

面向实体的空间数据模型

叶亚琴^{1,2}, 左泽均^{1,2}, 陈 波^{2,3}

1. 中国地质大学信息工程学院, 湖北武汉 430074

2. 武汉市中地数码科技有限公司, 湖北武汉 430074

3. 中国地质大学研究生院, 湖北武汉 430074

摘要: 空间数据模型是地理信息系统(GIS)研究的基础内容。为适应新一代大型 GIS 的数据量大, 拓扑关系复杂, 数据种类多等特点, 提出了一种新型 GIS 数据模型——面向地理实体的空间数据模型。它以实体为基本表达对象, 具有较强的空间和语义表现力, 支持语义表达和转换, 便于实现等优点。从空间数据管理, 概念数据模型等方面详细地介绍了该种数据模型。实践表明, 采用该数据模型, 系统对真实世界的实体表现力将大大增强, 数据组织更加合理。

关键词: GIS; 空间数据模型; 面向实体的空间数据模型。

中图分类号: P208; TP311

文章编号: 1000-2383(2006)05-0595-05

收稿日期: 2006-05-30

Orient-Entity Spatial Data Model

YE Ya-qin^{1,2}, ZUO Ze-jun^{1,2}, CHEN Bo^{2,3}

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

2. Wuhan Zondy Cyber-Tech Co., Ltd., Wuhan 430074, China

3. Graduate School, China University of Geosciences, Wuhan 430074, China

Abstract: The spatial data model is basic to the research of GIS. New generation large-scale GIS has high levels of data, complex topology relations and many data categories. In order to meet these requirements, we propose a new kind of GIS data model: the orient-entity spatial data model. An entity is the basic, representative object in this model. The advantages are a strong ability for representation and semantics, expression and transformation of semantics, and realization with ease. This paper introduces the data model from the sides of the spatial data management and the concept data model. A system with this model can strengthen its ability of representation to real world objects, and rationalizing data organization.

Key words: GIS; spatial data model; orient-entity spatial data model.

地理信息系统(GIS)从诞生以来一直是信息科学中研究与发展最活跃的部分, 它在空间信息的采集、建模、数据组织、可视化表示和分析等方面取得了大量的研究成果。其中空间信息的数据组织和管理是 GIS 研究的核心问题。空间数据类型繁多、异常复杂, 组织的好坏和检索的效率是衡量一种空间数据应用系统成败的关键, 对空间数据的处理及空间数据模型的研究是 GIS 区别于其他系统的特点之一(郭利川和郭建星, 2005)。

1 空间数据模型

地理信息系统软件已经历了 2 代空间数据模型, 即 CAD 数据模型和 Coverage 数据模型。

CAD 数据模型是通过点、线、面将地理数据存储在二进制文件中(陈静和张树文, 2003), 缺乏属性信息。由于它缺乏对属性数据的支持, 且空间数据不是存储在数据库中, 所以 GIS 软件开发和 GIS 数据共享都比较困难。

Coverage 数据模型(也称地理关系模型)比

CAD 数据模型有了较大的改进. 首先, 它支持属性数据, 将空间数据与属性数据有机地结合在一起; 其次, 它存储了矢量数据的拓扑关系. 因此 Coverage 数据模型提高了对地理空间信息的表达能力和数据的分析能力, 也提高了数据录入的准确性. 但其也有不足之处: (1) 它将空间数据和属性数据分开存储, 2 种数据的关联还需要大量的工作; (2) 它仅用点、线、多边形等几何元素描述空间实体, 从而缺乏对空间实体或现象的显示定义以及基础关系的描述.

随着面向对象的技术的成熟和广泛应用, 我们正在经历以 Geodatabase 为代表的面向对象的数据模型. Geodatabase 实际上是建立在 DBMS (database management system) 之上的统一的、智能化的空间数据库. 它采用面向对象技术将现实空间世界抽象为由若干对象类组成的数据模型, 每个对象类有其属性、行为和规则, 对象类间又有一定的联系. 用户可以在已有的空间数据模型之上, 建立符合应用需求的扩展模型. 因此, 具有较好的客户化能力和可扩展能力 (Michael, 1999).

