

基于序列模式的土地利用变化分析

——以广西壮族自治区为例

廖伟华¹, 聂鑫², 蒋卫国³

(1. 广西大学数学与信息科学学院, 南宁 530004; 2. 广西大学公共管理学院, 南宁 530004;

3. 北京师范大学地理科学部, 遥感科学国家重点实验室, 北京 100875)

摘要: 土地利用变化是一个长期的过程, 同时具有一定的复杂性。传统的转移矩阵方法只能在两期土地利用数据之间相互比较而不能总结长期整体的变化规律, 频繁项集不能导出变化序列。针对这些方法的不足, 本文提出了基于序列模式的土地利用变化序列分析模型。首先给出土地序列数据库的定义, 然后根据土地序列数据实际特点和垂直格式的序列模式 SPADE 算法, 给出土地变化序列模式的计算过程和方法。在以中国广西为实例的研究中, 计算了 1970s—2015 年共 7 期 22 种二级土地利用类型的变化序列。研究区土地利用变化主要发生在林地之间, 部分林地转换为厂矿、采石场、交通道路等建设用地类型; 城镇用地主要由旱地和水田转换而来; 在研究期内没有任何一个土地单元转化成水田。

关键词: 土地利用变化; 序列模式; SPADE 算法; 长期变化; 广西

土地资源是最基本的生产资料, 也是人类赖以生存的物质基础^[1]。土地利用是人类活动影响自然环境最直接的表现形式^[2], 是全球气候、环境变化和可持续发展研究的重要内容^[3]。由于土地利用具有不确定性和一定的复杂性, 政府管理者在对未来的土地利用进行规划和布局时, 无法直接根据过去和现在的土地利用状况进行推断, 需要应用一些科学方法进行评估和预测。土地变化模型是描述土地利用格局变化与动态分析的主要工具, 且一直以来都是土地利用和土地覆被变化研究领域的重点^[4,5]。科学地掌握土地利用结构变化特征, 能为区域土地规划提供依据, 而未来土地变化模型发展方向是在多尺度进行多方法耦合, 从而对土地变化过程进行更为明晰的刻画^[6]。

土地利用变化研究工作很多。罗娅等^[7]以花江、红枫湖和鸭池三个石漠化治理区为例, 运用“纵横向”拉开档次法, 开展了石漠化治理区土地利用变化安全性评价, 认为在后续石漠化治理和土地利用规划中, 应重点关注土地利用变化对区域生态安全的影响。权瑞松^[8]基于 2000 年、2003 年和 2006 年土地利用数据, 借助 Terrset CA-Markov 模型模拟预测 2030 年上海市土地利用结构, 并采用 SCS 模型探究了土地利用结构变化的水文效应。在一些著名的土地变化预测模型中, 如机器学习和统计模型、基于元胞自动机的 CLUE-S 模型^[9]和 CLUMondo 模型^[10]。包括土地所有者、家庭、农民、房地产开发商、集体农场、移民、管理者和政策制定者等任何与土地利用相关决策者的多智能体模型^[11]

收稿日期: 2019-03-15; 修订日期: 2019-05-29

基金项目: 国家自然科学基金项目 (71763001, 71973038); 国家重点研发计划专项项目 (2016YFC0503002); 广西重点研发计划项目 (桂科 AB18126007)

作者简介: 廖伟华 (1975-), 男, 湖南耒阳人, 硕士, 副教授, 主要从事空间数据挖掘研究。

E-mail: gisliao@163.com

通讯作者: 聂鑫 (1983-), 男, 湖北武汉人, 博士, 教授, 主要从事土地资源管理研究。E-mail: toefl678@163.com

等，土地利用变化都是重要的基础内容。在这些土地利用类型变化的计算步骤中，皆利用转移矩阵进行土地利用类型的规律发现和预测。

不管是采用趋势外推的土地变化预测模型，还是关于土地利用结构变化的研究，对于土地变化都是采用土地转移矩阵。如利用两期土地利用图得到转移矩阵，或是根据一个时间序列内的概率转移矩阵，计算研究期初各类型土地的流失去向以及研究期末各类型土地的来源与构成，对土地未来结构进行预测。土地转移矩阵能反映的是某一时间段各地类面积数据以及各地类在研究时期期初转出与期末转入的信息，而在一个长时间序列内，任何土地单元都是一个完整的变化事物序列，符合关联分析的研究范畴。关联分析源于1993年提出的关联规则问题^[12]，在很多行业都得到了应用。廖伟华等^[13]利用频繁项集和Apriori算法研究了南宁市服务业的空间关联与集聚问题，挖掘结果能为城市规划与管理者提供决策支持。基于频繁项集挖掘，一种洪水警报的识别与提取方法被提出，提出者通过实际工业数据的应用，验证了该方法的有效性^[14]。在地理学空间数据挖掘领域，研究者认为空间实体的频繁项集是一种空间同位模式，用于发现空间实体在空间分布上的关联与共生现象^[15]。通过同位模式分析华盛顿特区的空间食物环境，研究认为{便利店，快餐店，药房}在空间上集聚效应明显^[16]。本文利用6年的监测数据，采用两步聚类分析和Apriori算法，分析了两种典型的崩积滑坡的变形状态，为建立滑坡预警系统提供支持^[17]。

