

地理信息系统中由线段数字化 记录重建多边形的方法

周心铁

(中国科学院地理研究所资源与环境信息系统实验室)

1987年1月12日收稿

摘 要

本文研究由地图数字化线段记录重新形成多边形的方法。从结点、线段和多边形的拓扑关系出发,着重分析线段的连接、多边形组成线段的序列寻找、岛状多边形的处理、多边形属性编码等,为向量结构地理信息系统建立有关多边形的数据库记录提供算法基础。所述的方法无需规定线段数字化的方法与次序。

一、引 言

地理信息系统的数据内容涉及分布于地球表面的各种自然现象。如有关土地资源的土地类型、土壤类型、土地的利用条件(坡度、高度、土壤的侵蚀程度等);有关矿产资源的矿产分布、地质条件;有关水资源的流域范围、水系分布;有关环境特征的环境质量条件、污染分布等。也涉及社会人文、经济现象的空间分布,如行政管理范围、土地的利用现状、道路分布、人口分布、城镇结构、工业分布与规划等等。可以看出,这些事物与现象都与空间位置有关,我们称之为空间实体。这些空间实体可以用图形表示在一个二维平面上。用来表达它们的数据结构主要有两种:向量结构与格网结构。它取决于图形数字化的方式。本文仅讨论由向量结构来表示的地图元素数字化数据。基本的拓扑数据元素为结点、线段和多边形^[1]。由一对坐标 (x, y) 确定一点在平面中的位置。由若干点可构成对一条线段的模拟。点本身不具有属性意义。连结两点可构成一个向量,有方向性,它取决于数字化的方向。结点为特殊的点。它有以下两种情况:(i)特征点,表示点状空间实体,具有属性意义。如城市中心点、矿点、泉水点等。(ii)表示线性特征的两个端点,不具有属性意义。本文中涉及的结点仅指线段的两个端点,包括起始结点和终止结点。线段由两结点及结点间的一组有序点组成。它包含一个或若干个连续的向量。多边形表示面状空间实体的平面分布,由一个或若干线段组成的闭合范围,具有属性意义。在这里,线段是作为两个多边形的公共边界而存在的。从线由点构成、多边形由线构成这一角度看,这些实体间成为一种分级关系。但在实际上,它们之间应是一个网络关系^[2]。实体间空间关系的描述是通过结点、线段、多边形相互间的连接来实现的。当然,除了利用这种拓扑关系外,还需要补充实体的位置信息(组成线段之点的坐标记录)及属性信息。最终由数字化记录来恢复地图上的空间实体。

在数字化过程中，无须事先规定线段数字化的顺序和方向，它们可以是任意的。并且，每一线段都仅数字化一次。如何由这些看起来杂乱的线段把它记录下来构成图幅内的多边形是地理信息系统数据输入处理中的关键内容。

二、线段的连接

对于每个结点都可以确定一个或几个线段，其不同之处在于这些线段以不同的方向与此结点相连。这里有两方面的含义：一是对于这些线段来说，这一结点可以是起始结点 (FND)，或是终止结点 (TND)。二是各线段以不同的角度交于此结点，根据其方位角，它们以一定的次序构成连接于此点的线段序列 (图 1)。

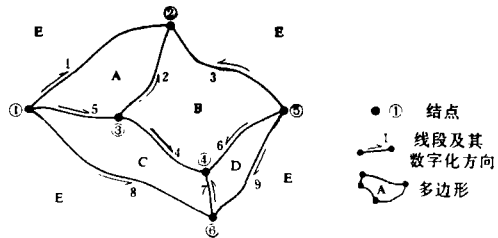


图 1 多边形、线段、结点的空间关系

Fig. 1 Relationship of Graphic Elements: Polygons, Lines and Nodes

在某一结点上的连接线段序列可表示为 $\{S_i\}$, $i = 1, 2, \dots, n$, n 为此结点上连接线段数目。设 $R(S_i)$ 为各线段与此结点相接的向量之方位角 (以正北方向为基准)。令 $R(S_i) > R(S_{i+1})$, $1 \leq i < n$ 。其意义是，在各结点上均以逆时针方向构成连接线段序列。以图 1 为例，各结点对应的连接线段序列见表 1。关于结点上各连接线段之构成、方位角及结点对于各线段的起止关系 (起始结点或终止结点) 等均可在数字化过程中自动记录。

