量(驱动因素)通过一系列自然和社会经济因素来描述。假设某一个栅格单元出现为城市用地类型的概率为 P , 这种概率可以用 Logistic 函数计算表示, 其表达式为:

收稿日期: 2010-12-04; 修回日期: 2011-02-06.

基金项目: 福建省自然科学基金项目(2009J01213); 中国科学院城市环境与健康重点实验室(城市环境研究所)开放基金课题(KLUEH201002); 福州大学科技项目基金(022349)资助。

作者简介: 杨云龙(1982-), 男, 广西桂林人, 硕士研究生。主要研究方向为环境与资源遥感应用研究。

E-mail: yangyunlong9643@163.com

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)} \quad (1)$$

Logistic 函数是协变的非线性函数,为求回归系数,对(1)式进行 Logit 变换,得到一个线性公式:

$$Y = \ln[P/(1-P)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (2)$$

其中, P 表示事件发生的概率, x_1, x_2, \cdots, x_k 为自变量,参数 $\beta_0, \beta_1, \cdots, \beta_k$ 为待求的回归系数。基于传统 Logistic 回归模型方法的漳州城市扩展模拟模型就是根据公式(2)的 Binary Logistic 回归方程构建而成的。

式(2)中如果 Y 不但与变量因子 X 有关,而且受 Y 本身周围地块分布的影响。则(2)应改写为:

$$Y = X\beta + \rho WY + \epsilon \quad (3)$$

式中, ρ 为空间自相关参数, W 是 $m \times m$ 空间权重系数矩阵。

由于空间权重系数 W 矩阵数据量通常很巨大,例如,预测一个大小为 $N=723 \times 398$ 的区域,将需要构造一个元素大小为 8.28×10^{10} 的权重系数矩阵。这说明在模型中直接考虑空间自回归结构是很困难的,必须考虑其他的办法,采用空间滤波技术^[7,9]可以近似计算空间 logistic 回归模型,将(3)式改写为:

$$Y = (I - \rho W) - 1X\beta + (I - \rho W) - 1\epsilon = (QX)\beta + Q\epsilon \quad (4)$$

预测因子从 X 到 QX 可以看成是空间滤波操作,其中, $Q=(I-\rho W)-1$

这种方法通过引进一个隐含变量并改变模型中的输入变量来消除模型中的空间自相关性,它使在新的自变量与因变量基础上建立的因果关系中不存在空间自相关性,从而把空间 Logistic 回归模型转变为普通 Logistic 回归。采用空间平滑滤波操作分别得出输入变量 Y 、原始自变量 x_i 的隐含变量 Y^* 和 x_i^* 。最后,由 Y^*, x_i^* 和 Y 建立的关系式:

$$Y = X^* \beta + \alpha Y^* + \epsilon \quad (5)$$

式(5)中 $X^* = [x_1^*, x_2^*, \cdots, x_k^*]^T$, 则(5)式变量之间不再存在空间自相关,用普通逻辑回归方法即可计算。需要明确的是,为了使模型能进行 3 年以上较远时间情况的空间预测,则只能对 4 年以前状态的 Y_0 进行空间平滑,而不能对因变量本身滤波。

2.2 指数平滑法

指数平滑法是在移动平均法的基础上发展起

来一种特殊的加权移动平均法。常用的平均法进行预测时需贮存多个实际值,当预测项目很多时,就要占据相当大的预测空间;其次,往往对最近的几个实际值等值来看待,并对 $t-n$ 期以前的数据完全不考虑。为了弥补这些缺点,美国学者布朗提出了指数平滑法,它是通过对预测目标历史统计序列的逐层的平滑计算,消除由于随机因素造成的影响,找出预测目标的基本变化趋势并以此预测未来。指数平滑法分为一次指数平滑法和多次指数平滑法,它们的基本思想都是:预测值是以前观测值的加权,且对不同的数据给予不同的权,新数据给较大的权,旧数据给较小的权。本文中采用一次指数平滑法,公式为 $S_{t+1} = \alpha X_t + (1-\alpha)S_t$, 其中 S_{t+1} 是时间 $t+1$ 的预测值, S_t 是时间 t 的平滑值, X_t 是时间 t 的实际值, α 是平滑系数。使用一次指数平滑进行预测的关键两点是:第一,初始值的确定,即第一期的预测值;一般可以选用第一期的实际观察值或前几期实际观察值的均值作为初始值。第二,平滑系数 α 的选择;由于平滑系数 α 的大小对预测值影响很大,所以, α 的选择很关键,一个总的原则是使预测值与实际观察值之间的均方误差 MSE 最小。从理论上讲, α 取 0-1 之间任意数都可以,但为了提高预测精度可运用以下三种方法^[10]来确定其值:(1)理论计算法;(2)经验判断法;(3)试算法。无论如何确定 α 值,预测者都要对预测对象的变化规律作出定性判断且计算预测误差。在实际应用中还要考虑到预测灵敏度和预测精度是相互矛盾的,必须给予二者一定的考虑,采用折衷的 α 值。

