

土地利用数据综合中的聚合与融合

艾廷华¹ 刘耀林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: 针对土地利用图中多边形地块具有全覆盖、无重叠、语义上多层次的特点, 对土地利用数据综合中的地块合并给出了多边形聚合与多边形融合两种操作; 前者针对具有同属性的语义邻近地块的合并, 后者针对不同属性的拓扑邻近地块的合并; 两种操作的区分是基于同时顾及空间、语义特征的邻近分析, 算法的实现均建立在 Delaunay 三角网模型上由骨架线支持; 详细讨论了两种操作的算法过程并给出了实际数据的实验结果。

关键词: 土地利用; 空间邻近; 地图综合; Delaunay 三角网

中图法分类号: P283. 1; P273

随着 GIS 研究的深入, 针对特殊用途的专题性质的空间数据处理越来越受重视。在地图综合研究领域, 由过去面向通用几何类型数据的综合算子、模型研究转向面向专题地理特征的空间数据研究。同样是多边形数据, 目前综合研究重点则集中在建筑物、街区、湖泊及土地利用几个类型上(Bader, 1997; Oosterom, 1995; Ruas, 1997)。

土地利用数据是多边形类型图的典型代表(其他还有诸如森林植被、土壤、矿物分布等), 在空间分布上具有全覆盖、无重叠、无缝隙特征, 在语义上具有类型的层次化划分(Bader & Weibel, 1997)。对其综合包含着空间上几何特征简化处理和语义上类型层次的归并, 相邻地块的邻近关系既有几何上的距离影响也有专题属性上的语义相似度影响。综合的结果应达到类型减少(子类向父类归并), 几何形状简化, 同时保持全覆盖、无重叠特征, 以及统计上的各类型用地面积相对百分比不变。同时满足这些综合约束条件是十分困难的, 对多边形地块逐个化简使几何形状简化了, 但必然导致共享边界的不一致, 产生裂缝或重叠碎片。对小面积地块删除由相邻地块填充可能导致小面积地块所在类型的百分比严重下降, 破坏原有土地类型的相对平衡。本研究针对类型归并与几何化简给出多边形合并的两种操作: 同类型多边形的聚合和不同类型多边形的融合, 操作对

象为多边形目标整体。在过去的研究中已完成了针对多边形内部细节综合的单个多边形化简(艾廷华, 郭仁忠, 2001), 以及化简后矢量多边形共享边界的一致性改正(艾廷华, 毋河海, 2000)。

1 邻近分析

“将邻近的多边形地块合并”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。空间认知中, 邻近概念至少受拓扑的、几何的、语义的和 Gestalt 原则 4 个因素的影响, 邻近通常表现为 4 种因素综合影响的结果。如果两地块多边形共享公用边界, 在拓扑关系上它们互为邻近; 如果两地块在拓扑关系上相离, 但相互间距离小, 在距离关系上为邻近; 当相互间距离十分接近且难以区分时, 具有相同或相近语义特征的地块具有邻近关系; 当相互间距离、语义均难以区分时, 确定邻近关系判断的是方向、大小、形状上的相似的 Gestalt 识别原则(建筑群分布模式识别是一典型例子)。在实体世界中, 地理现象的分布通常具有不确定性边界, 尽管在数据表达时可以绘制一条确定分界线, 在邻近关系判断时, 仅仅通过拓扑关系判断邻近是不完备的。本研究将距离关系、语义相似度融入并结合拓扑邻接关系联合讨论土地利用多边形地块的邻近性, 而 Gestalt

