

海量空间数据组织与管理初探

刘纪平

(中国测绘科学研究院,北京 100039)

摘要 在对海量空间数据分析的基础上,提出了以图幅为基本存储单位,利用拓扑数据模型解决大区域海量空间数据组织与管理方法。通过对实验结果分析,表明文中提出的方法是可行的。

关键词 海量空间数据,关联索引,位置索引

1 引言

进入90年代,与地理信息产业紧密相关的应用领域如城市规划、设施管理、土地管理等,生产了大量的空间数据。尤其一些综合管理机构,其空间范围更广,相应的空间数据库规模也相当可观。并且随着空间数据比例尺的增大,空间数据量也不断增加。如美国USGS生产的1:200万数字制图数据库经过更新需制成一张光盘,而美国全国1:10万水系和交通数据,全套则达14张光盘^[1]。我国全国1:100万基础地理信息数据量近300MB,而全国1:25万基础地理信息数据估计达到8GB^[2]。为了使这些海量空间数据得到更好的应用,必须考虑采用更为有效的管理手段对海量空间数据进行管理。

解决大区域空间数据的统一管理是地理信息系统广泛应用面临的主要问题。美国ESRI公司90年代推出的空间数据引擎(SDE)提供了一种解决该问题的方案。它依托于关系数据库环境,采用高效空间目标模型(Efficient Spatial Object Model)对空间实体进行完整组织,这一方面避免了大数据量传统拓扑模型组织数据时需要的大量的硬盘访问操作,另一方面,空间信息已作为关系库中的成员,因此属性信息与图形信息完全一体化了,避免了通过标识码连接属性,大大提高了信息检索的效率^[3]。

但是,这种对大区域空间数据的处理方法并非无懈可击。首先,由于空间目标均是完整保存的,对

现在以图幅为单位进行空间数据采集、更新极为不利,当图形进行修改时对应的空间连接索引必须全库随时重构。其次,就大区域而言,要求空间信息建立的同步进行是非常困难的。此外,这种方法客观上造成了空间信息有多级产品,数据冗余较大,又给管理、维护带来许多问题。

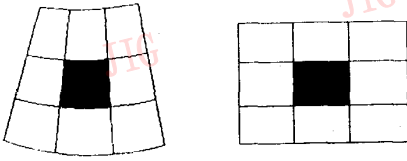
从当前空间数据的生产及数字化过程看,以图幅为单位进行,这一局面很难改变。美国USGS无论以前生产数字地图或现在生产的正射影像图均以图幅为单位进行。澳大利亚首都地区数字地块数据库近10万个地块,都是通过纸质图和平面图以图幅为单位生产的^[4]。数字化过程本身由于处理、加工工作量大,决定了加工过程中多次保存、读取必不可少,因此采用小数据量操作比较可行。

本文将讨论以图幅为基本存储单位,利用拓扑数据模型解决大区域空间数据的管理。

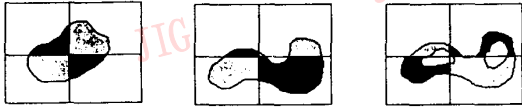
2 海量空间数据分析

按图幅将大区域空间数据进行分割,目前世界各国的一般方法是采用经纬线分幅或采用规则矩形分幅。因此一图幅最多可以与邻近的8个图幅相接如图1所示(见下页)。从几何形态上看不同图幅对同一空间实体具有以下几种分割方式:

简单分割为属于多个图幅的单一实体(一对一分割);分割为属于多个图幅的多个实体;复杂嵌套分割,如图2所示。



(a) 经纬线分幅 (b) 矩形分幅
图 1 图幅之间邻近关系示意图



(a) 一对一分割 (b) 一对多分割 (c) 复杂嵌套分割
图 2 图幅对空间实体的分割

因此,被分割的空间实体经合并后其空间维数保持一致,几何特征参数(包括长度、面积、周长等)进行累加,而质量特征参数(如分类、质地等)则保持不变。同样,分布于不同图幅上的同一个空间实体的各个组成部分具有共同的几何特征(点、线、面特征),同一个空间实体可能分布在一个或多个图幅上。

通过多个图幅相连接可以实现覆盖整个区域。每个图幅,只有那些分布于邻近 4 个边的空间实体才有可能与邻图幅中的相连实体连接,成为完整的同一个实体。不同图幅上,实体的总数也不相同,与其它图幅连接的空间实体数也不等,不同空间实体,被图幅切割的组成部分也不等。