2.3 技术路线

利用城市地块单元空间分布状况对城市扩展情况进行模拟预测,关键就是要计算得到预测年份的城市用地类型情况的概率分布图。因此,对城市扩展进行模拟预测,传统 Logistic 回归模型法是:①利用预测年份之前较近年份数据建立 Logistic 回归模型;②将预测年份影响力数据带入已建立好的 Logistic 模型中,计算出地块单元的空间分布概率图。本文以 2007 年的数据建立模型对 2009 年进行了预测。而如果在传统 Logistic 回归模型法中考虑空间自相关结构则成为空间 Logistic 模型预测法。基于时空 Logistic 回归模型的新预测方法是:①利用预测年份之前的多期数据分别建立不

同年份的子空间 Logistic 回归模型 M_i , 运用一次指数平滑法综合处理 M_i 构建出时空 Logistic 回归预测模型;②根据预测模型中的 M_i , 利用遥感处理软件 REDAS IMAGINE9.2 的空间建模工具分别模拟出各个时期的城市用地概率分布图;③将时空 Logistic 回归模型转换为对这些城市用地概率分布

图进行综合处理,直接平滑得到一个城市用地类型的预测概率分布图。

而后设定 $P=0.5$ 作为城市用地和非城市用地类型的阈值,对城市扩展预测。两种方法的技术路线见图 1,图 2。



图 1 传统 Logistic 回归模型预测方法的分析途径

Fig. 1 Traditional forecasting methods of logistic regression model

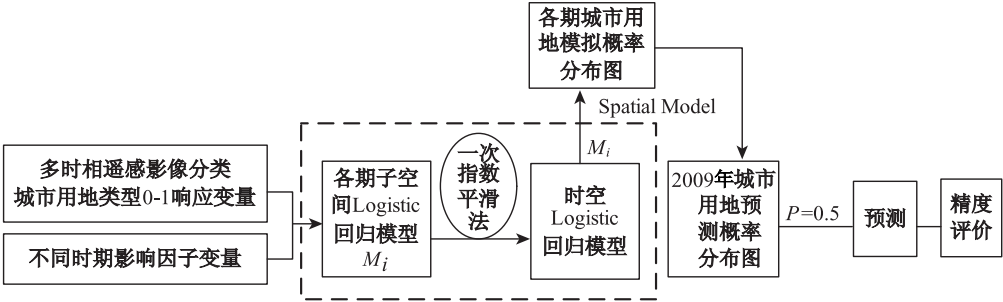


图 2 新预测方法的技术路线

Fig. 2 New forecasting methods of spatiotemporal logistic regression model

3 漳州市城区扩展的模型预测实例分析

漳州市位于福建省南部,地处东经 $117^{\circ} \sim 118^{\circ}$ 、北纬 $23.8^{\circ} \sim 25^{\circ}$ 之间,地形西北高东南低,其中,以山地丘陵为主,约占总面积的 70%,漳州市坐落在漳州平原,流经境内最大的河流是九龙江。漳州同厦门、汕头经济特区接壤,与厦门、泉州合称“闽南三角地区经济开发区”,是闻名于世的鱼米之乡、花果之城,也是著名的侨乡和台湾祖居地。漳州市城区辖两个区芗城区、龙文区,常住人口约为 50 万,面积约 47km^2 。本文以漳州市区作为研究区域。

