

非连续分布面状地理现象的 图形自动概括方法

齐清文 刘 岳

(中国科学院 地理研究所 北京 100101)
(国家计划委员会)

摘 要 本文首先阐述了图斑概括过程中应遵循的原则,即正确处理几何精确性与地理真实性、清晰性和详细性两对矛盾,继而提出了自动化制图综合的设计策略。然后探讨了非连续分布面状地理现象(即图斑)的自动取舍、化简、夸大和移位、以及合并等算法,介绍了上述算法的实现方法,并给出几种地物图形自动概括的应用技术流程。

关键词 面状地物 制图综合 图形自动概括

分 类 (中图法) K99 (科图法) 57 [187], 57.1059

在地理研究和地图制图过程中,制图综合问题一直是人们关注和研究的重点之一。该问题解决得成功与否,对制图精度,制图指标的分级、分类,表示方法,各地理要素在图上的合成与分解,相互协调,乃至空间数据模型和地理过程的模拟尺度等一系列问题都产生深刻的影响。在计算机制图、地理信息系统及专家系统等现代地理科学的技术手段得到空前发展和广泛应用的今天,人们也毫不迟疑地将自动化技术引入到制图综合研究领域。国内外已有不少专家、学者对自动制图综合进行了深入的、卓有成效的研究。不过,迄今为止的研究报告主要集中在点状和线状地物的自动概括方面,很少涉及面状地物的自动综合。其原因显然在于后者的复杂程度和难度较前者更大。本文提出的自动概括方法适用于面状地理现象中的一类——非连续分布面状地理现象。

“非连续分布面状地理现象”的概念与连续分布面状地理现象(包括用等值线法表示的地表起伏、气温、气压等现象和用质底法反映的土地利用、土壤类型等现象)相对,指在地面呈间断成片分布,在地图上能按比例尺表示的呈面状分布的地理现象,既指在地图上有实在分布形态的自然和人文现象,如植被、湖泊、水库、池塘、岛屿、居民地等,也包括较为抽象的或人眼看不见的现象,如民族分布区、某种语言分布区、重力异常区、地磁异常区等。这

类现象通常用范围法表示。其特点是轮廓线呈封闭的岛状,其中范围固定的,用封闭实线表示;范围不固定的,用虚线轮廓表示。本文称这些轮廓线围成的图形为“图斑”。

1 应遵循的原则和设计策略

1.1 正确处理两对矛盾

图斑概括的实质是解决缩小了的地图与地物原型之间的矛盾。具体来说,在图斑概括过程中应正确处理以下两对矛盾。一是几何精确性与地理真实性之间的矛盾,即既要尽量保持图斑的原轮廓形状和位置精度,又必须正确反映地物之间的拓扑关系和地物空间分布的总体布局,避免破坏地理规律。为突出主要特征,舍去次要细节,在对图斑进行选取、轮廓简化、移位等处理时,必须确定各图斑的重要程度,掌握图斑的主要轮廓特征,进而确定图斑、轮廓弧段选取(保留)的优先级和保持原位的优先级。而此优先级的确定须综合考虑各种因素:图斑个体的面积、图斑的密度格局、图斑的重要性、要素之间的相互关系等。二是清晰性与详细性之间的矛盾,即在小比例尺地图上力图保持原较大比例尺地图上信息载负量的同时,尽量使地图清晰易读。对此矛盾的处理原则是采用地图视觉负载量的概念,利用人对地图符号的视觉错觉原理,尽可能改变视觉变量参数(如图斑大小、色彩、晕线等),使在比例尺允许的情况下尽可能多地表示地物的同时,提高地图内容的清晰性。

