

无缝空间数据库的概念、实现与问题研究

朱欣焰¹ 张建超² 李德仁¹ 龚健雅¹

(1 武汉大学测绘遥感信息工程国家重点实验室,武汉市珞喻路129号,430079)

(2 黑龙江省水利厅,哈尔滨市南岗区文中街4号,150001)

摘要 分析了空间数据缝隙产生的原因,提出了空间数据库的逻辑无缝、物理无缝、逻辑查询无缝、逻辑接边、物理接边等概念,讨论了其实现途径,并阐述了基于关系数据库或对象关系数据库的物理无缝空间数据库实现中存在的存储、存取、无缝对象的分段等问题。

关键词 无缝空间数据库 逻辑无缝 物理无缝 逻辑查询

中图法分类号 P208;TP309.13

在现实世界中,地理空间是由地貌、地物组成的连续的表层空间^[1],地理信息则是有关地理空间的一切有用的知识。在计算机世界中,地理信息通过抽象、建模形成数字化的表示形式,通过空间数据库来进行表达、存储、存取和管理。空间地理数据的表示主要有栅格和矢量^[1]两种不同的形式。栅格形式是将地理表层空间划分为一系列网格,空间目标由这些网格的位置及其量化值来表示,矢量形式则是将地理空间的一切事物、概念进行抽象,形成点、线、面,由点、线、面来组成各类空间目标。由于数据获取、数据表达、地图投影等方面的原因,在计算机世界里,地理信息往往被表示成不连续的子空间,为了更好地模拟客观世界,人们需要处理成连续无缝的数据。

1 无缝的概念

1.1 缝隙的来源

空间地理数据缝隙的产生主要有数据的获取、表示与处理等方面。

在数据的获取方面,传统的纸质地图是一种基本的空间数据源。由于地图按图幅划分,按一定的标准,人为地将连续的地表空间划分成若干相互链接的子空间,使同一地物被划分成多个部分,分别落在不同的图幅中,从而产生缝隙,如图1(a)所示,有两个图幅A和B,面目标S被分割为两个子目

标 s_1 和 s_2 ,线目标L被分割为两个子目标 l_1 和 l_2 。另一种情况是在数字化时,由于设备的精度或采集的精度使图幅的接边处或在节点处产生缝隙。如图1(b)所示, N_1, N_2, N_3 是在图幅相接处产生的缝隙,这种缝隙一般通过接边来解决; N_4 是在图幅内部产生的缝隙,这种缝隙一般通过节点聚合或数字化时节点捕捉来解决。

在数据的表示和组织方面,不同的表示和组织方法会产生不同的缝隙。以经纬度表示的地理坐标系可以完整地覆盖整个球面,连续无缝。使用平面投影坐标来描述球面必然要遇到跨带投影的裂缝问题^[1]。在空间数据库组织与管理上,目前主要有文件型、文件与关系数据库混合型、全关系型以及对象关系型^[2]。传统的文件型空间数据库、文件与关系混合型空间数据库,按图幅或一定的区域范围以文件的形式来组织与存储空间几何数据,不同的图幅或区域之间存在缝隙。在文件与关系数据库混合型的空间数据库中,空间几何数据贮存在文件中,属性数据贮存在关系数据库中,属性数据和几何数据之间通过内部标识来链接,空间几何数据和属性数据之间存在缝隙。

在数据处理方面,数据处理过程中的容差或处理过程的不一致会产生缝隙。图2是由两层数据叠加后的某区域的数据,上一层是矢量等高线层,下一层是相应的DRG数据层。由原图扫描后得到DRG数据,进行矢量化后再进行几何纠



(a) 图幅的划分产生的缝隙

(b) 数字化过程中产生的缝隙

图 1 数据的获取产生的缝隙

Fig. 1 Seam or Gap Caused in Data Sources Acquisition

正 得到矢量数据再和 DRG 数据进行叠加 结果产生缝隙。产生这种缝隙的原因有 ①数据处理过程的顺序不一致 ②选择的处理参数不一致 ③数字化的精度不够。前两者可以通过选择适当的处理过程或处理参数来解决 而第三条原因则可以通过提高精度来解决。在图 3 中 有 a 、 b 两条相连的弧段 连接的端点分别为 N_1 、 N_2 设 N_1 的东西向坐标为 99.999 9, N_2 的东西向坐标为 99.998 5, N_1 、 N_2 的南北向坐标相同 系统的节点容差为 0.003。在数字化或在节点聚合过程中 由于两点间的距离在容差允许的范围内(图中的矩形部分),有些系统认为 N_1 、 N_2 是同一个点 而当数据精度要求更高时 或数据转到其他一些系统时 容易产生间隙。这种缝隙在数据处理时严格将 N_1 、 N_2 的坐标处理成同一个坐标值就可以解决。

