

空间数据库更新技术研究

学号：201028006010028
姓名：柏中强
单位：地理科学与资源研究所

目录

- 1 引言2
 - 1.1 空间数据库的背景和特点2
 - 1.2 空间数据库更新技术3
- 2 空间数据库更新存在的技术问题3
 - 2.1 并发编辑效率低下3
 - 2.2 数据库状态的被动式感知3
- 3 空间数据库更新的几种方法4
 - 3.1 基于版本的更新技术4
 - 3.2 基于增量的更新技术4
 - 3.3 主动式更新技术5
 - 3.4 多比例尺和多源空间数据库的更新5
- 4 实验评价和展望6
- 主要参考文献6

1 引言

1.1 空间数据库的背景和特点

数据库技术是对数据集合进行存储、管理和操作的技术。在二十世纪的后四十年，它飞速发展，经历人工管理阶段、文件系统阶段、数据库系统阶段，已经成为了一项理论成熟，应用极广的数据管理技术。各领域不仅可以借助数据库技术开发信息系统，而且发展了数据挖掘，专家知识决策等多种应用。关于数据模型，数据库技术应用方式研究的论文层出不穷，数据库技术已经是计算机科学与技术学科的重要组成部分。

地理信息系统（GIS）是对能够对地理信息进行存储，管理，显示和分析的系统。地理数据作为地理信息的载体，是 GIS 的基础和灵魂。遥感技术、全球定位系统和现代测绘技术制图技术的飞速发展，为 GIS 提供了海量的数据，计算机技术的发展使得地理学者能够对这些数据进行集成化存储、管理和应用，这里便要利用到数据库技术。地理数据大都是与地理位置有关系的数据，传统的关系数据库模型和技术不足以准确表达和存储这种信息，空间数据库应运而生。与普通数据库相比，它具有以下特点：

1) 存储的一般都是海量数据

由于地理学研究的地球区域表面的复杂性，描述这些信息需要的数据容量很大，常常可以达到 GB 级。图 1 所示为某市需要存储的地理信息。

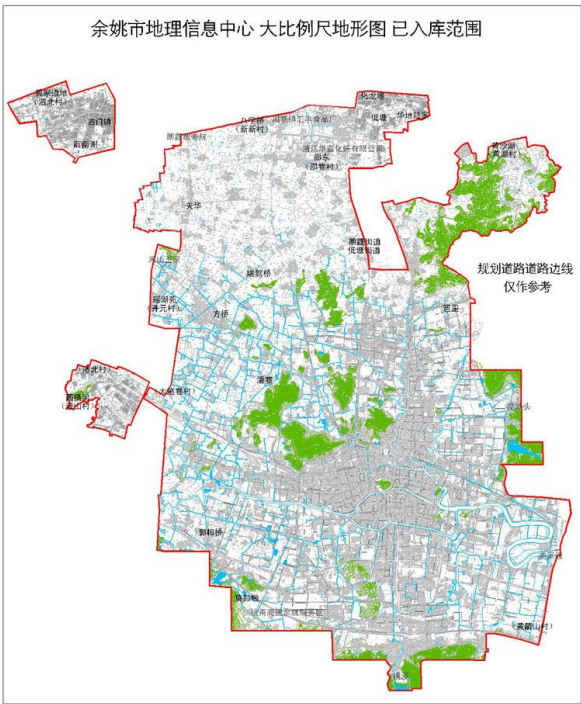


图 1 地理信息存储

2) 访问频率高，涉及到的交互操作多

制图功能能够很好地传达信息，即“一图胜千言”。空间分析功能是 GIS 区别于 MIS 的重要特征。这两项功能都需要快速地访问大量数据。

3) 空间数据模型复杂

空间数据大致可以分为三种类型：几何数据，描述地理实体的几何特征和位置信息；属性数据，描述地理实体的属性信息；关系数据，主要存储地理实体的拓扑关系。这三种数据结构不同，为了能够高效地存储和应用，需要设计独特的模型。

