

土地利用变化预测 CBR 方法的适应性分析

孙晔然^{1,2}, 杜云艳¹, 苏奋振¹, 周成虎¹

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101;

2. 中国科学院研究生院, 北京 100049)

摘 要:研究土地利用变化的方法很多,常见的有统计回归方法、概率统计方法、系统动力学方法、元胞自动机方法以及贝叶斯网络方法等。这些方法都各具特点,并得到了实际应用。虽然有些研究指出了一些方法的优点和不足,但是对其在土地利用变化问题上的适应性研究还比较有限。案例推理(Case Based Reasoning, 简称 CBR)作为研究土地利用变化的一种新方法,目前同样缺乏适应性方面的研究。因此,本文在土地利用的 CBR 方法的研究基础上,具体探讨该方法在土地利用变化预测上的适应性问题。研究分别从案例的选取方式、模型指标的选取以及指标权重的设定 3 个方面开展,通过对比试验以深入讨论 3 个方面的因素对预测精度的影响。研究结果显示,在历史案例丰富的情况下,通过选择重要的指标,并对其赋予较高权重,可以保证 CBR 方法在预测土地利用变化时的稳定性。此外, CBR 中“地理环境”组分的引入,有效地提高了土地利用变化预测的精度。研究表明, CBR 在解决土地利用变化问题上具有简单灵活、适用范围广、预测精度高以及保持形态稳定的特点,是一种解决地学问题的新方法。

关 键 词:案例推理(CBR);土地利用变化;三元组模式

1 引言

土地利用/土地覆被变化(Land Use/Land Cover Change, LUCC)问题已经成为全球环境变化的热点问题,国内外众多学者对此进行过探讨^[1-2]。近年来,随着 GIS 和 RS 技术的发展,出现了许多研究 LUCC 的模型和方法,常见的有统计回归、概率统计、系统动力学、元胞自动机以及贝叶斯网络等。史培军等^[3]对深圳市土地利用进行了测量,利用分类结果比较法得出了深圳市土地利用变化的转移矩阵,据此分析了深圳市 15 年来土地利用变化的空间过程;摆万奇^[4]采用系统动力学方法,将主要驱动力与土地利用类型置于统一的系统中,考察了深圳市土地利用的长期动态趋势;郭玉燕等^[5]利用马尔柯夫模型对南昌市的土地利用动态变化及演变趋势预测进行了分析,直观地反映出土地利用结构变化的程度;黎夏等^[6-8]利用 CA 模型模拟了广州地区的城市发展,通过多种方法来获得 CA 的转化规则,并将 CA 应用于城市的可持续发展规划。

上述方法具有各自的优点,但也存在着局限性。如贝叶斯网络方法计算开销较大,确定合理的前验密度比较困难,往往对如何判定实际问题是否

满足所要求的假设,缺乏现成的规则^[9]。与这些方法相比, CBR 方法具有简单灵活、适用范围广以及保持形态稳定的特点,一些学者对其在土地利用变化中的应用进行了探讨。Li 和 Yeh^[10]以珠江三角洲为例,利用 CBR 方法进行亚热带土地利用变化检测,并通过分层随机采样来收集光谱值不同的同类地物,以此克服遥感分类中的“同物异谱”现象。试验的结果显示出 CBR 方法分类的效果好于监督分类。Li 和 Liu^[11]在模拟珠江三角洲城市演变时,为了克服复杂区域中 CA 获取转换规则的困难,提出了基于 CBR 的 CA 模型,利用 CBR 来获取 CA 的转换规则,并对 k-邻近算法进行了改进,使其能反映转换规则的时空动态性。实验结果显示,与常规的基于规则的 CA 模型相比,基于 CBR 的 CA 模型,在复杂地区,可以获得更接近实际的模拟结果。Chen 等^[12]以北京为例,利用 CBR 方法对多时相 SAR 影像进行分类,以 2004 年的土地利用图构建案例库,对 2005 年的影像进行分类,得到 2005 年的土地利用图,整体的精度达到了 80%。针对传统的变化检测方法过于依赖于像元的光谱信息而忽视对象的空间信息的情况, Li 等^[13]以空间对象为基本单位进行土地利用的变化监测试验,研究中将面向对象分

收稿日期:2011-01; 修订日期:2011-04.

