

第二次土地资源调查数据建库中的 土地利用图综合缩编

艾廷华¹ 杨帆¹ 李精忠¹

(1 武汉大学资源与环境科学学院, 武汉市珞喻路 129 号, 430079)

摘 要:结合实际工作讨论了土地利用图综合缩编的技术路线,包括图斑综合前的最邻近地块探测、综合规则建立、图斑合并的方法、综合结果的面积平衡性评价等。
关键词:土地资源调查;地图综合;土地利用图;尺度变换
中图法分类号:P283.1; P273

随着第二次土地资源调查工作的展开,前期基于详查的成果 1:1 万土地利用现状图各地陆续完成,调查工作逐渐转移到后期的成果整理建库,由县级调查成果向地市级、省级、国家级成果库汇总。此时,要反映整个区域的土地利用状况,大比例地图是不能满足要求的,因此就需要将大比例尺土地利用图综合到小比例尺图,如市级单位的土地利用挂图比例尺为 1:5 万至 1:25 万等。与普通地形图相比,土地利用图的综合缩编不仅涉及降低空间分辨率的图斑形状化简,还包括降低语义分辨率的专题属性层次的归并,其数据综合既应该考虑专题属性上的语义相近度,也要考虑到空间上的几何邻近程度,要将这两方面因素结合起来综合处理^[1]。综合的结果不仅仅要达到地物目标数据个数减少、几何形状简化,同时保持其数据的全覆盖、无重叠特征,最重要的应在统计上保持原有各类类型用地面积的相对百分比大体上不变。本文结合实际工作讨论了土地利用图综合缩编的技术路线,包括综合规则的建立、邻近图斑的探测与合并、综合结果的面积平衡性评价等。

1 土地利用数据邻近分析

1.1 邻近关系的影响因素

将邻近的多边形图斑合并是土地利用数据综

合中的基本规则,然而如何判断邻近是一项复杂的决策,众多学者对该问题进行了研究^[1-5]。在土地利用数据中,相邻的图斑邻近程度不一定高,要结合专题属性以及语义邻近程度一同进行分析。语义邻近程度很高的,最终的邻近程度不一定最高。如图 1 所示,在该图中, A 代表有林地, B 代表村庄, C 和 D 代表旱地, E 代表果园(地块 A 、 B 、 C 、 D 和 E 同属于一个行政单位),其语义层次树如图 2 所示。与地块 A 有共享边的地块为 B 、 D 和 E 。虽然对于地块 A 自身来说,地块 E 具有最长的共享边边长,但是由于 A 、 B 和 D 在语义上同属于农用地,而且地块 B 与 A 的共享边边长是最小的,则地块 A 与 D 、 E 之间的邻近程度更高些。地块 A 、 D 和 E 分别属于林地、园地和耕地,在这种情况下,认为林地和园地在语义上的邻近程度要高些,即地块 A 与地块 E 在语义上的邻近度要高于 A 与 D 之间的邻近度,而且 A 与 E 之间的共享边边长要大于 A 与 D 之间的共享边边长,因此经过语义以及拓扑邻接上的考虑,地块 A 与 E 的邻近程度最高。地块 D 与其他三个地块都有邻接关系,但是地块 C 与 D 同属于旱地,认为与地块 D 邻近程度最高的地块为 C 。

1.2 语义邻近度计算

在第二次全国土地调查分类中的《土地利用现状分类》采用二级分类体系,一级类 12 个,二级类

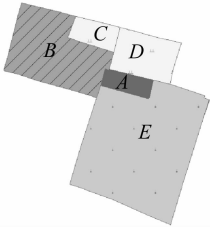


图 1 图斑邻近程度示例

Fig. 1 An Example of Neighborhood Degree

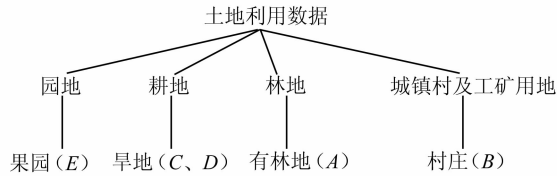


图 2 用地类型层次树

Fig. 2 Hierarchical Tree of Land Use Classification

57 个。在农村土地调查中,部分类型进行了归并,经过调整后,一级类有 8 类,二级类有 38 类。这些类型之间的相似性或语义邻近关系如何,是土地利用图综合前决策分析的基础,该问题属语义距离的量化计算,也是目前研究的热点问题。这里给出一种基于有序量的语义邻近度计算方法,参考地块与某地块的语义邻近度记为 $\text{sem}(X, Y_i)$ 。