4) 数据多源

遥感，全球定位系统，测绘数据，数字化地图等都是空间数据库的数据源，这也决定空间数据种类多样。

1.2 空间数据库更新技术

地理数据具有时空维度，要保证较高的一致性和现势性。同时，地理数据获取代价较大，根据邬伦教授的估算，地理信息系统应用中数据、软件、硬件之比是 100: 10: 1。地理数据的更新频度大概为一年一次，这种更新可能是局部的，也可能是全部的，数据多源和一致性检查使得更新的工作量变的更大，更加复杂。有时，这种更新需要若干小时或者几天的时间。GIS 空间数据库更新是一项国际难题。目前，有许多学者和相关领域的工作人员都在对它积极展开研究，科技部门和大学也都有专项基金支持研究。本文将对目前的研究成果进行概括和比较，并进行总结和展望。

2 空间数据库更新存在的技术问题

2.1 并发编辑效率低下

在数据库更新技术中，为了防止由于更新故障而丢失数据或产生数据不一致问题采用了锁的概念。锁是对数据操作的权限进行限制。空间数据库更新中存在的主要技术问题是并发编辑的效率很低。在管理空间数据时，通常采用“锁定-修改-解锁”的策略实现多个用户对数据库的编辑操作。如果某个时刻被用户 A 锁定的对象就不能再同时刻被用户 B 修改。这种方法的局限性在于随着并发用户的增加，系统的访问效率将会变的越来越低。在 GIS 领域中对多用户编辑问题的处理方式一般采用悲观锁和乐观锁技术。悲观锁是在事务处理开始前系统先锁定要被修改的数据，其它事务就无法同时修改这些数据，直到记录上的锁被释放为止。显然，对于多用户访问的空间数据库来说，如果前面的用户对某些数据进行修改并持续了很长时间，后面的用户将在很长时间都无法访问这些数据。在乐观锁技术中，各个用户对数据的修改继续在其事务内进行下去，直到事务提交时在检测修改结果间的冲突，如果发生了冲突，根据一定的规则，将某些修改丢掉，这对于那些花了很长时间做操作的用户来说，显然是不可接受的。大多数厂家把精力花在优化锁定粒度大小上，认为粒度越小越优秀。但是，随着并发用户的增多，这种优化的效果是非常有限的。

2.2 数据库状态的被动式感知

传统数据库的所有功能都有一个特性，就是本身是被动的。即只有用户或应用程序显式地发出要求时，它才能执行事务。在普通的数据库系统中，这并不是

一个问题。但这和 GIS 许多应用不符，例如交通系统，战斗管理系统中，数据库不仅存储状态（数据），还要存储规则，要在满足一定条件时，能自动适时地出发相应的活动。为了及时地探测到和响应相关的状态和条件，有两种解决方法，一是用户或应用程序频繁地询问，这样会引起系统抖动，浪费大量的系统资源，而且，询问的频度不好控制。二是在每次变更后就由应用程序检验监视状态和条件，从而决定是否执行活动。这种操作知识的过程嵌入破坏了软件的模块性和不利于应用开发，也加大了开发程序员的工作量。

3 空间数据库更新的几种方法

3.1 基于版本的更新技术

这是针对更新中的第一个问题设计的方法。其基本原理是，在锁机制的基础上建立的一种更高级的机制——在空间数据库的更新重引入版本的概念，各用户在编辑数据库前，先建立自己的数据库版本，然后各用户独立地在自己的数据库版本上进行编辑操作，而无需对其它用户同时访问的数据对象进行锁定。只有用户完成了他的事务后，系统才将他当时的数据版本合并到原来的数据版本中，同时进行数据的一致性处理。该机制能够让连接到空间数据库的用户独立作业而互不干扰。

版本是记录特定对象的各个可选状态的快照。在具体实现中，将初始版本的所有地物要素都详细地记录在一个表格中。而对于其下的所有版本数据库，只记录与初始版本的差异信息。每个版本数据库对应两张表格：“Adds”表和“Delete”表。前者记录该版本相对于初始版本的增量和修改记录，后者记录其相对于初始版本的删除要素记录。通过比较初始版本数据库的两张表，系统可以实时生成任何一个版本数据库的要素内容。另外，对于所有版本的属性信息（如权限，创建时间，作者等），也用一个表格记录。当一个新版本生成或者旧版本删除时，就会在该表中插入或删除一条记录。

该方法中，当多个用户分别通过自己的临时版本向父版本提交修改变化时，可能会发生冲突，这时只可能是多个用户对同一空间要素进行不同的修改。对于所有的冲突要素，把它放到一个图形窗口中进行交互式的处理。在该窗口环境下，用户可以通过手工的方式人工选择，也可以按照设置的参数进行自动选择。可以对冲突划分一定的类型，通过冲突要素的处理程序，将冲突要素的值修改一致的状态，最后通过版本提交来完成更新过程。