基金项目:国家 863 计划重点项目(2009AA12Z148);国家自然科学基金重点项目(088RA400SA)。

作者简介:孙晔然(1987-),男,硕士研究生,主要研究方向为 GIS 和空间数据挖掘。E-mail: sunyr@lreis.ac.cn

通讯作者:杜云艳, E-mail: duyuan@lreis.ac.cn

析和 CBR 方法进行有效的结合,通过对比验证,证明了该方法的有效性。

杜云艳等^[14-15]对传统的 CBR 模型进行了改进,提出了预测土地利用变化的 CBR 方法。该方法采用“问题-地理环境-结果”的三元组案例模式,定量地描述土地利用变化,利用相似性计算与推理预测土地利用的变化,并以珠海地区 1995-2000 年的土地利用变化为例开展了试验研究^[14-15]。从结果看, CBR 应用于土地利用变化的研究完全可行^[14-15];此外,研究中通过 CBR 与贝叶斯网络方法(BN)的对比分析发现, CBR 相对于 BN 方法具有简单灵活、使用方便、无需学习等独特优势^[14]。

但是,由于土地利用变化问题自身的复杂性,且受诸多因素的影响,研究^[14-15]无论是在方法本身还是在方法适用性上都有待于深入讨论。尤其需要深入地研究案例“三元组”定量表达指标的选取及权重大小设置的不同对最终结果的影响。再者,虽然目前国内外研究土地利用变化的模型和方法很多,都取得了不错的预测结果,但针对不同方法在解决问题的类型以及区域适应性方面还缺乏深入地讨论,亟待选择典型的方法进行分析研究。

因此,本文在研究^[14]的基础上,进行 CBR 方法在土地利用变化预测方面的适用性分析。通过案例的选取方式、“三元组”指标选取以及指标权重设置 3 个方面的对比试验,深入探讨该方法自身存在的规律及问题,进一步揭示 CBR 在解决土地利用变化预测问题上的适应性。

2 土地利用变化预测的 CBR 方法

2.1 方法概述

传统 CBR 的问题求解,是按照经典 4R 模型^[16],即检索(Retrieve)、重用(Reuse)、修改(Revise)和保存(Retain),由问题直接去寻找结果。杜云艳等^[14-15]针对土地利用变化问题的特点,提出了新的 CBR 模型。该模型对传统 CBR 进行了两点改进:①在土地利用变化案例的表达上,扩展了传统 CBR 二元组表达模式,引入了“问题-地理环境-结果”的三元组模式,增加了“地理环境”的描述,定量表达了空间环境对案例的影响^[14]。②在传统 CBR 方法进行相似性计算之前,增加了时空关系的抽取,即基于粗糙集进行历史案例库中空间关系的抽取,并把抽取到的空间关系作为问题求解的辅助信息^[14]。之

所以选择粗糙集的方法,主要考虑到它是一种研究不完整、不确定知识和数据表达、学习、归纳的理论方法,且不需要提供除问题所需处理数据集合之外的任何先验信息^[17];同时在保持分类能力不变的前提下,通过知识约简以导出决策或分类规则^[18]。

基于上述改进,图 1 给出了基于 CBR 方法进行土地利用变化的方法流程。首先需要针对土地利用变化问题进行案例界定与表达,明确三元组的具体内涵及描述指标;其次针对构建好的土地利用变化案例库进行案例之间时空关系规律的抽取;最后按照 CBR 的 4R 流程进行土地利用变化预测求解。问题的求解分为 3 步:①将待预测的图斑作为新案例,根据抽取的空间关系和案例相似性计算,从案例库中检索出相似的变化图斑;②使用相似图斑的变化结果对待预测图斑进行变化预测,进行案例重用(在这个过程中,需要进行相似图斑蕴含的时空关系规律约束);③根据预测后的变化图斑的特殊性,进行新案例的保存以扩充案例库。

2.2 方法实例

在研究^[14]中,以广东省珠海地区为例,开展过上述方法的应用试验研究。本研究在该试验基础上,深入进行 CBR 方法的适应性分析。本研究依然选定珠海地区(图 2)为研究区,进行对比试验。

参照研究^[14]中的试验步骤,选取与研究^[14]中不同的测试案例进行试验。具体的试验步骤如下^[14]:

(1) 案例表达与组织。针对该区域,首先对每个变化图斑进行三元组定量表达,如下:

“问题”:采用变化地块的图斑及对应于该图斑的周长(P)和面积(A)描述;

“地理环境”:选取 2 个拓扑相邻关系指标和 6 个空间距离关系指标描述。分别是:①该图斑 1995 年所邻接的土地利用类型(N_1);2000 年邻接的土地类型(N_2);②距最近的城镇距离(D_1)、距最近的建筑用地距离(D_2)、距最近的河渠距离(D_3)、距最近的水库的距离(D_4)、距最近的高等级公路的距离(D_5)、距最近低等级公路的距离(D_6);

