

城市基础空间数据质量检查技术研究

沈 涛, 李成名, 赵园春

(中国测绘科学研究院, 北京 100039)

【摘 要】 本文在现有城市空间数据质量检查与控制的理论上, 结合大比例尺数据生产的实际情况, 提出对空间数据进行几何误差、属性误差及逻辑误差进行整体检查的方法, 同时对实际接边误差进行研究, 提出了检查和处理的方法。

【关键词】 城市地理信息系统; 大比例尺; 质量检查; 拓扑一致性

【中图分类号】 P285

【文献标识码】 A

【文章编号】 1009-2307(2005)05-0048-03

1 引 言

利用现代 GIS 的先进技术, 用户可以把不同比例尺、不同精度和不同来源的城市空间数据进行叠加和分析, 应该说, 通过各种技术和软件的结合运用, 将多源和多元的空间数据进行融合和集成的技术已经日益成熟; 而随着航空遥感技术的发展, 空间数据的获取手段、方法和来源日趋丰富, 精度也日趋提高; 数字摄影测量技术的广泛应用和硬件设备性能的提高也使空间数据的采集从速度到精度不断得到提升。

随着 UGIS 在应用中所涉及到的空间数据的范围越来越广, 对数据产品的质量要求也越来越高, 因此很多国内外学者已经认识到了研究空间数据质量及进行质量检查与控制的重要性, 并开始致力于相关的研究工作。目前对空间数据的误差类型、来源^[1]及误差传播模型^[2]、空间位置不确定性、属性精度确定^[3,4]和质量控制^[4,5]都进行了相当深入的研究, 并取得了一定的成果。

但是对于空间数据质量的研究, 一直没有形成完整的理论与技术体系, 所有的研究与成果也基本上停留在其中的某一个方面, 而且研究主要着重于理论方面, 但是在具体如建设数字城市的过程中, 空间数据的质量控制是一个实际的问题, 为此, 我们需要一个易于理解且易实现的措施来保障空间数据质量。本文的研究主要是针对后者, 对城市基础空间数据的质量检查技术进行研究, 并通过软件基本上实现了主要的几种误差类型的自动检查与处理。

2 空间数据质量检查

2.1 必要性

目前, 城市空间数据的主要形式是数字地形图, 其生产已成为测绘部门一项日常工作。然而, 由于传统上测绘部门特别是城市测绘部门侧重于出图的需要, 因此在数据采集时过多地重视了数据获取、特征对象命名及文字表达等过程的精度和速度, 也就是说它们将一幅完整的地图作为首要的目标。其目的是生产可以达到国家标准及其它绘图标准的地形图, 而并不关心这些对象之间及对象与其属性的关系。因此, 目前城市基础空间数据的生产, 主要是针对 CAD 而进行设计的, 而不是用于满足后期 GIS 空间建库的需求。作为 UGIS 的基本信息, 数字地形图在入库前必须对其作进一步的处理、分类和拓扑化操作。而对其产品的检查也应不仅限于成图质量的好坏, 而应检查描述每个地理实体特征的各种几何、属性及拓扑关系的正确性。

但是目前空间数据采集的高效率与相对落后的数据质

量检查手段有较大反差。在 UGIS 数据采集和建库过程中, 质量控制的技术手段主要有手工校对检查和人机交互检查, 这些方法存在明显的缺点。

因此, 图形属性一体化的城市空间数据成果的质量已不是人工检查所能完全胜任的, 研究并建立实用的城市基础空间数据的质量检查技术, 对 GIS 数据进行全面质量检查与控制, 是完全必要的。同时, 在保证城市基础空间数据质量、提高质量检查与控制的工作效率、减轻劳动强度、促进质量控制的标准化等方面也具有相当重要的价值。

2.2 质量检查内容

对 GIS 空间数据的质量标准表述可以分为: 图形精度、逻辑一致性和完整性; 属性精度、逻辑一致性和完整性; 时间精度、逻辑一致性和完整性; 元数据精度、逻辑一致性和完整性^[6]。因此数据的误差主要有四大类, 即几何误差、属性误差、时间误差和逻辑误差。本文主要针对城市数字线划地图的几何误差和逻辑误差, 特别是对逻辑误差中的拓扑一致性进行了深入地研究。