原则在无规则多边形群邻近关系判断中影响不明显,同时该心理学识别原则用形式化模型表达具有相当的难度,在本研究中暂不考虑。

从土地利用多边形地块中选取某一类型的地块得到多边形子集 $\{c_i\}$, 子集内各多边形元素间相互分离, 相互间语义关系相同, 在子集内比较, 相互间距离越小, 邻近关系越强烈。另一方面, 任一多边形 p 有拓扑邻接邻域多边形群 $\{n_i\}$, p 与任一 n_i 的距离为零, 但相互间语义相近性不同。如图 1, 考虑地块 a 与 c 属于同一类型, 但在空间分布有距离, a 与 d 在空间分布上紧紧相连, 但类型上有差别。相对于 a 而言, 难以量化计算 d 与 c 哪个更适合作为 a 的邻近地块。Oosterom 在针对多边形群层次化剖分的 GAP 树结构中建立“邻近兼容度”概念 (Oosterom, 1995), 但语义层次树上任意两结点间的语义距离如何计算并没有给出方法。因此, 只能固定某一因素, 讨论单变量因素 (要么是距离差异, 要么是语义类型差异) 对邻近关系的影响, 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, 1m 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, 联合作用函数的形式, 可以定性地认为邻近强弱随着距离增大而减小, 随着语义层次树中结点距离增大而减小。代数中有众多这种递减函数, 选用何种函数更合适有待深入研究。

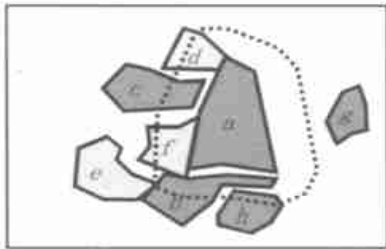


图 1 地块邻域示例

Fig. 1 A Neighborhood Cluster Example

要精确地确定某一地块与其他地块的邻近关系并对其强弱程度作完备的排序是困难的。然而在土地利用数据综合中的邻近概念只需定性讨论, 在一定条件下作邻近、非邻近的二值判断。为使问题简单化, 对某一地块 a , 按一定宽度作缓冲区, 与缓冲区有重叠的地块可归为 a 的几何邻近地块, 显然它包括了与 a 拓扑邻接的地块 (相互间距离零)。认为这些地块在空间距离关系上分别与 a 是平等的, 邻接的地块尽管相互间距离最小为零, 在邻近判断上也没有特权。这样, 邻近还是非邻近仅由缓冲宽度决定。

对拓扑相离目标间的邻近关系分析, 另一考虑的因素是上下文环境的影响。设定较大的缓冲宽度值, 相互间距离小于该阈值的目标 a 、 b 被认为是邻近的, 但是当第三者目标 c 出现在它们之间阻隔了其间的连通性时, a 、 b 便不再具有邻近关系 (尽管两者的空间位置均保持不变)。在对相离目标间的拓扑关系研究中, 陈军等 (1999) 建立了 K 阶邻域的概念。对土地利用地块中的邻近只考虑 1 阶的情形, 即相离中的最邻近 (immediate neighborhood)。上下文环境影响下的邻近分析可以基于 Voronoi 图实现, 每一多边形目标被一 Voronoi 元多边形包围, 元多边形间具有拓扑邻接关系时, 相应地被包围的目标多边形间便具有 1 阶邻域关系。基于以上讨论, 可以对邻近作如下定义。

对地块多边形群 $\{p_i\}$, 如果 p_1 与 p_2 满足下列条件之一, 则被认为是邻近的:

① p_1 与 p_2 是拓扑邻接的;

② p_1 与 p_2 具有相同语义属性, 相互间距离小于缓冲宽度阈值, 在拓扑关系上相离且在该类型下地块子集的 Voronoi 图中, VD 剖分多边形邻接。

条件①说明只要相互间距离为零, 不论语义特征差异如何均视为邻近, 条件②说明语义特征相同的地块需结合距离大小判断是否邻近, 同时还要满足相通性。为叙述方便, 根据控制邻近判断的主体因素, 将满足条件①的称为拓扑邻近, 满足条件②的称为语义邻近。

2 基于邻近分析的多边形合并

根据土地利用分布特征, 任一地块必然有至少一个拓扑邻近地块 (除非是孤岛且类型均一), 但不一定有语义邻近地块, 因为在规定的宽度阈值内作缓冲区不一定遇到当前地块的同类型邻近地块。综合过程主要是对小面积地块的处理, 小地块在综合中的归属可形象地理解为“生存竞争”的过程。当小地块 a 周围无同类型邻近地块时, 最大的可能性是被拓扑邻近地块吞并 (填充 a , 原有类型改变); 当 a 周围有同类型邻近地块时, 它有可能是联合语义邻近地块去吞并夹在其间的拓扑邻近地块, a 生存下来并同化了其他的地块, 该过程类似于围棋中黑白棋子的互包围竞争。