在语义层次树的某级类型划分中,设当前考察的用地类型为 X ,其他同级用地类型与 X 在语义上的邻近关系可以根据土壤成分差异、发育过程差异、功能差异作一个排序,得到其他用地类型与 X 的语义邻近等级。这里以耕地为例进行讨论,与耕地在语义上邻近程度由高向低排序为园地、林地、草地、城镇村及工矿用地、交通用地、其他土地、水域及水利设施用地。在二级类中,同样可以设置各一级类下属的用地类型子类的语义邻近程度并排序,以一级类耕地中的水田、水浇地、旱地 3 个二级用地类型为例,与水田在语义上邻近程度由高向低排序为水浇地、旱地;与水浇地在语义上邻近程度由高向低排序为水田、旱地;与旱地在语义上邻近程度由高向低排序为水田、水浇地。其他地类也如上述排序建立各类的语义邻近关系。

参考于某种用地类型,其他各用地类型排好序后,在“等级量”上确定了其间的邻近级别,但在“差异量”层次上的量化描述还有待于进一步加强。相对于 X 而言, Y_i 比 Y_j 在语义上更邻近,但邻近的差异程度是多少需要进一步计算。对于排好序的语义类型集 $\{Y_i\}$,相邻两元素间的语义距离规定为 1 个单位,则语义距离按自然数列 1, 2, 3, ..., i ... 递增,所有语义距离的累积和为 Σ ,则

排在第 i 个位置的语义邻近度为 $\text{sem}(X, Y_i) = 1 - i / \Sigma$ 。注意,距离越大,相近性关系越疏远。以耕地下属的 3 个二级类型为例,按照水田的语义邻近排列顺序(水浇地、旱地),将水浇地与水田的语义距离设为 1,旱地与水田的语义距离设为 2,则水浇地相对于水田的语义邻近度为 $2/3$,旱地相对于水田的语义邻近度为 $1/3$ 。

1.3 空间邻近度计算

对于空间位置上的邻近关系,通过拓扑邻近共享边的长短作为定量化计算指标,将其归一化到 $[0, 1]$ 上表达,可表示为 $\text{dis}(X, Y_i) = l_{Y_i} / C_X$,其中, l_{Y_i} 为 Y_i 与当前考察图斑 X 的共享边长度, C_X 为 X 的周长。如当两个图斑与当前考察小图斑的共享边边长分别为 2 cm 和 3 cm,则这两个图斑与小图斑的空间邻近度值分别为 0.4 和 0.6。

1.4 综合邻近度计算

土地利用图上两图斑多边形的邻近关系既要考虑语义上的邻近性,又要考虑空间位置上的差异性,如何将两者集成得到统一的邻近度是土地利用图综合的前提,也是一个难点。两种因素同时作用于邻近关系时,首先是如何确定两种因素的可比性,空间距离和语义差异是完全不同的变量域,图面上 1 cm 的空间差距相当于多少个语义上的等级差异难以回答。其次是联合作用函数为怎样的形式,定性上可以认为邻近强弱随着距离的增大而减小,随着语义层次树中结点距离的增大而减小,代数中有众多这种递减函数,选用何种函数更合适有待研究。本研究深入分析了空间距离与语义层次对邻近关系的影响后,提出了一种可操作的综合性邻近度计算方法,寻找当前考察图斑 X 的最邻近图斑 Y_k 。

定义图斑 Y_i 与 X 的综合邻近度为 $\text{adj}(X, Y_i)$,它由语义邻近与空间邻近联合决定,表示为 $\text{adj}(X, Y_i) = \text{sem}(X, Y_i) W_1 + \text{dis}(X, Y_i) W_2$ 。其中, W_1 和 W_2 分别为语义邻近和空间邻近对综合邻近度贡献作用的权重值。 X 的最邻近地块 Y_k 满足 $\text{adj}(X, Y_k) = \max\{\text{adj}(X, Y_i)\}$ 。