关联分析主要还是集中在基于Apriori算法的频繁项集研究，在地学空间数据挖掘方面也有成果，在土地利用变化方面也将有广阔的应用前景。基于频繁项集的土地利用变化分析，能挖掘出类似{水田，城镇用地}之类的频繁变化模式，但该模式会产生两条关联规则{水田→城镇用地，城镇用地→水田}，由于土地利用变化有时间序列属性和空间变化上的不可逆性，城镇用地→水田这条规则是不符合现实的。序列模式首先是由Srikant等^[18]提出，它所挖掘的对象以及结果都是有序的，旨在发现诸如用户先购买了《育儿指南》、然后购买婴儿床等这样带有一定因果性质的规律。序列模式自提出以来，有了一些研究和应用。针对时空数据的移动轨迹模式挖掘，本文利用序列模式提出了两种有效的轨迹模式挖掘算法VTPM-PrefixSpan和VTPM-GSP，用于挖掘具有时间间隔约束的变换轨迹序列^[19]。通过对合成真实数据集的综合性能研究，本文开发了一种提取多维连续模式的特征数据挖掘方法，并使用马里（西非）土地利用信息的真实数据集来评估模型^[20]。基于Apriori性质，Srikant等^[18]提出了三个候选码生成和测试挖掘算法：Apriori-Some、AprioriAll、DynamicSome，随后又提出了经典的GSP算法。Zaki^[21]提出了基于垂直格式的序列模式SPADE算法，该算法首先将序列数据库转换为记录项集位置的垂直格式数据库，然后动态连接挖掘频繁序列模式。

前人在空间关联、行业选址等方面应用了序列模式，但在土地利用变化的相关研究中还没有得到应用。已有关于土地转换规律的研究，基于转移矩阵建立两期之间变化表只能体现研究时期期初和期末各地类面积的数量转化关系，而任何区域的土地变化都是一个完整的序列。本文将引入数据挖掘中的序列模式，利用土地变化时间上的序列特点，分析长期土地利用类型变化序列，拓展长期土地利用变化研究方法。最后将利用中国广西1970s、1980s、1995年、2000年、2005年、2010年、2015年的土地利用数据，验证模型的适用性，并分析实验区的土地利用转换模式与特点。

1 研究方法与数据来源

1.1 土地序列数据库

土地利用长期变化是一种离散型的序列。设 $I=\{i_1, i_2, \dots, i_n\}$ 是所有土地类型项组成的集合，每种土地类型就是一个项。研究的第一步需要首先把多期土地利用变化数据构造土地变化序列数据库。例如图 1 中，有三期土地利用变化图 (s1, s2, s3)，假设 a 为旱地，b 为水田，c 为有林地，d 为城镇用地，则 $I=\{a, b, c, d\}$ 。根据序列模式和土地利用变化的实际特点，下面给出相关定义。

事件：事件是一个项集，在土地利用变化中，一个事件表示一个土地单元在一个时期的一种类型，表示为 (x_1, x_2, \dots, x_q) ，其中 $x_k (1 \leq k \leq q)$ 是 I 中的一个项。一个事件中所有项均不相同，每个事件有一个事件时间标识 EID，也可以表示为事件的顺序。如图 1 中的第一个土地单元在第一个时期土地类型为旱地，第二个时期土地类型为水田，则可以表示两个事件 {1, 旱地}, {2, 水田}。土地利用变化在一个时期只有一种土地利用类型，所以事件会由 1 项集组成。

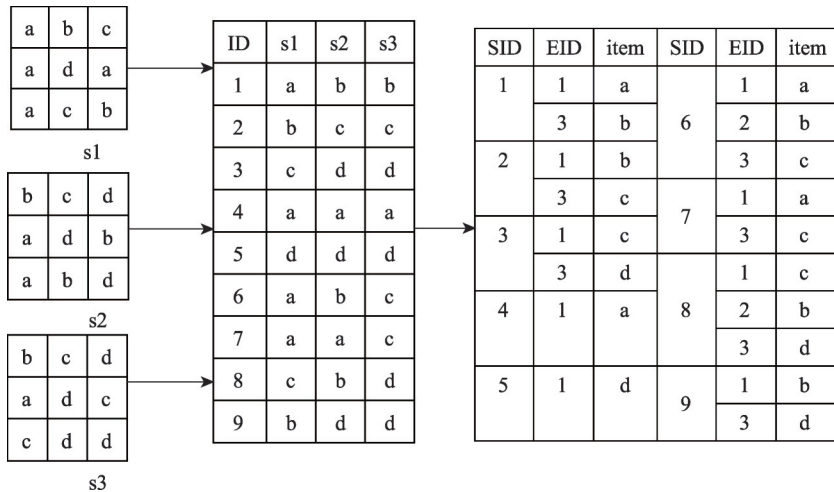


图 1 土地序列数据库生成

Fig. 1 Land sequence database generation

序列：序列 s 记作 $\langle e_1, e_2, \dots, e_l \rangle$ ，其中 $e_j (1 \leq j \leq l)$ 表示事件，也称为 s 的元素，序列是事件的有列表。

一个序列中的事件有时间上的先后关系，即 $e_j (1 \leq j \leq l)$ 出现在 e_{j+1} 之前。序列中事件个数是序列的长度，长度为 k 的序列称作 k -序列。在有些算法中，将含有 k 个项的序列称为 k -序列。如图 1 中的第一个土地单元则可以表示成 $\langle \{1, \text{旱地}\}, \{2, \text{水田}\} \rangle$ 序列，长度为 2，也就是 2-序列。

土地序列数据库：土地序列数据库是由元组 $\langle \text{SID}, s \rangle$ 组成的集合，其中 SID 表示序列编号， s 表示一个序列，每个序列由若干土地利用变化事件构成。在土地序列数据库中每个序列的事件在时间或空间上有序排列。

支持度：一个序列 x 的支持度计数是指在序列数据库 S 中包含 x 的序列个数。即：

$$\text{support}(x) = |\{(\text{SID}, s) | (\text{SID}, s) \in S \wedge x \}| \tag{1}$$

式中： $| \cdot |$ 表示土地序列数据库中出现的次数（次）。若序列 x 的支持度计数大于等于最小支持度阈值，即为频繁序列，频繁序列就是序列模式。如图1中的 a 在 $\{1, 4, 6, 7\}$ 中出现，则 a 的支持度为 $4/7$ 。