在地图综合基础算子研究中, 相同类型语义特征的多边形合并为聚合 (aggregation), 不同类型语义特征的多边形合并为融合 (amalgamation)

(McMaster & Shea, 1989)。根据这一划分, 语义邻近多边形的合并为聚合操作, 其间填充的间隙引入了异质类型的地块, 但当前参与操作运算的地块是同类型的, 填充部分可视为操作背景。拓扑邻近地块的合并为融合, 操作的对象本身就有类型差异。

2.1 多边形聚合

操作对象多边形类型相同, 但在拓扑关系上互相分离, 聚合的条件为上节中的条件②。操作的核心是对地块多边形群建立 Voronoi 图, 线群目标、面群目标的 Voronoi 图是 GIS 研究领域感兴趣的, 而计算几何中的 Voronoi 图是针对点群的 (Preparata & Shamos, 1985), 有关研究提出了基于栅格扩充的建立算法 (Li & Chen, 2000)。这里采用基于 Delaunay 三角网的矢量算法, 在选取的某类型的多边形群区域内建立约束 Delaunay 三角网, 只考察位于多边形外的三角形, 在这部分三角形区域上作骨架线便得到等剖分地块空白区域的 Voronoi 图, 具体算法见 (Ai & Jiang, 2001)。由骨架线穿越的一批三角形分别连接骨架线两侧的多边形地块, 各种三角形局域内邻近地块间的距离可表示为图 2 中的 $W_1 W_2$, 分别与骨架线垂直, 图中虚线为骨架线。从对骨架线穿越的三角形中选取距离 $W_1 W_2$ 小于距离阈值的三角形集, 并对其根据邻域关系聚类, 得到图 3 中深色区域所示的多边形地块, 可看作是邻近地块在瓶颈部位的“桥梁”。该“桥梁”区域可能连接两个或多个邻近地块。将“桥梁”区域多边形与其连接的邻近多边形作或运算, 并对结果多边形赋予邻近地块相同的类型值, 完成多边形聚合运算。该操作实质是对跨接邻近地块的桥梁区域作属性重置, 保持了边界的一致性。

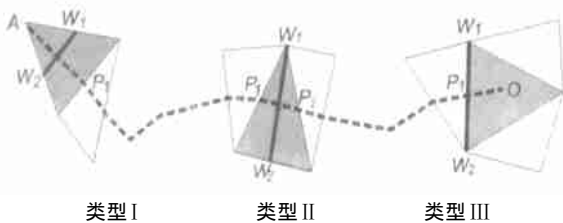


图 2 三种三角形的区分及局部距离 $W_1 W_2$ 的表达
Fig. 2 $W_1 W_2$ the Skeleton Width Representation of Three Types of Triangles

2.2 多边形融合

操作对象多边形的类型不同, 相互间具有拓扑邻接关系, 其实质是将考察的不重要的地块 (比如面积太小) 归并到其拓扑邻接的某个邻域多边形中。Oosterom 在 GAP 树综合策略中认为, 与

a 拓扑邻接的多边形中, 共享边界最长的邻域多边形 b 为 a 的最邻近地块, 综合时将 a 整个归并到 b 中 (Oosterom, 1995; Ai & Oosterom, 2001)。然而, 当与 a 邻接的多个多边形 $\{b_1, b_2, \dots\}$ 分别与 a 共享边界长度相近时, 从长度值数量比较上寻求最邻近地块已意义不大。如图 4, 在对小面积地块 p_0 的剖分填充中, 邻近地块 p_1, p_2, p_3 相近, 将 p_0 整个赋予 p_1, p_2 或 p_3 不合适。相对的解决策略是将 p_0 剖分为三块, 分别赋予与各块邻近的 p_1, p_2, p_3 地块。剖分线由三地块与 p_0 的邻近强弱决定, Bader & Weibel (1997) 在对多边形类型图综合中提出了骨架线剖分的思想, 但没有给出具体的算法。本研究提出了详细的剖分算法并实现了这一思想, 并基于加权剖分线方法将其推广到邻近度差异较大的剖分情形中。