3.1 数据的分析处理

(1)响应变量和影响因子变量提取

利用 Logistic 模型建立城市用地空间分布与影响因子的定量关系,势必涉及响应变量和影响因子变量,其中响应变量为城市用地,使用多时相遥感数据分类来获取,源数据包括 1989、1998 和 2000

年三期 TM 影像,2003 年 ASTER 影像,2005 年中巴资源卫星影像,2007 年和 2009 年两期 ALOS 影像。对此,为减少计算量,将分类精度统一调整为 $30\text{m}(723 \times 398)$ 。由于本文只考虑城市用地和非城市用地的空间分布,因此,将工业用地、居民用地、交通用地、开发建设用地、商业用地等用地类型都归为城市用地,用代码 1 标示(红色);而水体、植被、耕地、裸地(岩)、沙地(滩)等划分为非城市用地,用代码 0 标示(白色)。其中,1989、2009 年城市用地分类图见图 5(a)、5(b)。先通过人为选取合理的影响因子变量纳入模型中,影响因子变量,包括统计变量、距离变量和地形变量。变量因子及其描述如表 1 所示。人口密度制作主要通过格网法生成^[11];交通(包括主要道路、铁路)、河流矢量 shape-file 数据通过遥感影像提取,距离变量通过 ARC-GIS 平台 Distance 工具获取。影响因子变量与分类数据配准后,全部栅格化,分辨率为 $30\text{m} \times 30\text{m}$ 。

(2)影响因子空间滤波操作

空间数据转换时需先确定空间滤波窗口的大小,故此采用半方差的方法来自动确定移动窗口的

大小,并分别统计了 1989、1998、2000、2003 年各期的 Y 值在 30 个网格内的半方差值的分布情况,而后再用一个指数函数 $f(x)=b[1-\exp(-x/a)]$ 进行散点拟合,如图 3。以 1989 年为例,求出 $a=4.085, b=0.052$ 。因此,半径大小确定为 4,移动窗口大小为 9×9 。其他年份类似处理,各期 $a、b$ 值统计如表 2。

表 1 城市用地影响因子变量

Tab. 1 Impact factor variables of urban land use

类别	变量因子	变量描述
统计变量	Ppdens	人口密度
	Planning	虚拟变量,规划数据
距离变量	Adis_Dist	连续变量,到城市行政中心的距离
	Com_Dist	连续变量,到城市 CBD 的距离
	Htp_Dist	连续变量,到城市主要医院的距离
	Indus_Dist	连续变量,到工业中心的距离
	Park_Dist	连续变量,到城市公园的距离
	River_Dist	连续变量,到主要河流的距离
	Scl_Dist	连续变量,到主要学校的距离
	Trp_Dist	连续变量,到铁路和主要公路距离
地形因子	DEM	连续变量,地形高程

表 2 空间滤波窗口大小系数表

Tab. 2 Spatial filtering window size coefficients

年份	1989	1998	2000	2003
a	4.085	4.209	4.176	4.069
b	0.052	0.083	0.088	0.090

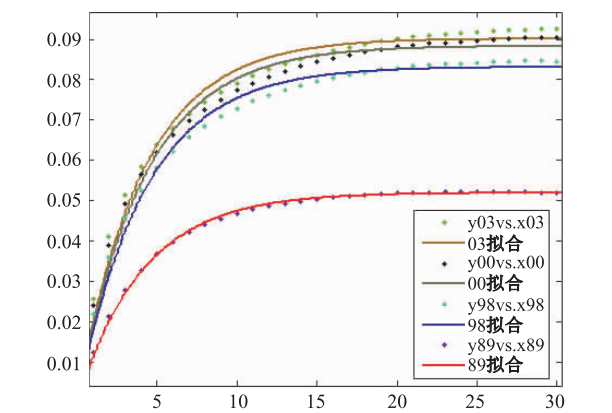


图 3 因变量 Y 的半方差及其拟合图

图 3 Semi-variance of the dependent variable Y and the fitting chart

确定了移动窗口半径后,计算出 \tilde{Y} 和 \tilde{x}_i ,并假设 $\rho=\mu_i$,则可以求出各期变化后的 X^* 和 Y^* 作为逻辑模型的输入变量,利用极大似然法解算各期逻辑回归模型。