随着比例尺的变化,一方面,同一空间实体的几何特征可以是一致的,也可以不一致。另一方面,图幅的范围也可以发生变化,图幅的负载量也在变化。

3 海量空间数据的组织与管理

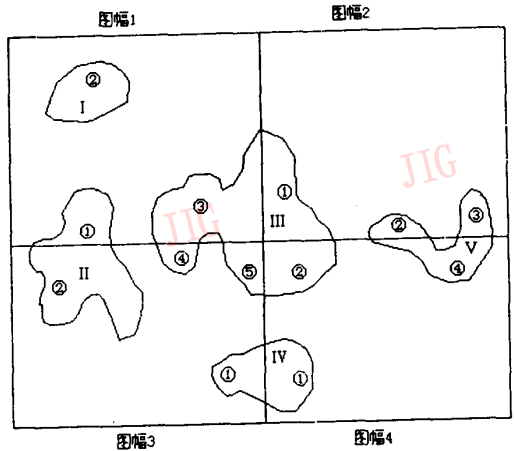
3.1 图幅数据的组织。

对单幅空间数据仍然以拓扑模型来组织,其基本单元为点、线及多边形。现实世界中,每个空间实体定义为一个空间目标,它是用户实际操作的最小单元。一个空间目标可以由一个或多个基本单元组成,它们共有一个标识码 Object-ID。具有相同质量、数量特征的多个空间目标构成一个要素层。多个要素层构成一个图幅。

3.2 图幅之间相同空间目标关系的建立:

(1) 不同图幅基本单元的连接关系

不同图幅上基本空间单元之间,因存在共为一个空间目标的可能,因此它们之间存在着多对多的对应关系,但这种关系仅限于相邻图幅之间,如图 3 所示。



注:③ 代表基本单元 II 代表空间目标编码
其中:

- I 由图幅 1 中②组成
- II 由图幅 1 中①及图幅 3 中②组成
- III 由图幅 1 中③,图幅 2 中①图幅 3 中④⑤及图幅 4 中②组成
- IV 由图幅 3 中①及图幅 4 中①组成
- V 由图幅 2 中②③及图幅 4 中④组成

图 3 不同图幅基本单元之间的连接

从以上的示意图可以看出:

- 每一个空间目标在不同图幅上其组成段(基本单元)总数不定,范围为 1~N,N 为有限数;
- 不同图幅上每个基本单元与其它基本单元连接的总数也为 1~N,是不定的;
- 每幅图中基本单元只与邻接最多 8 幅图中的基本单元连接;
- 当一幅图中图形更改后最多影响与其邻接的 8 幅图的连接。

(2) 不同图幅空间目标连接关系的建立,主要通过建立以下几种主要索引表实现

• 基本单元之间的索引建立

采用变长链表,记录每一基本单元与邻近 8 幅图上基本单元连接的图幅号及基本单元的序号,连接依据为它们是同一个要素并有相同的标识码

Object-ID, 8 幅图的排列序为从右邻接图起为 1 号, 按逆时针方法依次为 2 到 8 号, 0 号为本幅图, 就图 3 上的基本单元而言, 其索引表如图 4 所示。

索引头

基本单元编号	1	2	3	N
连接索引地址	0	-1	3	
图幅号、连接基本单元个数、及序号链表	③	1	2	②	1 1 ③ 2 4 5

注:索引不为 1 时,索引为链表的地址序号,链表中③为图幅号,其它依次为连接个数及基本单元序号。

图 4 基本单元之间索引表

• 空间目标全库索引建立

描述每一个空间目标在全库各图幅的分布,需要记录同一个空间目标在各图幅上的基本单元序号,每一个图幅包含的基本信息单元以空间目标 Object-ID 为序,记录为一个变长链表,就图 3 而言,全库索引表如表 1 所示。

表 1 全库索引表

空间目标	I	II	III	IV	V
图幅 1	1,②	1,①	1,③		
图幅 2			1,①		2,②③
图幅 3		1,②	②,④,⑤	1,①	
图幅 4			1,②	1,①	1,①

• 空间位置索引建立

记录每幅图在整个区域中的位置,以及基本单元在每幅图的位置分布。

3.3 不同图幅数据的调度、管理

大区域、多图幅空间数据的调度、管理主要依据用户工作区的变化进行,其基本思想是,首先根据空间位置索引与用户工作进行实时匹配,确定装载的图幅,为了获得系统较高的响应速度,图幅的装载与释放是动态进行的。其次,对装载的各图幅数据,分别按照图幅、图层、空间目标进行不同层次的管理图幅装载过程如图 5 所示。当进行图形检索属性时,对检索到的空间目标根据其连接索引将空间目标连接为一个整体。当进行属性检索图形分布时,根据全库索引搜索图形分布将覆盖整个区域,而不是单幅图。但就每一个单图幅而言,只需在本图幅查询目标表中记录本幅图查询到基本单元。当进行全库漫游显示时,通过每个空间目标位置与可视区域范围匹配确定其是否显示,对查询到的空间目标依据每幅图