图 3 一类土地利用地块分布及“桥梁”区域提取
Fig. 3 One Category Land-use Parcel Distribution and Bridge Region Extraction

在 Delaunay 三角网中作骨架线, 对不重要多边形 p_0 剖分, 这是算法的基本思想, 然而在几何层次上实现这一技术, 需要对 p_0 内的三角形作调整处理。为得到无缝剖分, 希望骨架线的端点刚好落在邻近多边形边界的连接结点上, 这包含两方面: ①骨架线的每一个端点刚好位于边界结点上; ②每一个边界结点刚好与一个骨架线端点吻合。假定在多边形边界组织中无伪结点, 即每个结点至少关联三条边, 是三个邻近多边形的汇合点。从图 4(b) 中可发现, 直接在多边形内的 Delaunay 网上提取骨架线显然不能满足这两个条件。

剔除那些三顶点位于同一条边界弧段的三角形, 如图 4(c) 所示, 对剩下的三角形 (连接不同的地块) 作骨架线, 这时可保证骨架线的端点刚好位于边界弧连接结点上, 不会出现像图 4(b) 中的位于边界弧的中间部位的情形。与骨架线的端点对应的三角形 t 只有一个邻域三角形 (根据邻域三角形数目, 该类三角形划分为类型 I, 如图 2 所

示), 且三角形 t 的三顶点在多边形边界上顺序相邻。假设某一骨架线终止于某边界弧的中部, 端点所在的三角形三顶点必属于同一边界弧, 因此

假设不成立。

从空间分布意义上讲, 剔除的三角形集区域

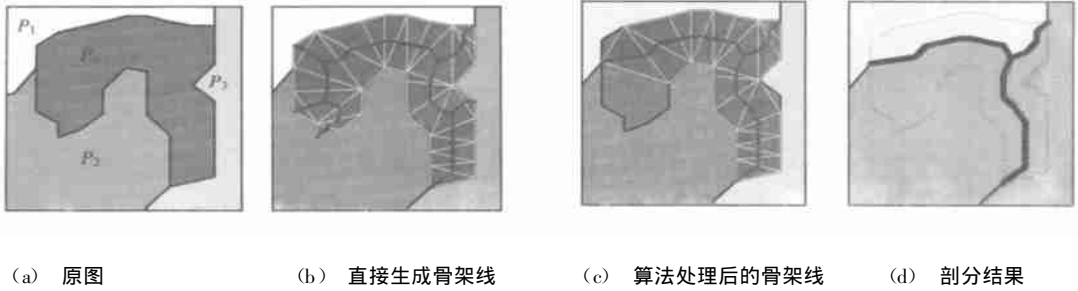


图 4 待融合地块多边形内骨架线生成及剖分结果

Fig. 4 Generation of Skeleton Edges Within One Insignificant Polygon

为邻近多边形的凹部, 该部分区域被邻近多边形强烈包围, 剖分时它不属于有争议的部位, 应完全归并到相应邻近多边形, 因此, 中剖分骨架线不应涉入该局部范围, 如图 4(c) 所示。

重要地块, 右图为被邻近多边形剖分的结果。

产生骨架线端点的 I 类三角形出现在边界弧段上相邻矢量点有明显转折角的情形, 如果相邻矢量点夹角接近 180° , 便无骨架线从此处产生, 尽管该矢量点可能为多个边界弧的交汇结点。这样, 根据正常的 Delaunay 三角网骨架线算法不能保证条件②, 即每个边界弧连接结点处产生一骨架线分支。如图 5 中的 E_1, E_2, E_3, E_4 为邻近多边形汇合点但没有 I 类三角形落在这些点上, 从而不能有骨架线分支产生。在图 5 中, 正常的三角网骨架线算法只能产生 E_0 到 E_5 单一曲线。需增加另外的骨架线, 方法是: 搜索出 E_1, E_2, E_3, E_4 这类孤立结点(无骨架线连接), 寻找每点距已产生骨架线(E_0E_5)最近的点, 根据 Delaunay 三角网的最邻近连接原则, 最短路径为某三角形边的一半。连接两点, 从而得到由孤立结点出发的新增骨架线, 原有骨架线被分解为两段。当某一邻近多边形要瓜分当前不重要多边形地块, 接受其剖分区域时, 两多边形应当有一段距离的共享边界。图 5 中的多边形 c 与当前多边形 a 只有一点共享, c 便不能从 a 中接收部分区域, 尽管在空间关系上 c 也为 a 的邻近多边形。