设与 X 存在拓扑邻近关系的图斑集为 $\{Y_i\} (i = 1, 2, 3, \dots, N)$,根据 $\{Y_i\}$ 中的图斑与 X 图斑的语义关系的层次差异,对该集合分解为几个子集的并集。首先定义两个函数: $f_2(X)$ 为返回图斑 X 的二级用地类型代码; $f_1(X)$ 为返回图斑 X 的一级用地类型代码。建立 $\{Y_i\}$ 集合的一个划分:

$$\{Y_i\} = \{Y_{2i}\} \cup \{Y_{1i}\} \cup \{Y_{0i}\}$$

其中, $\{Y_{2i}\}$ 为集合 $\{Y_i\}$ 中与图斑 X 为同一种二级

用地类型的图斑集合,满足 $f_2(X) = f_2(Y_{2i})$; $\{Y_{1i}\}$ 为集合 $\{Y_i\}$ 中与图斑 X 为不同二级用地类型,但是为相同一级用地类型的图斑集合,也就是同父亲类型,满足 $f_2(X) \neq f_2(Y_{1i})$, 且 $f_1(X) = f_1(Y_{1i})$; $\{Y_{0i}\}$ 为集合 $\{Y_i\}$ 中与图斑 X 既不同二级用地类型,也不同一级用地类型的图斑集合,即 $f_2(X) \neq f_2(Y_{0i})$, 且 $f_1(X) \neq f_1(Y_{0i})$ 。在图 3 中,考察当前地块 X 为水田,存在 6 个邻近地块,按照上述划分,有 $\{Y_{2i}\} = \{Y_1, Y_2\}$, $\{Y_{1i}\} = \{Y_3, Y_4\}$, $\{Y_{0i}\} = \{Y_5, Y_6\}$ 。这里对邻近地块分为三组, $\{Y_{2i}\}$ 与 X 在用地类型的语义关系上最密切, $\{Y_{1i}\}$ 次之, $\{Y_{0i}\}$ 再次之。寻找 X 的最邻近地块,按该顺序依次搜寻,分四种情形进行讨论。

1) 当 $\{Y_{2i}\} \neq \emptyset$, 即集合 $\{Y_i\}$ 存在与图斑 X 为相同二级用地类型的图斑,则最邻近地块 Y_k 只能在 $\{Y_{2i}\}$ 中产生,该子集中的地块与其他地块相比,在语义邻近上强烈占优,几乎可以不考虑空间邻近的影响。取 $\{Y_{2i}\}$ 中与 X 空间邻近最强的 Y_k , 满足 $\text{dis}(X, Y_k) = \max\{\text{dis}(X, Y_{2i})\}$ 。如图 3 所示,与图斑 X (水田)相邻的图斑有 Y_1 (水田)、 Y_2 (水田)、 Y_3 (水浇地)、 Y_4 (旱地)、 Y_5 (茶园)、 Y_6 (有林地),此时, $\{Y_{2i}\} \neq \emptyset$, 即存在图斑 Y_1 和 Y_2 与 X 为同一种二级用地类型的水田,则 X 的最邻近地块只能在 Y_1 和 Y_2 中产生,进一步比较 Y_1 与 X 的共享边长,得到 X 的最邻近地块为 Y_1 。

2) 当 $\{Y_{2i}\} = \emptyset$ 且 $\{Y_{1i}\} \neq \emptyset$ 时,即集合 $\{Y_i\}$ 不存在与图斑 X 有相同二级用地类型的图斑,但存在有相同一级用地类型的图斑。 $\{Y_{1i}\}$ 中的元素与 X 的语义邻近程度一般,在邻近度计算式 $\text{adj}(X, Y_{1i}) = \text{sem}(X, Y_{1i})W_1 + \text{dis}(X, Y_{1i})W_2$ 中,认为语义邻近和空间邻近两个因素具有同

