



云环境下基于 CA-Markov 的土地利用变化预测方法

康俊锋^{1,2} 李爽¹ 方雷^{3,4}

1 江西理工大学建筑与测绘工程学院,江西 赣州,341000

2 浙江大学地球科学学院,浙江 杭州,310027

3 复旦大学环境科学与工程系,上海,200438

4 香港大学深圳研究院,广东 深圳,518172

摘要:传统土地利用变化预测方法通常效率低,无法满足土地利用变化大数据分析和处理的需求。采用 MapReduce 编程模型对元胞自动机-马尔可夫模型进行并行化扩展,设计了基于 Hadoop 的土地利用变化预测方法(land-use change prediction method based on cloud computing, Cloud-CMLP),并选取杭州市进行实验,包括:①不同数据量下 Cloud-CMLP 核心算法的运行效率实验;②利用 Cloud-CMLP 方法模拟杭州 2013 年土地利用变化,并将模拟结果与 2013 年遥感影像解译结果进行对比分析,验证了预测方法的正确性;③杭州 2020 年的土地利用变化模拟预测分析,研究区中心地带受发达交通系统影响,建设用地面积整体呈快速上升趋势,且主要来源于农业用地的转换。

关键词:云计算;Hadoop;MapReduce;CA-Markov;土地利用变化

中图分类号:P208

文献标志码:A

土地利用变化与人类各类活动相互影响和作用,研究不同时空下的土地利用变化,预测未来的土地利用结构和空间布局,可为区域土地资源的科学利用、区域生态环境的保护与社会经济可持续发展提供科学参考^[1-2]。

很多学者提出了各自的土地利用变化模拟与预测模型,如 CLUE(conversion of land use and its effects)模型^[3]、CLUE-S(CLUE at small region extent)模型^[4]、元胞自动机(cellular automata, CA)^[5]、马尔可夫(Markov)链^[6]、系统动力学^[7]、SLEUTH^[8]、空间 Logistic 模型^[9]、CA-Markov 模型等。其中 CA-Markov 模型结合了 Markov 模型长期预测的能力和 CA 模型模拟复杂系统空间变化的能力,已被广泛验证^[10-13]。然而土地利用变化模拟预测方法涉及大量的数据和计算,传统的串行算法或者半手工计算方法效率较低,有学者设计了单机并行^[14]、MPI(message-passing-interface)^[15]、GPU(graphic processing unit)^[16-17]、

GPU/CPU(central processing unit)混合架构^[18]的并行 CA 模型,应用于城市变化模拟,并验证了其算法性能。但近年来,土地利用大数据的应用相比原有一城一地,在空间范围、时间跨度、空间分辨率和时间分辨率上都有巨大飞跃,目前,采用 MapReduce 对传统算法进行并行化处理以提高运行效率,已成为解决这一问题的主流方法和思路,并在很多方面得到了应用和验证^[19-20]。

本文首次采用 MapReduce 对基于 CA-Markov 的土地利用变化预测方法(land-use change prediction method based on cloud computing, Cloud-CMLP)进行改进,在兼顾预测正确性的基础上,提高了土地利用变化预测的效率,实现了云环境下的土地利用变化预测。

1 Cloud-CMLP 并行算法设计

1.1 算法概述

采用 MapReduce 编程模型对 CA-Markov 模

收稿日期:2019-06-20

项目资助:国家重点研发计划(2016YFC0803105);国家自然科学基金(41701462);国家留学基金(201808360065);江西省教育厅科学技术研究项目(3204704062)。

第一作者:康俊锋,博士,副教授,主要从事高性能 GIS 算法及应用研究。junfeng.kang@jxust.edu.cn

通讯作者:方雷,博士。fanglei@fudan.edu.cn

型进行并行化处理,设计云环境下基于CA-Markov模型的Cloud-CMLP,其可并行化部分主要包括两个算法:云环境下并行Markov算法(parallel Markov algorithm on cloud computing, Cloud-Markov)及云环境下土地利用变化综合评价算法(land use change comprehensive evaluation algorithm on cloud computing, Cloud-CELUC)。