“结果”:该图斑 1995-2000 年的土地利用变化情况。因此案例表达为:

$$\text{Case}=\{\text{ID}, P_i, A_i, N_{1i}, N_{2i}, D_{1i}, D_{2i}, D_{3i}, D_{4i}, D_{5i}, D_{6i}, \\ \text{Land}_{y1995}, \text{Land}_{y2000}\} \quad (1)$$

$$i=1,2,\cdots,k;$$

式中: ID 是案例在案例库中的唯一标识。

(2) 空间关系抽取及案例库建立^[14]。按照公式

(1)的表达模型,针对该区域 397 个变化的图斑进行上述 8 个空间关系指标的抽取,其中相邻关系的两个指标利用 GIS 空间数据库的拓扑关系描述获取,并采用了邻接度指数的算法获取所有的邻接类型,距离关系采用 GIS 空间距离分析算法获取,具体采用 ArcMap 的 VBA 按照上述算法编程实现。

选取 397 个变化图斑构成案例库(表 1)。从中选取 50 个案例作为测试案例(研究^[14-15]中选取了 30 个案例),将余下的 347 个案例作为历史案例,进行

变化预测,并给出精度评价。

(3) 相似性计算与推理^[14]。本试验中,首先进行案例间拓扑相邻关系的相似性计算。本试验直接采用研究中的权重大小,研究中则采取基于粗糙集自动学习方法获得各种特征属性权重大小^[19]。每个指标的具体权重可参见表 4 第 1 行。在此基础上,根据相似性计算的结果进行相似历史案例的抽取。本试验中没有进行案例间时空关系规律的抽取,因此直接选取出相似性计算中结果最大的历史

表 1 珠海区域 1995–2000 年土地利用变化案例库^[14]

Tab.1 The LUC case library for Zhuhai region from 1995 to 2000

ID	P	A	N ₁ ^注	N ₂ ^注	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	Land _{y1995}	Land _{y2000}
1	1071	32739	oja	jd*	9747	567	2357	1556	0	0	d	m
2	747	18567	oja	jd*	9483	582	2165	1759	0	0	d	m
3	2275	60565	oja	jd*	8943	363	1982	1317	0	207	d	m
4	5148	272201	nda	med	255.4	204	1929	682	1236	3822	g	c
5	25628	8E+06	qjf	***	9029	1256	1709	3069	9288	2714	i	*
6	2150	140977	qjf	***	13616	6006	5917	9088	13892	6592	i	k
7	2380	186535	ald	jkd	10788	793	43.3	948	7681	5674	j	p
8	1458	52556	od*	jd*	8404	1523	2772	709	0	1718	j	m
396	25628	8E+06	qjf	***	9029	1256	1709	3069	9288	2714	l	*
397	2150	140977	qjf	***	13616	6006	5917	9088	13892	6592	i	k

注: 该字段中的字母分别表示: a 平原水田; b 山地旱地; c 平原旱地; d 有林地; e 疏林地; f 其他林地; g 中覆盖度草地; h 低覆盖度草地; i 河渠; j 水库坑塘; k 滩涂; l 滩地; m 城镇用地; n 农村居民点; o 其它建设用地; p 高覆盖度草地; *: 海域; ***: 邻接类型不考虑(由于本研究考虑的邻接类型是陆地地块, 故海域的邻接类型在此不考虑)。