此外, 由于城市大比例尺数据的相邻图幅之间存在接边误差, 特别是进入数字化测图后, 接边问题所涉及的工作量占整个图形编辑工作量的 20%~30%, 因此接边处理也是城市大比例尺空间数据质量检查的重要内容。

2.3 质量检查方法

2.3.1 几何误差检查

几何误差检查包括几何粗差、平面位置精度及高程位置精度的检查。

对于空间数据的粗差处理原则是, 大于 3 倍中误差的空间数据应视为粗差, 不参与精度评定^[6]。

平面位置中误差的计算如下:

$$\left. \begin{aligned} m_x &= \sqrt{\frac{\sum (x'_i - x_i)^2}{2n}} \\ m_y &= \sqrt{\frac{\sum (y'_i - y_i)^2}{2n}} \\ m_p &= \sqrt{m_x^2 + m_y^2} \end{aligned} \right\}$$

式中: x'_i, y'_i 为检测点坐标值; x_i, y_i 为数字地图中同名点的坐标值; n 为检测点个数; m_x, m_y 以及 M_p 分别为在坐标轴方向上的中误差和点位中误差。

间距中误差的计算

$$m_s = \sqrt{\frac{\sum (S'_i - S_i)^2}{2n}}$$

式中: m_s 为地物点间间距中误差; S'_i 和 S_i 分别为相邻地物

点间的检测边长和在图上量取的同一边长值; n 为检测边数。

高程位置中误差的计算如下:

$$m_H = \sqrt{\frac{\sum (H'_i - H_i)^2}{2n}}$$

式中: m_H 为检测点高程中误差; H'_i 和 H_i 分别为检测点高程和图上同名点的高程; n 为检测点个数。

2.3.2 属性误差检查

属性误差检查主要包括空间数据对应的属性数据中要素分类与代码的正确性、要素属性值的正确性及与空间数据连接关系的正确性等。检查可通过回放图与原图套合或在屏幕上逐一显示要素,依据地图要素分类代码表抽样检查要素分类属性、代码的正确性,也可按各属性值取值调出图形元素检查各属性取值的正确性以及图形元素关系的正确性。同时对于要素代码的值域已定的属性,可以通过程序自动检查判断,从而保证属性数据的整体正确性。

2.3.3 逻辑误差检查

逻辑误差检查即空间数据之间关系正确性的检验,本文主要研究了空间数据拓扑一致性检查和接边误差检查。

2.3.3.1 拓扑一致性检查

拓扑一致性检查,即以拓扑规则为基础自动探测空间数据的几何误差。拓扑一致性检查主要针对点与点、点与线以及线与线的拓扑错误进行检查与处理,本文主要研究了重复点、伪节点、冗余点、短悬线误差、长悬线误差以及重复线等误差检测与处理方法。

1) 重复点判别

计算 A、B 两点的距离 s ,

① 如果 $s < \epsilon$,则点 A、B 重合;

② 如果 $s > \epsilon$,则点 A、B 不重合。(其中 ϵ 表示极小的正数,即容限值)

如果点 A、B 都是同一层的要素则表示该两点为重复点,处理时可以直接进行删除;如果点 A、B 不是同一层的要素则应该进行进一步的判断决定是否删除。

2) 伪节点判别

折线或线段的端点称为节点。当一条折线或线段的端点与另一条折线或线段的端点有重合,但重合处这两个端点同时存在(在允许的容限内),这样的点称为伪节点,一般来说伪节点是多余的点,在拓扑处理中应该删除掉。对于伪节点的处理应该是将两条折线或线段进行拓扑合并,使其成为一条折线或线段,两个伪节点就成为该折线或线段的结点。

3) 悬线误差判别

带有悬挂节点的折线或线段叫做悬线,按照悬线长度可以分为短悬线和长悬线两种。对于短悬线的处理主要是在容限范围内,去掉过头线,即称为去除短悬线;对于长悬线可以通过延伸的方法,即将长悬线延伸到另一条相邻

