

地理数据库版本驱动的离线编辑机制研究

杨林¹, 张宝才¹, 周顺平², 万波²

YANG Lin¹, ZHANG Bao-cai¹, ZHOU Shun-ping², WAN Bo²

1. 中国地质大学 研究生院, 武汉 430074

2. 中国地质大学 信息工程学院, 武汉 430074

1. The Graduate School, China University of Geosciences, Wuhan 430074, China

2. Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China

E-mail: yanglin_2002_wh@163.com

YANG Lin, ZHANG Bao-cai, ZHOU Shun-ping, et al. Research and realization of disconnected editing driven by geographical database version. Computer Engineering and Applications, 2008, 44(10): 88-90.

Abstract: In the domain of GIS, spatial data are frequently brought out to field or other data-collecting spot to update, this is called disconnected editing. Combined with the concept of MAPGIS7.0 version, a geographical-database version driven disconnected editing method is brought forward, the realization method is discussed with emphasis, and the key algorithm and points are also given. The strategy has been successfully implemented in MAPGIS7.0, and proved to be viable and right, and can satisfy the need of practical application of GIS remote working or field editing.

Key words: disconnected editing; check in; check out; version management

摘要:在 GIS 行业, 大量数据需要频繁地带到野外或其它数据采集现场进行更新, 即离线编辑。结合 MAPGIS7.0 版本管理, 提出了一种地理数据库版本驱动的离线编辑实现方法, 详细阐述了其实现策略, 并给出了实现的关键算法和注意要点。该方法已经被 MAPGIS7.0 平台采纳, 实现了地理数据库的离线编辑功能, 能够满足 GIS 远程或野外作业对离线编辑的实际需求, 证明了该机制的正确性和可行性。

关键词: 离线编辑; 检入; 检出; 版本管理

文章编号: 1002-8331(2008)10-0088-03 **文献标识码:** A **中图分类号:** TP311.52

1 前言

在 GIS 应用领域, 许多行业的机构设置都具有地域分布和行政分级的特点, 在多个城市和地区设有分机构, 按照机构级别又具有上下级的层次关系。分机构的人员常常需要通过城域网或局域网访问上级数据库的数据。对于这样的使用人员来说, 如何通过访问远方上级机构的数据库服务器来获取所需的数据, 是急需考虑的问题。一种解决方案是联机编辑, 即在同上级数据库服务器保持网络持续连接的情况下, 进行实时编辑和更新。这种模式受网络通信状况的影响, 加载大数据量及实时编辑的等待时间往往会令用户难以忍受。而且联机编辑会锁定资源, 影响到其他用户对上级数据库的访问和更新。另一种解决的方案是从远方上级数据库获取数据到本地, 然后断开与服务器的连接, 在离线的环境下继续进行工作, 即离线编辑。这样就避免了用户在保持与远方上级数据库连接过程中造成的巨大花费, 同时又不干扰用户的其它网络连接。另外离线编辑还能够满足去偏远地区或野外的用户, 在离线的环境下能够继续

使用和编辑数据的需求。很多移动设备的用户, 如实地测量工作人员, 他们往往需要长时间的与服务器断开连接。当进行某些特殊作业的时候, 获得的相应数据需要被传输到便携式的设备, 如笔记本电脑上。而这样的设备不得不离线工作, 使得工作人员可以进行独立的操作。当连接恢复后, 对数据所做的所有改变更新到上级数据库中。离线编辑允许组织机构发送他们的空间数据给各部门、代理商或是移动工作者, 并且保证数据的完整性和正确性^[1]。

虽然离线编辑模式可以满足远程客户端的编辑更新问题, 然而在众多的国产 GIS 平台软件中, 很少有看到提供离线编辑功能的产品。鉴于离线编辑对于 GIS 远程或野外作业的实用性, 本文结合国产大型 GIS 软件 MAPGIS7.0 平台, 对地理数据库的离线编辑机制进行了理论探讨和实践。

2 离线编辑原理

在离线编辑模式下, 用户可从一个空间范围很大的地理数

基金项目: 国家高技术研究发展计划(863)(the National High-Tech Research and Development Plan of China under Grant No.2003AA133010)。