如图1所示,利用Cloud-Markov对研究区进行土地利用类型变化总量预测,获得各地类数量变化;利用Cloud-CELUC来确定每个元胞对应各地类转变方向的评价值,得到所有评价值(即土地利用综合评价值统计总表),再结合第一部分Cloud-Markov的变化总量预测结果,判定所有元胞转换方向,从而实现土地利用变化空间分布预测。

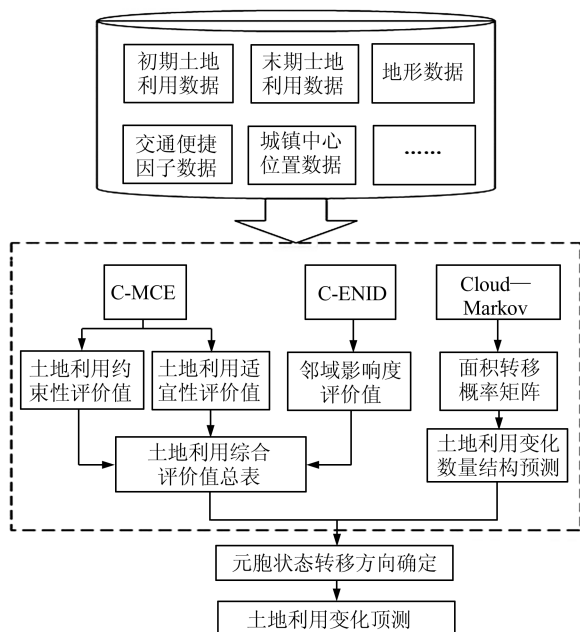


图1 基于MapReduce的CA-Markov并行算法

Fig.1 Land Use Change Prediction Algorithm of CA-Markov Based on MapReduce

其中,Cloud-CELUC包括云环境下邻域影响度评价(cloud-based evaluation of neighborhood influence degree, C-ENID)、云环境下多准则评价(cloud-based multi-criteria evaluation, C-MCE)及评价值综合计算3个部分。C-ENID利用MapReduce基于CA模型计算得到邻域影响度评价值, C-MCE利用MapReduce基于MCE计算得到土地利用约束性评价值(constraint evaluation value, CEV)与土地利用适宜性评价值(suitability evaluation value, SEV),评价值综合计算使用前两部分计算得到的约束性评价值与适宜性评价值、邻域影响度评价值计算得到每个元胞转换为每个地类的评价值。

1.2 Cloud-Markov算法设计

将相同空间位置不同时相的两幅栅格影像进行叠加分析,获取每个栅格的地类转换方向,并统计各转换方向的栅格数量,得到地类转换的面积转移矩阵,并计算得到土地利用面积转移概率矩阵。但通过该概率矩阵只能获得研究初期至末期时间间隔 m 年的各类土地利用面积,将该面积值除以 m ,得到每年各类土地利用转换面积,可以得到一年间隔的面积转移矩阵。研究给出的Markov模型的Map函数、Combiner函数和Reduce函数的定义公式为:

$$\begin{cases} \text{Map: } (N, (T1, T2)) \rightarrow \text{List}(C_{mk}, i) \\ \text{Combiner: } M(C_{mk}, \text{List}(i)) \rightarrow \text{List}(C_{mk}, s) \\ \text{Reduce: } L(C_{mk}, \text{List}(s)) \rightarrow \text{List}(C_{mk}, q) \end{cases} \quad (1)$$

式中, N 为输入行的行偏移量; $T1$ 表示土地利用初期数据栅格地类; $T2$ 表示土地利用末期数据栅格的地类; C_{mk} 表示栅格由 m 地类转换为 k 地类; i 表示栅格发生 C_{mk} 转换的个数; M 为节点本地数据有 M 种地类; s 为节点本地该种转换方向的栅格总数; L 为地类总数; q 为整个研究区发生 C_{mk} 转换的栅格总数与转移概率的组值。

