

国家科学技术学术著作出版基金资助
测绘科技专著出版基金资助

Voronoi 动态空间数据模型

陈 军 著

测 绘 出 版 社

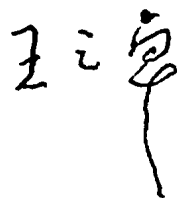
· 北 京 ·

序

该书作者陈军同志是“文革”后我招收的第二批(1979年)研究生,1981年初被国家选送到法国学习遥感技术,1982年底按时回国服务。在回国后的近20年时间里,他在遥感和地理信息系统(GIS)领域积极进取,在教学、科研和实验室建设第一线上踏实工作,锻炼、培养了很强的开拓精神,能敏锐地预见国际学术界的发展动向和提出国际学术前沿性的研究课题,在Voronoi空间数据模型、国家边界陆地边界空间数据库、城市图文办公系统、多维动态GIS等研究方向上取得了突出的成绩,形成了自己的特色。他于1987年被评为“国家级有突出贡献的中青年专家”,1993年被国务院学位办批准为博士生导师,先后入选国家教委跨世纪优秀人才培养计划、人事部“百千万工程计划”,获得了国家杰出青年科学基金,并担任了中国GIS协会会长、国际摄影测量与遥感学会(ISPRS)第二委员会主席。

本书是陈军同志根据其近年来的研究成果和国际研究动态,论述和介绍了Voronoi动态空间数据模型的基本原理、关键算法及应用模型,包括Voronoi空间数据模型的基本问题、空间目标的Voronoi图生成算法、基于Voronoi图的空间关系描述、Voronoi动态GIS的构建、基于Voronoi图的空间分析等。其内容丰富,结构严谨,文字流畅,是一本关于Voronoi动态空间数据模型的学术专著,有相当的理论深度和参考价值。其将为我国发展新一代动态GIS提供了新概念、新思路,对于发展我国新一代GIS基础软件平台和业务化GIS应用系统有重要的参考价值。

有鉴于此,值此书出版之际,特作此序,以示鼓励。



2001年9月5日于武汉大学

王之卓:中国科学院资深院士,前武汉测绘科技大学名誉校长。

前 言

空间数据模型是关于地理信息系统 (GIS) 空间数据组织和空间数据库设计的基础理论,由概念数据模型、逻辑数据模型和物理数据模型组成。其概念数据模型主要分为基于目标的矢量数据模型(object-based vector data model)和连续铺盖数据模型(tessellation-based data model)。前者是把现实世界的地理空间实体抽象地看作是点、线、面、体空间目标,显式地表达这些目标及部分空间关系(如相邻、包含、连通等);而后者是把整个空间用规则或不规则的铺盖(如方格、三角形、六角形、Voronoi 图等)覆盖,用一组铺盖单元记录或表达每一个地理实体的空间分布,并隐含地表达地理实体间的空间关系。

对于 Voronoi 图空间铺盖来说,其是将诸多地理空间实体作为生长目标,按距每一目标最近原则,将整个连续空间剖分为若干个 Voronoi 区(Voronoi polygon),每一个 Voronoi 区只包含一个生长目标。换言之,在 Voronoi 图中地理空间实体与 Voronoi 区一一对应,兼具了矢量数据模型和连续铺盖模型的基本特点,提供了诸如势力范围(influence region)、侧向邻近(lateral adjacency)、局域动态(local and dynamic)等重要特性,因而成为 GIS 空间数据建模和分析的一种重要工具。

80 年代后期,加拿大 Laval 大学 C. M. Gold 教授等率先将 Voronoi 图用于研究 GIS 动态空间数据模型。迄今为止,国际上对 Voronoi 动态空间数据模型的研究方兴未艾,主要研究热点包括基于 Voronoi 图的空间认知与空间建模,基于 Voronoi 图的空间关系描述、表达和推断,基于 Voronoi 图的空间分析, Voronoi 图的动态生成算法, Voronoi 动态 GIS 的基本结构和空间数据动态存取方法等。本书作者及其研究小组从 1992 年起开始了 Voronoi 动态空间数据模型的研究,先后获得了国家自然科学基金面上项目、跨学部重点项目和国家杰出青年科学基金的资助,取得了一定的研究成果。

为了便于 GIS 及相关领域同行了解 Voronoi 动态空间数据模型的研究进展,作者根据其研究小组近年来的研究成果和国际同行的研究动态,撰写了“Voronoi 动态空间数据模型”一书,介绍 Voronoi 动态空间数据模型的基本原理、关键算法和应用模型,包括 Voronoi 数据模型在 GIS 空间数据模型中的位置、空间目标的 Voronoi 图生成算法、基于 Voronoi 图的空间关系描述、Voronoi 动态 GIS 的构建、基于 Voronoi 图的空间分析等。其读者对象主要为从事 GIS、数据库及相关领域的学术研究、教学和软件设计的科技人员,测绘、地学、GIS、计算机信息处理等相关学科的硕士、博士研究生。由于 Voronoi 动态空间数据模型是国际 GIS 的前沿研究课题,而作者本人水平有限,加上时间仓促,书中定有错漏之处,敬请同行专家和读者批评指正。