根据以上定义和土地利用变化图的特点，在图1中，把左边三期土地利用类型图转换成栅格图，通过GIS的合并运算可以得到中间的三期土地利用变化事物数据库图，对土地变化事物数据库，利用SQL技术，消除重复事件，就可以得到右边的土地利用变化序列数据库。

1.2 序列模式计算

土地利用变化序列模式分析就是找出所有大于最小支持度的序列。本文采用序列模式中SPADE算法。SPADE是一种基于垂直数据格式的Apriori类算法，它的数据集由一系列序列标识符、事件标识符和项集组成。所有的频繁序列都可以通过搜索序列格的方式被挖掘出来，其基本运算是两个序列的组合。两个 k -序列组合的结果为 $(k+1)$ -序列的集合，它们在序列格中恰好处于这两个 k -序列的上一层，该运算用“ \vee ”表示。

前缀^[21]：前缀形式化定义为一个函数 $p: (S, N) \rightarrow S$ ，其中 S 是一个序列集合， N 是一个非负整数， $p(X, k)=X[1:k]$ ，换句话说， $p(X, k)$ 返回 X 的 k 长度的前缀。

在序列格 S 上定义一个等价关系如下： $X, Y \in S$ ，当且仅当 $p(X, k)=p(Y, k)$ ，也就是说这两个序列共享长度为 k 的前缀，则它们是 θ_k 等价的，记为 $X \equiv_{\theta_k} Y$ 。由 X 构成的等价类记为 $[X]_{\theta_k}$ 。只有同一个等价类中的两个 k -序列才能进行时态连接运算，并产生长度为 $k+1$ 的候选序列。

对于图2中所示的垂直数据库，首先要构建长度为1的序列的id-list表，1序列的id-list表由每个土地利用项出现的土地单元和出现土地利用单元时期事件构成。在1序列的基础上，利用并运算“ \vee ”生成2序列的id-list表，例如图2中 a, b 是两个1-序列，则它们的组合 $a \vee b = \langle \{a, b\}, \{ab\}, \{b, a\} \rangle$ ，由于土地利用变化过程中，任何一个土地单元都不可

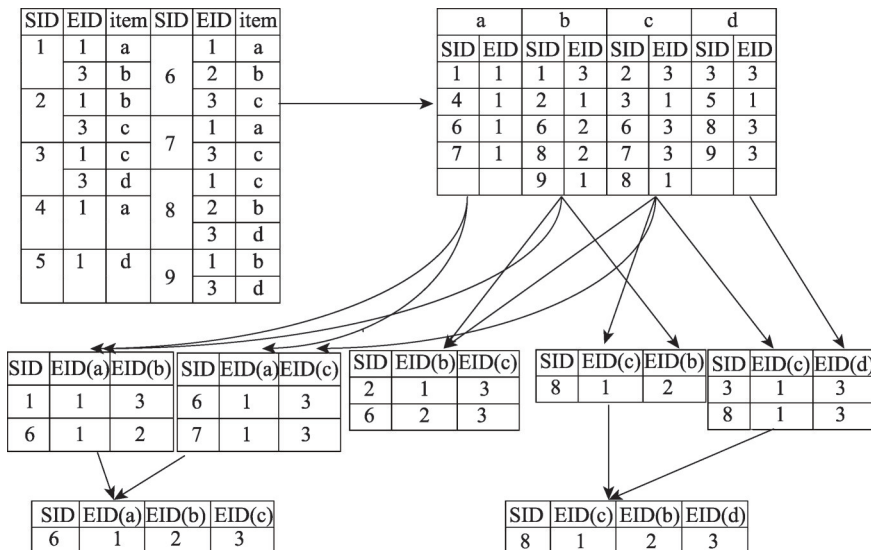


图2 土地利用变化序列模式计算过程

Fig. 2 Land use change sequence model calculation process

能同时存在两种土地利用类型, 所以 $a \setminus b = \langle \{a, b\}, \{b, a\} \rangle$ 。土地利用变化序列模式比频繁模式更能精确地体现土地利用变化的序列特点。以 b, c 是两个 1-序列生成 2-序列为例, 频繁项集会生出支持度为 0.5 的频繁项集 $\langle b, c \rangle$, 然后生成 $b \rightarrow c, c \rightarrow b$ 两条序列, 支持度也都为 0.5, 根据图 2 和序列模式定义可知, 序列 $b \rightarrow c$ 的支持度为 0.33, 序列 $c \rightarrow b$ 的支持度为 0.2, 更符合土地利用变化的实际。而 3 序列的 id-list 表由 2 序列的 id-list 表生成, 根据定义, 只有共享前缀的 2 序列项才能进行时态连接, 如图 2 中 $\langle a, b \rangle$ 能与 $\langle a, c \rangle$ 连接, 而不能与 $\langle c, d \rangle$ 连接。 $\langle a, b \rangle$ 与 $\langle a, c \rangle$ 进行连接得到 $\langle a, bc \rangle$ 、 $\langle a, b, c \rangle$ 或 $\langle a, c, b \rangle$, 根据土地利用变化特点, 不会在同一时期存在 $\{bc\}$ 两种类型, 所以连接结果为 $\langle a, b, c \rangle$ 和 $\langle a, c, b \rangle$, 序列 $\langle a, b, c \rangle$ 代表了一种土地利用变化序列 $\langle a \rightarrow b \rightarrow c \rangle$ 。为了挖掘尽量多的序列模式, 本文把最小支持度设为 0, 具体的计算过程总结如图 2。