表 1
Table 1

结点号	连接线段序列
1	8, 5, 1
2	1, 2, 3
3	5, 4, 2
4	4, 7, 6
5	3, 6, 9
6	7, 8, 9

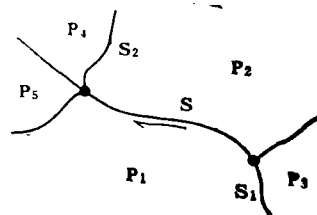


图 2 线段 S 的邻接线段

Fig. 2 Two Lines (S_1, S_2) Adjacent to Line S, the Common Border of Polygon P_1 and P_2

对于一线段 S ，设其左多边形为 P_1 ，右多边形为 P_2 。确定线段连接关系的方法是，在起始结点的连接线段序列中寻找出 P_1 与其它多边形的公共边界线段 S_1 ，称其为对于多边形 P_1 ，线段 S 的邻接线段。在终止结点的连接线段序列中寻找出 P_2 与其它多边形的公共边界线段 S_2 ，称其为对于多边形 P_2 ，线段 S 的邻接线段 (图 2)。

这样,我们就有可能对于图幅内的各多边形,以其任意一个构成线段开始,以一定的方向序次搜索相邻接的线段,最终建立每个多边形的组成线段序列。这里,线段数字化方向的任意性对此并无妨碍。因为在线段的两个结点上都已同时确定了与左右两侧多边形有关的邻接线段。需要考虑的只是搜索的方向与线段数字化的方向是否一致。如果一致,则在线段终止结点上寻找,如果相反,则在其起始结点上寻找。

1. 确定线段在两个结点上的邻接线段

设线段号为 S , 起始结点为 f , 终止结点为 t , 先在结点 f 上确定其邻接线段。 f 上的连接线段序列 $\{Sf_i\}$, $i = 1, 2, \dots, n_f$, n_f 为 f 上连接线段数。在 $\{Sf_i\}$ 中可以找到 $Sf_K = S$, $1 \leq K \leq n_f$ 。其邻接线段 SF 按下面原则确定:

(i) 若对于线段 Sf_{K+1} 来说, 终点 f 是其起始结点, 则有

$$SF = \begin{cases} Sf_{K+1} & \text{当 } K+1 < n_f \\ Sf_1 & \text{当 } K+1 = n_f \end{cases}$$

(ii) 若对于线段 Sf_{K+1} 来说, 结点 f 是其终止结点, 则有

$$SF = \begin{cases} -Sf_{K+1} & \text{当 } K+1 < n_f \\ -Sf_1 & \text{当 } K+1 = n_f \end{cases}$$

随后,可用同样的方法在终止结点 t 上确定线段 S 的邻接线段 ST 。

2. 线段连接关系表的建立

对图幅内的所有线段找出邻接线段后,可以构成线段连接关系表,形式如下:

S_1	SF_1	ST_1
S_2	SF_2	ST_2
	\vdots	
S_n	SF_n	ST_n

在图 1 中各线段所建立的连接关系见表 2。

表 2

Table 2

线段号	起始结点上邻接线段	终止结点上邻接线段
1	8	-2
2	-5	-3
3	6	-1
4	2	-7
5	1	4
6	9	-4
7	-8	-6
8	5	-9
9	3	7