等重要的作用,设权值为 $W_1 = W_2 = 0.5$ 。取 $\{Y_{1i}\}$ 中与 X 综合邻近最强的 Y_k , 满足 $\text{adj}(X, Y_k) = \max\{\text{adj}(X, Y_{1i})\}$ 。如图 4 所示,与图斑 X (水田)相邻的图斑有 Y_1 (水浇地)、 Y_2 (旱地)、 Y_3 (茶园)、 Y_4 (有林地),此时, $\{Y_{2i}\} = \emptyset$ 且 $\{Y_{1i}\} \neq \emptyset$, 即不存在图斑与 X 为同一种二级用地类型,但存在 Y_1 和 Y_2 与 X 为同种一级用地类型。 X 的最邻近地块只能在 Y_1 和 Y_2 中产生, Y_2 与 X 的语义邻近度低,但空间邻近度高, Y_1 与 X 的语义邻近度高,但空间邻近度低,按上述公式综合计算得到 X 的最邻近地块为 Y_2 。

3) 当 $\{Y_{2i}\} = \emptyset$ 且 $\{Y_{1i}\} = \emptyset$, 而 $\{Y_{0i}\} \neq \emptyset$ 时,即集合 $\{Y_i\}$ 不存在与图斑 X 有相同二级用地类型的图斑,也不存在有相同一级用地类型的图斑,这是最一般性的情形。由于 $\{Y_{0i}\}$ 中的地块与 X 的语义邻近度很弱,综合邻近度的计算主要取决于空间邻近,在 W_1 、 W_2 权重值的分配上应减小语义邻近所占的权重,使得 $W_1 < W_2$, 如取值 $W_1 = 0.3$, $W_2 = 0.7$, 代入公式计算后,取 $\{Y_{0i}\}$ 中与 X 综合邻近最强的 Y_k , 使其满足 $\text{adj}(X, Y_k) = \max\{\text{adj}(X, Y_{0i})\}$ 。如图 5 所示,与图斑 X (水田)相邻的图斑有 Y_1 (茶园)、 Y_2 (有林地),既不存在图斑与 X 有同一种二级用地类型,也不存在与 X 有同种一级用地类型的图斑,按上述公式计算出 X 的最邻近地块为 Y_2 , 这里 Y_1 、 Y_2 相对 X 的空间邻近关系相同,由于语义邻近上 Y_2 强于 Y_1 , 最后得到最邻近地块为 Y_2 。

4) 当 $\{Y_{1i}\} = \emptyset$, 且 $\{Y_{2i}\} = \emptyset$, $\{Y_{3i}\} = \emptyset$ 时,表明图斑 X 为孤立地块,这样的地块按需要进行取舍,根据选取资格限值决定,该情形如图 6 所示。

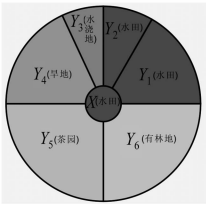


图 3 $\{Y_{2i}\} \neq \emptyset$
拓扑邻近示例
Fig. 3 An Example of Neighborhood Relationship with $\{Y_{2i}\} \neq \emptyset$

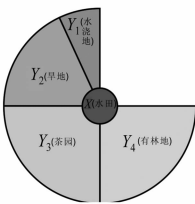


图 4 $\{Y_{2i}\} = \emptyset$ 且 $\{Y_{1i}\} \neq \emptyset$
的拓扑邻近示例
Fig. 4 An Example of Neighborhood Relationship with $\{Y_{2i}\} = \emptyset$ and $\{Y_{1i}\} \neq \emptyset$

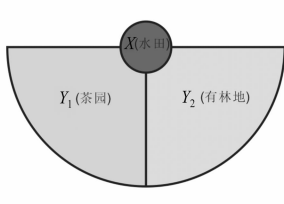


图 5 $\{Y_{2i}\} = \emptyset$, $\{Y_{1i}\} = \emptyset$ 且 $\{Y_{0i}\} \neq \emptyset$ 的拓扑邻近示例
Fig. 5 An example of Neighborhood Relationship with $\{Y_{2i}\} = \emptyset$ and $\{Y_{1i}\} = \emptyset$ and $\{Y_{0i}\} \neq \emptyset$



图 6 孤立图斑
的示例
Fig. 6 An Example of No Neighbor

2 图斑综合化简

土地利用图综合在操作上表现为小图斑删

除,寻找其最邻近地块将小图斑合并。在图斑合并的过程中,由小图斑归并生成的新图斑还要判断是否符合选取资格的要求,如果不符合,则继续将这个新图斑也加到待删除小图斑集合中。因