1.2 逻辑无缝与物理无缝

相邻的图幅之间的裂缝(图 1(b)) 经过接边后可以使图幅之间得到很好的链接(图 1(a)) 这种接边称为逻辑接边。在文件系统或文件与关系混合型中 不同的图幅或区域 用不同的文件来存放。在图形显示时 两个图幅或区域的内容看上去不存在裂缝。笔者将图 1(a)所示这种图幅之间链接上的无缝称为逻辑无缝。但逻辑上无缝的相邻图幅 由于在物理存储上分离 使得在现实世界中本应是一个整体的地物目标被分割成了多个对象 存储在不同的文件中 从物理存储上看是有缝的 即物理有缝。物理有缝是文件系统难以解决的问题。在图 1(a)中 面目标 S 被分割为两个子目标 s_1 和 s_2 线目标 L 被分割为两个子目标 l_1 和 l_2 。由于这种缝隙的存在 在查询时必须将 s_1 和 s_2 、 l_1 和 l_2 同时选中 才能正确地反映查询结果 在系统的处理上带来了许多不便。

如果将图 1(a)所示的两个图幅合并成一个更大的区域 AB 对原来两个图幅之间被分割开的对象也进行合并(包括几何合并和对应的属性合并) 如图 4 所示。这时在物理存储上也合成了一个文件 两个图幅之间在物理上真正变成无缝。将这种物理上真正的无缝称为物理无缝 相应的

对象合并过程称为物理接边。

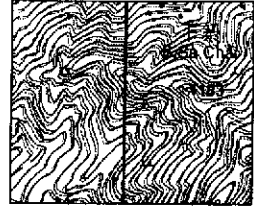


图 2 数据处理不一致产生的缝隙

Fig. 2 Gaps Caused by Different Data Processing Orders

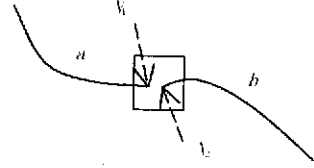


图 3 系统容差产生的裂缝

Fig. 3 Gap Caused by System Tolerance

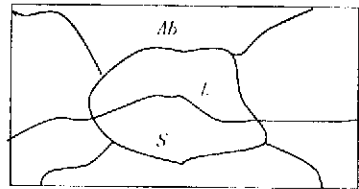


图 4 物理无缝

Fig. 4 Physical Seamlessness

1.3 逻辑无缝查询

在文件系统或文件与关系混合型系统中 对于图 1(a)的逻辑无缝情况 在查询时会产生裂缝。例如 根据点查询面 S 当查询点落在图幅 A 的 s_1 范围内 查到的结果为子目标 s_1 当查询点落在图幅 B 的 s_2 范围内 查到的结果为子目标 s_2 但笔者所希望的结果是整个目标 S 而且查询所得的属性列表应为一记录 不是多条记录 如图 5 所示 图中显示的查询结果实际上就像是物理无缝情况下得到的查询结果一样。这种在逻辑无缝情况下能够得到物理无缝同样正确结果的查询称为逻辑无缝查询 简称逻辑查询。

1.4 属性数据与几何数据之间的无缝

如前所述,在文件与关系混合型系统中,属性数据与几何数据之间存在着裂缝,这种缝隙是由数据模型决定的。属性数据与几何数据之间的裂缝会给系统的几何目标和属性之间带来不一致性。例如,当几何对象被修改后,其面积或周长可能会产生变化,如果不对其属性表中的面积或周长作相应修改,则会产生不一致。又如,某一几何对象被删除后,如果对应的属性记录没有被删除,则由属性查几何对象时,根据属性表中的对象标识找不到几何对象。实际上,系统实现时要将属性和几何对象的处理作为事务来处理,以保证系统数据的正确与一致性。要做到属性与几何数据之间无缝,则必须将它们作为一个整体来表示。