1.3 Cloud-CELUC模型设计

1) 元胞邻域读取处理

邻域影响度评价值的计算需要邻域元胞的状态数据,常用的元胞邻域有Von Neumann型和Moore型邻域,本文算法采用Moore 3×3 型邻域进行设计,每个中心元胞都以4周8个方向的元胞为邻居。以 m 行 n 列的栅格影像为例,各元胞是按行顺序依次读取,并记录每个元胞的序号 K_x (其中 K 为行顺序号,下标 x 为列顺序号),因此 K_x 的邻域元胞为 $(K-1)_{x-1}$ 、 $(K-1)_x$ 、 $(K-1)_{x+1}$ 、 K_{x-1} 、 K_{x+1} 、 $(K+1)_{x-1}$ 、 $(K+1)_x$ 、 $(K+1)_{x+1}$ 。

基于MapReduce计算CA模型时,需要查找各元胞的邻域元胞,因此为更好地减少MapReduce计算过程中节点间的数据交换,元胞读取时是以行为单位转换为一维数组,同时保留了各元胞的行列号,因此可快速地获取任意元胞的邻域元胞。

2) 多准则评价因子确定

研究利用MCE计算约束性评价值与适宜性评价值,MCE是可在具有相互冲突、不可共度的有限(无限)方案集中进行选择的方法,其目的在于确定最优决策方案^[21]。评价标准分为适宜性因子和约束性因子两种。适宜性因子是将影响因素标准化为可度量的连续值,进而评价区域并进行选择。约束性因子是将分析要素空间化处理为布尔值,标准只有“0或者1”两种选择,没有

中间灰色的地带。

通过层次分析法并结合专家打分,判定不同适宜性因子的权重。另外,无论上述适宜性分析结果如何,有元胞所对应的土地类型都不会发生变化,则转变概率为 0。约束性因子为 0 或 1,不受约束性因子影响的其值为 1,否则为 0,例如坡度大于 25° 的山体,水域,生态保护区的约束性因子设置为 0。

3) Cloud-CELUC 算法

C-MCE 的计算需要每个元胞的位置与各因子数据位置一一对应,按相应的行号将两者合并。获得邻域影响度评价、约束性评价与适宜性评价后,进行综合评价计算。所有节点计算出所有元胞的地类转换综合评价,即为土地利用综合评价统计总表。给出 C-ENID 与 C-MCE 的 Map 函数定义如下:

$$(N, (i, H1, H2, H3 \cdots Hm)) \rightarrow ((i, j), (L1, L2 \cdots Lm)) \quad (2)$$

式中, N 为输入行的行偏移量; i 表示影像中行号; $H1$ 为需计算的元胞行所有元胞状态值; $H2$ 与 $H3$ 表示 $H1$ 的上下行元胞状态; $H4, H5 \cdots Hm$ 等为该元胞行相对应的各种约束性因子数据与适宜性因子数据; (i, j) 为元胞在影像中的行列号; Lm 为该元胞转换方向与对应转换方向综合评价 CE-LUC 的组合值。

1.4 元胞地类转换

采用多目标土地利用竞争方法实现元胞土地类型转换,该方法是解决同一元胞发生土地利用类型转换方向冲突时的抉择方法。以土地利用综合评价统计总表为基础,并以土地利用变化面积转移矩阵为约束,进行各元胞转换方向确定。其基本算法如下:

1) 忽略元胞地类转换方向,计算全体元胞空间中评价最大值,记录其元胞位置以及地类转换方向。

2) 该种地类转换的土地面积是否到达 Cloud-Markov 预测的转换上限。

3) 若已经达到上限,标记该元胞该地类转换评价值为 0,即丢弃该种转换方向,返回步骤 1); 若未达到上限,将该元胞记入该地类,丢弃该元胞其他类型转换评价,返回步骤 1)。

重复上述过程,直到全体元胞均发生转换,最终获得土地利用变化分布预测。

2 实验结果与分析

2.1 研究区数据

以中国科学院地理研究所数据中心(<http://www.gscloud.cn/>)提供的研究区 2006 年 Landsat

TM 与 2013 年 Landsat 8 分辨率为 30 m 的影像为基础,利用 ENVI 5.3 和 ArcGIS 10.2 对影像进行预处理与解译,获得两期研究区土地利用图。其他数据包括杭州市 30 m 分辨率数字高程模型、路网、行政区划等矢量数据。

