

文章编号: 1007-4619(2005)04-0343-06

GIS 中地理边线不一致性的处理

邓 敏¹, 刘文宝², 冯学智¹

(1. 南京大学 城市与资源学系, 江苏 南京 210093; 2. 山东科技大学 地球信息科学与工程学院, 山东 泰安 271019)

摘 要: 地理边线的不一致性是 GIS 多源空间数据集成、建库和分析中的常见问题。首先分析了地理边线不一致性的类型, 并归结为顶点数目完全和不完全匹配两种情况, 其中前者可以直接利用结点广义捕捉算法处理。为了匹配后者的顶点数目, 提出了投影配点的标准化处理, 并导出了投影点的坐标计算和误差估计公式。此外, 还导出了一致化改正后地理边线长度的误差估计公式。最后通过算例说明所提出方法的实现过程。

关键词: 边线匹配; 不一致性; 误差估计

中图分类号: P208 **文献标识码:** A

1 引 言

空间数据质量是 GIS 用户评价数据适用性 (Fitness of use) 的重要依据^[1, 2]。国际公认的空间数据质量评价指标包括: 位置精度、属性精度、渊源信息、完整性、逻辑一致性、语义精度和时态信息^[3]。以往 GIS 数据质量方面的研究重点集中在位置精度和属性精度, 而对不一致性问题关注甚少^[4, 5]。

随着 GIS 在各行业应用的兴起, 多源数据的集成和融合问题日益增多, 空间数据建库和分析中的不一致性问题也越来越突出^[6]。在许多 GIS 软件 (如 ARC/INFO) 中, 这种不一致性是通过数据编辑中的捕捉 (Snapping) 功能来处理的。其实质是通过预先设置一个模糊容限值, 使在此范围内的点自动捕捉为同一点^[7]。这种处理方法虽然简单, 但却导致空间目标定位精度损失很大, 并带来随后精度评价方面的困难^[8]。通常, 位置不一致性的处理可以视为一个坐标参数的直接观测平差问题^[9]。基于这种思想, 文献^[10]提出了一种结点广义捕捉算法, 其优点是可以利用所有捕捉点中的位置信息和统计信息来生成一个带有误差指标的统计最优新

结点。本文进一步探讨将广义捕捉算法应用于处理地理边线不一致性时产生的有关理论和技术问题, 例如顶点数目匹配的投影方法、投影点的计算和地理边线长度的误差估计。

2 地理边线不一致性的类型

2.1 直线段情况

如图 1, 对模糊容限内的两条线段 l_1 和 l_2 存在两种可能的不一致性关系: (a) 有一个顶点叠合; (b) 没有顶点叠合。这时, 处理不一致性的最简单方法是删除其中的一条线段 (不妨设为 l_2)。注意, 图 1(a) 中只需删除 l_2 与 l_1 中不重合的那个端点, 而 (b) 中则需要删除 l_2 的两个端点。

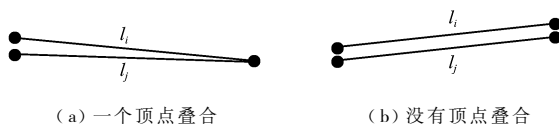


图 1 直线段的不一致性

Fig 1 Inconsistency between two line segments with in a fuzzy tolerance

收稿日期: 2003-07-29; 修订日期: 2004-06-10

基金项目: 国家自然科学基金项目 (编号 40471109)、测绘遥感信息工程国家重点实验室开放研究基金项目 (编号 WK1040304)、基础地理信息与数字化技术山东省重点实验室开放研究基金项目 (编号 SD040201)。

作者简介: 邓敏 (1974-), 男, 江西临川人。1997 年在山东科技大学分别获学士、硕士学位; 2003 年在武汉大学获博士学位; 现为南京大学博士后。主要研究领域为遥感与 GIS 空间数据质量、地理信息的不确定性分析与建模, 以及空间关系理论与应用, 已发表学术论文 30 余篇。

2.2 折线段情况

如图 2,对模糊容限内的两条折线 l_k 和 l_m ,存在两种可能的不一致性关系:(a) l_k 上的每个顶点都能在 l_m 上找到相匹配的顶点;(b) l_k 上的顶点不一定能在 l_m 上发现相匹配的点,反之亦然。这种地理边线的不一致性问题可以出现在当同一地理边界特征存在不同数据采集来源的数据集时或其空间叠置分析中^[11]。通常,位于模糊容限内的不一致折线具有不相等的顶点数。这时,需要通过垂直投影法对两条折线段进行标准化处理,即将两个折线的顶点向对方作垂直投影。如果投影产生的顶点与原有的顶点在模糊容限内,则将投影所得顶点删除。经过这种处理后,两条折线上相应的顶点数会相同。