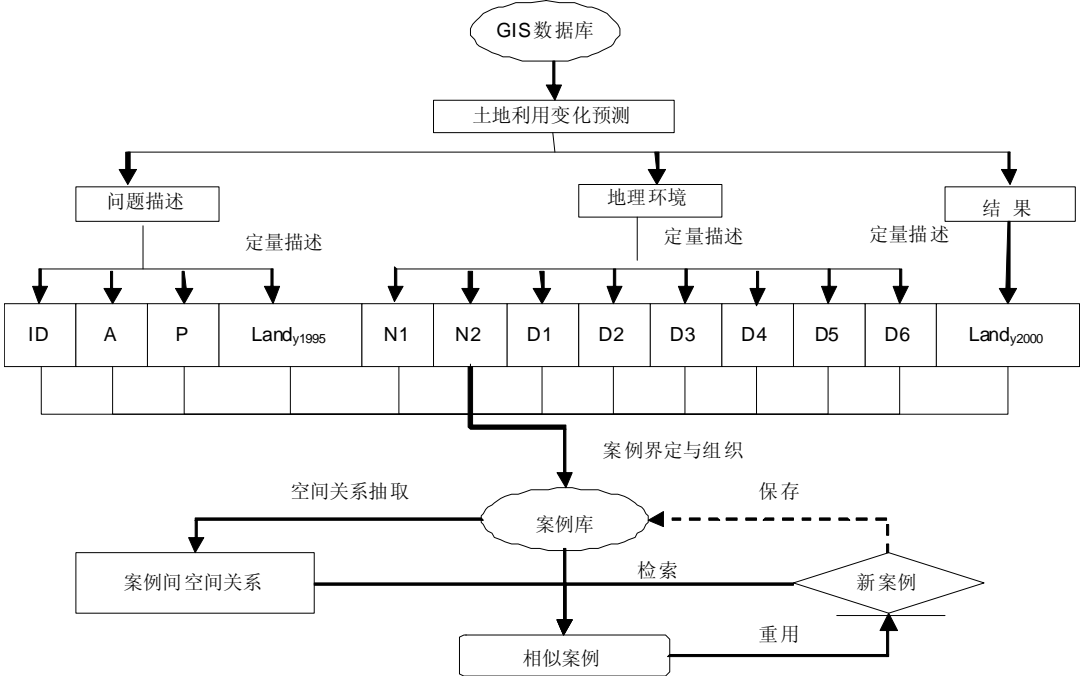


图 1 CBR 预测土地利用变化的方法流程

Fig.1 The flow chart of CBR method in the prediction of LUC

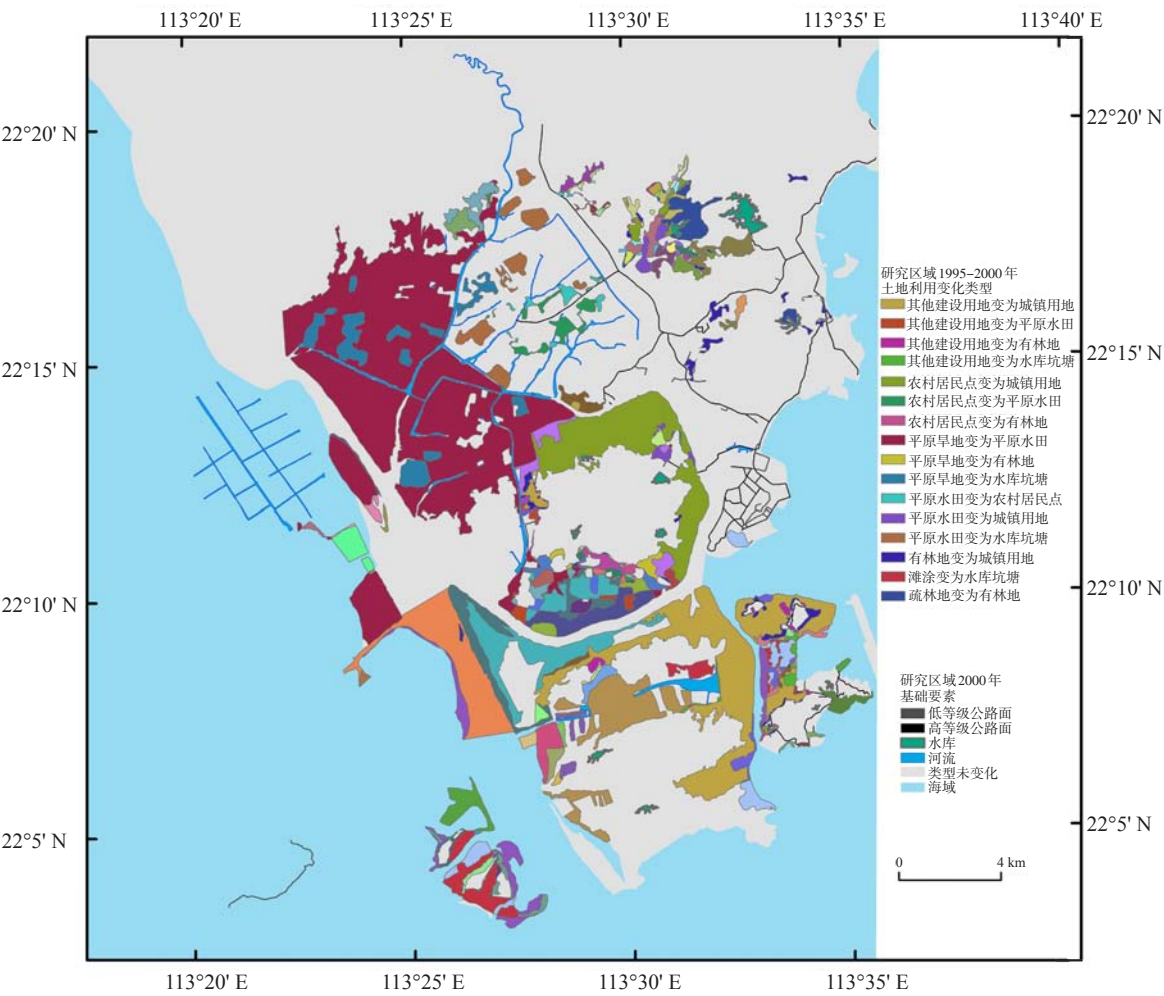


图2 试验区 199-2000年的土地利用变化空间分布图(参见文献[14])

Fig.2 The spatial distribution of LUC during 1995-2000 in the experimental region