1.2 图斑自动概括的设计策略

目前在设计中有四种思路:一是从人在自动化制图综合过程中所起的作用角度来考虑,可设计交互式处理和批量处理两种形式。二是从对制图对象的处理流程、顺序角度考虑,可分别设计综合型方案(又称“自上而下”的概括方案)与分解式方案(又称“自下而上”的概括方案)。三是从自动综合软件的通用性和专用性角度来考虑,一些学者强调其通用性和标准化,还有相当一批制图学家则倾向于研究和开发出“面向区域”(如面向黄土高原、岩溶丘陵、冰川地貌区)、或“面向要素”(如面向植被、水系、交通网络、等高线)、或“面向空间单元类型”(如面向点、线、面、体)的专用软件。四是从自动综合软件的智能化角度考虑,当前提得较多的是建立制图综合专家系统的思想。从理论而言,我们认为最理想的方案是通用性、标准化的软件设计思想,批量处理的过程,“自上而下”的实现顺序,最终发展到智能化的专家系统。然而就目前的现实条件和对问题的研究深度而言,只能采取另一种策略,即:①通用性与专用性相结合,采用模块化或面向对象的方式设计一批基本“操作”组成通用“工具包”,而对于特殊区域或专门要素的处理,则从工具包中提取单项操作并予以合理“装配”,形成专用型制图综合模型;②“自下而上”的设计顺序,即把制图综合问题分解为一个个子问题。子问题可按地理现象类别来分(如分为土地利用、城市居民地、植被等),也可以按制图对象的空间分布形式(点、线、面)来分,或者二者的结合。一次处理一类要素或现象的制图综合问题,最后再解决图幅的全盘问题。这一途径较容易实现;③人机交互式处理过程。因为在智能化问题没有很好解决之前,批量处理方式难以完成所有的制图综合问题,得不到令人满意的结果;④关于制图综合专家系统的建立,目前无论从理论研究还是在实践上条件都不成熟,因为制图专家的制图综合知识与经验还未达到为专家系统推理机能够接受的,具有

严格性、准确性和系统性的水平，但这是未来发展的趋势。

2 算法研究

2.1 图斑取舍

在制图综合的取舍算法中，较成熟的是定额选取模型，即解决“取（舍）多少”的数量指标模型，缺乏成熟的、但经实践检验较为实用的结构选取模型，即解决“取（舍）哪些”的模型。我们认为：对于非连续分布面状地理现象（即图斑）的选取，后者更重要。本文提出的算法将两者有机地结合起来。基本思想是：首先选用定额选取模型计算出新图上应有的图斑数量（用总面积表示），然后计算原图上每个图斑的综合评分，而该评分由面积评分、属性特征评分、位置特征评分三项因子经标准化后相加得出。解决“选取哪些图斑”的问题就以综合评分为依据。将各图斑按综合评分由大到小排序，并计算出自上而下的面积累积值。最后将理论上的图斑选取总面积与此面积累积值相比较，累加值达到理论面积时的图斑即为选取下限，即从此图斑开始向上的所有图斑都被选取，其余舍去。现对定额选取模型和综合评价的具体算法说明如下：

2.1.1 定额选取模型 这里采用 Topfer 的方根模型，即新图与原图上的图斑总面积之比等于新图与原图比例尺之比平方的开方，即：

$$G_t = G_a \sqrt{\left(\frac{M_a}{M_t}\right)^2}$$

式中， G_a 、 G_t 分别为原图、新图上的图斑总面积（ mm^2 ）； M_a 、 M_t 分别为原图、新图的比例尺分母，其中，

$$G_a = \sum_{i=1}^n G_i$$
$$G_t = \sum_{j=1}^m G_j$$

式中， G_i 和 G_j 分别为原图和新图上单个图斑的面积（ mm^2 ）。

2.1.2 计算各图斑的综合评分 综合评价中考虑了影响各图斑取舍的三项因子：面积、属性特征和位置特征。

a. 面积评分（C）：以图斑的面积值代替。面积越大的图斑，其重要性越大，因而评分越高。

b. 属性特征评分（A）：按表 1 的标准评分。在表中，按属性特征可将图斑分为：
①重要的——行政等级高或人口数量多的居民地，或有多条公路、河流通过的居民地；经济价值高的湖泊（如矿泉湖、缺水区的淡水湖等）和有旅游价值的湖泊、水库（如风景区的喀斯特湖等）；战略地位重要的岛屿等等。“重要的”中又分为一、二、三级。
②特殊的——类型特别、种类稀少的地物，如特（稀）有种群的植被区，大型独立建筑物，作为河源的小湖，稀有的少数民族分布区等。
③一般的——除上述以外的其余图斑。