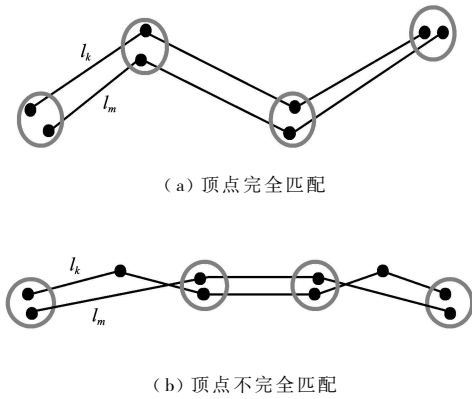


图 2 两折线间的拓扑不一致性

Fig 2 Topological inconsistency between two poly-lines

于是,地理边线不一致性处理的主要流程为:(1)判断两条折线段是否在容差范围内,若是则转(2);(2)判断两条折线段顶点是否相互匹配,若是则转(4),否则转(3);(3)折线段标准化处理;(4)利用广义捕捉算法对模糊容限内的顶点对进行捕捉处理,生成新的顶点;(5)删除原有的折线和顶点,并利用新生成的顶点创建一条新的折线。

3 点的坐标计算与误差估计

3.1 投影点

如图 3,根据解析几何理论可以计算投影点 $z(x_t, y_t)$ 的坐标为

$$\begin{cases} x_t = x_2 + \frac{n}{m} \Delta x_{21} \\ y_t = y_2 + \frac{n}{m} \Delta y_{21} \end{cases} \quad (1)$$

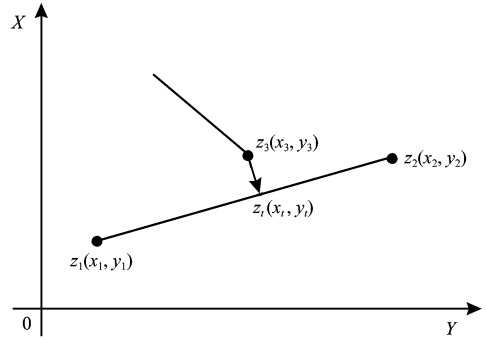


图 3 投影点坐标的计算

Fig 3 Computation of coordinates of the projected point

式中: $m = \Delta x_{21}^2 + \Delta y_{21}^2$, $n = \Delta x_{21} \Delta x_{32} + \Delta y_{21} \Delta y_{32}$ 。对式(1)进行线性化,求得各系数为

$$\begin{aligned} a_1 &= \frac{\partial x_t}{\partial x_1} = \frac{2n\Delta x_{21} - m\Delta x_{32}}{m^2} \Delta x_{21} - \frac{n}{m}, \\ a_2 &= \frac{\partial x_t}{\partial y_1} = \frac{2n\Delta y_{21} - m\Delta y_{32}}{m^2} \Delta x_{21}, \\ a_3 &= \frac{\partial x_t}{\partial x_2} = 1 + \frac{m(\Delta x_{32} - \Delta x_{21}) - 2n\Delta x_{21}}{m^2} \Delta x_{21} + \frac{n}{m}, \\ a_4 &= \frac{\partial x_t}{\partial y_2} = \frac{m(\Delta y_{32} - \Delta y_{21}) - 2n\Delta y_{21}}{m^2} \Delta x_{21}, \\ a_5 &= \frac{\partial x_t}{\partial x_3} = \frac{\Delta x_{21}}{m}, \\ a_6 &= \frac{\partial x_t}{\partial y_3} = \frac{\Delta y_{21} \Delta x_{21}}{m}, \\ b_1 &= \frac{\partial y_t}{\partial x_1} = \frac{2n\Delta x_{21} - m\Delta x_{32}}{m^2} \Delta y_{21}, \\ b_2 &= \frac{\partial y_t}{\partial y_1} = \frac{2n\Delta y_{21} - m\Delta y_{32}}{m^2} \Delta y_{21} - \frac{n}{m}, \\ b_3 &= \frac{\partial y_t}{\partial x_2} = \frac{m(\Delta x_{32} - \Delta x_{21}) - 2n\Delta x_{21}}{m^2} \Delta y_{21}, \\ b_4 &= \frac{\partial y_t}{\partial y_2} = 1 + \frac{m(\Delta y_{32} - \Delta y_{21}) - 2n\Delta y_{21}}{m^2} \Delta y_{21} + \frac{n}{m}, \\ b_5 &= \frac{\partial y_t}{\partial x_3} = \frac{\Delta x_{21} \Delta y_{21}}{m}, \\ b_6 &= \frac{\partial y_t}{\partial y_3} = \frac{\Delta y_{21}^2}{m}, \end{aligned}$$