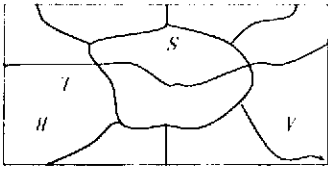


图5 逻辑无缝查询

Fig.5 Logical Querying Seamlessness

1.5 空间数据库不同层次与无缝的关系

如果将空间数据库按物理层、概念层和用户层3个层次^[2]来划分,则物理无缝在物理层上实现,逻辑无缝、逻辑无缝查询在用户层上实现,属性数据与几何数据之间的无缝在概念层和物理层上实现,如图6所示。

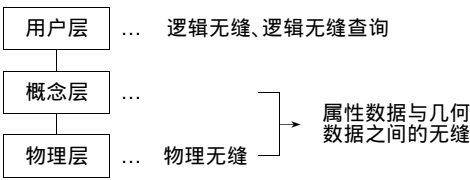


图6 数据库不同层次与空间数据无缝关系
Fig.6 Relation Between Database Levels and Seamless Spatial Data Levels

2 无缝空间数据库的实现

2.1 逻辑无缝与物理无缝的实现

逻辑无缝通过逻辑接边来实现,目前的大部分GIS软件都提供了逻辑接边的功能。在文件型系统或文件与关系混合型系统中,物理无缝通过物理接边来实现,将不同图幅或区域合并成更大的区域,用一个文件来存储。许多GIS软件(如Arc/Info、GeoStar)都在特定的条件下提供了

将多个目标合并的功能,但物理接边完全自动化实现难度大,实现的算法还有待进一步研究,用这种区域合并的方法实现物理无缝,在区域很大时,存储的文件会变得很大:①文件的大小受操作系统的限制;②文件很大时,会给系统的效率、索引和处理上带来很多问题。因此,这种方法并不可取。另一种实现物理无缝的方法是扩大区域的范围,将跨图幅或跨区域的地物目标在数字化或物理接边时整体放在其中的一个区域,每个区域一个物理文件,区域之间有部分相互重叠。这样,地物目标能够保证在存储一级上的物理无缝。如图7所示,图1(a)中的 s_1 和 s_2 、 l_1 和 l_2 经物理接边后分别变成整体地物目标S和L,扩大区域A的范围,S和L完全落在A区域内,图中虚线表示的地物落在B区域内,区域A和B之间有相互重叠的部分。这种物理无缝的实现方法,不需要将整个空间目标存储在一个物理文件中,可以解决部分问题。但当跨多个图幅的地物目标很多时,区域的重叠很多,区域的划分复杂,划分方法不确定,因此在管理上也带来了困难。

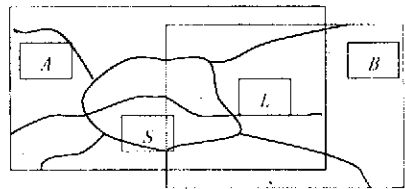


图7 文件系统、混合型系统中物理无缝的实现

Fig.7 Implementation of Physical Seamlessness in Spatial Database Based on File System or Mixed System

目前,真正较好地解决物理无缝问题是用全关系型或对象关系型数据库^[2]来存储和管理空间数据。在传统关系数据库的基础上,扩充变长字段(BLOB字段)或增加空间对象数据类型后(如Oracle 8i Spatial),关系数据库能够存储和管理空间数据。由于不受图幅或空间区域的物理存储限制,复杂的空间对象可以被完整地作为一条记录来存储,从而真正实现物理无缝。除了实现物理无缝外,用关系数据库来管理空间数据库的优点是可以较好地解决空间数据访问的并发控制问题和数据安全性问题。然而,并非使用了全关系型或对象关系型数据库来存储和管理空间数据库就等于实现了物理无缝。例如,如果按图幅将空间数据存入Oracle数据库,这时的数据仍然是物理有缝的,只有对它们进行物理接边后,才可以实现物理无缝。物理接边可以通过两种方式来实现:①在入库之前,利用目前编辑软件提供的空间