此,小图斑的处理是一个反复的过程,直到图斑面积均满足选取指标。

2.1 各用地类型选取指标设置

在土地利用数据的图斑处理中,选取指标限值有较大的差别。按照二调的要求,城镇村及工矿用地的选取指标为 4.0 mm^2 ,耕地、园地为 6.0 mm^2 ,林地、草地等其他地类为 15.0 mm^2 。这些选取指标应根据各个地区具体情况的不同作适当的调整,如在山区,耕地很少,为了在图面上表现出这些耕地地块,需要相应地减小耕地的选取指标。

2.2 小图斑排序与合并

在进行图斑合并操作之前,要对小图斑集 $\{X_i\}$ 进行排序,其原因是地块合并的先后顺序对最后地类类型面积的统计有影响。如地块 X_1 为水田, X_2 为有林地,如果 X_1 先进行合并,且在所有地块中, X_2 与 X_1 的邻近程度最高,则 X_1 将合并到 X_2 中,形成一个新的地块 X (类型为有林地)。反之,如果 X_2 先进行合并,且在所有地块中, X_1 与 X_2 的邻近程度最高,则最后生成的新地块用地类型为水田。这两种合并顺序产生的效果如图 7 所示。

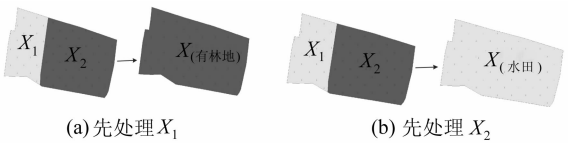


图 7 地块合并示例
Fig. 7 An Example of Combination

对一个地区来说,最不重要的用地类型就应该先进行处理,如裸地就要先进行图斑的合并,果园、水田等就要后处理。各个地区根据自身的具体情况,也可以相应作出调整。综合的基本过程为:查询出不满足选取指标的小图斑 $\{X_i\}$,逐一扫描,寻找当前考察小图斑的最邻近图斑,然后将小图斑合并到该图斑中,如果小图斑的用地类型和将要合并的图斑的用地类型相同,则属性保持不变,如果用地类型不同,则新生成的图斑属性将以合并到的图斑为主。

2.3 相近邻近度图斑的剖分

上述寻找最邻近地块的结果是惟一的,但是对于 1.4 节中的情形 3), 当与 a 邻接的多个多边形 $\{b_1, b_2, \dots\}$ 分别与 a 共享边界长度相近,这时从长度值数量比较上寻求最邻近地块已意义不大,可以采取中剖分的办法,将次要地块均分后分

别归并到邻近地块中。具体的剖分算法与处理过程见文献[1,7]。

3 综合后评价

土地利用图综合的决策与实施主要是基于小图斑的处理,是微观层次上的数据加工,完成的结果在宏观层次上的数据分布特征是否得以保持,需要进行后续评价,其中一个主要的评价内容便是各类用地面积平衡的变化情况。综合操作中,各类用地图斑的生存是一个相互竞争的过程,某类型下的部分地块为其他地块合并,同时部分地块也合并其他地块,总面积分别呈减小、增大的变化趋势,理想的情况是两种趋势抵消,面积保持不变。实际情况取决于各类用地图斑的大小与分布模式,往往出现某类用地为散落分布的小地块被其他地块合并的机会更多,使得该类地块综合后,面积严重减小。通过土地用地面积平衡评价,发现变化显著的类型,然后适当修改综合规则,调整选取指标,使得土地用地面积平衡。

以 1:1 万到 1:5 万的土地利用图综合为例,用笔者研制的软件对某县的数据进行综合^[8]。原始数据中,有 69 226 个图斑,按照前文提到的方法,综合到 1:5 万之后,剩下 45 361 个图斑,再进行邻近图斑合并操作后,剩下 30 269 个图斑。这里说的邻近图斑合并指的是在同一个行政区范围内(如村),对于相邻的图斑,如果用地类型、权属性质等属性都相同,则将其合并在一起。图 8 是综合前后的对比效果截图。