令 $\eta = \begin{bmatrix} a_1, a_2, a_3, a_4, a_5, a_6 \\ b_1, b_2, b_3, b_4, b_5, b_6 \end{bmatrix}$ 由方差 协方差传播律,

即可计算投影生成点 z_t 的协方差阵:

$$\Gamma_{z_t} = \begin{bmatrix} \sigma_{x_t}^2 & \sigma_{x_t y_t} \\ \sigma_{x_t y_t} & \sigma_{y_t}^2 \end{bmatrix} = \eta \Gamma_{\eta} \eta^T \quad (2)$$

如果各顶点坐标之间是相互独立的,则有

$$= \begin{bmatrix} \Gamma_1^* & & & \\ & \Gamma_2^* & & \\ & & \ddots & \\ & & & \Gamma_n^* \end{bmatrix}$$

而 $K=(\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \alpha_n, \beta_n)$ 中各元素分别是对式 (6) 进行线性化所得的系数, 具体数值如下:

$$\begin{aligned} \alpha_1 &= -(x_2^* - x_1^*) / z_1^* z_2^*, \quad \beta_1 = -(y_2^* - y_1^*) / z_1^* z_2^*, \\ \alpha_2 &= (x_2^* - x_1^*) / z_1^* z_2^* - (x_3^* - x_2^*) / z_2^* z_3^*, \\ \beta_2 &= (y_2^* - y_1^*) / z_1^* z_2^* - (y_3^* - y_2^*) / z_2^* z_3^* \\ &\dots \dots \\ \alpha_{n-1} &= (x_{n-1}^* - x_{n-2}^*) / z_{n-2}^* z_{n-1}^* - (x_n^* - x_{n-1}^*) / z_{n-1}^* z_n^*, \\ \beta_{n-1} &= (y_{n-1}^* - y_{n-2}^*) / z_{n-2}^* z_{n-1}^* - (y_n^* - y_{n-1}^*) / z_{n-1}^* z_n^* \\ \alpha_n &= (x_n^* - x_{n-1}^*) / z_{n-1}^* z_n^*, \quad \beta_n = (y_n^* - y_{n-1}^*) / z_{n-1}^* z_n^* \end{aligned}$$

于是, 式 (7) 可具体表达为:

$$\sigma_{\Gamma^*}^2 = \sum_{i=1}^n (\alpha_i^2 \sigma_{x_i^*}^2 + 2\alpha_i \beta_i \sigma_{x_i^* y_i^*} + \beta_i^2 \sigma_{y_i^*}^2) \quad (8)$$

5 算 例

当同一地理边线在不同的两个数据源中时, 其通常表现为不一致性的两条数字折线段。如图 4, 由于数据质量间的矛盾使得两邻接多边形的共享边

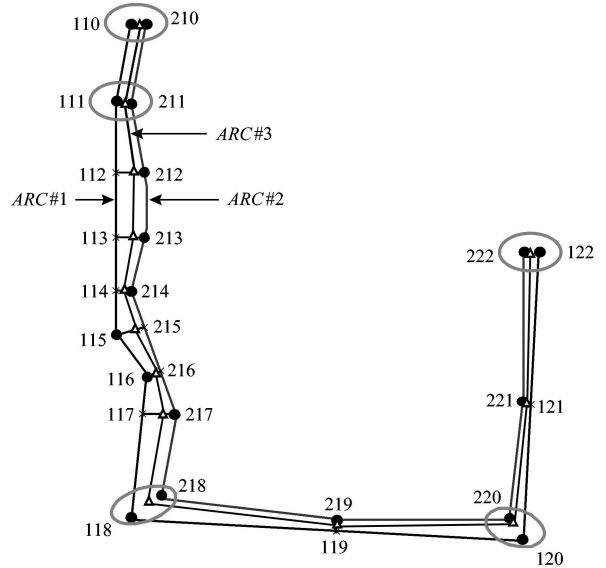


图 4 共享地理边线的不一致性

Fig 4 Two inconsistent geographic lines and the inconsistency handling

存在不一致性。其中 ARC#1 是一条由两个结点和 5 个中间顶点构成的折线, 而 ARC#2 则由两个结点和八个中间顶点构成。这里, 空间坐标数据的精度参照 1:24000 比例尺的土地利用图精度标准设计^[12]。表 1 列出了两折线的顶点坐标和精度信息, 其中编号按图 4 设置。