对于图廓线上起始或终止的线段需要作一些特殊的处理。设一图廓结点 a , 联结于此点的线段序列为 $\{S_{ai}\}$, $i = 1, 2, \dots, n_a$, n_a 为线段数。有 $R(S_{ai}) > R(S_{ai+1})$, 当 $1 \leq i < n_a$ 时, S_{a1} 和 S_{ana} 被图廓线分隔。为求其邻接线段, 扩展 S_{ai} 的范围。由图 3 所示的图廓线上按逆时针方向寻找结点 a , 前后相邻的两个结点 b 和 c 。有相应的线段序列 $\{S_b\}$ 和

$\{S_{ck}\}, j = 1, 2, \dots, n_b, k = 1, 2, \dots, n_{co}$

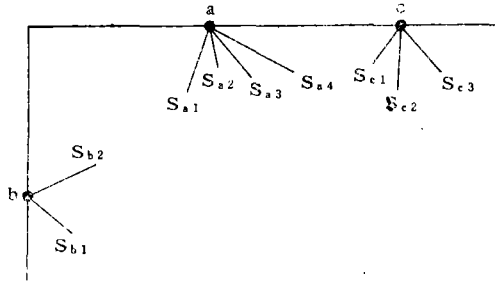


图 3 图廓结点连接的线段

Fig. 3 Line Segments Linked to Frame Nodes

可对图廓结点 a 构成一个新的线段序列 $\{S_{a'i}\}, 1 \leq i \leq n_a + 2$ 。有以下几种情况:

(1) $S_{a1} \neq S_{bnb}$, 此时 $i = 1, 2, \dots, n_a + 2$, 并有

$$S_{a'i} = \begin{cases} S_{bnb} & \text{当 } i = 1 \\ S_{a_{i-1}} & \text{当 } 2 \leq i < n_a + 2 \\ S_{c1} & \text{当 } i = n_a + 2 \end{cases}$$

在这一情况下,如果 $S_{ana} = S_{c1}$, 则将有 $S_{a'na+1} = S_{a'na+2} = S_{c10}$

(2) $S_{a1} = S_{bnb}$, 此时, $i = 1, 2, \dots, n_a + 1$, 并有

$$S_{a'i} = \begin{cases} S_{a_i} & \text{当 } 1 \leq i < n_a + 1 \\ S_{c1} & \text{当 } i = n_a + 1 \end{cases}$$

以图 4 为例, 图幅中各图廓结点的连接线段序列分别为, 结点 1: S_1, S_2, S_3 , 结点 2: S_7, S_6, S_1, S_1 , 结点 4: S_4, S_7, S_6 , 结点 6: S_4, S_7 , 结点 7: S_2, S_3, S_4, S_4 。对图廓结点上起止线段建立连接关系的原理与前面所述的完全一样。

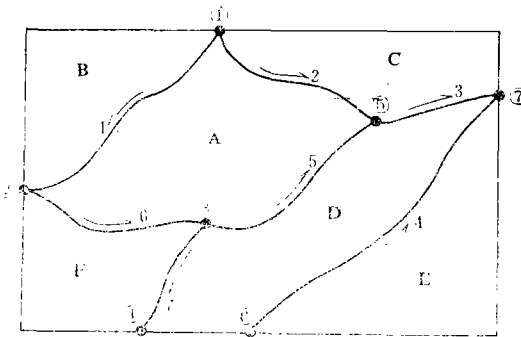


图 4 图廓起止线段与其它线段的空间关系

Fig. 4 Linkage of Line S Terminated or Started on Frame Nodes with Other Lines

三、多边形的恢复

1. 多边形组成线段的序列寻找

在这里,序列寻找是指在数字化图幅范围内,逐个地对多边形按一定方向寻找其组成的线段。序列方向可以规定为顺时针方向。如果一多边形以线段 S_i 为起始线段,从前面的分析可以知道,在线段连接关系表中 SF 和 ST 两列的值 $SF(i)$ 和 $ST(i)$ 分别为 S_i 相应于左侧多边形 P_L 与右侧多边形 P_R 的邻接线段。可先确定 P_L 的组成线段序列,以 $S_{PL}(K)$ 表示, $K = 1, 2, \dots, n_l$ 。 n_l 为组成此多边形的线段数。 $S_{PL}(K)$ 可确定为