表 1 属性特征评分标准

Tab. 1 Assessment Criteria of the Attribute Feature of Area Patches

重要的			特殊的	一般的
一级	二级	三级		
10	7	4	7	1

c. 位置评分 (B): 位置特征的评价原则是: 当某图斑与另一图斑的距离小于某一阈值时, 位置越近评分越大; 当距离大于阈值时, 则距离距远评分越大。其理论依据是: 一组相互间靠得很近或离得很远的图斑比其它图斑 (即相距不远、不近的图斑) 更有机会保留, 因为它们能反映区域的总体地理格局。具体评分步骤如下:

第一, 计算阈值。按照视觉分辨能力, 原图上的阈值定为 2mm, 则在新图上阈值为:

$$S = 2 \cdot \sqrt{\frac{M_a}{M_1}}$$

第二, 计算某图斑到距其最近的图斑的距离 d_i 。经搜索后, 寻找到距某图斑距离最近的图斑, 它们的距离为:

$$d_i = \sqrt{(X_i - x_i)^2 + (y_i - y_i)^2}$$

第三, 判断并计算赋值 B_i :

$$B_i = \begin{cases} \sqrt{\frac{S}{d_i}} & \text{当 } d_i \leq S \\ \sqrt{\frac{d_i}{S}} & \text{当 } d_i > S \end{cases}$$

d. 归算处理和计算综合评分: 为了使三项因子的评分具有相同的量度, 采用普遍标准化方法对三项因子的评分进行归算化处理, 设处理后的值为

$$x'_i = \frac{x_i}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}}$$

式中, x_i 为某变量值 (这里为评分值)。故综合评分值为:

$$F_i = \frac{C_i}{\sqrt{\frac{1}{n} \sum_{i=1}^n C_i^2}} + \frac{A_i}{\sqrt{\frac{1}{n} \sum_{i=1}^n A_i^2}} + \frac{B_i}{\sqrt{\frac{1}{n} \sum_{i=1}^n B_i^2}}$$

2.2 图斑轮廓化简

图斑轮廓化简可归结为线性符号 (曲线) 的化简。目前已有不少曲线化简方法, 本文采用 Douglas-Peucker 方法, 其原理是: 在图 1-a 中某条曲线的支撑点 (曲线的顶点) 用记号 \bigcirc 标出, 首先通过起点 P_1 和终点 P_n 作一条弦, 并找出距弦最远的支撑点 P_k , 若这个距离大于规定阈值, 则在 P_1 和 P_k 以及 P_n 和 P_k 之间做新的弦并重复找出这两条弦的两个最大支撑点, 这样得到点 P_i 。继续这个过程直至所有余下的点子至新弦的距离都不超过规定阈值为止,

舍去那些剩余的支撑点,曲线就由所作的弦的一个个端点表示。该方法的特点是用逐步筛选的方法保留曲线的部分支撑点,因而具有相对最大的信息量。使用上述方法之结果是否令人满意,取决于阈值的确定。不同的曲线和同一曲线的不同段往往具有不同复杂程度,要使阈值适应各种复杂程度是十分困难的。本文提出的方法是按不同复杂程度的曲线段来计算阈值。此方法能够兼顾“保留多少弯曲”和“保留哪些弯曲”两个方面。图斑轮廓化简的具体步骤如下:

a. 将每个图斑的轮廓线按其复杂系数分为若干段。复杂系数用单位长度内的弯曲个数表示,即单位长度内的弯曲个数越多说明其复杂性越大。

b. 用方根规律计算出新图上各曲线段应保留的理论弯曲个数,即

$$D_i = N_n \sqrt{\frac{M_u}{M_i}}$$

式中, N_u 、 N_i 分别为某曲线段在原图、新图上的弯曲个数; M_u 、 M_i 分别为原图、新图的比例尺分母。

c. 计算每个弯曲的深度,并将该曲线段的所有弯曲按弯曲深度由大到小排序。

d. 以上述弯曲深度为阈值,由大到小依次代入前述 Douglas-Peuker 算法中,统计各次计算结果中所保留的弯曲数目,直到弯曲数符合理论数为止。这个弯曲深度便是符合要求的阈值,即此弯曲以上的所有弯曲的尺寸都超过阈值,应予以保留。