(1) 把土地利用变化序列数据库转化为垂直表示格式的 id-list 表。

(2) 第一次扫描产生 1-序列。

(3) 第 2 次扫描生成 2-序列, 同步构建格, 把相同前缀项的序列建在同一格内。

(4) 第 3 次扫描, 动态连接产生所有的序列, 枚举所有频繁 k -序列。在扫描过程中, 随着序列长度的增长, 表越来越小, 表连接的次数和候选序列的产生大大减小。

1.3 研究区概况

广西位于中国西南部, 是中国西南的出海通道, 具有沿边、沿海、沿江的区位优势。广西土地类型齐全, 地形四周高中间低。四周多被山地和高原环绕, 中间山地、丘陵、台地和平原交错, 南部沿海。进入新时代, 随着新型工业化、城镇化建设的深入推进, 广西的耕地面积逐年减少。本文将在以往的土地利用变化数据中, 利用序列模式发现研究区长期土地利用变化规律。

1.4 数据来源

本文数据来源于中国科学院资源环境科学数据中心 (<http://www.resdc.cn/Default.aspx>) 数据集目录中的中国土地利用遥感监测数据广西土地利用数据部分, 研究数据包含了 1970 年末、1980 年末、1995 年、2000 年、2005 年、2010 年、2015 年共 7 期土地利用类型数据, 空间分辨率为 1 km, 土地利用类型包括 22 个二级土地利用类型。广西在研究期内, 耕地面积呈现减少趋势, 仅在 1995—2000 年间出现小幅度增加, 林地面积变化有波动, 而建设用地逐年增加, 在 2010—2015 年间增加幅度尤其明显。具体各类型土地面积见表 1。本文以广西出现的 22 个二级土地利用类型进行土地利用类型变化序列模式分析, 验证模型的适用性, 探寻广西的土地利用变化规律。

2 结果分析

根据 7 期土地利用类型数据, 利用 GIS 的合并计算, 结合频度处理计算, 可以得到图 3 广西土地利用变化频度图, 图中一次变化区域对应序列模式的 2-序列集, 二次变化区域是土地变化序列模式的 3-序列集。图 3 能看出广西土地利用变化频次, 但还是计算不出研究区的土地利用类型变化去向和来源。在图 3 中, 叠加其他专题图, 广西土地利用变化区主要分布在道路、城镇和居民点周边。在表 1 中, 能看出各种土地类型的面积变化情况, 得到水田和旱地的面积在逐年减少、居民点和城镇用地逐年增加等规律, 也不能计算出各种土地类型的变化去向与来源。

本文首先做出了研究区所有相邻两期的二级土地利用变化转移矩阵。转移矩阵在两期之间的转换计算和量化方面非常清晰,但不能完整地求出多期长序列变化。利用提出的方法,分别对广西土地利用7期数据进行了7种项集种类的序列模式分析。在1970s—2015年间,共有15569个土地单元有土地利用类型变化,占全区土地的6.58%,并没有发生土地利用类型变化的占93.42%,研究土地利用类型变化序列只在发生了土地类型的变化数据中进行。在图4的1-序列计算结果中,支持度前5的分别是{有林地}、{其他林地}、{疏林地}、{高覆盖度草地}和{旱地},支持度分别为0.5109、0.3810、0.2333、0.2101和

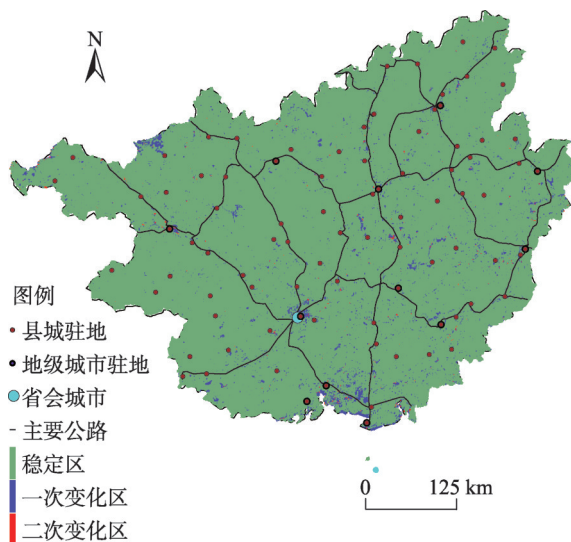


图3 广西土地利用变化区域分布

Fig. 3 Regional distribution of land use change in Guangxi

表1 1970s—2015年广西7期土地面积

Table 1 Land area of Guangxi during 1970s-2015

(km²)

土地利用类型	1970s	1980s	1995年	2000年	2005年	2010年	2015年
水田	25533	25379	25232	25225	25189	25131	24922
旱地	26415	26345	26411	26656	26533	26409	26151
有林地	85569	86045	86677	86275	85593	85513	84933
灌木林	37039	37066	36729	36720	36674	36614	36637
疏林地	29037	28692	29148	28826	28739	28539	28423
其他林地	3711	3565	3060	3598	4612	5094	5089
高覆盖度草地	18397	18602	18016	17990	17836	17699	17989
中覆盖度草地	3001	2957	3014	2960	2934	2890	2877
低覆盖度草地	259	110	112	118	110	108	110
河渠	1643	1651	1621	1608	1603	1622	1618
水库坑塘	1451	1482	1584	1593	1593	1702	1749
滩涂	192	150	96	321	311	316	312
滩地	287	275	235	293	343	294	290
城镇用地	412	552	789	811	913	986	1078
农村居民点	3256	3345	3388	3390	3385	3386	3416
其他建设用地	196	247	258	266	286	349	1046
沙地	8	8	8	8	8	7	7
盐碱地	6	3	3	3	3	3	7
湿地	3	3	3	3	2	2	2
裸地	14	14	14	14	14	14	17
裸岩石质地	15	11	11	11	11	11	11
海洋	210	213	309	26	33	28	28