在本书撰写的过程中,博士研究生赵仁亮、乔朝飞、曹菡、赵学胜和韩刚等协助完成了插图整理、初稿校对等工作,香港理工大学李志林教授、国家基础地理信息中心蒋捷博士等提出了宝贵的意见。对此,作者向他们表示衷心的感谢。

值此成书之际,作者要感谢导师王之卓院士、杨凯教授的培养教育;感谢徐冠华院士、陈俊勇院士、宁津生院士、李德仁院士、张祖勋教授等专家的提携和鼓励;感谢国家测绘局历任领导金祥文先生、陈邦柱先生等的一贯支持;感谢国家基础地理信息中心和原武汉测绘科技大学领导和同事们的帮助;感谢我的历届博士、硕士研究生为实践我的研究思想而做出的不懈努力;感谢 C. M. Gold 教授、杨维平博士等学术界朋友的真诚帮助;最后还要感谢我的家人多年如一日的理解和支持。

目 录

第一章 GIS 空间数据模型概述	(1)
1.1 GIS 空间数据模型的重要性	(1)
1.1.1 GIS 空间数据模型的三个层次	(1)
1.1.2 GIS 空间数据模型的重要作用	(2)
1.2 GIS 空间概念数据模型	(4)
1.2.1 平面图数据模型	(4)
1.2.2 基于连续面片的铺盖模型	(5)
1.2.3 两种基本空间概念数据模型的讨论	(7)
1.3 GIS 空间逻辑数据模型	(8)
1.3.1 结构化模型	(8)
1.3.2 关系数据模型	(8)
1.3.3 扩展的网络模型	(10)
1.4 GIS 空间物理数据模型	(11)
1.4.1 数据库物理组织的一般问题	(12)
1.4.2 空间数据库物理组织的三种基本方法	(13)
1.4.3 空间目标的分解问题	(16)
1.5 GIS 空间数据模型的若干学术前沿	(17)
1.5.1 空间关系	(17)
1.5.2 三维空间数据模型	(18)
1.5.3 时空数据模型	(19)
1.5.4 分布式空间数据模型	(20)
1.6 Voronoi 动态空间数据模型	(20)
第一章参考文献	(21)
第二章 空间目标 Voronoi 图的生成	(24)
2.1 Voronoi 图的基本概念	(24)
2.1.1 Voronoi 图的定义	(24)
2.1.2 Voronoi 图的若干重要性质	(25)
2.2 Voronoi 图的矢量生成方法	(28)
2.2.1 生长元的距离计算	(28)
2.2.2 矢量 Voronoi 图生成的基本方法	(30)
2.2.3 基于半线的 VORDLL	(32)
2.3 Voronoi 图的栅格生成方法	(38)
2.3.1 传统距离变换算法	(38)
2.3.2 动态距离变换	(41)
2.3.3 面条生长元的 Voronoi 图	(43)
2.3.4 Rvkit-栅格 Voronoi 工具	(45)
2.3.5 讨 论	(46)
2.4 顾及障碍物的 Voronoi 图生成	(47)
2.4.1 顾及障碍物的 Voronoi 图类型	(47)