2.2 算法效率分析

实验用 5 台机器搭建 Hadoop 云环境。其软件环境为 CentOS 7.1.1503 系统,Java 版本 1.8.0_112, Hadoop 版本 2.7.3;硬件环境为 CPU4 核 2.4 GHz,内存 4 GB。串行算法测试环境与 Hadoop 云节点相同。

Markov 串行算法与 Markov 并行算法的运行结果及加速比如图 2、图 3 所示,单机单次 CE-LUC 与 Cloud-CELUC 运行结果及加速比如图 4、图 5 所示,其中横坐标表示元胞数量(以 n 为单位,1 n 约为 9 000 000 个),纵坐标表示运行时间(以 s 为单位)。

从实验结果可知,Cloud-Markov 相对串行 Markov 模型加速比为 3.27 时增长趋向于平稳,为 3.43 时达到最高,Cloud-CELUC 相对于串行 CELUC 为 1.77 时增长趋向于平稳,为 1.86 时达到最高。

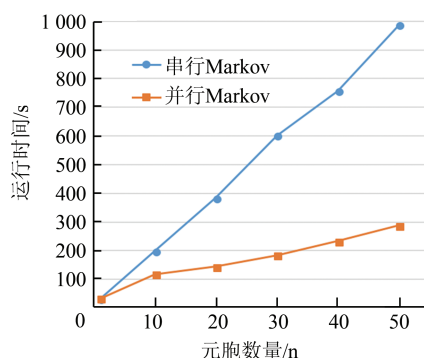


图2 Cloud-Markov 与串行 Markov 效率对比

Fig.2 Running Efficiency Between Cloud-Markov and Serial Markov Algorithms

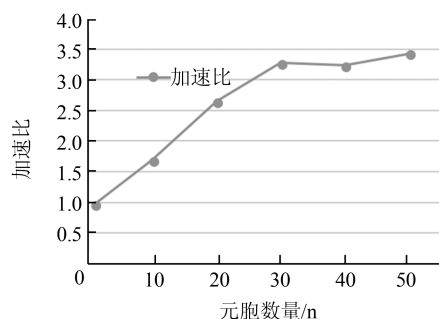


图3 Cloud-Markov 与串行 Markov 加速比

Fig.3 Speedup Ratio of Cloud-Markov and Serial Markov Algorithms

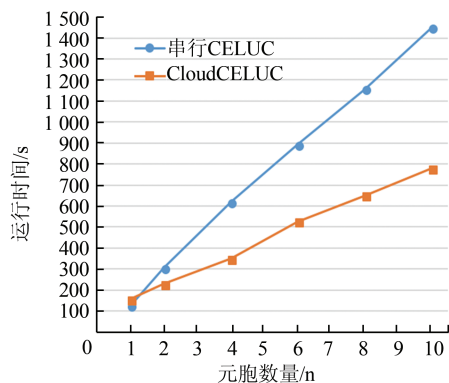


图4 Cloud-CELUC与串行CELUC效率对比

Fig.4 Running Efficiency Between Cloud-CELUC and Serial CELUC Algorithms

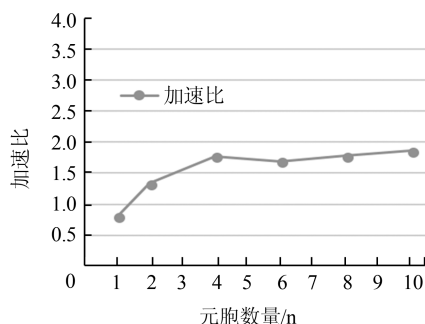


图5 Cloud-CELUC与串行CELUC加速比

Fig.5 Speedup Ratio of Cloud-CELUC and Serial CELUC Algorithms

通过对实验过程中各节点的运行情况进行分析,Cloud-Markov运行效率提高的原因是Map-Reduce分散了两期数据元胞同位置匹配与统计数量的工作量;Cloud-CELUC在Map阶段运行速度较快,而在对Map阶段运算得到的输出进行Reduce合并时,数据迁移占用较长的运行时间,对运行效率有一定影响。未来将考虑设计数据分布优化策略提高Reduce阶段的合并效率。

2.3 精度评价与结果分析

本文采用基于2006年与2013年土地利用数据计算面积转移概率矩阵,以2006年作为初始年份,水域设为不变,对研究区2013年土地利用数据进行模拟,面积转移矩阵与面积转移概率矩阵如表1、表2所示。