0.1718。支持度和排序与频繁项集分析一致。依据定义可知, 1-序列支持度高, 说明它们在长期的土地利用变化中变化比较频繁, 参与变化其他土地利用类型的比例高, 是长期土地利用变化的主要类型, 在土地利用变化2-序列和3-序列等序列中也会是主要的变化参与类型。支持度高说明在广西近35年的土地利用变化单个类型中, 变化最大的是有林地, 其次是其他林地, 几种林业用地变化活跃。

在2-序列中, 支持度前4的分别是{有林地, 其他林地}、{其他林地, 有林地}、{疏林地, 有林地}和{高覆盖度草地, 有林地}, 其中支持度最高的2-序列{有林地, 其他林地}的支持度为0.2040, 说明在近35年广西土地利用类型有变化的土地单元中, 有20.4%的土地单元变化是由有林地变化成其他林地。2-序列{其他林地, 有林地}说明在近35年广西土地利用类型有变化的土地单元中, 又有6.99%的土地单元变化从其他林地变化成了有林地, 有林地与其他林地的相互转换较为频繁, 这条序列在频繁项集分析中是无法实现的。{有林地, 其他林地}和{其他林地, 有林地}两种2-序列的支持度之和与频繁项集2-项集{其他林地, 有林地}相等, 证明了计算方法的正确性, 同时又能挖掘出频繁项集不能挖掘的轨迹转换。在广西近35年的土地利用中, 由于鼓励桉树等成片有林地的种植, 把郁闭度为10%~30%的疏林地以及覆盖度>50%的高覆盖度草地种植成了用材林和经济林。在城镇化过程中, 为了满足城镇人口增长及城市扩张等需求, 需要将其他土地类型转换成城镇用地。2-序列{旱地, 城镇用地}和{水田, 城镇用地}的支持度分别为0.0283和0.0234, 两个2-序列的支持度排序分别为9和12, 说明城镇用地也是一种比较主要的土地利用转化类型, 也说明广西近35年的城镇化建设比较活跃。结合序列模式的时序特点, 说明广西城镇用地主要由旱地和水田、高覆盖度草地转换而来。由于广西以水稻为主要粮食作物, 出于人口需求与粮食安全的需要, 城镇用地的转换相对较多地源于旱地。农田开发为城镇用地后, 难以恢复为原质量耕地, 因此广西城镇用地主要由旱地和水田单向转换而来且不能相互变化。{旱地, 有林地}的支持度为0.0167, 说明广西有改变耕地用途、发展桉树等有林地行为。考虑到这一特性, 要求广西各级政府切实保护耕地, 未经批准不能改变土地用途, 加强用地评价和建设项目选址, 减少对耕地特别是优质农田的侵占, 严格落实占一补一、占优补优、占水田补水田的要求。

在长期的土地利用变化中, 2-序列一些变化成3-序列, 一些维持2次变化状态不变。广西土地利用变化类型中, 没有4-序列以上的变化单元, 说明在7期土地利用变化数据中, 没有任何土地单元有4期以上都存在变化的剧烈区域, 3-序列代表了研究期内的最终序列。结合3-序列支持度排序结果, {有林地, 其他林地}和{其他林地, 有林地}两个支持度高的2-序列的前缀按支持度分别是疏林地、高覆盖度草地、灌木林等, 说明林地之间以及林地和草地之间变化比较频繁, 广西在发展林业的过程中, 主要利用草地和林地之间的相互转换, 由水田、旱地和海洋等土地类型转换而来的支持度很低。{有林地, 其他

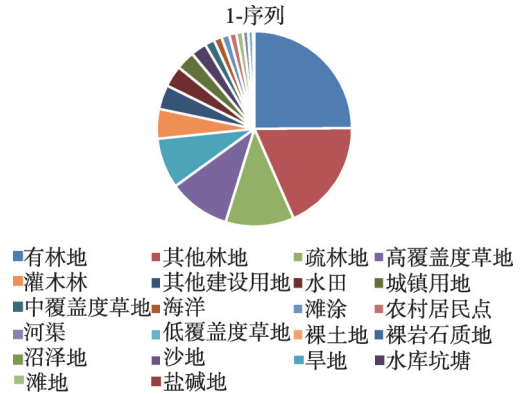


图4 广西土地变化1-序列

Fig. 4 1-sequence of land change in Guangxi

林地,其他建设用地}在3-序列中支持度排序第17,支持度为0.0007。由于广西矿产比较丰富,在广西林地区域有部分用地的变化轨迹为有林地→其他林地→其他建设用地,表明为了充分利用矿产资源,林地已经转换为厂矿、大型工业区、采石场等用地以及交通道路、机场及特殊用地类型。为了实现绿色可持续发展,需要广西全区各林区对废弃土地及重大建设项目临时用地组织实施土地复垦重大工程,有效增加林地面积和改善土地生态,防止土地利用的生态型问题。{水田,旱地,城镇用地}在包含有城镇用地的3-序列中排序最高,进一步说明广西城镇用地主要由耕地类型转换而来,并且从时间变化上看,水田先转换成旱地,旱地又转化为城镇用地。限于篇幅,在表2中只列出了广西土地利用类型变化支持度前40的2-序列和3-序列。

序列模式不但能分析出区域土地变化的整体情况,还能计算每种土地利用类型变化。本文计算了每种土地利用类型的变化序列,限于篇幅,仅列出水田的变化序列(表3)。从表3中可以看出,水田作为广西的一种主要的耕地类型,在近35年的变化中,都是作为一种主要的土地变化的流出类型,并没有任何一个土地单元变化成水田,这一点更加为广西耕地逐年减少提供了理论依据,同时反映出耕地利用类型一旦被改变就难以恢复。在长期的变化中,水田更多地变化为城镇用地、其他建设用地、农村居民点等城乡、工矿、居民建设用地类型,说明广西在有限的耕地资源条件下,还没有解决建设用地供需矛盾,由此反映出广西在农用地转用审批制度的执行、建设用地的供给途径,对耕地保护的强度以及对闲置土地的利用效率等方面存在一定问题。水田的流出变化类型还有很多其他土地类型,需要政府坚持最严格的耕地保护制度,守住耕地保护红线和粮食安全底线,确保广西现有耕地数量保持稳定。