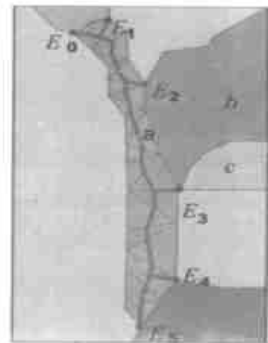


图 5 骨架分枝追加

Fig. 5 Adding A Additional Skeleton Edges

经过以上两步骤的处理, 严格保证本节开始所讨论的多边形边界结点与骨架线端点吻合匹配关系, 选取骨架线分支及多边形共享边界弧段, 通过多边形重组便可完成次要多边形地块的分解, 分解的各部分区域分别沿共享边界弧段赋予邻近多边形, 修改其原类型值。图 6 是为本算法完成的多边形与邻近地块多边形融合处理的结果。图 6 中左图粗线划显示边界的多边形为小面积非

以上剖分骨架线的生成是假定邻近多边形有相同的空间竞争剖分能力, 属中剖分, 当进一步考虑邻近关系有强弱程度之分时, 可对骨架线作加权改进。定义函数 $\text{collapse}(a, b)$ 为次要多边形 a 瓜分邻近多边形 b 的剖分能力表达, 该函数将受以下因素的影响: ①类型兼容度: $\text{type-compatibility}(a, b)$; ②共享边界长: $\text{length}(a, b)$; ③ b 的重要性: $\text{imp}(b)$ 。函数的形式为:

$$\text{collapse}(a, b) = \text{type}(a, b) * \text{length}(a, b) * \text{imp}(b)$$

计算出 $\text{collapse}(a, b)$ 值后, 得到邻近多边形的剖分能力, 这时 a 内部的骨架线可根据三角形边跨接邻近多边形的剖分能力的对比确定, 如图 7, a 的邻近地块 b, c, d 分别有剖分能力值 7、3、5。 b 与 c 夹住的骨架线调整为如图 7 所示的情形, 即 c 剖分 a 的区域占 30%, b 剖分 a 的区域占 70%。调整后的骨架线称为加权骨架线, 在几何算法上加权骨架线由三角形边的二等分变为比例分段即可。然而, 类型兼容度和邻近多边形的重要性定量计算需要结合专题数据的性质、综合的目的及应用对象详细讨论, 本文不深入研究。