3.2 基于增量的更新技术

这里主要针对的是矢量数据，对于矢量地图数据而言，更新的实质可以归结为三种操作类型：增加，删除，修改。“增量”指的是具体操作所引起的变化信息。由于修改具有较高的计算复杂性，不便于控制，增量理论对其不适应。为了建模和操作的简单性，又将增量分为“增加”和“删除”。

为了提高更新操作的效率，保证数据的安全性，需要建立数据库缓存方案，用于转存更新范围的局部数据。具体操作是建立一套相同结构的图层集合，采用一对一的映射方式为每一个图层设计一个缓存表。在数据需要更新时，将更新的局部范围数据暂存到这些缓存表当中，然后再输出到本地，提供给外业人员使用。

在修测完成后, 通过缓存表来间接更新主数据库。

增量识别是该种技术方法的关键技术。前面已经讲增量类型划分为增加和删除两类, 在具体的操作中, 将更新前的数据称为集合 A, 更新后的数据称为集合 B, 集合 C 表示二者的并集, 则增加的内容可以表示为:

$$M=\{a/a \in C, a \in B, a \notin A\}$$

删除内容的集合可以表示为:

$$N=\{b/b \in C, b \in A, b \notin B\}$$

M 和 N 为增量内容。将这两部分内容存入到删除表和增加表中。在后续的增量融合中, 增加地物只需将增加表里的要素转移至目标图层即可。删除地物算法相对复杂些, 林娜等提出了三个条件递进判别算法, 即 1) 图形坐标点数目一致; 2) 最小外接矩形一致; 3) 坐标值一致。依次判断, 前一条件满足在进行后续条件判断, 全部满足则执行删除。

3.3 主动式更新技术

主动数据库 (active database) 早在 1983 年即被提出来了。数据库的主动功能是要要求数据库系统在某些事件发生时主动地执行由用户预先指定的动作或加工处理。数据库系统在运行应用程序的过程中“在什么时刻主动地执行什么动作”是由一些“事件”的发生而引发的, 即“事件驱动”。

主动式更新技术是对触发器控制技术的扩展。触发器是一条语句, 当对数据库修改时, 它自动地被系统执行。要设置触发器机制, 必须要满足两个条件。第一, 指明什么条件下触发器被执行; 第二, 指明触发器执行时的动作。触发器模型也被称为触发器的事件—条件—动作模型。在空间数据库中, 这种触发器机制会有专门的事件探测程序、信号激发程序, 事务调度和处理程序。系统的程序使主动更新更加安全和高效。这种主动式更新技术能够使空间数据库更好地应用于面向外部的情形监控, 报警, 支持推理, 数据交换及配置管理策略的施加, 实时处理及支持合作协同的工作。对于数据库本身, 它能更好地进行完整性控制, 安全性控制, 导出数据, 数据模型定义和继承, 性能测度。

3.4 多比例尺和多源空间数据库的更新

多比例尺空间数据库对空间地理对象通常采用了多重表达, 即同时存储多个尺度的空间数据, 不同尺度的数据之间存在映射关系, 数据更新时先更新较大比例尺的数据, 在根据这种映射关系对较小比例尺的数据进行快速级联更新。在实际生产中, 常见的更新方法包括属性标记法 (基于要素的更新)、区域替换法和基于空间叠加分析法。在具体操作中, 地理对象关联技术和更新数据识别方法是关键技术。前者涉及到地图综合, 数据传输性能等问题, 可以通过设置实体表和关联表来解决。不同的比例尺之间这两张表并不同, 这需要地图数学模型的知识来做指导。后者要对目标要素和源要素进行空间匹配和属性比较, 这方面可以借用 GIS 空间分析中的大量算法。

多源空间数据库更新是 GIS 领域内研究具体的问题, 有专门的自然科学基金项目支持。在国家基础地理信息中心和各省基础地理信息中心, 这项应用也最为广泛。主要有以下技术及进展:

1) 多源遥感影像与矢量化地图数据的自动配准 包括遥感影像与矢量地图的自动配准及不同时期的遥感影像的配准。张祖勋等将物理意义上点线扩展到数学

意义上,并建立共线方程,为自动配准建立了基础。通过点线混合连接策略和平差计算过程,实现了最小二乘法目标约束下的整体最优影像外参数计算。对于与不同时期的遥感影像,利用二维松弛匹配算法。如果比例尺不一致并存在噪声,提出了基于小波分解和多重数据不同时相影像匹配算法。这些方法能达到较高的匹配精度。

2) 地理要素变化检测和智能化提取 根据地图要素实体的分层特征,对不同的地物类型开发了不同的算法。对居民地是基于纹理的先验知识贝叶斯自动检验方法,对于水系是 C 均值法模式分类,对于道路是基于感知编组和视觉分割相结合的道路网提取算法。

4 实验评价和展望

操震洲等在 C/S 模式构建的 GIS 系统中,以 Oracle 10g 为 DBMS, ArcGIS Engine 来开发客户端软件的基本功能,以 ArcSDE 作为连接客户端与数据服务器的功能服务器。采用版本管理技术更新系统的空间数据库,数据的入库,编辑,查询,输出等采用版本管理的模式进行,其中版本管理模块集成有版本的创建,删除,协调,冲突处理,提交等功能。版本冲突处理采用窗口交互模式进行。实验结果证明:在并发编辑用户增多的情况下,采用该技术能简化系统的更新流程,明显地提高更新效率。实验取得了预期的效果。林娜等利用某市 1:1000 地形图作为实验数据,对增量建模和更新技术进行了实验,主要包括 3 个环节:数据出库和入库,数据转移和增量处理。实验算法,系统的增量识别算法能够对地物要素编辑结果进行准确判读,增量融合后,数据的完整性得到了保证。但是,一些数据裁剪、迁移操作等将会破坏数据一致性,需要人工进行一致性维护。

版本更新技术操作更安全,但是,它的无用开销依然很多,与之相比,增量式更新技术更有效,但其算法复杂,对系统性能的要求也更加苛刻。主动式数据库更新技术已经成为空间数据库的基本模块之一。目前,前两种技术方法已经广泛应用于各省市基础地理信息数据库和地籍数据库的更新中,取得了良好的应用效果。多比例尺数据库和多源数据库的更新问题较为复杂,在大多数研究成果转化为技术的背景下,不断有新的优化算法提出。随着人们对空间定位精度要求的提高及数据获取的多样性和科学性的发展,空间数据库的更新的问题将集中于多比例尺和多源数据融合上面。这里既涉及到对历史数据的管理和利用,也包括对现有数据的存储和分发。数据挖掘和知识发现是其中的热点和难点。分布式空间数据库是空间数据库的热门选择。可以肯定的是,无论何时,数据的现势性和完整性始终是空间数据更新的最终要求,为了达到这个目标,人们需要开展更高效更安全的空间数据库更新技术。

主要参考文献

- [1] 操震洲,李清泉.空间数据库的更新技术研究.[J].测绘通报.2007 年第 11 期.
- [2] 陈军,刘万增,张剑清,潘励.GIS 数据库更新模型与方法研究进展.[J].地理信息世界.2008 年 6 月.

- [3] 潘瑜春, 钟耳顺, 赵春江. GIS 空间数据库的更新技术. [J]. 地球信息科学. 2004 年第 1 期.
- [4] 林娜, 王斌. 空间数据库增量式更新机制的研究[J]. 测绘科学. 2008 年 5 月.
- [5] 陈珉. 分布式空间数据库主动数据更新研究. [D]. 武汉大学博士学位论文. 2004 年 4 月.
- [6] 高翔, 袁超, 瞿晓雯, 张红文. 多源数据更新空间数据库的方法研究. [J]. 城市勘测. 2009 年.
- [7] 张锦, 董晓媛, 金雁中. 多源数据更新地理空间数据库的理论与关键技术. [J]. 科技导报. 2005 年第 8 期..
- [8] 傅仲良, 吴建华. 多比例尺空间数据库更新技术研究. [J]武汉大学学报. 信息科学版. 2007 年 12 月.
- [9] 王磊. 面向对象特征的城市空间数据库动态更新机制研究与实现. [J]. 测绘工程. 2009 年 6 月.
- [10] Abrahman Silberschatz 等. 数据库系统概念. [M]. 第五版. 机械工业出版社.