支持度前10的2-序列后缀中,有3种有林地,2种其他林地,2种疏林地,1种高覆盖度草地,1种旱地,1种城镇用地。3-序列支持度前40的后缀中,有12种有林地、6种其他林地、5种疏林地、3种高覆盖度草地、3种其他建设用地,说明在近35年中,广西土地变化去向大多为有林地等林地类型。考虑到广西曾经有桉树种植、果树、苗圃等一系列林业种植补贴政策,从整体序列分析,广西土地变化的主要驱动因素是林业政策因素。借助转移矩阵,广西林业用地在2000—2005年间最为剧烈,其次为2005—2010年。

3 结论与讨论

3.1 结论

本文在土地利用变化分析方法和研究区土地利用变化结果有三个主要结论:

(1) 给出了土地利用变化序列数据库定义和基于序列模式的土地利用变化序列模型计算方法。该方法能解决多期土地利用变化整体序列分析,从而科学地找到研究区域的土地变化规律。由于土地利用变化的不可逆性,序列模式还能解决频繁模式不能导出具有时间先后顺序的土地利用变化序列的弊端。土地变化序列模式不但能分析研究区域土地利用的整体变化情况,还能计算每种土地类型的变化序列。在土地利用变化期越多、土地利用类型越多的研究中,土地变化序列模式越能体现出优越性。

(2) 近35年来,广西有6.58%的区域发生了土地利用类型变化,变化区域主要集中在道路及城镇、居民点周边。广西土地变化1-序列支持度前4的分别是{有林地}、{其他林地}、{疏林地}和{高覆盖度草地},支持度分别为0.5109、0.3810、0.2333和0.2101。2-序列支持度前4的分别是{有林地,其他林地}、{其他林地,有林地}、{疏林地,有林地}

表2 广西土地利用变化支持度前40的2-序列和3-序列

Table 2 The top 40 2-sequence and 3-sequence of land use change support degree in Guangxi

2-序列	支持度	3-序列	支持度
有林地, 其他林地	0.2040	疏林地, 其他林地, 有林地	0.0065
其他林地, 有林地	0.0699	疏林地, 有林地, 其他林地	0.0044
疏林地, 有林地	0.0640	高覆盖度草地, 有林地, 其他林地	0.0031
高覆盖度草地, 有林地	0.0529	有林地, 其他林地, 疏林地	0.0027
疏林地, 其他林地	0.0444	高覆盖度草地, 其他林地, 有林地	0.0025
有林地, 疏林地	0.0420	灌木林, 其他林地, 有林地	0.0024
有林地, 高覆盖度草地	0.0405	其他林地, 疏林地, 有林地	0.0013
灌木林, 旱地	0.0349	高覆盖度草地, 疏林地, 有林地	0.0011
旱地, 城镇用地	0.0283	有林地, 其他林地, 高覆盖度草地	0.0011
高覆盖度草地, 疏林地	0.0282	灌木林, 疏林地, 其他林地	0.0010
旱地, 其他建设用地	0.0259	高覆盖度草地, 疏林地, 其他林地	0.0010
水田, 城镇用地	0.0234	中覆盖度草地, 高覆盖度草地, 有林地	0.0009
高覆盖度草地, 其他林地	0.0199	其他林地, 有林地, 疏林地	0.0009
其他林地, 疏林地	0.0187	有林地, 其他林地, 灌木林	0.0009
灌木林, 其他林地	0.0174	水田, 旱地, 水库坑塘	0.0008
海洋, 滩涂	0.0169	高覆盖度草地, 疏林地, 旱地	0.0008
疏林地, 高覆盖度草地	0.0169	有林地, 其他林地, 其他建设用地	0.0007
旱地, 有林地	0.0167	其他林地, 有林地, 灌木林	0.0007
水田, 其他建设用地	0.0155	有林地, 河渠, 水库坑塘	0.0006
旱地, 疏林地	0.0126	高覆盖度草地, 有林地, 旱地	0.0006
低覆盖度草地, 高覆盖度草地	0.0125	疏林地, 其他林地, 灌木林	0.0006
有林地, 灌木林	0.0120	有林地, 疏林地, 高覆盖度草地	0.0006
有林地, 其他建设用地	0.0118	疏林地, 有林地, 高覆盖度草地	0.0006
水田, 农村居民点	0.0104	高覆盖度草地, 灌木林, 有林地	0.0005
灌木林, 有林地	0.0099	低覆盖度草地, 高覆盖度草地, 有林地	0.0005
高覆盖度草地, 灌木林	0.0097	高覆盖度草地, 有林地, 疏林地	0.0005
水田, 水库坑塘	0.0091	灌木林, 其他林地, 疏林地	0.0005
高覆盖度草地, 旱地	0.0088	疏林地, 灌木林, 其他林地	0.0005
疏林地, 其他建设用地	0.0086	疏林地, 其他林地, 其他建设用地	0.0005
旱地, 其他林地	0.0085	高覆盖度草地, 其他林地, 疏林地	0.0004
水库坑塘, 滩地	0.0078	高覆盖度草地, 灌木林, 其他林地	0.0004
高覆盖度草地, 其他建设用地	0.0078	有林地, 疏林地, 其他建设用地	0.0004
其他林地, 高覆盖度草地	0.0078	滩涂, 其他建设用地, 海洋	0.0004
其他林地, 灌木林	0.0076	水田, 旱地, 城镇用地	0.0004
旱地, 农村居民点	0.0073	低覆盖度草地, 灌木林, 中覆盖度草地	0.0003
旱地, 水库坑塘	0.0069	疏林地, 高覆盖度草地, 有林地	0.0003
有林地, 旱地	0.0063	旱地, 其他林地, 有林地	0.0003
其他林地, 旱地	0.0062	灌木林, 旱地, 有林地	0.0003
滩涂, 海洋	0.0062	灌木林, 高覆盖度草地, 有林地	0.0003
中覆盖度草地, 高覆盖度草地	0.0058	有林地, 疏林地, 水库坑塘	0.0003