2.3 夸大和移位

图斑轮廓的夸大算法可以归结为移位问题的算法,因为夸大可看作是轮廓线某点向外移位或整体向四周移位;本意的“移位”则是指图斑整体向某一侧移动。

2.3.1 问题的关键

a. 哪些图斑或图斑的哪些点需要夸大和移位?这是必须首先明确的问题。为确定移位对象,可制定图斑的“保持原位优先级”。而此优先级应与图斑的重要性相吻合,因而本文 2.1.2 中确定的图斑综合评分可应用在这里。综合评分越高,图斑的保持原位优先级越高。图斑需要夸大的弯曲点用数据库检索技术确定。

b. 向什么方向移位?两个图斑应沿它们的质心连线向相背的方向移动(见图 2)。

c. 移位量多大?此问题有两层含义。一是每个图斑移位多少。这与问题 a. 紧密相关,保持原位的优先级越高,其移位量应越小,因此还是归结为综合评分的问题,即每个图斑的位

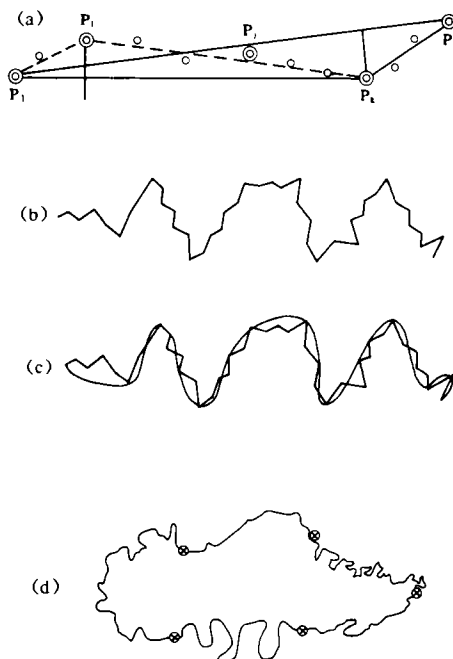


图 1 图斑轮廓化简示例

Fig. 1 Example of area patch simplification

a-c 用 Douglas-Peuker 方法化简曲线的过程

d 图斑轮廓按复杂系数分段 (⊗为分段点)

移量与其综合评分值成反比。二是总位移量多大,即移位结束时图斑间应保持多大的距离。这取决于视觉可识别的地图上的最小间距,一般取 0.2—0.4mm。

2.3.2 移位的算法 位移算法的基本思路是:在确定带有夸大或移位特征的点后,将此点与其前后两点组成三角形,算出三角形顶角(需夸大或移位的角)平分线的长度,然后在此长度靠顶角一端 P_i 点加一个定值 T ,得到 Q_i 点,再算出 Q_i 点的坐标,并用它取代 P_i 点,这就使图斑的局部点得到夸大(见图 3-a)。若使顶角向内移动,只要 T 取负值即可,也就是在顶角平分线长度靠 P_i 点一端减去 T 值(见图 3-b)。不同点的连续夸大处理就可构成移位(见图 3-c)。根据需要,闭合图斑的移位有多种情况。例如,整条线向同一方向移位,此时其形状不变,但中心位置

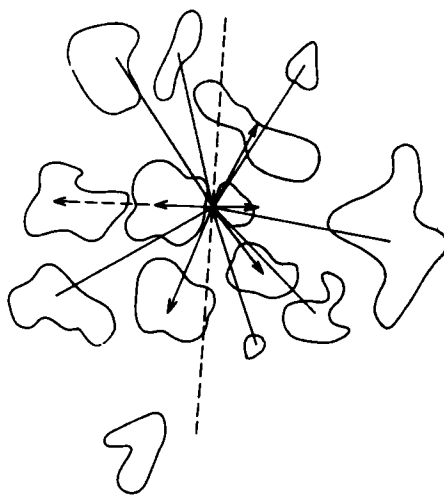


图 2 图斑移位原则

Fig. 2 Principle of area patch displacement

发生移动;整条线向不同方向移位,其形状变化而中心位置基本不变;整条线向四周移位,其中心位置未变但图形放大;整条线向中心移位,其中心位置不变而图形缩小;一条线一部分移位,一部分不移位,其形状和中心位置均发生改变。