案例,将它的“结果”作为新案例的“结果”。本实验中具体的相似性计算公式和案例间蕴含的时空关系规律的抽取方法,参见文献[14,15]。

(4) 结果说明。试验结果为:针对 50 个测试案例,其总体预测精度为 82%,所有测试案例在初始相似度阈值为 75%情况下都有相似的历史案例。

3 测试案例选取方式对 CBR 预测精度的影响分析

本小节主要从测试案例选取方式对 CBR 预测精度的影响进行对比分析,总结规律。为了便于对比,把 2.2 中的试验称为“参考试验”,本文中的其他试验全部在“参考试验”基础上进行并与之对比。

参考试验中,案例的选取方式是针对性选取。测试案例选取之前,先对案例库按照变化的类型进行统计分析可知,该区域 397 变化图斑中,共有 93

种变化类型,按照每种变化类型的图斑个数进行排序,以有选择地选取图斑个数较多的变化类型作为测试案例选择的依据,选取 50 个作为测试案例。

本节的对比试验中,案例的选取方式完全是随机的,即不做任何分析,随机从 397 个变化图斑中抽取 50 个作为测试案例。按照 2.2 中的实验步骤,采用这 50 个随机抽取的案例作为测试案例,进行对比试验。

对比试验的总体预测精度为 56%,明显低于参考试验。根据测试案例中各种变化类型历史案例个数的不同,对预测结果进行统计汇总,表 2 是汇总后的结果。表 2 第 2 列和第 4 列表示的是:在针对性选取的 50 个测试案例中,有 24 个案例是属于历史案例个数大于 10 的变化类型,其中 71% 的案例预测正确;另外 26 个案例则属于历史案例个数小于 10 的变化类型,其中 35% 的案例预测正确。由表 2 可知:针对历史案例丰富(历史案例数大于 10)的

变化类型,两种选取方式下,整体预测精度差异较小(分别是71%和91%);针对历史案例匮乏(历史案例数小于10)的变化类型,两种选取方式下,整体的预测精度差异较大(分别是35%和71%)。产生这种较大的差异可能是因为在随机性选取中,选取了一些历史案例极少(小于5个)的变化类型的案例作为测试案例,其中有几种变化类型的历史案例只有1-2个,这种类型的案例难以保证预测的准确性;而在针对性选取中,考虑到CBR对历史案例的依赖性,避开了那些只有1-2个历史案例的变化类型,从历史案例相对多(5-9个)的变化类型中选取测试案例,而这些类型的案例的预测精度通常较高一些。

表3是按照具体的变化类型进行的详细的精度统计分析,这里只统计了历史案例个数大于10的变化类型。由表3可知,针对具体的变化类型,如果该类型历史案例丰富,在两种选取方式下的预测结果一般没有太大差异,基本上都比较理想。

由上述结果的分析可知:在历史案例丰富的情况下,案例选取方式对CBR方法的预测效果没有大的影响;在历史案例匮乏的情况下,案例选取方式对其有很大影响。

4 三元组指标选取对CBR推理精度的影响分析

研究^[14-15]从地理案例特殊性出发,提出了“三元组”案例表达模型,给出案例表达的一个通用模式。针对特定问题,如何明确其各自内涵及选用相应的表达指标,不同的研究群体由于研究切入点的不同会有不同理解和看法。因此,“三元组”中每个组分的定量表达指标的选取会对CBR预测精度有着很大的影响。此外,在土地利用变化研究中,当可获取的数据有限时,如何选取有效的指标进行CBR预测,就显得十分关键。

本小节从三元组指标的选取对CBR预测精度影响进行对比分析,探讨规律。对比试验采用与2.2中的参考试验完全相同的测试案例和历史案例,以参考试验中选取的指标为基

础,在维持其他指标不变的基础上,依次舍去各种类型指标进行对比试验。分别将“问题”组分中的4个指标、“地理环境”组分中的2个拓扑相邻关系指标和“地理环境”组分中的6个距离关系指标删除。针对删除之后新的指标组合,依次进行试验,共进行了7次对比试验,表4给出了试验结果。表4中数字第1行表示的是参考试验选取的指标和预测结果,前12列的数字表示的是对应指标被赋予的权重大小;其余每行对应一次指标改变后的试验结果,数字0表示对应的指标在该次试验中不予考虑,空格表示与第一行的权重大小相同。

由表4可知,当忽略“问题”组分的4个指标其中之一时,预测精度没有受到明显影响;当忽略“地理环境”中的6个空间距离关系指标时,预测精度也没有发生明显变化,甚至得到了一定程度的提高;当忽略“地理环境”中的2个拓扑相邻关系指标时,预测精度则受到了很大的影响。结果显示,拓扑相邻关系指标对精度的影响明显大于其他指标。另一方面,“问题”组分的4个指标对精度的影响相当,究其原因,主要是由于所选取的变化斑块的定量描述间存在着变化一致性,比如通常面积大的斑块,周长也较大;本文中的案例在数据库中是按照土地利用变化类型排列的,ID与土地利用类型有着一致性。再者,表4还显示空间距离关系指标对预测结果并没有起到促进作用,这与当初的预想结果不同,一定程度上反映了在珠江三角洲这种社