和{高覆盖度草地, 有林地}, 支持度分别为0.2040、0.0699、0.0640和0.0529。土地利用变化序列从长期序列变化的角度说明广西土地变化主要发生在林地类型之间, 部分林地转换为厂矿、采石场、交通道路等建设用地类型; 城镇用地主要由旱地和水田转换而来; 在研究期内没有任何一个土地单元转化成水田。

(3) 从变化去向看, 广西2-序列变化去向频次高低依次为有林地、其他林地、疏林地、高覆盖度草地、旱地和城镇用地。3-序列变化去向频次高低依次为有林地、其他林地、疏林地、高覆盖度草地和其他建设用地, 说明广西近35年土地变化的驱动首要因素是林业政策因素, 其次为城镇化建设因素。

3.2 讨论

土地利用变化序列模式解决了长期土地变化的序列问题, 并没有实现对研究区域土地变化的预测问题, 下一步的工作, 将基于土地变化序列模式计算土地转换的概率, 实现对研究区域土地变化的预测。本文提出的基于序列模式的土地利用变化分析方法, 从整体上能计算出土地变化的序列特点, 该条序列具体在哪个时间段变化最剧烈等细节还要借助土地变化转移矩阵, 更好的分析方法应该是两种方法以及与其他方法的融合, 这样才能全面地分析一个地区的土地变化问题。在土地利用变化中, 人文与政策是两种主要的因素, 下一步将充分考虑这两种因素对土地变化的影响, 结合经典的CLUMondo模型和MAS等模型, 实现研究区的土地利用变化情景模拟分析。

根据研究结果, 提出以下政策建议:

(1) 加强对生态环境的保护。广西林地资源丰富, 生态环境脆弱, 要完善林地管理及保护等相关制度, 通过技术手段、法律手段等多种方式强化对林地的监管, 取消桉树等速生林种植补贴。严格执行生态环境保护规划, 积极开展岩溶山区土地石漠化治理, 减缓水土流失。

(2) 加强土地利用规划和监管。严格评估建设用地选址, 合理规划, 将土地利用规划、环境保护规划、林地与耕地保护规划、水资源规划等有机结合, 利用遥感技术实时

表3 广西水田变化轨迹

Table 3 The changing track of paddy fields in Guangxi

序列	支持度
水田, 城镇用地	0.02339
水田, 其他建设用地	0.01553
水田, 农村居民点	0.01042
水田, 水库坑塘	0.00909
水田, 其他林地	0.00246
水田, 旱地	0.00208
水田, 高覆盖度草地	0.00114
水田, 有林地	0.00104
水田, 旱地, 水库坑塘	0.00076
水田, 滩地	0.00057
水田, 河渠	0.00057
水田, 海洋	0.00057
水田, 疏林地	0.00047
水田, 旱地, 城镇用地	0.00038
水田, 中覆盖度草地	0.00028
水田, 海洋, 水库坑塘	0.00028
水田, 灌木林	0.00028
水田, 其他建设用地, 城镇用地	0.00019
水田, 其他林地, 有林地	0.00009
水田, 滩涂	0.00009
水田, 海洋, 滩涂	0.00009
水田, 水库坑塘, 滩地	0.00009
水田, 高覆盖度草地, 水库坑塘	0.00009
水田, 水库坑塘, 其他建设用地	0.00009
水田, 城镇用地, 其他建设用地	0.00009
水田, 裸土地	0.00009
水田, 其他林地旱地	0.00009
水田, 水库坑塘, 海洋	0.00009
水田, 水库坑塘, 高覆盖度草地	0.00009
水田, 水库坑塘, 城镇用地	0.00009

监测, 严格控制城镇无序扩张和乱占用农用地。明确各部门责任, 适当纳入绩效考核, 对于失责行为要采取一定的惩罚措施。

(3) 完善各项制度。进一步完善耕地占补平衡制度、土地征收制度、农用地转用审批制度、建设用地的供给制度和生态补偿制度, 开展清理擅自改变耕地用途种植桉树的行动。全方面保障生态文明建设与高质量城镇化的推进。

(4) 保障粮食安全。守住耕地红线不动摇, 提高耕地的利用效率及农业生产效率, 保证人口增长对粮食的需要。通过生物、工程等技术对土地展开复垦, 提高耕地质量, 增加后备耕地数量, 加大力度治理非法占用耕地的行为, 严格限制农用地转为非农用地。

参考文献(References):