目标合并的功能对图幅进行物理接边。接边可以分层进行,以免接边后生成的文件过大^②。在数据入库后进行,按一定的约束条件进行自动接边,或提供可视化界面通过手工和自动化相结合的办法进行接边,具体算法将另文讨论。

2.2 逻辑查询的实现

逻辑接边后,文件系统、文件与关系混合型空间数据库,或全关系型、对象关系型数据库(未进行物理接边)是逻辑无缝的。逻辑无缝空间数据库在用户层上实现逻辑查询无缝,可以部分地解决逻辑无缝带来的查询、显示不一致的问题。

逻辑查询无缝实现的主要思想是:设空间目标对应的属性表中有一个或多个属性项是相关联的,根据这些属性项的相关条件,找出被分割的各个子对象。进行查询时,只要其中一个子对象符合查询条件,则查出被分割的所有子对象,进行属性合并和几何合并后,作为整体表现在界面上。对于图1的逻辑无缝情况,假设S面所在的层对应下面属性表,其中OID=100代表 s_1 ,OID=101代表 s_2 ,Name为相关属性项,即 s_1 和 s_2 在Name上的值相等,如表1所示。

表1 属性表
Tab.1 Attribution Table

OID	Name	Area	...
000	Sname	500.0	
101	Sname	600.0	
...	

1)图形查属性。考虑按点查询。在图上选一个点,查出相应对象的属性,并显示相应的图形。如果查询点落在 s_1 的范围内,先得到 s_1 的OID=100,其属性项Name的值为'Sname',通过属性表得到所有Name值等于'Sname'的子对象的OID,本例为100和101。将OID为100和101的几何对象作为一个整体显示到图形界面上,必要时可对它们进行合并处理,将OID为100和101的属性记录进行合并处理并进行显示,本例中如果Area中的值代表每个子对象的值则合并后的值为1100.0,合并的结果只用于显示,可以不进行物理存储。

2)属性查图形。查Name='Sname'的面对象所在的地理位置,并显示相应的属性值。先从属性表中查出Name='Sname'所有子对象的OID,本例为100和101。将OID为100和101的子对象进行几何合并后作为一个整体显示到图形界面,将OID为100和101的属性记录进行合并处理并进行显示。

按矩形、圆、多边形等方式查询时,先按不同查询方式得到结果集,再根据相关属性项对结果集分组,然后对每组进行合并,合并方法与按点查询类似。

2.3 属性数据与几何数据之间无缝的实现

属性数据与几何数据之间无缝要求将属性数据和几何数据完整地集成在一起。用全关系型或对象关系型数据库来存储和管理空间数据时,可以将同一目标的属性数据和几何数据存放在同一条记录中,几何数据用BLOB字段或几何类型字段来存放。这样,属性数据和空间几何数据就可以完整地集成在一起。然而,在全关系型或对象关系型数据库中,属性数据和几何数据也可以分开用不同的表来存储(这样存储也带来了许多灵活性)。属性和几何数据即使存储在一个表中,对关系表的操作(如投影操作)也可以对属性和几何数据实际上分开使用,因此,此时属性数据和几何数据的一致性的维护也需要做特殊处理。

更有效的属性数据与几何数据之间无缝的集成方法是建立纯的面向对象的数据库管理系统^{3]}。面向对象的方法要求将对象的属性及其方法进行封装,由对象本身来维护自身的状态。对于空间对象,属性和几何数据一起封装在对象中,系统对对象的操作和处理以对象为单位进行。因此,属性和几何数据得到了很好的统一。

3 有关问题的研究与讨论

3.1 存储问题

在关系数据库中,存储管理以页(Pages)或块(Block)为单位^[4,5],页(或块)的大小一般为2K,一条记录如果不足一页,则用一页的空间来存储。用关系数据库来存储和管理空间数据时,每一个空间目标存储在一条记录中。如果空间目标的数据量远远小于一个存储页,则会有很多冗余,对于大比例尺的情况尤为如此。比如一个房屋用4个坐标点,如果用double型来表示,需要64bytes存储空间,相对于一个存储页有很多剩余。这就是将一个较小的空间数据文件转到关系数据库中会占用大得多的数据库空间的原因。由此可见,空间数据管理上,对同样的空间数据量,关系数据库会占用比文件系统更多的存储空间,大比例尺的空间数据冗余比小比例尺大。