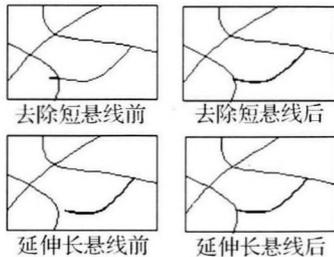


图 1 悬线判别

线的临近节点上或线段中间处(见图 1)。

4) 冗余点判别

如果在折线上某点附近存在两个点号不同的顶点,且两个顶点之间的距离小于或等于 ϵ 容限,则这两个顶点就构成冗余顶点。识别并去掉冗余顶点的操作被称为去除冗余点。检查并剔除冗余点对于空间数据是非常有意义的

操作,它可以有效减少数据量,提高数据显示和计算的速度(图 2)。

5) 重复线判别

如果两个线对象包括节点在内的全部顶点两两重叠,则称为重合线对象。重合线对象的判断不考虑方向。重合线对象往往由部分重合的两个线对象相交后产生。

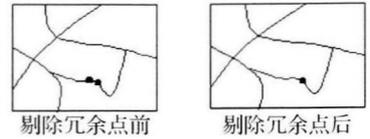


图 2 冗余点判别

对于重合线的处理,如果两条线段都是同一层的要素,则可以直接删除其中一条线段;如果两条线段不是同一层的要素则应该进行进一步的判断决定是否删除,因为对于空间数据特别是城市大比例尺数据常有不同空间要素共用同一线段的情况,如道路边线与花坛的边线共线等,需要进行判断。

2.3.3.2 接边误差处理

数字化时大多是分幅进行的,数字化后相邻图幅边界的几何位置与属性会出现不吻合,因而产生图形的接边误差。由于接边误差产生的种类比较多,情况比较复杂,因此对于接边误差的处理本文采用了多种接边处理算法:

1) 平均法

平均法是取图廓边两边待接点的坐标均值作为接边后的点的坐标,点 $p_1(x_1, y_1)$ 和 $p_2(x_2, y_2)$ 是待接点,点 $p_0(x_0, y_0)$ 的坐标为:

$$x_0 = \frac{1}{2}(x_1 + x_2)$$

$$y_0 = \frac{1}{2}(y_1 + y_2)$$

该方法简单易行,适用于接边误差在精度范围内的各种直线、多义线类的接边处理,容易实现接边的自动批量处理。

2) 强制法

强制法是为主动图幅,将副图幅的一条待接边的待接点强制附合到主图幅另一条待接边的待接点上,以点 p_1 所在的图幅为主图幅,将点 p_2 坐标所在图幅内的坐标强制附和上去,则点 p_0 的坐标为:

$$x_0 = x_1 \quad y_0 = y_1$$

3) 优化法

无论采用了平均法还是强制法,在边界处都要产生一个拐点,如果待接边明显是一条直线边的话,这个拐角既影响图形精度也影响图形的显示,因此我们不但要在图边界上把两个点接在一起,还要考虑到接边后的待接线段在一条直线上,这种方法就是优化法,点 $p_0(x_0, y_0)$ 是接边点,除了要满足将 p_1, p_2 两点结合外,还要保证:

$$\begin{cases} k_1 = \frac{y_3 - y_0}{x_3 - x_0} \\ k_2 = \frac{y_0 - y_4}{x_0 - x_4} \\ k_1 = k_2 \end{cases}$$

3 实 现

本文在以上研究的基础上,采用 VC++6.0 研制了空间数据质量检查软件 NewmapQC,同时使用该软件对西宁市市区的 2775 幅 1:500 及 254 幅 1:2000 地形图数据进行了整体的质量检查,并成功地将其转换为 GIS 数据,最后以检查后的数据为基础建立了西宁市地理信息系统并投入使用,在实践中检验了以上理论的正确性和可行性。

(下转第 64 页)

Voronoi 多边形的建立无需在线进行, 当有新的连锁店建立或老的商店撤销时, 只需要利用 GIS 平台重新计算一次 Voronoi 多边形服务域, 并更新居民点所属的服务商店的属性值就可方便地完成, 系统的服务器和客户端程序无需任何修改, 提高了系统的可维护性。