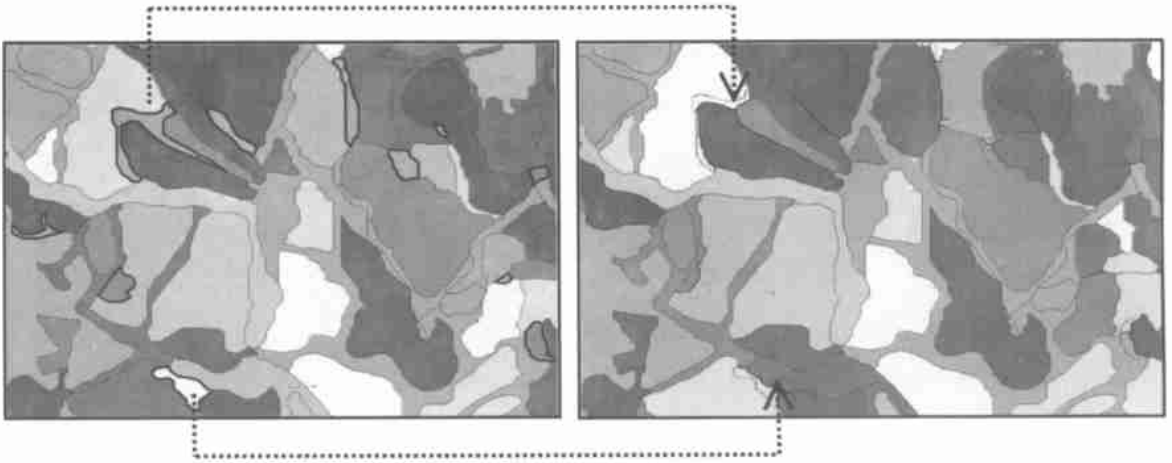


图6 小面积次要地块与拓扑邻近地块的融合实验

Fig. 6 An Experiment of Amalgamating Insignificant Parcels Based on A Adjusted Skeletons

3 控制流程与实验

对小面积次要多边形地块进行综合, 面临融合与聚合两种操作。那么, 以何种条件启动某种操作要由一定条件来控制, 控制流程如图8所示。该控制流程是渐近式反复综合过程, 控制参量有两个: 判定为次要地块的面积阈值 λ_1 和语义邻近缓冲区宽度阈值 λ_2 。按面积由小到大, 逐步进行次要地块的综合。该控制流程实质是以语义