$$S_{PL}(K) = \begin{cases} SF(i) & \text{当 } K = 1 \text{ 时} \\ ST[S_{PL}(K-1)] & \text{当 } K \neq 1, \text{ 且 } S_{PL}(K-1) \text{ 为正值时} \\ SF[|S_{PL}(K-1)|] & \text{当 } K \neq 1, \text{ 且 } S_{PL}(K-1) \text{ 为负值时} \end{cases}$$

这里显然有 $S_{PL}(n_l) = S_i$ 。

若 P_R 的组成线段序列为 $S_{PR}(K)$, $K = 1, 2, \dots, n_r$, 则 $S_{PR}(K)$ 可确定为

$$S_{PR}(K) = \begin{cases} ST(i) & \text{当 } K = 1 \text{ 时} \\ ST[S_{PR}(K-1)] & \text{当 } K \neq 1, \text{ 且 } S_{PR}(K-1) \text{ 为正值时} \\ SF[|S_{PR}(K-1)|] & \text{当 } K \neq 1, \text{ 且 } S_{PR}(K-1) \text{ 为负值时} \end{cases}$$

并有 $S_{PR}(n_r) = S_i$ 。

在实际处理中,多边形组成线段的寻找在线段连接关系表中从上向下进行。图 1 中各多边形的组成线段序列见表 3。表中线段号为负值表明线段数字化方向与多边形顺时针线段序列方向相反,正值则表明一致。表 3 中所列多边形 E 的组成线段 $S_8, -S_9, S_3, -S_1$, 为逆时针序列,它们构成岛状多边形的外围,将在下面详细说明处理的方法。

表 3

Table 3

多边形	多边形组成线段序列
A	-2, -5, 1
B	-3, 6, -4, 2
C	-7, -8, 5, 4
D	9, 7, -6
E	8, -9, 3, -1

关于多边形特征的记录文件将由其组成线段序列中各线段数字化记录包含的拓扑信息与形态信息进一步建立起来。

2. 岛状多边形的处理

岛状多边形指多边形位于一较大多边形之中,与其无连接,呈孤岛状。它可能是复合型的或是单一型的。前者指孤岛由一个以上多边形组成,后者指孤岛由单个多边形组成(图 5)。

(i) 岛状多边形的确定

构成岛状各组多边形之组成线段序列中必有唯一的一组逆时针方向序列,如图 5(b) 中的 $-S_1$ 和图 5(a) 中的 $S_8, -S_9, -S_3, -S_1$ 序列。它们构成岛状的外围,即岛状与其外围多边形的边界,而不构成多边形实体。对于单一型岛状来说,另一组同样线段组成的顺时针方向序列则构成岛状多边形本身。而对于复合型岛状来说,其余各组顺时针方向线段序列则构成岛状内各多边形实体。由此可以反推,图幅内若干组多边形组成线段

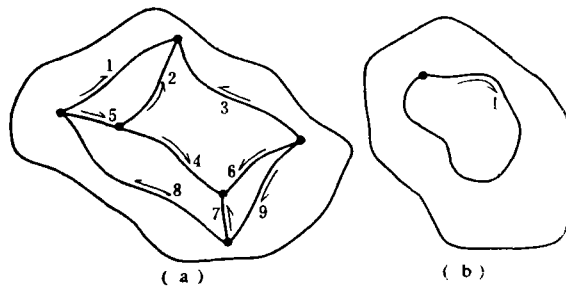


图 5 岛状多边形的组成

Fig. 5 Components of an Island

序列中逆时针方向序列必构成岛状多边形之外围。根据这一特点来确定岛状多边形的外围边界线段。

多边形的面积可由下式确定，

$$S = -\frac{1}{2} \left\{ \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \dots + \begin{vmatrix} x_n & y_n \\ x_1 & y_1 \end{vmatrix} \right\}$$