2.4 合并

合并是面状地物内部结构化简的主要方法。图斑的合并必须满足下列条件:轮廓线间的最小距离不能大于定值;合并后的图斑外形要与原图形基本保持相似,并保证不破坏原图斑的拓扑关系;合并前后的图斑分散程序

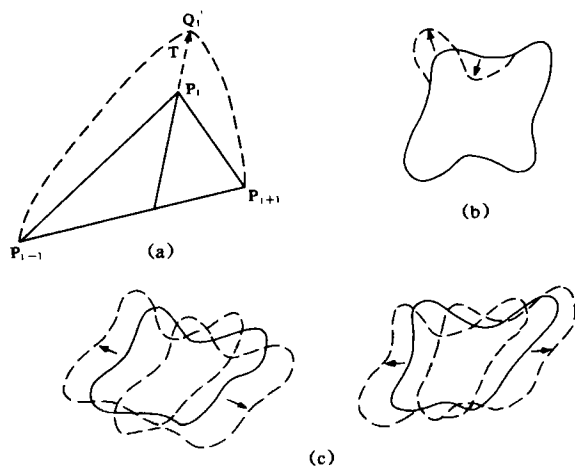


图 3 图斑夸大和移位示例

Fig. 3 Example of area patch exaggeration and displacement

不能相差太大;合并前后的大小图斑对比格局基本保持一致。

图斑合并的方法要点是:利用数据库检索功能找出相距较近的图斑,判断哪两个(或几个)合并较为合适,计算可合并的数量。当能满足合并条件时,计算合并后的轮廓点位,实现合并。具体方法和步骤如下:

a. 判断图斑之间的距离是否小于定值, 这要计算图斑间的距离。由于符号的面积有大有小, 可以是两个小图斑合并, 也可以是一大一小合并, 甚至是两个大图斑合并, 因而距离的算法应当有所区别。如果是两个小图斑, 可计算它们中心之间的距离; 如果是一个小图斑和一个大图斑, 则应计算小图斑中心到大图斑轮廓线的最短距离; 如果是两个大图斑, 应计算两图斑轮廓线上相距最近的两点间的距离。

b. 确定合并后的轮廓点, 简单的方法是将二个图斑的原始轮廓点合二为一, 然后找出 x, y 中的极大值和极小值, 将四个极值点顺次编号, 作为合并后新图斑的轮廓点。但此方法不能反映图斑之间形状的差异, 因而不适用于轮廓比较复杂的图斑。本文采用下述方法:

如果两个图斑互不包含, 则先找了两个图斑各自的极值点 X_{max} 、 Y_{max} 、 X_{min} 、 Y_{min} 、即图 4-a 中 i 区的 1、2、3、4 和 k 区的 $1'$ 、 $2'$ 、 $3'$ 、 $4'$, 然后找出这两个图斑之间最近的两个极值点, 如 2、 $4'$, 再将此两极值点到各自前后极值点之间的数据删去, 最后将两个图斑剩余的数据点按绘制的顺序拼接起来, 例如 1, $1'$, ..., $2'$, ..., $3'$, 3..., 4, ..., 1。用光滑程序连接诸点即可 (图 4 中的粗线)。如果两个图斑中的一个被另一个包含, 则先找出包含图斑中所有的极值点 (如图 4-b 中的点 1、2、3、4、...10) 和被包含图斑的四个极值点 (如 $1'$ 、 $2'$ 、 $3'$ 、 $4'$ 点); 然后找出包含图斑包含处的二个连续凸向极值点, 如 7、9, 并删去点 7 和点 9 间的所有点; 再分别找出点 7、点 9 和被包含图斑极值点的距离, 保留被包含图斑上距离最小的点, 如 $1'$ 和 $4'$, 删去 $1'$ 、 $4'$ 之间的其它点; 将被包含图斑上保留的点顺次插入包含图斑; 最后用光滑程序顺次连接, 就可得到合并后的图形。