2 面向实体的空间数据模型

目前广泛应用的各种地理信息系统中所采用的空间数据模型基本上都沿用了图库、图幅、图层、地理对象、几何对象这样的组织结构, 从本质上是面向纸质地图的数字化处理的, 具有比较突出的缺点: (1) 水平分幅、垂直分层的数据组织模型不能直观反映人们对现实世界的感知. 这种以图层作为空间数据处理的基本单位的数据组织模型虽然方便数据采集、组织、存储和管理, 但却人为地割裂了地理空间, 不能直接反映人们的感知. 而且分层的数据组织不能表达不同类型实体之间的空间和语义关系; 分幅割裂了同一个实体的空间连续性, 并削弱了 GIS 的处理分析能力, 尤其是对大范围空间的数据处理能力较弱. (2) 忽视地理实体之间的语义拓扑关系. 例如居民区与飞机场之间的关系、河流与桥梁之间的关系因为分层都难以表达; 这就使 GIS 成了功能层次较低的空间数据存贮和管理系统; 不能满足空间决策分析的需要; 缺乏对规则、定义域、行为的定义; 只能表达狭义的实体关系. (3) 传统数据模型不足以表达一些较特殊的地理现象. 例如地质单元或植被单元之间没有明确的界限, 还有连续的场有时被突然地不连续中断. 精确的对象模型和连续的场

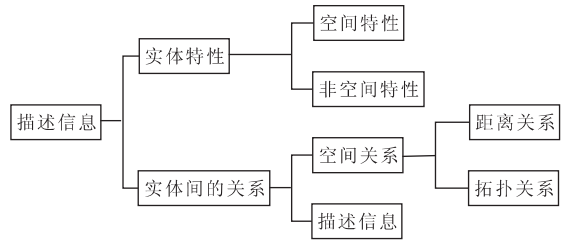


图 1 面向实体的空间数据模型的地理信息的组成

Fig. 1 Components of orient-entity spatial data model's geographic information

模型是现实世界的 2 个极端抽象, 传统的数据模型无法表达, 所以模型表达能力不够. (4) 空间数据模型标准化工作较少.

这样, 在大型的分布式地理信息系统中的空间数据模型就需要突破现有空间数据模型的限制, 具备准确描述地理实体、实体之间的语义关系和标准化程度高的特征. 新一代的空间数据模型的要求具体体现在以下几个方面: (1) 能比较真实地模拟现实世界. 为了达到这一目的, 面向地理实体空间数据模型是根据语义而不是根据几何表示的复杂性来划分实体. 能全面描述地理实体的空间数据模型, 必须要能存储地理实体的空间特征, 非空间特征以及实体之间的空间关系和非空间关系. 信息组成如图 1 所示. (2) 容易为人所理解, 支持语义表达和转换. 首先, 其数据组织和分类符合人类的思维习惯; 第二, 支持语义关系, 以便于语义的表达和转换. 例如沉积岩和砂岩就是类与子类的关系, 而塑料管是水管的子类型等. 重要的是对空间数据中所包含的意义的共同理解. 系统应该提供一种语义的表达和转换机制, 使得具有不同应用背景的系统 and 用户能够方便地交换空间数据而不会造成语义上的误解和损失; 第三, 支持规则. 规则的类型包括空间规则、连接规则、属性规则和关系规则等. (3) 便于在计算机上实现. 该模型既要便于在关系性数据库中还要便于在文件系统中的实现. (4) 便于空间信息图示化. (5) 支持时态数据管理, 具有时序分析功能. (6) 标准化程度较高, 具有强大的互操作性. 基于对上述几个方面的考虑而应运而生的面向实体空间数据模型首先强调人对地理世界的理解, 其次才是如何将人对地理世界的理解图示化. 这种模型将地理世界分解为实体, 通过描述实体的特性和实体间的关系, 建立观察范围内的地理世界的视图, 通过定义与实体特性、实体关系相关的操作, 模拟人类理解地理世界

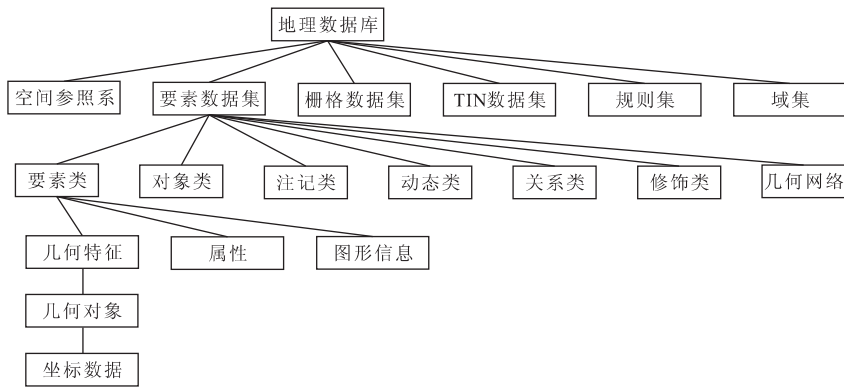


图 2 面向实体的空间数据模型的概念层

Fig. 2 Concept model of orient-entity spatial data model

的语义环境。

在面向实体空间数据模型中,非空间实体被抽象为对象,空间实体被抽象为要素,相同类型的对象构成对象类,相同类型的要素构成要素类,若干对象类或要素类组成要素集,若干要素集构成地理数据库(吴信才,2004)。