作者简介: 杨林(1982-), 女, 博士研究生, 研究方向: GIS 理论及软件开发、空间数据库; 张宝才(1983-), 男, 硕士研究生, 研究方向: 空间数据库、长事务处理、历史数据管理; 周顺平(1967-), 男, 教授, 博士, 研究方向: GIS 理论及软件开发、空间数据库; 万波(1975-), 男, 讲师, 博士研究生, 研究方向: GIS 理论及软件开发、空间数据库。

收稿日期: 2007-08-10

修回日期: 2007-10-11

据库中取出(检出)一小块任意区域的数据,置于本地的个人空间数据库中,在离线编辑会话中对其进行更新。这个更新可持续几小时、几天或几星期,最后用户再将本地编辑的个人空间数据库提交(检入)到主地理数据库^[6]。在这个过程中,其他用户对该区域数据的访问不受影响。一般的离线解决方案都需要对取出的对象进行不同程度的锁定,然而地理数据库版本管理提供了一种不需要锁定对象实现离线编辑的方案。笔者结合地理数据库版本管理从实现原理上进行了深入地分析。

2.1 地理数据库版本管理

地理数据库版本管理可以实现多个用户并发地编辑某个地理数据库而不用明确地锁定要素或者复制数据。地理数据库版本管理基于状态和版本这两个概念。状态是地理数据库变化过程中某一瞬间的标识。任何改变地理数据库的操作都产生新状态。地理数据库中的这些状态可以组织成一棵树,在这棵树的线性结构中描述了各个状态的父子关系。版本则是一个命名的状态。地理数据库中的每一个版本都明确指向一个具体的状态^[4]。这样,在进行地理数据库编辑时,可以为同时进行分工合作的人员定义各自的版本,每个人都在自己的版本空间下工作,不受其他人编辑的干扰,完成编辑后,再进行版本的合并。图1中每个版本指向一个具体的地理数据库状态。版本管理原理在文献^[4]中有详细阐述。

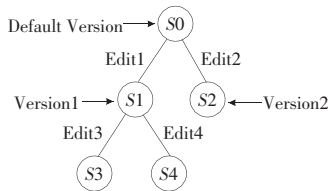


图1 状态树与版本对应关系图

2.2 基于版本的离线编辑机制

在引入了地理数据库版本和地理数据库状态的概念后,本文将基于地理数据库版本提出一种对地理数据库进行离线编辑的方法。离线编辑主要有3个动作:检出数据、离线编辑、检入数据。从一个地理数据库中检出数据的动作,就是拷贝用户选中的空间数据并将它们传递到另一个地理数据库,建立检出地理数据库的过程;离线编辑动作则是在此检出地理数据库上进行添加、修改和删除操作的过程;检入数据的动作,就是将检出地理数据库离线编辑过程中编辑了的部分检回初始地理数据库的过程。

那么地理数据库版本和状态是如何为离线编辑提供支持

的呢?基本的思路是将地理数据库中选中的所有空间对象置于一个独立的特定版本中,客户端在这个独立的版本中进行离线编辑,离线编辑过程中发生的添加、删除和更新操作都利用版本管理自身机制来维护,检入时依据地理数据库中的版本和状态查找出这个独立版本产生的所有变化,最后将这些变化检回原始地理数据库。版本为检出的地理数据库提供了载体,并且提供了一套完善的管理机制来维护在检出阶段进行的所有离线编辑操作,借用这个概念在很大程度上简化了离线编辑的实现思路。需要注意的是这个独立的特定版本指的是基于关系数据库存储的地理数据库,它提供了版本管理的功能。当然这个独立的特定版本也可以是基于文件存储的地理数据库,由于其不支持版本,此处暂不做讨论。

在这个基本思路扩展,更进一步地分析。离线编辑的过程中存在两个地理数据库。用户检出数据的源地地理数据库,称为主地理数据库,除此之外,还需要一个供用户离线编辑的临时地理数据库,称为检出地理数据库。在检出动作发生时,将主地理数据库中被选数据及其相关联对象复制到指定的检出地理数据库。考虑为主地理数据库创建一个主检出版本,主检出版本作为检出动作发生时这个编辑状态缺省版本的子版本被创建,检入时可以用来接收来自检出版本的变化,在数据未检入前,主检出版本不可编辑,但可以满足日常的查询作业。同时考虑为检出地理数据库创建两个版本,记为参照版本和检出版本。参照版本作为这个检出地理数据库缺省版本的子版本被创建,它反映检出时地理数据库的状态,设定这个版本不可编辑,仅作参考之用。检出版本作为这个参照版本的子版本被创建。检出版本是用户进行离线编辑的场所,设定只有对这个版本的编辑才能被检回到主地理数据库,通过对检出版本与参照版本的比较可以获取在离线编辑中发生的所有更改动作,数据检入时主地理数据库的主检出版本接收这些改变,数据一旦被检入,检出地理数据库就不再与主地理数据库相关联,整个离线编辑过程完成。离线编辑过程中各个时刻发生的动作如图2所示。地理数据库与版本对应关系的详细描述如表1所示。