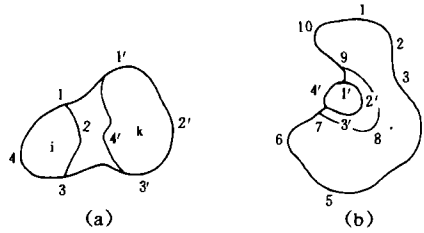


图 4 图斑合并示例

Fig. 4 Example of area patch aggregation

3 图斑自动概括方法的实现

3.1 软硬件环境

非连续分布面状地理现象的图形自动概括方法在下列开发环境中实现:

硬件: 主机以 VAX 小型机为主。辅以微机 (AST386, 兼容 486), 外围设备有图形终端 TEKTRONIX4208, 手扶跟踪式数字化仪 CALCOMP9100。

软件: 操作系统为 VMS 和 DOS6.0, 关系数据库管理系统采用 INFO、FOXBASE2.1, 程序开发工具为 ARC/INFO 中的 AML 语言并嵌入 C 语言。

上述环境能够满足面状制图综合要求的图形分辨率、数据库中的图形一属性库和交互编辑功能, 以及模块化软件开发性能等技术指标。

3.2 技术系统

假定原图已经以数字地图的形式存于大型 INFO 数据库中。本试验的目的是试图将已编

制完成的各单项操作软件组合成一个能与地图数据库交互操作和通讯的集成式系统,以便于用户使用(见图5)。其中用户界面设计成多级菜单式,用户可以根据其需要设置各种参数,选择不同的“操作项目”,完成各种复杂的处理。

3.3 不同要素的图形自动概括流程

在对非连续分布面状地理现象实施图形自动概括时,我们注意到,由于不同要素各自的特殊性,其概括过程和“操作”项目也有所不同。例如,植被图往往可以合并,也可以删除,但湖泊、岛屿图斑一般只能删除而不能合并。因此对不同要素应设计不同的技术流程。下面选两类代表性要素及多要素迭置图说明之。

3.3.1 植被图斑的图形自动概括流程

这是假定结果图上只有植被一

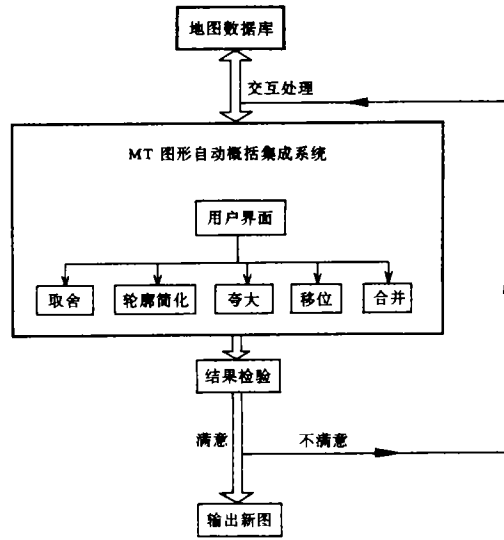


图5 非连续分布面状地理现象的图形自动概括技术系统

Fig. 5 Technical system of the automated graphic generalization of discontinuous phenomena

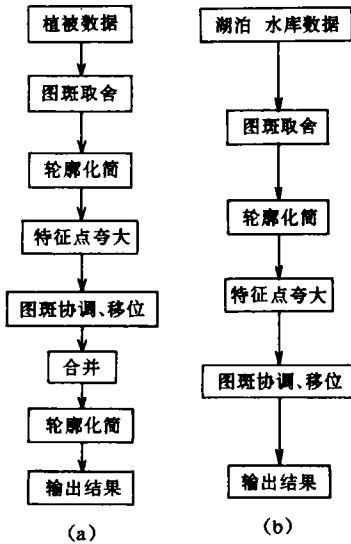


图6 单要素图斑自动概括流程

Fig. 6 A workflow of the automated generalization on single feature area patches

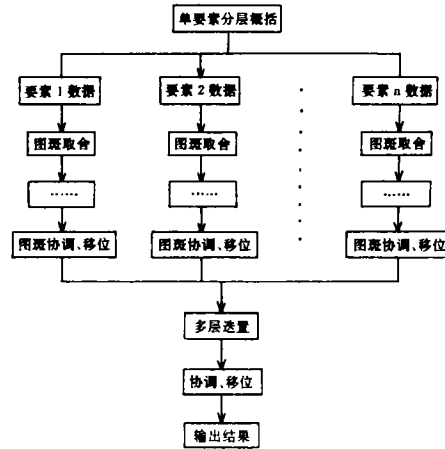


图7 多要素迭置图的图形自动概括流程

Fig. 7 A workflow of the automated graphic generalization on the map with multi-feature area patches