表2 不同选取方式不同历史案例数下的预测精度
Tab.2 Accuracy estimation under different selection approaches and different numbers of the cases

历史案例个数	测试案例个数 (随机性选取)	测试案例个数 (针对性选取)	准确率(随机性选取)	准确率(针对性选取)
>10	24	23	0.71	0.91
<10	26	27	0.35	0.70
总计	50	50	0.56	0.82

表3 不同选取方式下历史案例丰富的变化类型的预测精度(历史案例个数大于10)

Tab.3 Comparison of accuracy estimation of the LUC categories possessing abundant historical cases with different selection approaches

1995年 土地利用类型	2000年 土地利用类型	测试案例个数 (随机性选取)	测试案例个数 (针对性选取)	历史案例个数	准确率(随机性选取)	准确率(针对性选取)
有林地	城镇用地	3	6	33	0.67	1
平原水田	水库坑塘	2	2	16	0.5	1
平原水田	农村居民点	3	2	15	1	1
平原旱地	平原水田	1	2	15	0	0
农村居民点	平原水田	2	3	14	1	1
农村居民点	有林地	3	1	12	1	1
其他建设用地	有林地	2	1	12	0	1
平原旱地	水库坑塘	0	2	12	-	1
滩涂	海域	3	1	10	0.67	1
其他建设用地	城镇用地	3	1	10	0.67	1
平原水田	城镇用地	2	2	10	1	1
总计		24	23	159	0.708	0.913

会经济比较发达的区域,土地利用变化主要驱动力并不是来自道路和水系的牵引。这样的研究结论与目前该区域内国内其他同行的研究结论呈现出很大的一致性。如史培军等^[3]在研究深圳市土地利用变化机制的时候,通过回归分析,研究土地利用变化与各个驱动因素之间的关系,表明邻接土地利用类型的相关系数大于各种空间距离变量,表现出更强的相关性。樊风雷等^[20]和王云鹏等^[21]在研究珠江三角洲的土地利用变化机制时,发现道路和水系的影响主要作用在城市用地的扩张,即其他用地向城镇用地的转变,对其他的土地变化类型的影响较小。在本研究中,转变为城市用地的案例较少,因此,道路和水系的影响作用较弱,邻接土地利用类型起着主要作用。

因此,在保证重要指标选取的前提下,次要指标的选取对 CBR 预测精度没有大的影响。对于珠江地区,在数据有限情况下,应当优先选择拓扑相邻关系指标,进行土地利用变化的预测。另外,试验中拓扑相邻关系指标对预测精度的重要性也在一定程度上反映了土地利用类型在空间上的分布关系及土地利用变化的作用规律。

5 指标权重对 CBR 推理精度的影响分析

传统 CBR 中,指标权重研究是研究较多且比较热的领域之一;此外,土地利用变化不仅受指标选取,同样也受指标权重大小的影响,因此,本小节从三元组指标的权重大小对 CBR 预测精度的影响进行对比试验,寻找规律。

2.2 的参考试验中各指标权重采取基于相似粗糙集的自动学习方法获得^[14-15],具体大小见表 5 第 3

行。对比试验完全采用与参考试验相同的测试案例和历史案例,采取“先降后升”的策略:以“参考试验”中的指标权重为基础,在维持其他指标权重大小不变的基础上,依次降低和升高各个指标的权重进行试验。首先分别改变“问题描述”的 4 个指标属性的权重,再改变“地理环境”中的 2 个拓扑相邻关系指标的权重,最后改变“地理环境”中的 6 个空间距离关系指标的权重,依次进行了 14 次对比试验。表 5 给出试验的结果,表中数字第一行表示的是参考试验的指标权重大小和试验结果,下面每一行代表一次对比试验的指标权重大小和试验结果。

由表 5 可知,“问题”组分的 4 个指标权重改变对预测精度影响较小,“地理环境”中的空间距离关系指标其次,拓扑相邻关系指标的影响最大。

其中,“问题”组分的 4 个指标权重改变对预测精度影响呈现出了一定的一致性,还是由于所选取的变化斑块的定量描述间存在着变化一致性。面积(A)和周长(P)权重变化对预测精度的影响比较一致,主要是由于两者之间存在变化的一致性。面积(A)和周长(P)无论在几何学上还是地学上往往呈现出一致的变化规律,本研究中的土地利用图斑,在形态上大致都接近圆形、三角形和矩形,面积大的图斑通常周长也较大。ID 和 1995 年土地利用类型的权重改变对预测精度的影响也比较一致,这种一致性也很好地反映了数据库中案例排序的原则——按照 1995 年土地利用变化类型进行排序。“地理环境”组分中的两大类指标(拓扑关系类和距离关系类)对预测精度的影响较大,特别是拓扑相邻关系指标的权重下降时,预测精度下降了 20%。