其中, $x_1, y_1, \dots, x_n, y_n$, 为多边形各线段数字化坐标记录, 它们是多边形的各个顶点。

如果 S 为负值, 则说明这一线段序列构成的多边形为一岛状多边形, 但其本身不成为一实体。它代表的多边形的面积应是直接包围此岛状多边形组成线段序列所计算得到的面积值减去 S 的绝对值。如果 S 为正值, 表明线段序列构成多边形实体, S 值直接代表多边形的面积。

(ii) 外围多边形的寻找

在不确定的条件下, 对于一个岛状多边形可能找到若干同心圆式的外围多边形。这里所指的是直接外围多边形。在同一编码层次中, 它是唯一的。

分两步确定外围多边形。

(a) 缩小寻找的范围

设岛状多边形边界线段坐标的最小 x, y 值与最大 x, y 值分别为 $\text{Min}x_D, \text{Min}y_D$ 和 $\text{Max}x_D, \text{Max}y_D$ 。其外围多边形的最小 x, y 值和最大 x, y 值分别为 $\text{Min}x_W, \text{Min}y_W$ 和 $\text{Max}x_W, \text{Max}y_W$ 。它们必须满足以下关系:

$$\begin{cases} \text{Min}x_W < \text{Min}x_D \\ \text{Min}y_W < \text{Min}y_D \\ \text{Max}x_W > \text{Max}x_D \\ \text{Max}y_W > \text{Max}y_D \end{cases}$$

符合这一条件的多边形可能不止一个。这样, 可能的外围多边形还应当进一步确定。根据下式:

$$\begin{aligned} \text{Min}x_W &= \text{Max} \{ \text{Min}x_{W1}, \text{Min}x_{W2}, \dots \} \\ \text{或 } \text{Min}y_W &= \text{Max} \{ \text{Min}y_{W1}, \text{Min}y_{W2}, \dots \} \\ \text{或 } \text{Max}x_W &= \text{Min} \{ \text{Max}x_{W1}, \text{Max}x_{W2}, \dots \} \\ \text{或 } \text{Max}y_W &= \text{Min} \{ \text{Max}y_{W1}, \text{Max}y_{W2}, \dots \} \end{aligned}$$

由上述条件所求得的多边形与岛状多边形之间关系仍不是唯一的。它可能有并列和包含这两种情况(图 6)。很显然,只有属于包含关系的多边形才是所要寻找的外围多边形。

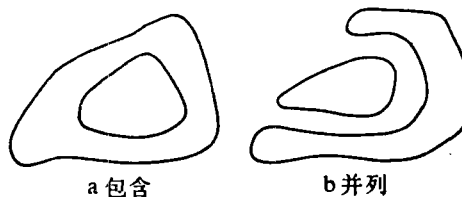


图 6 多边形的空间关系

Fig. 6 Spatial Relationship between Polygons

(b) 包含关系的确定

已求得可能的外围多边形后确定其是否包含此岛状多边形,可通过判别该岛上任意一点是否在这一多边形之内实现。可由岛状多边形边界上一点沿水平方向或垂直方向向外引直线。如果与这多边形有奇数个交点,则说明此岛状被包含在多边形之内,此多边形为其外围多边形。如交点数为偶数,则说明此岛状与此多边形为并列关系。

对于确定的外围多边形来说,它的面积必须扣除包含在内的岛状多边形的面积。

3. 包含图廓线之多边形的处理

如果一多边形以图廓线上一结点起始或终止,则必定还有另一图廓结点连接该边界线段或一其它线段,成为此多边形的边界。在确定该多边形组成线段序列以后,另一个需要解决的问题是进一步确定多边形内这两个图廓结点之间所包含的图廓线范围,将它作为此多边形一个新的组成线段,从而构成一个完整的多边形。