表 1 不一致性处理前 ARC#1 和 ARC#2 中特征点的坐标及精度

Table 1 Coordinates and accuracy of base points in ARC#1 and ARC#2

ARC#	Point No	x/m	y/m	σ_x /m	σ_y /m	σ_{xy} /m ²
1	110	766232.57	2933428.62	20.00	20.00	0
	111	766173.48	2933016.78	20.00	20.00	0
	115	766174.71	2931923.57	20.00	20.00	0
	116	766291.43	2931718.65	20.00	20.00	0
	118	766231.51	2931107.12	20.00	20.00	0
	120	767812.56	2930937.27	20.00	20.00	0
	122	767855.46	2932281.72	20.00	20.00	0
2	210	766238.42	2933430.93	10.00	10.00	0
	211	766178.36	2933015.96	10.00	10.00	0
	212	766186.02	2932676.00	10.00	10.00	0
	213	766186.35	2932340.91	10.00	10.00	0
	214	766180.48	2932103.01	10.00	10.00	0
	217	766282.63	2931530.97	10.00	10.00	0
	218	766246.00	2931116.43	10.00	10.00	0
	219	767021.01	2931027.25	10.00	10.00	0
	220	767804.62	2930948.78	10.00	10.00	0
	221	767846.12	2931575.24	10.00	10.00	0
	222	767847.28	2932281.16	10.00	10.00	0

和 (2) 计算投影点的坐标及其精度, 结果列于表 2; (iii) 对所有的匹配点对分别进行捕捉处理, 并利用式 (4) 和 (5) 计算捕捉处理后新点的坐标及其精度, 结果列于表 3; (iv) 为所有新生成点建立连接关系; (v) 利用式 (6), (7) 和 (8) 分别计算新生成弧 线

(用 ARC#3 表示) 的长度及其方差, 并与现有的结点捕捉 MOVE 方法相比较, 结果列于表 4。其中 MOVE 方法是将容限范围内精度较低的点移位到精度最高或确认为正确的点上, 并且捕捉后所得点的精度将继承它原有的精度。

表 2 标准化处理后 ARC#1 和 ARC#2 中投影点的坐标及精度

Table 2 Coordinates and accuracy of projected points in normalizing ARC#1 and ARC#2

ARC#	Point No	x/m	y/m	σ_x /m	σ_y /m	σ_{xy} /m ²
1	112	766173.86	2932675.99	15.11	10.01	0
	113	766174.24	2932340.90	14.53	10.00	0
	114	766174.51	2932103.00	17.04	10.00	0
	117	766273.13	2931531.90	15.15	10.05	0
	119	767020.48	2931022.36	10.60	14.10	0
	121	767832.93	2931575.66	14.16	10.01	0
2	215	766211.35	2931930.11	8.32	19.73	0
	216	766250.42	2931711.33	8.27	19.74	0

表 3 捕捉处理后新生成点的坐标及其精度

Table 3 Coordinates and accuracy of the new points generated in snapping

ARC#	Point No	x/m	y/m	σ_x /m	σ_y /m	σ_{xy} /m ²
3	310	766237.25	2933430.47	8.94	8.94	0
	311	766177.38	2933016.12	8.94	8.94	0
	312	766182.46	2932676.00	8.34	7.07	0
	313	766182.45	2932340.91	8.24	7.07	0
	314	766178.95	2932103.31	8.62	7.07	0
	315	766205.94	2931926.88	7.68	14.05	0
	316	766256.41	2931714.94	7.64	14.05	0
	317	766279.75	2931531.16	8.35	8.98	0
	318	766243.10	2931114.57	8.94	8.94	0
	319	767020.76	2931025.61	7.27	8.16	0
	320	767806.21	2930946.48	8.94	8.94	0
	321	767841.73	2931575.38	8.17	8.16	0
	322	767848.92	2932281.27	8.94	8.94	0

表 4 不同处理方法下地理边线长度的精度

Table 4 Accuracy of lengths of geographic lines under different approaches

处理方法	ARC#3 长度 l/m	标准差 σ_1 /m	相对误差 σ_1/l
MOVE 法	5231.05	24.82	5/1000
本文方法	5239.26	22.17	4/1000
差值	-8.21	2.65	1/1000