利用模拟结果对C-MCE的各类因子权重参数进行精度评价校正,以提高模拟精度,实现准确预测。目前常用的精度评价方法有目视比较法、分维度检验法、像元对比检验法以及Kappa系数检验法,其中Kappa系数检验法是最常用的定量检验方法。对比计算2013年土地利用模拟数

据与2013年土地利用实际数据,得到Kappa系数计算结果如表3、表4、表5所示。

表1 2006—2013年面积转移矩阵/ km^2 Tab.1 Area Transfer Matrix in 2006—2013/ km^2

地类	农用地	建设用地	自然保留地	总计
农用地	1 282.95	409.71	95.67	1 788.33
建设用地	210.94	1 381.69	12.52	1 605.15
自然保留地	93.39	50.79	4 409.60	4 553.78
总计	1 587.28	1 842.19	4 517.79	7 947.26

表2 2006—2013年转移概率矩阵/%

Tab.2 Transition Probability Matrix in 2006—2013/%

地类	农用地	建设用地	自然保留地
农用地	71.74	22.91	5.35
建设用地	13.14	86.08	0.78
自然保留地	2.05	1.12	96.83

表3 自然保留地Kappa系数检验表/ km^2 Tab.3 Kappa Coefficient of Natural Reserve/ km^2

地类	自然保留地	非自然保留地	总计	准确度/%	Kappa系数
自然保留地	4 221.35	288.47	4 509.82	93.60	0.86
非自然保留地	296.44	3 431.50	3 727.94	92.05	
总计	4 517.79	3 717.97			

表4 建设用地Kappa系数检验表/ km^2 Tab.4 Kappa Coefficient of Construction Land/ km^2

地类	建设用地	非建设用地	总计	准确度/%	Kappa系数
建设用地	1 391.41	452.77	1 844.18	75.45	0.68
非建设用地	450.78	5 942.80	6 393.58	92.95	
总计	1 842.19	6 395.57			

表5 农用地Kappa系数检验表/ km^2 Tab.5 Kappa Coefficient of Agricultural Land/ km^2

地类	农用地	非农用地	总计	准确度/%	Kappa系数
农用地	1 152.64	427.17	1 579.81	72.96	0.66
非农用地	434.65	6 223.30	6 657.95	93.47	
总计	1 587.29	6 650.47			

自然保留地、建筑用地、农用地对应的Kappa系数分别为0.86、0.68和0.66,说明2013年模拟结果精度较高,可以推测对未来土地利用预测可信度较高。2020年土地利用变化预测实验基于2013年土地利用变化数据与表6面积转移概率矩阵对2020年土地利用进行预测,预测结果如图6所示。

从预测结果分析可知,研究区内建设用地面积整体呈上升趋势且主要来源于农业用地转换,

农用地面积呈下降趋势,自然保留地面积变化比例不大。建设用地总体增长面积较大,尤其各县市区中心因受路系与公交系统发达的影响,建设面积增速较快,其中西湖区、拱墅区、下城区、滨江区、萧山区建设用地的增长突出,其余县市区因受地形边界限制,建设用地保持基本稳定。

表 6 2013—2020 年土地利用面积转移预测矩阵/ km^2

Tab.6 Area Transfer Matrix in 2013—2020/ km^2

地 类	农用地	建设用地	自然保留地	总计
农用地	1 133.35	361.94	84.52	1 579.81
建设用地	242.35	1 587.45	14.38	1 844.18
自然保留地	92.49	50.30	4 367.03	4 509.82
总计	1 468.19	1 999.69	4 465.93	7 933.81