3 离线编辑机制的实现

依据基于版本管理的离线编辑实现思路,笔者进一步思考了离线编辑的各个环节在具体实现过程中会遇到的问题。主要有检入过程中如何查找地理数据库中发生变化的空间数据、如何实现变化数据与主地理数据库主检出版本的融合、检入动作如何驱动等关键问题。

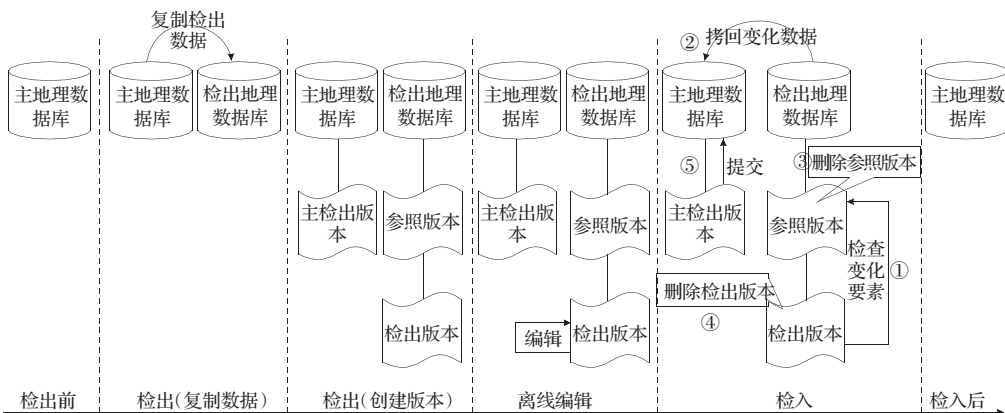


图2 离线编辑不同时刻发生的动作

表1 地理数据库与版本关系表

地理数据库	对应版本	作用
主地理数据库	主检出版本	(1)检出时作为当前连接版本的子版本被创建,检入完成时不被自动删除
		(2)检入时接收来自检出版本的变化
检出地理数据库	参照版本	(1)检出时被创建,不可编辑
		(2)检入时与检出版本做比较,确定发生变化的数据,检入完成后自动删除
	检出版本	(1)记录检出地理数据库中用户对数据的编辑情况
		(2)检入完成时被自动删除

3.1 检入过程

检入过程的主要任务是确定检出地理数据库中的哪些数据发生了变化,并将这些变化检回到主地理数据库。关键就是如何查找变化了的空间数据,以及如何实现变化要素与主地理数据库主检出版本的融合。

检出地理数据库通过版本机制自动维护,查找变化要素可以依赖版本管理中状态与版本之间的关系进行查询。由于版本管理是通过增加A表和D表来记录变化的。添加操作时,在A表增加一条记录,删除操作时,在D表增加一条记录。更新操作时在A表和D表同时增加记录。因此可通过对A、D表的联合查询得到发生变化的要素。版本管理实现原理在文献[3]中有详细阐述。

查找变化要素的算法步骤如下:

(1)获取检出地理数据库检出版本和参照版本当前各自指向的地理数据库状态ID,假设分别为STATE_CHECKOUT和STATE_REF;

(2)在状态树中查询这2个状态的状态分支号;

(3)查询这个状态分支上在STATE_REF和STATE_CHECKOUT之间的地理数据库状态列表SL;

(4)对A表D表进行联合查询,状态满足状态列表SL,在A表中存在,D表中不存在的要素为添加的要素,该要素集合记为CA_i:

$$CA_i = \{F | (F.State \in SL) \text{ and } (F \text{ Exsist in } A) \text{ and } (F \text{ not Exsist in } D)\}$$