要素在某个空间参照系中的几何特征被抽象为几何元素,几何元素由任意的点状、线状或面状几何实体组成,几何实体通过几何坐标点表达(吴信才,2004)。该空间数据模型的概念层次如图 2 所示。地理数据库是用于存储地理数据的数据库,提供管理地理数据服务,这些服务包括有效性规则,关系和拓扑关联。

3 数据结构

实体是现实生活中的地理特征和地理现象,可根据各自的特征加以区分。实体的特征至少有空间位置参考信息和非空间位置信息 2 个组成部分。空间特征描述实体的位置、形状,在模型中表现为一组几何实体;非空间特征描述的是实体的名字、长度等与空间位置无关的属性。

地理实体在模型中表示为要素。要素是由几何实体和属性组成的。它包括简单类型,例如一个界址点、一个行政界线、一块土地;它们的几何形态分别为简单点、简单线和简单区。还有一些复杂类型的实体,例如一个河流的流域。它的几何特性对应的是多种形态的几何实体,所以它的几何特性是一个复合类型。换句话说,通过原子几何实体(点、线、区)的任意组合可表达和描述任意几何复杂度的实体。

几何实体是地理对象的外观特征或可视化形状。地理实体可以用 3 种几何实体表示在地图上:点、线、多边形。继续细分下去,几何形态包括单点、多点、单弧段、多弧段、多边形等。所有的几何实体都是由更为基本的点和弧段组成的。我们将点和弧段叫做空间数据。也就是说几何实体是由空间数据和图形信息组成的。

模型中实体的空间数据是如下表示的。

3.1 要素的表示结构

要素由几何形态和属性组成。几何形态表现为一组几何实体。图 3 引用数据的 ID 包括 2 个部分:几何实体所在的要素类的 ID 和几何实体 ID。要素的类型包括:简单点、简单线、简单区、复杂点、复杂线、复杂区以及复合类型(图 3)。

3.2 几何的表示结构

要素类中所有的几何实体形成了该要素类的几何实体集。所有类型的几何实体都是引用空间数据中的弧段或点。几何实体存储形式为“弧段号”或“点号”的数组和相应的“图形显示参数”。

(1)点几何实体。包括单点、多点;有特定的位置,维数为 0 的空间组分。点几何实体引用的是空间数据的点。引用数据存放的是组成该点几何实体的点号序列;每个点几何实体都有相应的图形信息(图 4)。

(2)线几何实体。包括单弧段、多弧段;由一系列

fID	要素类型	显示优先级编码	时间戳	矩形范围	几何实体个数(N)	几何实体组成:①	属性
-----	------	---------	-----	------	-----------	----------	----

图 3 要素信息结构

Fig. 3 Feature information structure

①代表 fclsID, gid1, inf1, ..., fclsID, gidN, infN

geomID	引用数据的个数(N)	引用数据: (pntNo1, pntNo2, ..., pntNon)	点图形信息
--------	------------	--	-------

图 4 点几何实体信息结构

Fig. 4 Point geometry information structure

geomID	引用数据的个数(N)	引用数据: (pntNo1, pntNo2, ..., pntNon)	线图形信息
--------	------------	--	-------

图 5 线几何实体信息结构

Fig. 5 Line geometry information structure

geomID	引用数据的个数(N)	引用数据: ①	区图形信息
--------	------------	---------	-------

图 6 区几何实体信息结构

Fig. 6 Region geometry information structure

①代表(arcNo1, arcNo2, 0, arcNo3, arcNo4, 0, ..., arcNon)

坐标表示的 1 维的空间组成。线几何实体由空间数据中的弧段组成。它的存储结构记录为弧段序列。弧段方向与子线方向一致用正数, 否则用负数(图 5)。

(3)区几何实体可以包含单环组成的区、多环组成的区。每一个环可能由一条弧段或者多条弧段组成; 是二维空间组分。区几何实体也是由空间数据中的弧段组成的。该区几何实体的边界各环之间用 0 隔开; 在数据中与区方向一致的线用正数表示, 与区方向相反的线则用负数表示(图 6)。

4 空间数据

空间数据集存储与要素类相关的空间坐标数据及其拓扑数据, 空间数据包括点和弧段 2 种类型。根据是否依赖于其他要素类中的空间数据, 点和弧段又可以是“假点”和“假弧段”, “假点”和“假弧段”记录了要素类之间空间数据的依赖关系, 是实现空间边界共享的机制。如图 7 所示, 省界要素类中的元素 B 依赖于河流要素类中的元素 A, 在面向实体的空间数据模型中, 省界要素类中的元素 B 可记录为“假弧段”, 并将该“假弧段”指向元素 A。这样, 河流在地质运动下产生变化时, 省界也会随着变化。