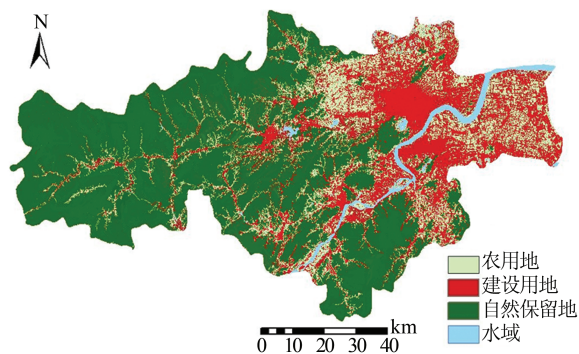


图 6 2020 年研究区土地利用变化预测图

Fig.6 Land Use Change Prediction of Research Area in 2020

研究区西北部与北部农用地与建设用地大片面积交织错落,但远离市区中心且交通系统欠发达,农用地面积也基本不变。其余地区农用地受到城市影响辐射、路系与公交系统的影响,较大面积转换为建设用地。自然保留地由于多在西部与南部丘陵山区,受约束性因子坡度的影响较大,且交通欠发达,基本保持稳定。

3 结 语

本文利用 MapReduce 编程模型将 CA-Markov 模型进行并行化设计,包括云环境下 Markov 算法以及云环境下土地利用变化综合评价算法。通过选取实验区并搭建实验环境,验证了方法的可靠性、正确性和运行效率,并模拟分析了 2020 年杭州市的土地利用变化预测结果。实验表明,将云计算技术与 CA-Markov 模型结合模拟预测土地利用变化从技术上是可行的。

本文首次将云计算应用于土地利用变化预测,并且详细设计了云环境下的并行土地利用变

化预测算法,但依然存在一些不足,主要有:土地利用变化除了自然条件的限制,还受到政治、经济、人口等复杂因素的综合影响,由于时间关系以及数据源的限制,研究主要针对交通、地形、区位因素进行建模,未来研究中将考虑更多的社会经济、政策及人口等因素,如国家规定的耕地红线、生态红线对于城市扩张边界的限制,完善土地利用变化空间格局预测模型。

参 考 文 献

- [1] Li Xiubin. A Review of the International Researches on Land Use/Land Cover Change [J]. *Acta Geographica Sinica*, 1996, 51(6): 558 - 565(李秀彬. 全球环境变化研究的核心领域—土地利用 / 土地覆被变化的国际研究动向[J]. *地理学报*, 1996, 51(6): 558-565)
- [2] Wijesekara G N, Farjad B, Gupta A, et al. A Comprehensive Land-Use/Hydrological Modeling System for Scenario Simulations in the Elbow River Watershed, Alberta, Canada[J]. *Environmental Management*, 2014, 53(2): 357-381
- [3] Verburg P H, Overmars K P. Combining Top-Down and Bottom-Up Dynamics in Land Use Modeling: Exploring the Future of Abandoned Farmlands in Europe with the Dyna-CLUE Model [J]. *Landscape Ecology*, 2009, 24(9): 1 167
- [4] Luo G, Yin C, Chen X, et al. Combining System Dynamic Model and CLUE-S Model to Improve Land Use Scenario Analyses at Regional Scale: A Case Study of Sangong Watershed in Xinjiang, China [J]. *Ecological Complexity*, 2010, 7(2): 198-207
- [5] Clarke K C, Hoppen S, Gaydos L, et al. Modifying Historical Urbanization in the San Francisco Bay Area [J]. *Environment and Planning B: Planning and Design*, 1997, 24(2): 247-261
- [6] Fu X, Wang X H, Yang Y J. Deriving Suitability Factors for CA-Markov Land Use Simulation Model Based on Local Historical Data[J]. *Journal of Environmental Management*, 2018, 206: 10-19
- [7] Le Q B, Park S J, Vlek P L G, et al. Land-Use Dynamic Simulator (LUDAS): A Multi-Agent System Model for Simulating Spatio-Temporal Dynamics of Coupled Human-Landscape System. I. Structure and Theoretical Specification [J]. *Ecological Informatics*, 2008, 3(2): 135-153
- [8] Dietzel C, Clarke K C. Toward Optimal Calibration of the SLEUTH Land Use Change Model [J]. *Transactions in GIS*, 2007, 11(1): 29-45
- [9] Hu Z, Lo C P. Modeling Urban Growth in Atlanta