(2) 一致化改正后新得到的地理边线的位置精度不仅与原始数据集的位置精度有关, 还与所采取的处理方法密切相关。

(3) 与现有 GIS 软件中提供的一致性处理方法相比, 本文方法不仅严密, 能大大提高处理精度, 而且为一致化改正后的结果提供精度信息。

参 考 文 献 (References)

[1] Burrough PA, McDonnell RA. Principles of Geographical Information Systems [M]. Oxford: Oxford University Press 1998

[2] Veregin H. Data Quality Parameters [A]. Longley P A, Goodchild M F, Maguire D J, et al. Eds. Geographical Information System [C]. 2nd Eds. New York: John Wiley & Sons 1999

6 结 论

(1) 一致化改正后新得到的地理边线位置介于被捕捉的地理边线之间, 而且新地理边线的空间位置更接近于捕捉前具有较高位置精度的地理边线。

- [3] Guptill S C, Morrison J L. Elements of Spatial Data Quality [M]. New York: Elsevier Science, 1995.
- [4] Deng M. Extended Models on Topological Relations in Vector GIS: Theories and Methods [D]. Wuhan: Wuhan University, 2003. [邓敏. 矢量 GIS 中拓扑关系的扩展模型: 理论和方法 [D]. 博士学位论文, 武汉: 武汉大学, 2003.]
- [5] Servigne S, Ubeda T. A Methodology of Spatial Consistency Improvement of Geographic Databases [J]. GeoInformatica, 2000, 4(1): 7-34.
- [6] Ai T H, Wu H H. Consistency Correction of Shared Boundary between Adjacent Polygons [J]. Journal of Wuhan Technical University of Surveying and Mapping, 2000, 25(5): 426-431. [艾廷华, 毋河海. 相邻多边形共享边界的一致化改正 [J]. 武汉测绘科技大学学报, 2000, 25(5): 426-431.]
- [7] ESRI Arcedit Users Guide: Interactive Graphics Editor [Z]. California: Environmental Systems Research Institute, 1988.
- [8] University Consortium for Geographic Information Science Research Priorities for Geographic Information Science [J]. Cartographic and Geographic Information System, 1996, 23(3): 115-127.
- [9] Mikhail E M. Observations and Least Squares [M]. New York: A Dun-Donnelley Publisher, 1976.
- [10] Liu W B, Xia Z G, Cui X G. A New Generalized Algorithm of Node Snapping and a Universal Model of Error Propagating [J]. Acta Geodaetica et Cartographica Sinica, 2001, 30(2): 140-147. [刘文宝, 夏宗国, 崔先国. GIS 结点捕捉的广义算法及误差传播模型 [J]. 测绘学报, 2001, 30(2): 140-147.]
- [11] Harvey F. Defining Unmovable Nodes/Segments as Part of Vector Overlay [A]. In: Advances in GIS Research [C]. Taylor and Francis, London, 1994.
- [12] Horn R. Land-use Map Accuracy Criteria [J]. Photogrammetric Engineering and Remote Sensing, 1976, 42(5): 671-677.

Dealing with Inconsistency between Digital Geographic Lines from Multi-data Sources in GIS

DENG Min¹, LIU Wen-bao², FENG Xue-zhi¹

(1. Department of Urban & Resources Sciences, Nanjing University, Nanjing 210093, China;

2. College of GeoInformation and Science, Shandong University of Science and Technology, Tai'an, Shandong 271019, China)

Abstract: Data quality in spatial databases has been widely recognized as one of the most critical issues in GIS. Its key dimension is assessing data suitability, including data accuracy, precision, uncertainty, compatibility, consistency, completeness, availability and timeliness recorded in the data lineage. Over decades, data quality has attracted much attention in the GIS community, and various aspects of data quality have been investigated. However, research on topological consistency remains very weak, though it is crucial for GIS spatial analysis. This paper proposes a new technique on handling topological consistencies as well as tracing error propagation. First, inconsistencies of digital geographic lines are categorized into two groups with or without equal numbers of vertexes. Then a standard formalizing approach has been introduced to make numbers of vertexes matched in the two poly-lines so that the generalized algorithm of node snapping can be used to remove the inconsistencies between the two poly-lines. Third, the equations of estimating the coordinates of the projected points and their errors in the process of standard formalization have been derived. The formulae of calculating the length, together with its accuracy, of new generated poly-lines have also been obtained. Finally, simple examples are provided to demonstrate how to implement the proposed technique into practice.

Key words: geographic line matching; inconsistency; error assessment