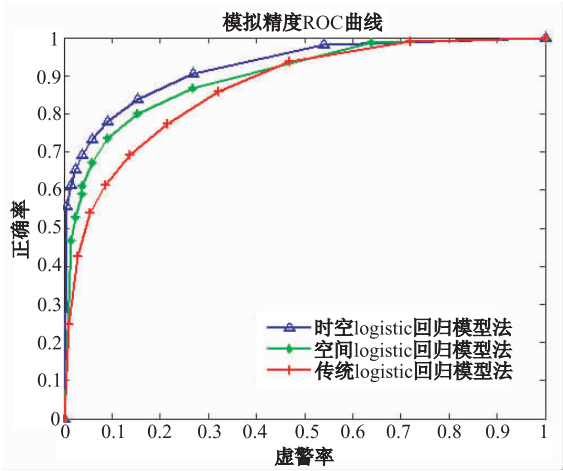


图 4 比较 3 种方法的 ROC 曲线

Fig. 4 ROC curve comparison of the three methods

(3)影响因子数据归一化处理

将各期影响因子数据用 min-max 方法进行归一化处理。

(4)随机点数据提取

由于 Logistic 回归模型的结果与样本选择关系很大,因此,需要保证数量足够多,模式多样化的训练样本。采用分层随机取样的模式选取了 22000 个样本(其中 10000 个为 1,12000 个为 0)进行模型拟合。这种取样方式能够保证变化与不变化大致相同的预测精度。

(5)影响因子多元共线性检验

由于逻辑模型对多元共线性非常敏感,因此,首先需要选择部分良好的因子变量消除自变量之间的多元共线性。通常采用指标 $(TOL)=1-R_{i2}$ 和 $(VIF)=1/TOL$ 来确定变量之间的线性关系。如果某个变量的 $VIF>5.0$ 或者 $TOL<0.2$,则表明该变量与其他变量之间存在重要的多元共线性^[12]。利用 SPSS 统计软件计算各期的影响因子,VIF 统计结果如表 3,删除 $VIF>5.0$ 的变量,保留其他变量作为下一步拟合的影响因子。

3.2 模型的建立

利用 SPSS 中 Binary Logistic 将各期响应因子与经过筛选的影响因子进行拟合计算,以显著性 $\text{sig}<0.05$ 为标准,删除不符合要求的变量,各期最终保留变量 Logistic 回归模型(“—”表示该变量对没有通过多元共线性和显著性检验,对模型无效)如表 4。

表 4 Binary Logistic 回归分析与评价结果 (Variables in the Equation)
Tab. 4 Binary Logistic regression analysis(Variables in the Equation)

年份	1998		2000		2003		2005		2007	
	β	sig	β	sig	β	sig	β	sig	β	sig
Ppdens	6. 775	0. 0	7. 521	0. 0	3. 612	0. 0	3. 422	0. 0	3. 881	0. 0
Planning	—	0. 0	1. 501	0. 0	1. 438	0. 0	1. 263	0. 0	1. 604	0. 0
Adis_Dist	—		—		−0. 732	0. 001	—		—	
Com_Dist	—		—		—		—		—	
Htp_Dist	−3. 387	0. 0	−3. 45	0. 0	—		—		—	
Indus_Dist	—		—		—		—		−1. 794	0. 0
Park_Dist	—		—		—		—		—	
River_Dist	1. 267	0. 0	0. 796	0. 0	0. 985	0. 0	1. 858	0. 0	2. 240	0. 0
Scl_Dist	—		—		—		—		—	
Trp_Dist	−1. 134	0. 0	—		−2. 618	0. 0	−3. 709	0. 0	−2. 884	0. 0
DEM	−3. 756	0. 0	−5. 678	0. 0	−12. 353	0. 0	−8. 763	0. 0	−7. 353	0. 0
Y*	7. 939	0. 0	7. 861	0. 0	9. 693	0. 0	7. 202	0. 0	7. 638	0. 0
常量	−1. 27	0. 0	−1. 299	0. 0	−2. 299	0. 0	−1. 104	0. 0	−1. 257	0. 0