- Using Logistic Regression[J]. *Computers, Environment and Urban Systems*, 2007, 31(6): 667-688
- [10] Memarian H, Balasundram S K, Talib J B, et al. Validation of CA-Markov for Simulation of Land Use and Cover Change in the Langat Basin, Malaysia[J]. *Journal of Geographic Information System*, 2012, 4(6): 542-554
- [11] Liu X H, Andersson C. Assessing the Impact of Temporal Dynamics on Land-Use Change Modeling [J]. *Computers, Environment and Urban Systems*, 2004, 28: 107-124
- [12] Keeley J E, Zedler P H. Large, High-Intensity Fire Events in Southern California Shrublands: Debunking the Fine-Grain Age Patch Model[J]. *Ecological Applications*, 2009, 19(1): 69-94
- [13] Araya Y H. Analysis and Modeling of Urban Land Cover Change in Setúbal and Sesimbra, Portugal [J]. *Remote Sensing*, 2010, 2(6): 1 549-1 563
- [14] Guan Q F, Clarke K C. A General-Purpose Parallel Raster Processing Programming Library Test Application Using a Geographic Cellular Automata Model [J]. *International Journal of Geographical Information Science*, 2010, 24 (5): 695-722
- [15] Li X, Zhang X H, Ye H A, et al. Parallel Cellular Automata for Large-Scale Urban Simulation Using Load-Balancing Techniques[J]. *International Journal of Geographical Information Science*, 2010, 24 (6): 803-820
- [16] Li Dan, Li Xia, Liu Xiaoping, et al. GPU-CA Model and Large-scale Land Use Change Simulation[J]. *Chinese Science Bulletin*, 2012(11): 959-969(李丹, 黎夏, 刘小平, 等. GPU-CA模型及大尺度土地利用变化模拟[J]. 科学通报, 2012(11): 959-969)
- [17] Wang H J, Zhang A Q, Zhang W T, et al. A High-Performance Cellular Automata Model for Urban Simulation Based on Vectorization and Parallel Computing Technology [J]. *International Journal of Geographical Information Science*, 2018, 32(2): 399-424
- [18] Guan Q F, Shi X, Huang M Q, et al. A Hybrid Parallel Cellular Automata Model for Urban Growth Simulation over GPU/CPU Heterogeneous Architectures [J]. *International Journal of Geographical Information Science*, 2016, 30 (3): 494-514
- [19] Yao Xiaochuang, Yang Jianyu, Li Lin, et al. Parallel Algorithm for Partitioning Massive Spatial Vector Data in Cloud Environment[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43 (7): 1 092-1 097(姚晓闯, 杨建宇, 李林, 等. 云环境下海量空间矢量数据并行划分算法[J]. 武汉大学学报·信息科学版, 2018, 43(7): 1 092-1 097)
- [20] Almeer M H. Cloud Hadoop Map Reduce for Remote Sensing Image Analysis [J]. *Journal of Emerging Trends in Computing and Information Science*, 2012, 3(4): 637-644
- [21] Wu Yiping. The Evaluation of the Multicriteria Evaluation Methods[J]. *Journal of Decision and Decision Support Systems*, 1994(1): 74-80(吴翼平. 多准则评价方法的评价[J]. 决策与决策支持系统, 1994(1): 74-80)

Land Use Change Prediction Method Based on CA-Markov Model Under Cloud Computing Environment

KANG Junfeng^{1,2} LI Shuang¹ FANG Lei^{3,4}

1 School of Architectural and Surveying & Mapping Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China

2 School of Earth Sciences, Zhejiang University, Hangzhou 310027, China

3 Department of Environmental Science and Engineering, Fudan University, Shanghai 200438, China

4 Shenzhen Institute of Research and Innovation, the University of Hong Kong, Shenzhen 518172, China

Abstract: Traditional land-use change prediction methods are usually implemented by serial algorithms or semi-manual methods and they were often inefficient. This paper develops a parallel land-use change prediction method based on cloud computing (Cloud-CMLP), the map reduce programming model is used to parallelize and extend the cellular automata(CA)-Markov model. Taking Hangzhou as a study area, the experiments are conducted as follows: ① Efficiency tests are conducted to compare the core algorithms of Cloud-CMLP under the different number of data. ② The Cloud-CMLP method is used to simulate the land-use change in 2013, and the simulated results are compared with the 2013 remote sensing image classification results to verify the validity of Cloud-CMLP method. ③ The land-use change in 2020 is predicted and

(下转第 1034 页)