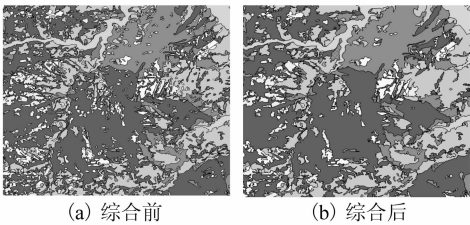


图 8 综合前后的截图
Fig. 8 An Example of Comparison Before and After Generalization

经过综合操作后,各类用地类型的面积变化不大,各类用地类型之间的面积比例基本保持平衡,而且图斑数量有大幅度的减少。其中,田坎和沙地变化达 100%,这是因为田坎作了中轴化操作,而沙地只有一个图斑,且面积较小,属于正常的变化。

4 结 语

土地利用图的综合需同时兼顾空间特征与语义特征,其核心是邻近关系的判断。本文基于空间邻近与语义邻近分析,提出了兼顾两种影响因素的邻近度计算方法。对于土地利用图的综合操作,在前期的研究基础上给出了小图斑的合并和剖分方法,针对实际土地利用数据进行了算法实验。在综合结果评价上,通过土地用地各类型面积平衡的计算发现不合理的情形,通过修改综合规则,使各类面积达到平衡。

第二次土地资源调查后期的成果整理建库需要基于 1 : 1 万土地利用图综合缩编小比例尺(1 : 5 万、1 : 25 万等)的大挂图,由县级调查成果向地市级、省级、国家级成果库的汇总也涉及到图形信息、专题属性信息的综合缩编。传统的手工作业综合缩编不仅耗时,综合结果与大比例尺原图相互独立,缺乏一致性保证,综合质量与作业员的水平、经验、工作态度有关,不确定性因素太多。在数字技术环境下,将计算机制图综合引进到土地利用数据的综合上,通过制图软件完成土地利用图的缩编是本次土地调查的迫切需求。

参 考 文 献

[1] 艾廷华,刘耀林.土地利用数据综合中的聚合与融合[J].武汉大学学报·信息科学版,2002,27(5):

486-492

[2] 毋河海.地理信息自动综合的基本问题研究[J].武汉测绘科技大学学报,2000,25(5):377-386

[3] Bader M, Weibel R. Detecting and Resolving Size and Proximity Conflicts in the Generalization of Polygonal Maps[C]. The 18th ICC, Stockholm, Sweden, 1997

[4] van Oosterom P. The GAP-tree, An Approach to On-the-Fly Map Generalization of An Area Partitioning[M]. London :Taylor & Francis, 1995:120-132

[5] McMaster R, Shea K. Cartographic Generalization in a Digital Environment: When and How to Generalize [C]. The 9th International Symposium on Computer-Assisted Cartography, Washington D C, 1989

[6] Jones C B, Bundy G L, Ware J M. Map Generalization with a Triangulated Data Structure[J]. Cartography and GIS, 1995, 22(4): 317-331

[7] 艾廷华,郭仁忠,陈晓东. Delaunay 三角网支持下的多边形化简与合并[J]. 中国图像图形学报,2001,6(7):701-709

[8] 艾廷华,郭宝辰,黄亚峰. 1 : 5 万地图数据库的计算机综合缩编[J]. 武汉大学学报·信息科学版, 2005,30(4):297-302

第一作者简介:艾廷华,教授,博士,博士生导师,研究方向为地图综合、空间数据挖掘。
E-mail:tinghua_ai@163.net

Land-use Data Generalization for the Database Construction
of the Second Land Resource Survey

AI Tinghua¹ YANG Fan¹ LI Jingzhong¹

(1 School of Resource and Environment Sciences, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: The resulted data handing and database construction of the second land resource survey includes the land-use data generalization and map design of different administrative levels. From the perspective of practical works, we explore a technical method of land-use data generalization, including the detection of polygons with closest relationship from neighborhood candidates, the establishment of the generalization rules, the aggregation and amalgamation methods for polygon combination and the evaluation of the generalization results.

Key words: land resource survey; map generalization; land-use map; scale transform

About the first author: AI Tinghua, professor, Ph. D, Ph. D supervisor, majors in map generalization and spatial data mining.
E-mail: tinghua_ai@163.net