需要进一步指出的是: 建立在最短距离意义上的商品配送系统实现了完全自动化的任务分配, 其不足之处是没有考虑商品服务需求的空间平衡, 即可能存在这样的情况, 某些居民点的购物需求旺盛, 而距离这些位置较近的商店有数家, 根据 Voronoi 多边形算法所确定的配送商品的连锁商店固定在一家。为了避免某一商店配送商品任务集中的情况, 我们还提供缓冲区分析的方法, 即以消费者所在的居民点为中心, 在某一距离内分布的商店都可能是执行配送任务的商店, 而具体的配送任务由业务人员人工指定, 从而使系统具有充分的灵活性。

参考文献

- [1] 陆大道. 区位论及区域研究方法 [M]. 北京: 科学出版社, 1991.
[2] 周明德. 计算几何—算法分析与设计 [M]. 北京: 清华大学出版社, 2000.

- [3] Calciu M, Salerno F. A New Approach to Spatial Management of Retail Networks, based on German School's Central Place Theory [A]. In: Application to Bank Location, In H. Mühlbacher and J-P Flipo (Eds.) [C], Advances in Services Marketing, Gabler, Wiesbaden, 1997.
[4] NAKANISHI M, COOPER L G. Parameter Estimate for Multiplicative Interactive Choice Model: Least Squares Approach [A]. In: Journal of Marketing Research [C], 1975, (11): 303-311.
[5] 霍亮. GIS 系统集成策略探讨 [J]. 测绘科学, 2004, 29 (6).



作者简介: 王远飞(1963-), 男, 博士, 副教授, 主要从事 GIS 应用, 地理数学建模等方面的研究工作。
E-mail: yfwang@geo.ecnu.edu.cn

(上接第 49 页)

4 结论

城市大比例尺空间数据的质量问题是摆在数字城市建设面前的基本问题, 虽然目前已有不少专家和学者关注或致力于这方面的研究, 但仍有不少的关键问题期待解决, 特别是缺乏将理论与实践结合, 并能够投入实际生产应用的相关软件, 这也是目前城市数据生产单位所面临的紧迫问题。

本文首先分析了城市空间数据质量检查的必要性, 并从以下三个方面介绍了质量检查的内容与方法, 包括空间数据的检验、属性数据的检验以及逻辑一致性与完整性的检查。重点对逻辑检查中的拓扑一致性检查和接边误差处理进行了较深入地研究。在拓扑一致性检查中, 本文主要进行了重复点判别、伪节点判别、冗余点判别、悬线误差判别和重复线判别五种拓扑关系检查。对于接边问题的处理, 论文主要试验了三种方法, 即平均法、强制法和优化法。目前这些研究都已经通过质量检查软件实现并投入实际的生产应用。

参考文献

- [1] 史文中. 空间数据误差处理的理论与方法 [M]. 北京: 科学出版社, 1998.

- [2] 刘大杰, 刘春. GIS 空间数据不确定性与质量控制的研究现状 [J]. 测绘科学, 2001, (3).
[3] 史文中, 王树良. GIS 数据之属性不确定性研究 [J]. 中国图像图形学报, 2001(9).
[4] 刘大杰, 刘春. GIS 数字产品质量抽样检验方案探讨 [J]. 武汉测绘科技大学学报, 2000(8).
[5] 杜道生, 王占宏, 马聪丽. 空间数据质量模型研究 [J]. 中国图像图形学报, 2000, 5A(7): 559~562.
[6] 魏克让, 江聪世. 空间数据的误差处理 [M]. 北京: 科学出版社, 2003.
[7] 赵园春. 城市空间数据质量综合评价研究 [D] [硕士学位论文]. 北京: 中国测绘科学研究院, 2004.



作者简介: 沈涛(1976-), 男, 江苏省苏州市人, 博士研究生, 主要从事地理信息系统及数字城市领域的研究。
E-mail: shentao@casm.ac.cn

名人名言

我希望大家多注意物理现象, 你对现象接触得多了, 就可能产生新的直觉的东西, 而这种直觉往往是科学研究工作中的重要环节。费米在做慢中子实验时, 正是一种下意识的直觉, 使他取得了成功。这种直觉存在的基础, 是广深的知识和丰富的经验。

[美籍华人] 理论物理学家 杨振宁