根据已确定的多边形组成线段序列,以顺时针方向找到线段终止的结点,并确认为图廓结点。以此作为新组成线段(图廓线段)的起点。正如上面所说,必有另一个图廓结点为此线段的终点。下一步则是根据这两个图廓结点在四条图廓线(左、右、上、下)的不同部位,确定起点与终点之间所包含的图廓顶点的个数与位置。仍以图 4 为例说明。对多边形 C 来说,已找到其组成线段为 $-S_3$, $-S_2$ 。由线段数字化记录及上述原则不难确定结点 1、图廓顶点 (x_1, y_1) 及结点 7 相联结的向量可形成多边形 C 的一个新组成线段,与 $-S_3$ 、 $-S_2$ 共同构成多边形 C。

四、多边形的属性编码

多边形的属性编码是指在数字化地图图幅范围内根据各面状空间实体所属类型(如土壤类型、植被类型)对多边形分别标记相应的属性代码。通常可以通过两种途径来实现编码。一是在多边形的空间形态确定以后进行。按照多边形编号次序从上而下分别指定属性码。另一种则通过构成多边形的组成线段序列完成编码,它需要在线段数字化过程中同时输入线段左右两侧多边形之属性码。这对数字化的速度有所妨碍。在前一种情况

下,对多边形属性编码后还要根据多边形组成线段的空间关系对每一线段的数字化记录补充其左右两侧多边形的属性码信息。

若一多边形属性码为 A ,此多边形组成线段序列为 $S_p(I), I = 1, 2, \dots, n$ 。用 C_{LK} 和 C_{RK} 分别表示线段序列中某一线段 $K (1 \leq K \leq n)$ 的左、右侧多边形的属性码,有

$$\begin{cases} C_{RK} = A & \text{当 } S_p(K) \text{ 为正值时} \\ C_{LK} = A & \text{当 } S_p(K) \text{ 为负值时} \end{cases}$$

由前面描述的多边形组成线段序列的表示及数字化方向的意义可以很容易说明这一确定的原则。

五、结 语

在地学数据分析处理中多边形是经常遇到的空间实体形态。本文所述由线段记录恢复多边形的方法基于地图图形元素数字化方向与次序的不确定性。这样对于数字化操作提供了很大的便利,从而可以提高数字化数据输入的速度。经过数字化、多边形形成的与属性编码等过程所建立的关于空间实体的结点、线段、多边形拓扑信息、位置坐标信息和属性信息的数据文件是地理信息系统中空间数据库的主体。

由于地理信息系统用户指定的查询空间范围根据应用的需要而定,有一定的随机性系统随时需要根据新的空间范围进行线段的剪辑,并重新形成多边形。为此,有必要对空间数据进一步进行有效的组织,如四分树(Quadtree)等方法,以提高寻找多边形组成线段的速度。

参 考 文 献

- [1] Kurt E. Brassel, "A topological data Structure for multi-element map processing", Harvard papers on Geographic Information System, 4, Addison Wesley, Reading MA. 1978.
- [2] Bruce G. Cook, "The Structural and algorithmic basis of a geographic database", Harvard papers on Geographic Information Systems, 4, Addison Wesley, Reading MA. 1978.

Reconstruction of Polygons by Digitized Line Segments for a GIS

Zhou Xintie

(Lab. REIS, Institute of Geography, Academia Sinica)

Abstract

This paper describes a method of reconstructing polygons by digitized line segments in a map coverage. The order and direction of digitization are arbitrary. Based on the analysis of topological relationships of the basic graphic units: Polygon, Line and node, the emphasis of the discussion is placed on the linkage of Lines, the sequential searching for Lines which the polygon consists of, transaction of islands, and attribute coding of polygons. As a result, the series records related to the polygons are then produced for vector-based geographic information systems.