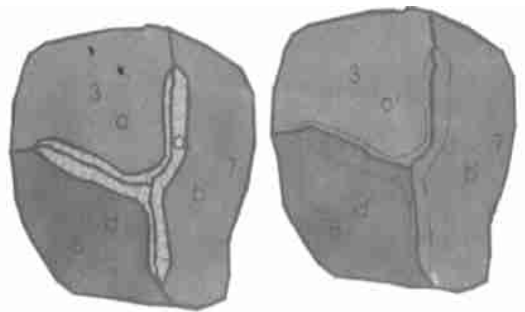


图7 基于加权骨架线的融合操作

Fig. 7 The Generation of Weighted Skeleton

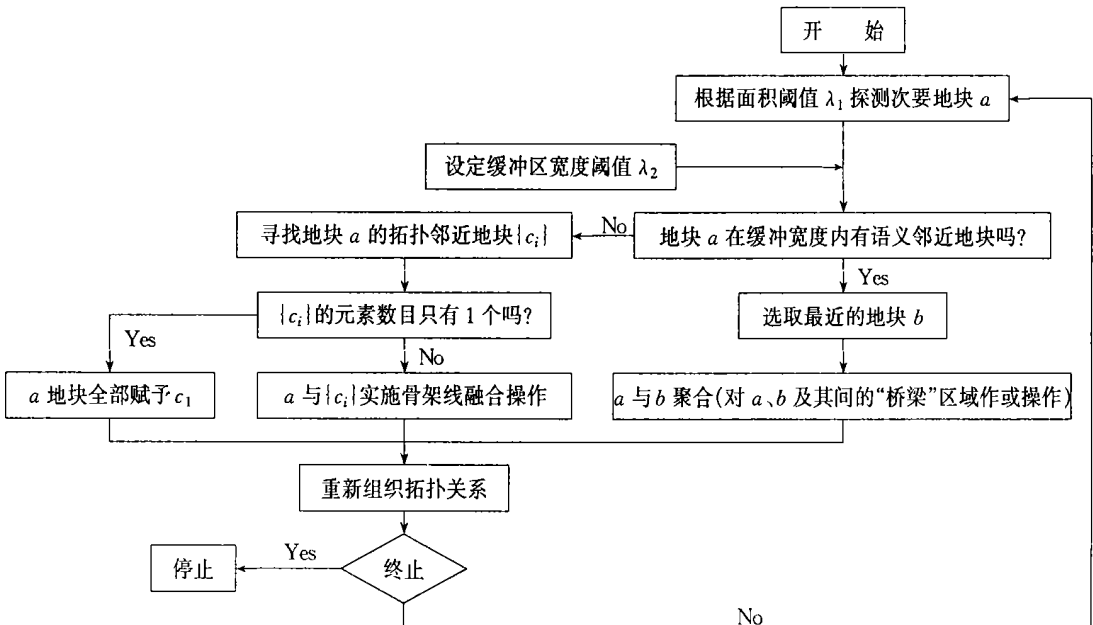


图8 融合、聚合操作的控制流程

Fig. 8 The Workflow of the Operations Aggregation and Amalgamation

邻近地块的聚合为优先, 一旦发现有同类型的语义邻近地块, 便对其聚合, 否则将当前地块融合到拓扑邻近地块中。当拓扑邻近地块只有 1 个时,

不存在骨架线剖分的问题。

根据这一算法, 笔者在自行开发的交互式地图综合软件 AutoMap 上用 VC 6.0 程序进行了实

验, 实验的算法对上述控制流程作了简单调整。上述流程每一步只进行一个次要地块的综合, “操作分辨率”高。当地块 a_1 与 a_2 语义邻近, 面积均很小, 聚合结果 a_3 仍小于面积阈值 λ_1 时, a_3 马上又将被融合, 没有生存下来。这样不如对 a_1 、 a_2 直接融合, 一步完成该综合操作。具体地规定以下条件, 当地块 a 满足时将对其融合:

1) 无“桥梁”区域连到同类型的语义邻近地块, 且面积小于阈值;

2) 有“桥梁”区域连通, 但“桥梁”区域的面积大于面积阈值, 而 a 的面积小于阈值;

3) 有“桥梁”区域连通, 但连通的语义邻近地块面积和“桥梁”区域面积之和小于面积阈值。

图9为实验样图, 图10为对应的某一类型地块群 Voronoi 图, 其中深色“桥梁”区域表明其连接的地块将作聚合。Voronoi 图多边形包围的白色区域表明将被拓扑邻近地块融合, 它要么是孤立的, 要么与语义邻近地块合起来面积太小, 要么“桥梁”区域面积太大。被融合的地块如果只有一个拓扑邻近地块(通常作为其岛屿), 将直接将其删除, 由拓扑邻近地块直接填充。

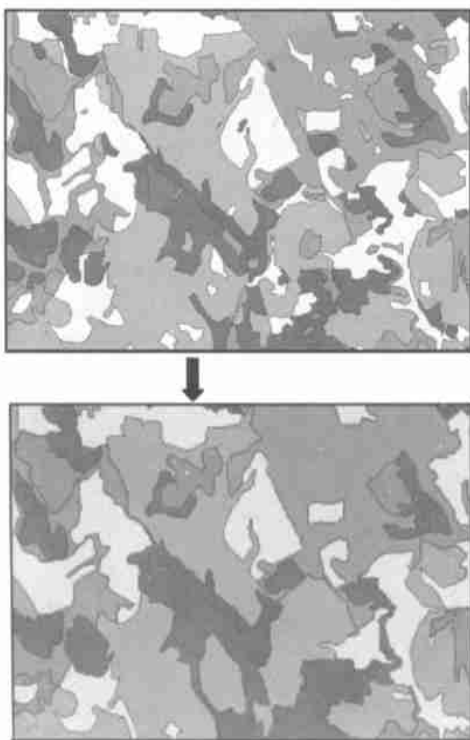


图9 土地利用数据综合实验(聚合与融合)

Fig. 9 The Generalization of Land use Parcel based on Aggregation and Amalgamation

4 结论

土地利用图的综合是多边形目标综合中的难

点, 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。本文基于多边形群的 Voronoi 图并结合语义类、特征类的划分提出了拓扑邻近与语义邻近的概念, 并分别给出对应的综合操作算法: 融合与聚合, 针对实际土地利用数据进行了算法实验。以下问题有待进一步研究: 拓扑邻近与语义邻近可比性研究纳入到语义层次树体系中, 加权骨架线剖分中邻近地块剖分能力函数的计算, 顾及各类用地百分比保持平衡的综合策略。