Logistic 回归模型评价		1998 年	2000 年	2003 年	2005 年	2007 年
HL	Chi-Square	355. 115	356. 647	136. 880	101. 970	118. 666
	d. f.	8	8	8	8	8
	significance	0. 000	0. 000	0. 000	0. 000	0. 000
2LL		14075. 961	14327. 629	7133. 967	13294. 180	13130. 988

从表 4 中可以看出,首先,不同子模型显著性(significance)都为零,表明模型与各系数统计显著,因而,各子空间 Logistic 模型都有效。其次,各个时期影响因子不尽相同,其中,人口密度和地形高程影响力较大,且人口密度系数为正,说明人口密度是影响城市用地分布的主要正因素。地形高程因素为负因素,说明城市用地主要集中在地势比较平坦的地区,这符合漳州城市在漳州平原发展的实际。另外,城市规划因子为正值,说明它是城市扩展的正驱动力,规划区域相对其他地方会优先发展。所有交通距离为负因子,表明距离主要道路铁路的交通距离越近,路上交通运输成本越低,越容易促进该地区向城市用地转换。此外,到河流距离为正驱动力,说明距离河流越远地块转化为城市用地的可能性越大,因为城市要距河流一定距离以避免洪水的影响。

其分别用 M_{98} 、 M_{00} 、 M_{03} 、 M_{05} 、 M_{07} 表示 1998 年、2000、2003、2005、2007 年各年份的子模型,以

1998 年为例(其他类似处理),建立模型的结果如下:

$$M_{98} = \text{EXP}(-1.27 - 3.756 * \text{DEM} - 3.387 \times \text{Htp_Dist} + 6.775 \times \text{Ppdens} + 1.267 \text{River_Dist} - 1.134 \times \text{Trp_Dist} + 7.939 \times Y^*) / (1 + \text{EXP}(-1.27 - 3.756 \times \text{DEM} - 3.387 \times \text{Htp_Dist} + 6.775 \times \text{Ppdens} + 1.267 \times \text{River_Dist} - 1.134 \times \text{Trp_Dist} + 7.939 \times Y^*))$$

得到不同时期的子空间 Logistic 回归模型 M_{98} 、 M_{00} 、 M_{03} 、 M_{05} 和 M_{07} 后,利用一次指数平滑法,以 M_{98} 为初始值、 α 为平滑系数,就可建立预测 2009 年的时空 Logistic 回归模型,记为 M_{09} ,则有:

$$M_{09} = \alpha * M_{07} + \alpha * (1 - \alpha) * M_{05} + \alpha * (1 - \alpha)^2 * M_{03} + \alpha * (1 - \alpha)^3 * M_{00} + (1 - \alpha)^4 * M_{98}$$

时空 Logistic 回归预测模型中每一个 M 实质

能计算栅格单元在一定影响因子数据下呈现的相对概率,这一过程可通过 ERDAS IMAGINE 软件中的空间建模工具 Spatial Modeler 建模运算实现。利用 M_i 模拟出不同年份城市用地概率分布图 P_i (分别记为: $P_{98}, P_{00}, P_{03}, P_{05}, P_{07}$), 这样时空 Logistic 回归预测模型转换为直接对城市用地的空间分布概率图 P_i 的平滑。平滑得到预测 2009 年的城市用地空间分布概率分布图, 记为 P_{09} :

$$P_{09} = \alpha * P_{07} + \alpha * (1 - \alpha) * P_{05} + \alpha * (1 - \alpha)^2 * P_{03} + \alpha * (1 - \alpha)^3 * P_{00} + (1 - \alpha)^4 * P_{98}$$

对此,问题的关键是平滑系数的选择:本文以 1998、2000、2003、2005 年的城市用地概率分布图,分别选用 0.2 - 0.8 平滑出一个 2007 年预测图,然后与实际计算相比较,得到均方误差 MSE (Mean Squared Error) 的情况(如表 5)。

表 5 不同 α 值的均方误差(SME)

Tab. 5 Mean square errors (SME) for different α values							
	0.2	0.3	0.4	0.5	0.6	0.7	0.8
MSE	0.0292	0.0226	0.0177	0.0141	0.0114	0.0091	0.0073