的查询目标表进行显示。当对空间目标进行修改时,如仅仅是几何位置变化,组成结构和连接关系不变,只需重建图幅的空间索引,否则需重建本幅图的连接索引以及全库索引。

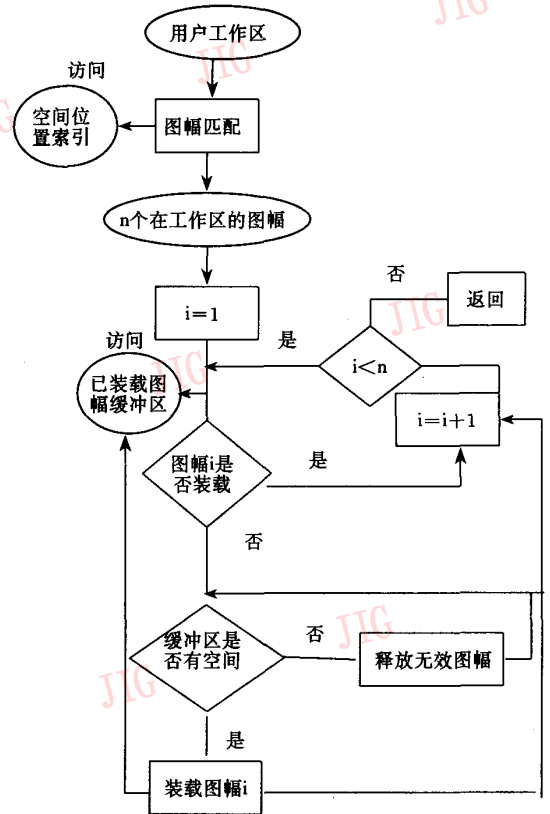


图 5 图幅装载框图

4 实验及分析

实验数据采用全国 1 : 100 万基础地理信息数据,包括境界、公路、铁路、水系、居民地等 15 个图层,涉及 70 个图幅。实验过程中通过建立位置索引、连接索引及全库索引后,将 70 幅图连接为一体,可以在全国范围进行大范围漫游、检索。通过实际检验,表明本文所提出的方法可以用于对大区域、多图幅海量空间数据进行有效的组织、管理。通过对实验过程及结果分析,可以得出以下结论:

1. 以图幅为单位,采用矢量拓扑数据结构可以对大区域海量空间数据进行组织、管理。
2. 通过建立连接索引及空间位置索引,不仅可以保证空间构体的完整,同时,通过搜索这些索引提高了空间数据组织的效率,满足多图幅调度的需要。
3. 由于主要索引是以图幅为单位建立的,因此

局部图形更新时仅仅需要为局部图幅的索引进行更新,比较适于空间数据的更新。

4. 用于空间检索的空间数据与维护、更新的空间数据基本一致,易于管理。

5. 空间数据的存储结构直接影响各图幅调度的效率

可以看出,本文的提出的方法在很大程度上已解决了以图幅为单位海量空间数据的管理,但是还有以下几个问题需进一步完善:

1. 目前的存储结构是以图幅为单位,这样的存储范围,对于动态调度冗余仍然较大,存储结构没有与空间位置索引紧密结合。因此,需要进一步研究与空间位置配合的结构,保证最大限度地降低数据实际装载量。

2. 图幅调度的预处理十分必要,无论计算机速度多么快、图幅切割的多么小,临时加载数据总存在一定滞后,因此根据若干图幅操作,预先判断出所需要加载的图幅,并预先调入内存,无疑是解决这一问题的有效手段。

3. 多比例尺空间数据相关配合才可以解决用户

对大区域空间数据、多尺度的需求。因此,同一地区不同比例尺的空间数据间需要建立连接关系,既要处理几何特征相同空间实体间(分布于不同比例尺上同一实体)的连接,也要处理几何特征不同空间实体间的连接。

参考文献

- 1 蒋景瞳. 美国地质测量局 GIS 技术最新发展. 地理信息世界 95 (3).
- 2 张家庆. 大区域分布式 GIS 软件设计的研究, '94 GIS 学术讨论会论文集.
- 3 Spatial Database Engine (SDE). An ESRI White Paper
- 4 蒋景瞳,唐棣等. 咄咄逼人的地理信息系统世界,1995.



刘纪平 副研究员,1992 年毕业于武汉测绘科技大学制图系,获硕士学位。1992 年 5 月至今在中国测绘科学研究院工作,主要从事 GIS 的研究、开发及应用。

Discussion on Organization and Management for Large Seamless Spatial Data

Liu jiping

(Chinese Academy of Surveying & Mapping, Beijing 100039)

Abstract Based on analysis of large area spatial data, a method of organization and management for large area seamless spatial data is presented in this paper. The proposed method supports topological data model and spatial area is splitted and stored by many sheets. The experimental results have proven that the proposed method is efficient.

Keywords Large seamless spatial data, Relational index, Positional index.

欢迎订阅

《中国图象图形学报》月刊 国内外公开发行 国际标准刊号 ISSN 1006-8961, 国内统一刊号 CN11-3758/TB, 国际流行开本, 96 页正文, 彩封彩页, 印刷精美。

全年订价: 120 元, 每期定价: 10 元

1998 年由邮局发行, 代号: 82—831

也可来函来电编辑部索取样刊和订阅单(或直接汇款, 款到, 发票与过刊立即寄出)

编辑部地址: 北京海淀花园路 6 号

电话: (010)62328784, 62014411—2503 邮编: 100088