因此,在保证重要指标具有较大权重的前提下,次要指标的权重对 CBR 预测精度没有大的影响。因此,对于珠江口地区,在明确“地理环境”组份中拓扑关系的重要性之后,应当赋予拓扑相邻关

表 4 不同指标选取下的预测精度
Tab.4 Accuracy estimation under different selections of variables

问题描述			地理环境			求解			总权重			预测精度	
ID	A	P	Land _{y1995}	N1	N2	D1	D2	D3	D4	D5	D6	Land _{y2000}	
2	1	1	2	3	3	3	2	1	2	4	3	27	0.82
				0	0							21	0.46
						0	0	0	0	0	0	12	0.90
				0	0	0	0	0	0	0	0	6	0.42
												25	0.80
												26	0.80
0	0	0	0									26	0.80
												25	0.80

注:表中“0”表示该指标的权重为 0;空格表示相对于“参考试验”权重大小保持不变。

表5 不同指标权重下的预测精度

Tab.5 Accuracy estimation with different weights of variables													总权重	预测精度
问题描述				地理环境								求解		
ID	A	P	Land _{y1995}	N1	N2	D1	D2	D3	D4	D5	D6	Land _{y2000}		
2	1	1	2	3	3	3	2	1	2	4	3		27	0.82
↓													26	0.78
↑													29	0.80
	↓												26.5	0.80
	↑												28	0.76
		↓											26.5	0.80
		↑											28	0.78
			↓										26	0.78
			↑										29	0.80
				↓	↓								24	0.62
				↑	↑								33	0.86
						↓	↓	↓	↓	↓	↓		19.5	0.86
						↑	↑	↑	↑	↑	↑		42	0.72
						↓	↓	↓	↓	↓	↓		16.5	0.80
						↑	↑	↑	↑	↑	↑		48	0.82

注：表中第一行的数字是 参考试验 各个指标的权重， ↓ 表示权重减小一半， ↑ 表示权重增加一倍，空格表示相对于 参考试验 权重保持不变。

系指标较大的权重,进行土地利用变化预测。

具有可靠性。

6 结论

本文分析了目前研究土地利用变化的多种定量分析方法存在的不足,在土地利用变化的CBR模型的基础上,探讨了CBR在土地利用变化预测上的适用性问题,并以广东省珠海地区为实例进行了多次的试验对比。基于“问题-地理环境-结果”的三元组模式构建地理案例,利用粗糙集抽取空间关系,作为辅助信息,通过经典的4R过程,预测土地利用的变化。通过案例选取方式、指标的选取以及指标的权重3个方面探讨了CBR方法在解决土地利用变化问题上具有的一些规律和特点。

试验的结果显示,在历史案例丰富的情况下,通过选择重要的指标,并对其赋予较高权重,可以保证CBR方法在预测土地利用变化时的稳定性,显示了该方法良好的适用性。另外,拓扑相邻关系在土地利用变化预测中的重要性,说明在CBR方法中考虑“地理环境”对解决地学问题的重要意义。不过,土地利用变化是个复杂的地学问题,受到众多的自然因素和人文因素影响,本文考虑的影响因素还十分有限。一些常见的因素,如土壤、高程和降水等自然因素,还有人口、经济和政策等人文因素都没有考虑,这些都需要在今后的研究中继续进行深入探讨。

此外,纵观整个研究过程,CBR方法在预测土地利用变化时,显示出了简单灵活的特点,可以在对问题理解不断深入的基础上进行快速地调整和

扩展,以满足更进一步的预测需求,且具有预测精度高和保持变化斑块形态稳定的特点,是一种解决该问题的有效方法。另外,在使用CBR方法研究复杂的地理现象时,一定要考虑到CBR方法自身的特点,考虑到CBR方法在解决地理问题中的适应性。通常只有在历史案例相对丰富的情况下,CBR方法的预测结果才

参考文献

[1] 张华,张勃. 国际土地利用/覆盖变化模型研究综述. 自然资源学报, 2005, 20(3): 422-431.

[2] 黄秋昊,蔡运龙. 国内几种土地利用变化模型述评. 中国土地科学, 2005, 9(5): 25-30.

[3] 史培军,陈晋,潘耀忠. 深圳市土地利用变化机制分析. 地理学报, 2000, 55(2): 52-62.

[4] 摆万奇. 深圳市土地利用动态趋势分析. 自然资源学报, 2000, 15(2): 112-116.

[5] 郭玉燕,欧名豪. 马尔柯夫模型在土地利用结构预测中的应用. 广东土地科学, 2004, 3(6): 37-41.