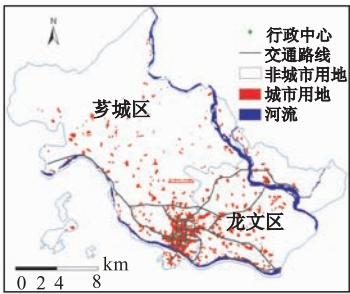
经过比较分析,当 $\alpha=0.8$ 时 MSE 最小,所以 α 的最佳值为 0.8。所以预测模型为:

$$P_{09} = 0.8 * P_{07} + 0.8 * (1 - 0.8) * P_{05} + 0.8 * (1 - 0.8)^2 * P_{03} + 0.8 * (1 - 0.8)^3 * P_{00} + (1 - 0.8)^4 * P_{98}$$

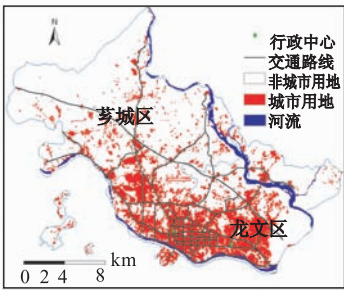
3.3 预测结果与精度评价

3.3.1 漳州城市扩展预测结果分析

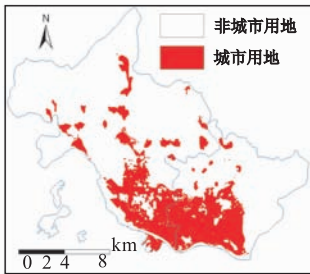
利用传统 Logistic 回归模型方法、空间 Logistic 回归模型法和时空 Logistic 回归模型方法,对漳州市城区 2009 年城市扩展情况进行了模拟预测,预测结果见图 5。



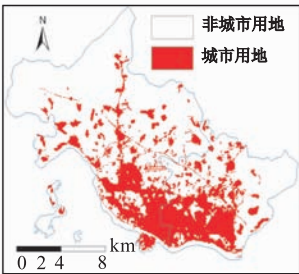
(a) 1989年实际城市用地(分类图)



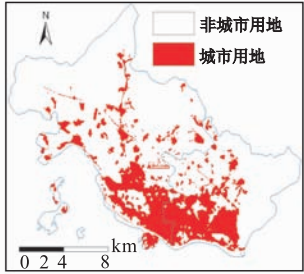
(b) 2009年实际城市用地(分类图)



(c)传统Logistic回归法预测09年结果图



(d)空间Logistic回归法预测09年结果图



(e)基于时空Logistic回归模型法预测09年结果图

图 5 城市扩展预测结果比较

Fig. 5 Comparison of urban growth prediction

a: 1989 urban land use classification map; b: 2009 urban land use classification map; c: Prediction for 2009 by traditional logistic regression model method; d: Prediction for 2009 by space logistic regression model method; e: Prediction for 2009 by spatiotemporal logistic regression model method.

图 5 可见,从 1989 - 2009 年,漳州城市扩展迅猛,面积增加了好几倍,其扩展特点^[13]: (1)城区主要沿九龙江河岸东西方向扩展,且东部龙文区扩展

面积明显大于芗城区,这是由于漳州市 1996 年新增龙文区,为城市化发展提供了充足的土地。(2)重点依托交通用地沿线的扩展,随着漳龙高速公路

的修建,以及市区内外多条道路的修建,降低了生活生产(运输)成本,使得沿交通干线的地块更易转为城市用地,说明交通的迅速发展是带动漳州市扩展的重要因素。(3)规划区域扩展,据“漳州市城市总体规划(2000年至2020年)”,漳州市的金峰工业区、蓝田工业区、朝阳工业区建设从2000年开始规划后就如火如荼地开展,市区主要是房地产业的填充式开发。(4)扩展受人口因素影响很大,1989年漳州市区人口为33.34万人,随后人口逐年增加,1998、2003、2005和2008年人口分别增加到49.05万、51.87万、53.36万和54.60万人。人口的增加,为城市建设带来了大量的劳动力,非农人口比例由1989年的53.35%增加到2003年的59.51%,推进着城市化的进程,导致大量农业耕地转变为城镇用地,从而使城市规模不断扩大。