“假点”和“假弧段”不能修改其空间位置, 对应的点位置、弧段位置只能由拥有“真实点”和“真实弧段”的要素类修改。

(1)点: 单个坐标点。①一般点: 只存有空间坐标信息的点。②实结点: 实结点具有拓扑连接意义, 是

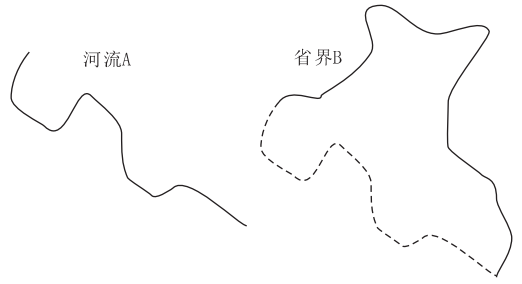


图 7 空间边界共享机制

Fig. 7 Share of spatial boundaries

ID	点类型	引用标志	X坐标	Y坐标	Z坐标	被引用点的ID: (refFclsID, refDotID)
----	-----	------	-----	-----	-----	-----------------------------------

图 8 点信息结构

Fig. 8 Point information structure

ID	弧段类型	维数	Z值	坐标点数	坐标数据+解析线参数	拓扑数据	被引用的弧段的ID
----	------	----	----	------	------------	------	-----------

图 9 弧段信息结构

Fig. 9 Arc information structure

结点, 同时也表示该结点具有地理含义(如界碑)。系统进行自动拓扑处理时不能修改和删除它。例如, 在结点平差时不能移动实结点, 而是应该移动其他的虚结点向实结点靠齐。实结点与拓扑信息一一对应。③虚结点: 虚结点仅仅具有拓扑连接意义, 系统可以自动对它进行修改和删除。每个虚结点也有对应的拓扑信息。即虚结点没有地理意义, 且精度低于实结点。④假点: 引用别的要素类中的一般点或者实结点的点。点类型包括一般点、实结点、虚结点。图 8 中引用标志 0/1/2 分别表示未被引用的真点、假点、被引用的真点; 被引用点的 ID 包括 2 部分: 被引用点所在的要素类 ID+被引用点的 ID(图 8)。

(2)弧段: 具有拓扑连接意义, 可以构成线几何实体和区几何实体。弧段从表示形式上可以分为折线和解析弧段。图 9 中引用标志 0/1/2 分别表示未被引用的真弧段、假弧段、被引用的真弧段; 被引用弧段 ID 同样包括 2 个部分。

5 结论

在新一代 GIS 的发展过程中, 要求数据管理系统能够支持统一的海量数据的存储、查询和分析处理。包括支持 TB 级以上的空间数据存储; 有效的空

间、属性一体化管理、查询机制等(东凯和方裕, 2004)。面向实体的空间数据模型是针对海量 TB 级数据的大型 GIS 的数据模型。它以实体及其关系的描述为基本表达,具有较强的空间和语义表现力,支持基于结构和基于空间规则的空间拓扑关系,全面支持继承、组合等非空间关系,适应复杂地学应用,同时兼顾制图和 CAD 方面的需求。利用要素、几何、空间数据 3 个层次,该空间模型还解决了空间边界共享的机制。目前国内著名的 GIS 软件 MAPGIS 已率先采用面向实体的空间数据模型,并在 MAPGIS7.0 中得到了实现,这也说明了面向实体的空间数据模型的可行性。

References

- Chen, J., Zhang, S. W., 2003. The third spatial data model and its implement. *Surveying and Mapping in Shanxi*, 10(2):7-8(in Chinese with English abstract).
- Dong, K., Fang, Y., 2004. Concept of spatial database model

and its architecture research. *Geomatics World*, 2(2):15-16(in Chinese with English abstract).

- Guo, L. C., Guo, J. X., 2005. Study on GIS data model. *Geospatial Information*, 3(1):37(in Chinese with English abstract).
- Michael, Z., 1999. Modeling our world; The ESRI guide to geodatabase design. ESRI Inc., [s. n.].
- Wu, X. C., 2004. New generational MAPGIS. *Geomatics World*, 2(2):4-5(in Chinese with English abstract).

附中文参考文献

- 陈静,张树文,2003. 第三代空间数据模型及其实现. 三晋测绘, 10(2):7-8.
- 东凯,方裕,2004. 空间数据库模型概念与结构研究. 地理信息世界, 2(2):15-16.
- 郭利川,郭建星,2005. 浅谈地理信息系统中的空间数据模型. 地理空间信息, 3(1):37.
- 吴信才,2004. 新一代 MAPGIS. 地理信息世界, 2(2):4-5.