[6] 黎夏,伍少坤. 面向对象的地理元胞自动机. 中山大学学报: 自然科学版, 2006, 45(3): 90-94.

[7] 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.

[8] 黎夏,叶嘉安. 约束性单元自动演化CA模型及可持续发展形态的模拟. 地理学报, 1999, 54(4): 289-298.

[9] 林士敏,田凤占,陆玉昌. 贝叶斯网络的建造及其在数据挖掘中的应用. 清华大学学报, 2001, 41(1): 49-52.

[10] Li X, Yeh A G O. Multitemporal SAR images for monitoring cultivation systems using case-based reasoning. Remote Sensing of Environment, 2004, 90(4): 524-534.

[11] Li X, Liu X P. An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region. International Journal of Geographical Information Science, 2006, 20: 1109-1136.

[12] Chen F L, Wang C, Zhang H et al. SAR images classification using case-based reasoning method//IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), Barcelona, Spain, 2048-2051, July 23-27, 2007.

- [13] Li X, Yeh A G O, Qian J P, et al. A matching algorithm for detecting land use changes using case-based reasoning. *Photogrammetric Engineering & Remote Sensing*, 2009, 75(11): 1319-1332
- [14] 杜云艳, 王丽敬, 季民, 等. 土地利用变化预测的案例推理方法. *地理学报*, 2009, 64(12): 1421-1430.
- [15] 杜云艳, 温伟, 曹峰. 空间数据挖掘的地理案例推理方法及试验. *地理研究*, 2009, 28(5): 1285-1296.
- [16] Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 1994, 7: 39-59.
- [17] 王国胤. *Rough集理论与知识获取*. 西安: 西安交通大学出版社, 2001: 1-25.
- [18] 张文修, 吴伟志, 梁吉业. *粗糙集理论与方法*. 北京: 科学出版社, 2001: 1-39.
- [19] 周平, 丁进良, 岳恒, 等. 基于相似粗糙集的案例特征权值确定新方法. *信息与控制*, 2006, 35(3): 329-334
- [20] 樊风雷, 邱茂慧, 王云鹏. 珠江三角洲经济核心区土地利用遥感监测及其转化机理. *东华理工学院学报*, 2007, 30(3): 252-256.
- [21] 樊风雷, 王云鹏. 基于遥感数据的珠江三角洲核心区土地利用时间序列变化分析. *农业现代化研究*, 2008, 29(2): 223-226.

Study on the Suitability of CBR Method in the Estimation of Land Use Change

SUN Yeran^{1,2}, DU Yunyan¹, SU Fenzhen¹, ZHOU Chenghu¹

(1. State Key Laboratory of Resources and Environmental Information System,
Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: There are various approaches used to study land use change (LUC), such as regression analysis, probability statistics, system dynamics, cellular automata and Bayesian network. These approaches have their specific characteristics and practical applications in the LUC. Although there were some researches revealing the advantages and disadvantages of some approaches, there were relatively few studies on the suitability of these approaches. This study focused on the suitability of CBR approach for LUC estimation, on the basis of the CBR model for the LUC estimation. The comparison experiments were conducted from three aspects, selection approach of the test cases, selection of variables and weights of the variables, to explore the influences of these factors on the estimation accuracy of LUC. The land use changes in Zhuhai region, China during 1995-2000 were used as a case study to conduct the comparative experiments. The concrete comparison strategies include: (1) To choose the test cases by selective approach and stochastic approach to explore the effects of the selection approach on the LUC estimation accuracy; (2) to neglect different variables in turn representing three categories of impacts respectively to explore the effects of the neglect of the variables on the estimation accuracy; (3) to change the weights of variables in turn to explore the effects of the weights of specific variables on the estimation accuracy. The experimental results are shown as follows. Firstly, the selection approach of test cases has insignificant effects on the LUC estimation accuracy under the circumstance that the historical cases are abundant. Secondly, the neglect of the ordinary variables has insignificant influences on the estimation accuracy on the condition that vital variables are selected. Thirdly, the weights of the ordinary variables have insignificant effects on the estimation accuracy in the event that greater weights are assigned to the vital variables. These results demonstrate that CBR is an effective method for solving LUC problems with the advantages of simple construction, wide application, high accuracy and stable pattern. The stability of the LUC estimation accuracy based on CBR approach can be kept on the condition of plentiful historical cases when vital variables are selected and higher weights are assigned to them. In this case, CBR method shows a good suitability for LUC estimation. In addition, the incorporation of the new component “geographic environment” into the CBR model efficiently improves the estimation accuracy of LUC.

Key words: case-based reasoning (CBR); land use change (LUC); three-component model

本文引用格式:

孙晔然, 杜云艳, 苏奋振, 等. 土地利用变化预测 CBR 方法的适应性分析. *地理科学进展*, 2011, 30(7): 912-919.