有的关系数据库(如Oracle)可以在数据库创建时设置块(页)的大小,根据管理的空间数据的比例尺不同,可以适当调整设置值来解决部分问题。

3.2 存取问题

在关系数据库中,物理接边后空间数据达到物理无缝。对于空间覆盖范围大的空间对象(比如京广线如果用一个对象来表示),每次存取时会作用于整个对象,当用户处理的范围是其中很小的一部分时,也需要取出全部对象,这样会增大网络的流量,同时会影响系统的效率。由于小比例尺的空间数据、空间对象的覆盖范围较大,存取问题比大比例尺数据突出。

3.3 无缝对象的分段管理

对于一个无缝的对象,有时需要对其中的不同部分进行分段表示和管理。比如,一条公路表示成一个无缝对象,但不同的路段有不同的路基、等级、宽度等信息,实际应用时需要对这些信息进行描述和管理,并用于分析和决策中。在逻辑无缝系统中,可以将它处理成多段,而在物理无缝的系统中,必须用动态分段管理技术。

3.4 面向对象的空间数据管理方面问题

面向对象的数据库可以较好地解决物理无缝问题、属性数据与几何数据之间无缝的问题,但纯面向对象的系统也存在效率问题^[6]。在较好地解决属性数据与几何数据之间无缝问题的同时,面向对象方法也会带来数据更新的复杂性。比如,在有缝的情况下,可以根据几何数据与属性数据的关键字来进行链接,从而得到与几何数据相

关的更多的属性信息。但如果将这些属性信息全部集成在对象中 ①当几何数据进行更新时,属性数据也需要做相应处理,增加了更新的复杂性 ②给属性数据的共享带来了一定的难度,在 MIS 与 GIS 结合的应用中尤为明显。因此,在实际系统中的属性数据与几何数据的无缝集成方面,要具体分析,不必一味地追求高集成度。

参 考 文 献

- 1 李爱勤.无缝空间数据组织及多比例尺表达和处理研究 [博士论文].武汉:武汉大学,2001
- 2 龚健雅.地理信息系统基础.北京:科学出版社,2001
- 3 边馥苓,孙红星.基于组件的 GIS 与 MIS 集成系统的实现.武汉测绘科技大学学报,2000,25(增刊):1~5
- 4 Leverenz L. Oracle8 Concepts. <http://www.oracle.com>, 1997
- 5 Durbin J. Oracle8 Utilitiess. <http://www.oracle.com>, 1997
- 6 Zhu X Y ,Gong J Y ,Zhang X D ,et al. Object-oriented Spatial Data Oragnization and Management in GeoStar. Geo-spatial Information Science ,1999, 2(1) :49~54

作者简介:朱火焰,教授。现主要从事空间数据库、GIS、Web GIS、空间信息移动用户应用服务等方面的研究与开发。代表成果:地理信息系统中多用户环境下数据共享一致性问题研究;GeoStar 空间数据组织与管理。

E-mail zxy@rcgis.wyusm.edu.cn

Concepts , Implementation and Problems of the Seamless Spatial Database

ZHU Xinyan¹ ZHANG Jianchao² LI Deren¹ GONG Jianya¹

(1 National Laboratory for Information Engineering in Surveying , Mapping and Remote Sensing , Wuhan University ,129 Luoyu Road ,Wuhan ,China 430079)

(2 The Department of Water Resources of Heilongjiang 4 Wenzhong Road ,Nangang District ,Harbin ,China ,150001)

Abstract :This paper analyzes three aspects of the reasons ,the spatial data source ,the spatial data representation and organization and the spatial data processing. On the basis of these analyses ,we advanced some new concepts about the seamless spatial database ,such as logical seamlessness , physical seamlessness ,logical querying ,physical merge. We also proposed the ways to implement the seamless spatial database in different levels. The problems of storage access and dynamic segment in the physical seamless spatial database are also discussed in detail.

Key words :seamless spatial database ;logical seamlessness ;physical seamlessness ;logical querying

About the author :Zhu Xinyan ,professor. His main research orientations are GIS ,Web GIS ,spatial Database and Spatial information Services for mobile uses. And his typical achievements are study on Data Consistency in Spatial Database System. Spatial Data Organization and Management in Geostar ,etc.

E-mail zxy@rcgis.wyusm.edu.cn