3.3.2 预测精度评价

表 6 三种预测方法预测精度比较(概率阈值为 0.5)
Tab. 6 Prediction accuracy of the three methods and their comparison, with a probability threshold of 0.5

预测方法	混淆矩阵			
传统 Logistic 回归	1-1=62892	0-1 = 36007	灵敏度= 0.6359	特异度=0.9014
模型法预测	1-0 =18612	0-0= 170243	总精度= 0.8102	Kappa 系数= 0.5608Z
空间 Logistic 回归	1-1= 76982	0-1 = 37317	灵敏度= 0.6735	特异度= 0.9467
模型法预测	1-0= 9239	0-0 = 164216	总精度= 0.8382	Kappa 系数= 0.6474
时空 Logistic 回归模型的	1-1= 72532	0-1 = 26367	灵敏度= 0.7334	特异度= 0.9415
新方法预测	1-0= 11040	0-0 = 177815	总精度= 0.8700	Kappa 系数= 0.7008

表 6 统计了 3 种方法的预测精度,可以看出,时空 Logistic 回归模型的新方法预测精度最好。城市用地灵敏度的精度从 0.6359 提高到 0.6735 和 0.7334,总精度从 0.8102 提高到 0.8382 和 0.8700。kappa 系数也从 0.5608 提高到 0.6474 和 0.7008。这表明,基于时空 Logistic 回归模型的新方法比传统普通 Logistic 模型法和空间 Logistic 回归模型法预测精度有明显提高。ROC 曲线下面积(AUC)的大小作为模型预测准确度的衡量指标,其取值范围为[0, 1],值越大表示模型判断力越强。图 4 给出了 3 种预测方法的 ROC 曲线,经计算,相比传统 Logistic 回归模型预测方法和空间 Logistic 回归模型法,时空 Logistic 回归模型新方法 ROC 曲线下的面积 AUC 分别从 0.826、0.883 提高到 0.924,由此可知,新方法的预测精度最好。

预测实验效果评价包括目视效果评价和统计评价。对比漳州市区 2009 年实际城市用地扩展(图 5(b)),从目视效果看,传统 Llogistic 回归模型方法预测结果图总体上虽呈现了城市形态,但大都呈团状,细节区分能力较差;空间 Logistic 回归模型法有一定的细节区分能力,但部分地区预测结果多于实际扩展情况,这可能是在设置概率阈值 $p=0.5$ 偏小的结果;而基于时空 Logistic 回归模型的新方法预测结果图与实际最接近,且具有较好的细节区分能力。统计评价包括预测的位置、数量准确率等方面。位置精度的评价通常采用形状指数、空间相关系数、KAPPA 系数,以及 ROC 曲线等^[14-15]。利用预测的 2009 年城市用地与遥感影像实际分类的结果(图 5(b))逐个网格相比较,统计出预测正确与错误的网格数目。

4 结 论

传统 Logistic 回归模型在城市扩展预测中,一方面未考虑空间关系,会降低模型的解释能力,进而影响到预测精度;另一方面,模型预测也受到预测年份影响因子数据难以获取的限制。而加入空间自相关关系的 Logistic 回归模型,则更符合影响城市用地空间分布的实际,因而将会使模型的预测能力大幅提高。城市发展过程中,不同时期影响的主要因素也有所不同,通过建立不同时期的子空间 Logistic 回归模型,再利用指数平滑法综合这些时间序列的模型建立时空 Logistic 回归模型的新方法。其对城市扩展预测,一则使其不再受到预测年份影响因子数据的限制;另外,使它更接近城市发展的本身特点,故此提高了预测精度。

参考文献:

- [1] 姚士谋,帅江平. 城市用地与城市生长[M]. 合肥:中国科学技术大学出版社,1995.
- [2] Lopez E, Bocco G, Mendoza M, Duhau E. Predicting Land Cover and Land Use Change in the Urban Fringe: A Case in Morelia City, Mexico[J]. Landscape and Urban Planning, 2001, 55 (4): 271 - 285.
- [3] Li X and Yeh A G O. Neural-network-based Cellular Automata for Simulating Multiple Land Use Changes Using GIS[J]. International Journal of Geographical Information Science, 2002, 16(4): 323 - 343.
- [4] Clarke K C, Gaydos L J, Hoppen S. A Self-modified Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area[J]. Environment and Planning B, 1997, 24:247 - 261.
- [5] Verburg P H, van Eck J R R, de Nijs T C M, *et al.* Determinants of Land-use Change Patterns in the Netherlands[J]. Environment and Planning B: Planning and Design, 2004, 31:125 - 150.
- [6] Verburg A, Soepboer P H, Veldkamp W, *et al.* Modeling the Spatial Dynamics of Regional Land Use: The CLUE-SModel[J]. Environmental Management, 2002, 30(3):391 - 405.
- [7] Bo Huang, Li Zhang, Bo Wu. Spatiotemporal Analysis of Rural-urban Land Conversion[J]. International Journal of Geographical Information Science, 2009, 23 (3): 379 - 398.
- [8] Wu F, Yeh A G. Changing Spatial Distribution and Determinants of Land Development in Chinese Cities in the Transition from a Centrally Planned Economy to a Socialist Market Economy: A Case Study of Guangzhou [J]. Urban Studies, 1997, 34(11):1851 - 1879.
- [9] 吴楷钊,吴波. 基于空间相关的逻辑回归模型的城市扩展模拟[J]. 河南大学学报(自然科学版), 2010, 40 (3):267 - 273.
- [10] 张德南,张心艳. 指数平滑预测法中平滑系数的确定[J]. 大连铁道学院学报, 2004, 25(1):79 - 80.
- [11] 王英安,岳天祥. 基于格网生成方法的山东省人口密度空间分布模拟[J]. 曲阜师范大学学报, 2004, 30 (3):91:93.
- [12] 王济川,郭志刚. Logistic 回归模型方法与应用[M]. 北京:高等教育出版社,2001,33 - 55.
- [13] 周小成,汪小钦,吴波,等. 城镇扩张的多源遥感图像动态监测分析[J]. 地球信息科学, 2008, 10 (3): 332 - 337.
- [14] Pontius R G Jr, Schneider L C. Land-cover Change Model Validation by an ROC Method for the Ipswich Watershed, Massachusetts, USA[J]. Agriculture, Ecosystems and Environment, 2001, 85,239 - 248.
- [15] Lantz C A, Nebenzahl E. Behavior and Interpretation of Factors Affecting the Performance of Climatic Envelope Models[J]. Journal of Clinical Epidemiology, 1996, 49:431 - 434.

Urban Expansion Prediction for Zhangzhou City Based on GIS and Spatiotemporal Logistic Regression Model

YANG Yunlong, ZHOU Xiao Cheng, WU Bo

(Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou 350002, China)

Abstract: We start this study aimed at building a new method of spatiotemporal logistic regression model to predict urban expansion. This method first established a space Logistic regression model by adding autocorrelation structure based on the traditional logistic regression model, then built the multiple sub-space Logistic regression model M_i of urban growth simulation of different stages by Zhangzhou City's nearly 20 years (from 1989 to 2009) data. After this work, a spatiotemporal logistic regression model which took into account the spatial complexity and temporal complexity was constructed by using single exponential smoothing to treat these time series sub-model synthetically. On one hand, this novel method has over-

come the traditional shortcoming that the influencing factor data is difficult to obtain in the prediction year, on the other hand, the model considers the complexity of urban growth in the long time series, that is a combination of urban expansion in different periods of different factors situation, bring it closer to the actual urban expansion, which will improve the prediction accuracy. Zhangzhou City of Fujian Province was taken as an example in the study, and the urban expansion in 2009 was forecast by using three methods, i. e. traditional logistic regression model, space logistic regression model and spatiotemporal logistic regression model. The results showed that the prediction accuracy of the new method based on spatiotemporal logistic regression model was more better than others, for which the overall prediction accuracy were 81.02%, 83.82% and 87.0% respectively, and the sensitivity of urban land use prediction increased from 63.59% to 67.35% and 73.3%. Area AUC under the ROC curve raised in size from 0.826 to 0.883 and 0.924.

Key words: Spatial Logistic Regression Model; extended forecast; exponential smoothing; spatial autocorrelation; Zhangzhou City