种要素的情况。其概括流程见图 6-a。在图上,植被数据从数据库中被提取出来后,首先进行图斑取舍处理,然后经过轮廓化简和特殊点夸大处理,一些由于弯曲夸大而造成相互间距离太近(以至于不能分辨)的图斑要实施移位,发生部分重叠的图斑则予以合并。合并后的图斑轮廓可能有少许改变,可视情况再次实施轮廓化简处理。最后,检查无误后输出结果图。

3.3.2 湖泊、水库图斑的图形自动概括流程 与上例相仿,这里也是假定结果图上只有湖泊或水库一种要素的情况。其概括流程见图 6-b,与植被图斑概括不同,它们一般不进行合并,因而也就没有“再化简”的步骤。

3.3.3 多要素迭置图的图形自动概括流程 对于多要素并存的情况,处理流程是先进行单要素分层概括,然后将各层迭置,着力处理好要素间的协调和移位,最后输出结果图(图 7)。

4 结语

(1) 本文提出的方法只尝试解决面状地物的一类——非连续分布面状地理现象的图形概括(即取舍、化简、夸大、移位、合并)问题,未涉及制图综合的另两个问题:质量概括和数量概括。原因是后两个问题对于点、线、面状各种地物的制图综合在方法上共性较多。换言之,非连续分布面状地物在质量概括和数量概括方面较之其它地物特殊性较少,其处理方法也不如本文涉及的几个方面的概括问题那么复杂。此外,数量概括和质量概括的实质是制图对象分级、分类的概括,而对于分级、分类问题,地图学界已有相当深入的研究。这也是本文没有涉及的原因。

(2) 本文论述的“自动概括”方法是以交互编辑为前提而设计并实现的一些算法和软件。如前所述,真正的“自动概括”应该是以批处理式设计为起点的,集专家知识和经验与数学模型于一体的智能型专家系统。这是制图综合自动化的真正出路所在。

参 考 文 献

- 1 F. 特普费尔. 制图综合. 北京:测绘出版社,1982.
- 2 Muller J C. and Wang Zeshen. Area-patch generalization; a competitive approach. The Cartographic Journal, 1992, Vol 29, 137-144.
- 3 Lee, Dan. Cartographic Generalisation. Unpublished Technical Report Intergraph Corporation, USA.
- 4 Cromley R G and Campbell G M. Integrating quantitative and qualitative aspect of digital line simplification. The Cartographic Journal, 1992, Vol 29, 25-30.
- 5 Muller. J C. Optimum point density and compaction rates for the representation of geographic lines. Preceedings of Auto-Carto 8, 221-230
- 6 Douglas and Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographic, 1973, 10 (2)
- 7 王长耀等. 普通地图制图综合原理. 北京:测绘出版社,1993.
- 8 祝国瑞,徐肇忠. 普通地图制图中的数学方法. 北京:测绘出版社,1990.
- 9 毋河海. 地图数据库系统. 北京:测绘出版社,1990.
- 10 胡友元,黄杏元. 计算机地图制图. 北京:测绘出版社,1987.

11 廖克. 试论自然轮廓界线的特点与绘制方法. 地理集刊, 1963, 14.

AN AUTOMATED SOLUTION TO THE GRAPHIC GENERALIZATION OF DISCONTINUOUS AREA PHENOMENA

Qi Qingwen Liu Yue

(Institute of Geography, Chinese Academy of Sciences, Beijing, 100101)

Abstract

In this paper, some principles of automated generalization on the discontinuous area phenomena were discussed, followed by the strategies of auto-generalization. Furthermore, the authors expounded the algorithms of every operation of the figure auto-generalization of area-patch, i. e., selection/elimination, simplification, exaggeration, displacement, and aggregation. Finally, the implementation methods of the above algorithms, and the operational workflows of a few cartographic features, were reported.

Key words Area phenomena, Generalization, Figure Auto-generalization