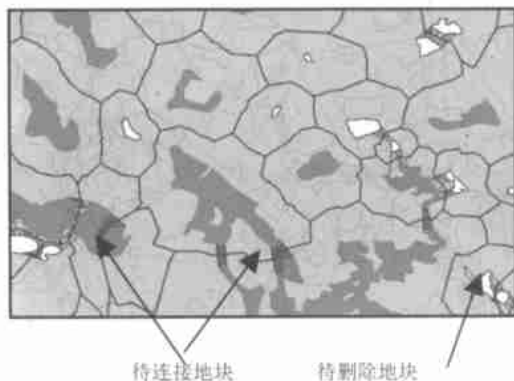


图10 基于 Voronoi 图判断次要地块的邻近性, 决定聚合、融合操作的选取

Fig. 10 VD Model Constructed in Different Category Parcels to Detect Whether a Parcel Has Geometric Neighbors or Its Isolated

参 考 文 献

- 艾廷华, 郭仁忠. 支持地图综合的面状目标约束 Delaunay 三角网剖分. 武汉测绘科技大学学报, 2000, 25(1): 35~41
- 艾廷华, 毋河海. 相邻多边形共享边界的一致化改正. 武汉测绘科技大学学报, 2000, 25(4): 426~431
- 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. 中国图像图形学报, 2001(7): 701~709
- Ai T H, Guo R Z, Liu Y L. A Binary Tree Representation of Bend Hierarchical Structure Based on Gestalt Principles. The 9th International Symposium on Spatial Data Handling, Beijing, 2000
- Ai T H, Jiang W. Distribution Analysis and Automatic Generalization of Urban Building Cluster. The 3rd DMGIS-ISPRS, Bangkok, 2001
- Ai T H, van Oosterom P. A Map Generalization Model Based on Algebra Mapping Transformation. The 9th ACM-GIS, Atlanta, 2001
- Bader M, Weibel R. Detecting and Resolving Size and Proximity Conflicts in the Generalization of Polygonal

- Maps The 18th ICC, Stockholm, 1997 (3): 1 525 ~ 1 532
- 8 Jones C B, Bundy G L, Ware J M. Map Generalization with a Triangulated Data Structure. *Cartography and GIS*, 1995, 22(4): 317~331
- 9 Li C, Chen J, Li Z L. A Raster-based Algorithm for Computing Voronoi Diagrams of Spatial Objects Using Dynamic Distance Transformation. *International Journal of Geographic Information Sciences*, 1991, 13(3): 209~225
- 10 van Oosterom P. The GAP-tree: An Approach to On-the-Fly Map Generalization of an Area Partitioning. In: Muller J C, Lagrange J P, Weibel R eds. *GIS and Generalization: Methodology and Practice*. London: Taylor & Francis, 1995
- 11 van Oosterom P. *Reactive Data Structure for Geographic Information Systems*. Oxford: Oxford University Press, 1994
- 12 陈 军, 赵仁亮. GIS 空间关系的基本问题与研究进展. *测绘学报*, 1999, 28(2): 95~102
- 13 McMaster R, Shea S. *Cartographic Generalization in Digital Environment: When and How to Generalize*. Proc. Autocarto, 1989
- 14 Ruas A, Mackness W A. *Strategies for Urban Map Generalization*. The 18th International Cartographic Conference, Stockholm, 1997

作者简介: 艾廷华, 博士, 副教授。研究领域: 地图综合, 空间数据挖掘。

E-mail: aith@wuhan.cngb.com

Aggregation and Amalgamation in Land-use Data Generalization

AI Tinghua¹ LIU Yaolin¹

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

Abstract: The land-use map is a typical kind of categorical theme and has the properties covering the whole area with neither gap nor overlapping area. Its generalization has to consider geometric simplification, semantic hierarchy abstraction and statistical properties maintenance. This paper focuses on the combination of neighbor parcels giving two operations: aggregation and amalgamation. The previous aims at the polygon objects with homogeneous semantics, and the latter aims at those with heterogeneous semantics. The decision of neighborhood relationship is the basis of two operations. This study considers both spatial and semantic facts in the judgment of neighborhood cognition based on Voronoi diagram model. The work flow of two operations is presented in land-use data generalization with the experiment result of real data.

Key words: land-use; spatial neighbor; map generalization; Delaunay triangulation

About the author: AI Tinghua, Ph. D. associate professor. His research interests include map generalization and spatial data mining.
E-mail: aith@wuhan.cngb.com