(5)对A表D表进行联合查询,状态满足状态列表SL,在A表中不存在,D表中存在的要素为删除的要素,该要素集合记为CD_i:

$$CD_i = \{F | (F.State \in SL) \text{ and } (F \text{ not Exsist in } A) \text{ and } (F \text{ Exsist in } D)\}$$

(6)对A表D表进行联合查询,状态满足状态列表SL,在A表中存在,D表中也存在的要素为更新的要素,该要素集合记为CU_i:

$$CU_i = \{F | (F.State \in SL) \text{ and } (F \text{ Exsist in } A) \text{ and } (F \text{ Exsist in } D)\}$$

(7)在离线编辑中所有变化的要素集合 C=(CA_i)∪(CD_i)∪(CU_i)。

有了变化要素结果集,如何使主地理数据库的主检出版本接收这个结果集中所有的变化信息呢?图3描述了各个地理数据库版本状态的变化关系。检入的过程如下:

(1)为主地理数据库主检出版本创建一个新的状态SN,其父状态为检出时刻的状态SO,主地理数据库主检出版本在离线编辑过程中不可编辑,此过程没有新状态产生,检入的时刻

仍维持着检出时刻的状态SO,因此新状态SN与原状态SO一定是连续的,在这个状态下添加记录能够保证整个地理数据库的逻辑上的一致性。

(2)通过变化要素查找算法找出发生变化的数据记录集合,将这个数据记录集导入主地理数据库的A表和D表中。复制动作完成后将A表和D表中这些新导入的所有记录的状态ID都指向步骤(1)中创建的新状态SN,并将主检出版本的状态也指向新状态SN。检入过程完成。

上述动作等同于在主地理数据库的主检出版本上进行了客户端的所有编辑更新操作,使其来到了一个新的地理数据库状态SN,从而轻松地实现了主地理数据库的主检出版本与检出地理数据库检出版本的融合,完成了从检出地理数据库到主地理数据库的检入。检入完成后,数据管理者有责任清除临时地理数据库的痕迹,此时将检出地理数据库中的参照版本和检出版本的删除。

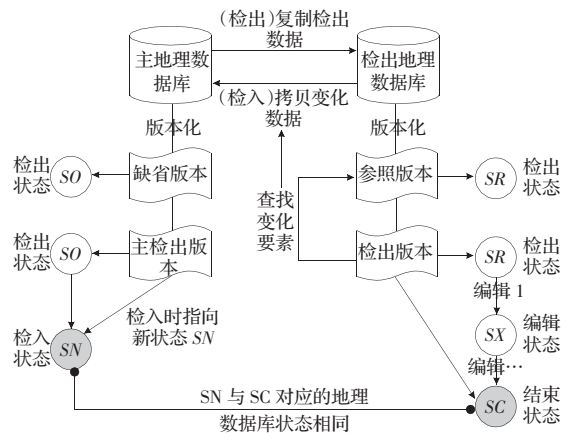


图3 离线编辑状态图

3.2 检入方式

实现模型中的检入动作需要通过合适的方式激发。本文设计了两种检入方式:拉动方式和推动方式。拉动方式由主地理数据库发起,推动方式由检出地理数据库发起,用户可以自由选择检入方式。当有多个检出需要检入到一个主地理数据库时,可以选择拉动方式,打开一个主地理数据库的连接,依次定位每个检出地理数据库,这种方式下检出地理数据库要定位为检入过程的一部分,使得数据的变化主动由检出地理数据库传递到主地理数据库;当结束对检出地理数据库中数据的编辑时,适合使用推动方式,只需获取与主地理数据库的连接,检出地理数据库就可以把改变推入到主地理数据库。

3.3 检入后对主检出版本的处理

主检出版本接收了离线编辑中的所有更改后,用户可以选择对主检出版本的处理方式,是否在主地理数据库中将主检出版本提交于其父版本。如果不选择提交,检入完成后将保留主检出版本,供用户以后编辑;如果选择提交且没有检测到冲突,说明离线编辑修改的部分与源数据没有冲突,可以合并,主检出版本将被删除;如果选择提交但是检测到冲突,说明在离线编辑的时间段,主地理数据库缺省版本和检出地理数据库检出版本对相同的数据都进行了编辑,这时需要根据实际对这些冲突进行协调。协调完成后对主检出版本进行处理。

由于为主地理数据库建立了两个版本。缺省版本和主检出