- [1] VITOUSK P M, MOONEY H A, LUBCHENCO J, et al. Human domination of earth's ecosystems. *Science*, 1997, 277(5325): 494-499.
- [2] LIU J Y, KUANG W H, ZHANG Z X, et al. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *Journal of Geographical Sciences*, 2014, 24(2): 195-210.
- [3] STERLING S M, DUCHAME A, POLCHER J. The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change*, 2012, 3(4): 385-390.
- [4] BROWN D G, VERBURG P H, PONTIUS R G, et al. Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Environmental Sustainability*, 2013, 5(5): 452-457.
- [5] ROBINSON D T, VITTORIO A D, ALEXANDER P, et al. Modelling feedbacks between human and natural processes in the land system. *Earth System Dynamics Discussions*, 2017, 2: 1-47.
- [6] VERBURG P H, SOEPBOER W, VELDKAMP A, et al. Modeling the spatial dynamics of regional land use: The clue-model. *Environmental Management*, 2002, 30(3): 391-405.
- [7] 罗娅, 熊康宁, 李永奎, 等. 石漠化治理区土地利用变化安全性评价: 以花江、红枫湖、鸭池石漠化治理区为例. *自然资源学报*, 2019, 34(3): 600-612. [LUO Y, XIONG K N, LI Y Y, et al. Evaluation on security of land use change in rocky desertification control areas: Taking Huajiang, Hongfenghu and Yachi as examples. *Journal of Natural Resources*, 2019, 34(3): 600-612.]
- [8] 权瑞松. 基于情景模拟的上海土地利用变化预测及其水文效应. *自然资源学报*, 2018, 33(9): 80-90. [QUAN R S. Prediction of land use change and its hydrological effect in Shanghai based on scenario simulation. *Journal of Natural Resources*, 2018, 33(9): 80-90.]
- [9] VERBURG P H, DEARING J A, DYKE J G, et al. Methods and approaches to modelling the Anthropocene. *Global Environmental Change*, 2016, 39: 328-340.
- [10] OMETS Muller C, VERBURG P H, HEINIMANN A. Scenarios of land system change in the Lao PDR: Transitions in response to alternative demands on goods and services provided by the land. *Applied Geography*, 2016, 75: 1-11.
- [11] BROWND G, ROBINSON D T. Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology & Society*, 2006, 11(1): 709-723.
- [12] AGRAWAL R. Mining association rules between sets of items in large databases. *Acm Sigmod Record*, 1993, 22(2): 207-216.
- [13] 廖伟华, 聂鑫. 基于大数据的城市服务业空间关联分析. *地理科学*, 2017, 37(9): 25-32. [LIAO W H, NIE X. Spatial association analysis for urban service based on big data. *Scientia Geographica Sinica*, 2017, 37(9): 1310-1317.]
- [14] HU W, CHEN T, SHAH S L. Detection of frequent alarm patterns in industrial alarm floods using itemset mining methods. *IEEE Transactions on Industrial Electronics*, 2018, 65(9): 7290-7300.
- [15] HUANG Y, PEI J, XIONG H. Mining co-location patterns with rare events from spatial data sets. *GeoInformatica*, 2006, 10: 239-260.
- [16] LESLIE T F, FRANKENFELD C L, MAKARA M A. The spatial food environment of the DC metropolitan area: Clustering, co-location, and categorical differentiation. *Applied Geography*, 2012, 35(1-2): 300-307.
- [17] WU X, ZHAN F B, ZHANG K, et al. Application of a two-step cluster analysis and the apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China. *Environmental Earth Sciences*, 2016, 75(2): 146.
- [18] SRIKANT R, AGRAWAL R. Mining quantitative association rules in large relational tables. //Proceedings of 1996 AC-

- MSIGMOD International Conference on Management of Data (SIGMOD'96), Montreal, Canada, 1996, 1-12.
- [19] WANG L, HU K, TAO K, et al. Mining frequent trajectory pattern based on vague space partition. Knowledge-Based Systems, 2013, 50: 100-111.
- [20] PITARCH Y, LENCO D, VINTRON E, et al. Spatio-temporal data classification through multidimensional sequential patterns: Application to crop mapping in complex landscape. Engineering Applications of Artificial Intelligence, 2015, 37(37): 91-102.
- [21] ZAKI M J. Spade: An efficient algorithm for mining frequent sequences. Machine Learning, 2001, 42(1-2): 31-60.

Analysis of land use change based on sequence model: Taking Guangxi Zhuang Autonomous Region as an example

LIAO Wei-hua¹, NIE Xin², JIANG Wei-guo³

(1. College of Mathematics and Information Science, Guangxi University, Nanning 530004, China; 2. School of Public Administration, Guangxi University, Nanning 530004, China; 3. State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China)

Abstract: Land use change is a long-term process with certain complexity. The traditional transfer matrix method can only compare the two phases of land use data, but cannot summarize the long-term overall change law. The frequent itemsets method cannot derive the sequence trajectory. To overcome the shortcomings of these methods, this paper proposes a land use change model based on sequence model. According to the actual characteristics of land sequence data and the vertical format sequence pattern SPADE algorithm, taking Guangxi, China as an example, we calculated the sequence of 22 secondary land use types in 7 periods from 1970s to 2015. The results show that, in the 35 years, land use types changed in 6.58% of Guangxi, and the change areas were mainly concentrated in roads, towns and settlements; the top 3 types of land use change 1-sequence support degree in Guangxi are {wood land}, {other wood land} and {sparse forest land}, and the support degree values are 0.5109, 0.3810 and 0.2333, respectively. The top 3 types of the 2-sequence support degree are {wood land, other wood land}, {other wood land, wood land} and {sparse forest land, wood land}, and the support degree values are 0.2040, 0.0699 and 0.0640, respectively. The top 3 types of the 3-sequence support degree are {sparse forest land, other wood land, wood land}, {sparse forest land, wood land, other wood land} and {high coverage grassland, wood land, other wood land}, and the support degree values are 0.0065, 0.0044 and 0.0031, respectively. The land use change 3- sequence {with forest land, other woodland, and other construction sites} has a support degree of 0.0007. The land use change in the study area mainly occurs in forest land, and some forest land is converted into construction land types such as mines, quarries and traffic roads; urban land is mainly converted from dry land and paddy fields; there is no land unit converted into paddy fields during the study period. The land use change sequence analysis model proposed in this paper can calculate the sequence of multi-period land use change as a whole, and make up for the shortcomings of studying land use change from a long-term scale.

Keywords: land use change; sequence model; SPADE algorithm; long term change; Guangxi