2.4.2 基于 VORDLL 的可视最短路径 Voronoi 图生成	(48)
2.5 球面 Voronoi 图的生成	(53)
2.5.1 球面 Voronoi 图的定义与性质	(54)
2.5.2 O-QTM 的球面三角网划分	(54)
2.5.3 基于球面三角格网的距离变换	(57)
2.5.4 球面 Voronoi 图的生成算法与误差分析	(59)
第二章参考文献	(61)
第三章 基于 Voronoi 图的空间关系描述	(63)
3.1 基于点集拓扑的交叉模型	(63)
3.1.1 从 4 元组到 9 元组	(63)
3.1.2 9 元组模型存在的若干问题	(68)
3.2 基于 Voronoi 图的 9 元组描述框架	(70)
3.2.1 V9I 的基本框架	(70)
3.2.2 V9I 元素值计算和空间关系语义判断	(73)
3.2.3 基于 V9I 的空间关系操作	(75)
3.3 基于 V9I 的空间关系描述	(76)
3.3.1 面\面关系	(76)
3.3.2 面\线关系	(78)
3.3.3 线\线关系	(80)
3.3.4 点\面、点\线、点\点	(81)
3.4 基于 Voronoi 距离的 k 阶邻近	(82)
3.4.1 空间邻近关系与 Voronoi 距离	(82)
3.4.2 基于 Voronoi 距离的 k 阶邻近	(83)
3.4.3 基于 k 阶邻近的邻域	(84)
第三章参考文献	(85)
第四章 基于 Voronoi 图的动态 GIS	(87)
4.1 Voronoi 动态 GIS 的基本思想	(87)
4.1.1 拓扑关系的显式表示	(87)
4.1.2 拓扑数据组织与维护的问题	(89)
4.1.3 Voronoi 动态 GIS 研究方向	(90)
4.2 蕴含邻近关系的 Voronoi 空间数据模型	(91)
4.2.1 顾及侧向邻近关系的 VECR 模型	(91)
4.2.2 用 Voronoi 图扩展 MapInfo 拓扑功能的初步研究	(97)
4.3 基于 Voronoi 图的空间关系推断	(99)
4.3.1 根据 Voronoi 图直接判断内点和岛屿	(99)
4.3.2 基于内外边界的 4 邻近推断模型	(100)
4.3.3 用邻近关系判断拓扑关系	(101)
4.3.4 基于 Voronoi 图的 MapInfo 拓扑分析功能的实验	(103)
4.4 基于球面 Voronoi 图的层次数据模型	(106)
4.4.1 球面数据模型的层次结构问题	(106)
4.4.2 基于 Voronoi 图的球面层次数据模型	(107)
4.5 Voronoi 数据的树型层次管理	(111)
4.5.1 全局 Voronoi 图的平面层次分割	(111)
4.5.2 VMO-树:树型层次 Voornoi 数据结构	(112)

第四章参考文献.....	(113)
第五章 基于 Voronoi 图的空间分析	(115)
5.1 空间内插	(115)
5.1.1 基于 Voronoi 图的自然邻近内插	(116)
5.1.2 基于 Voronoi 图的最近邻近内插	(117)
5.1.3 基于 Voronoi 图的 DEM 内插	(119)
5.2 基于 Voronoi 图的设施选址分析	(123)
5.2.1 设施选址的优化问题	(123)
5.2.2 基于 Voronoi 分区的 minisum 选址	(124)
5.2.3 基于最小覆盖圆的 minimax 选址	(125)
5.2.4 基于最大空心圆的 maxmin 选址	(126)
5.3 基于 Voronoi 图的空间认知	(127)
5.3.1 Voronoi 图中的空间定性概念	(127)
5.3.2 Voronoi 图在地图数据综合中的应用	(128)
第五章参考文献.....	(130)

第一章 GIS 空间数据模型概述

地理信息系统(Geographic Information System, 简称 GIS)是一种专门用于采集、存储、管理、分析和表达空间数据的信息系统。其既是表示、模拟现实空间世界和进行空间数据处理分析的工具,也可看作是人们用于解决空间问题的“资源”,同时还是一门关于空间信息处理分析的“科学技术”[陈述彭,1991;何建邦,钟耳顺,1993;陈军,1995]。

就其“工具”特性来说,GIS 为人们采用数字形式表示和分析现实空间世界提供了一系列空间操作和分析方法,包括综合地存储管理人们研究和解决空间问题所需的各种空间数据;根据用户的要求,查询检索有关的空间分布信息,进行各种统计量算、列表制图;根据规划、管理、生产的需要,进行多因素的综合研究、决策方案的模拟优化等。在 GIS 提供的空间数据输入、数据库管理、查询检索、分析、显示输出、联网分发等诸多功能中,空间数据库管理系统是按一定方式组织和存储管理空间数据,具有较高的程序和数据独立性,能以较少的重复为多个用户或应用程序服务,因而是整个 GIS 的核心[Laurini and Thompson, 1992]。

“资源”特性是指一个具体应用部门设计和建立的 GIS 将单一分散的数据资料积累集成起来,成为研究和解决本部门(或本单位)的空间问题所必须的综合信息资源,亦可用于为社会服务。例如,国家测绘局建立的 1:100 万、1:25 万基础地理信息系统,构成了国家基础地理信息资源[杨凯,1997]。

从“科学技术”属性的角度看,GIS 主要是研究现实世界空间实体及其相互间关系的描述和表达,以及在计算机环境下的空间数据组织、存取、分析、可视化,应用系统的设计、数据集成和业务化运作等[Worboy, 1995]。在 1970-1990 的 20 年里,国内外 GIS 的发展都主要是靠“应用驱动”和“技术导引”[徐冠华, 1998]。为了给 GIS 应用与产业化发展提供更多的理论支持,1990 年以来国际学术界加强了对 GIS 基础理论问题的研究,研究重点包括空间关系、空间数据模型、空间认知、空间推理、地理信息机理(产生、施效和人机作用等)、地理信息不确定性等[NCGIA, 1989; 李德仁,1996]。其中空间数据模型是关于 GIS 中空间数据组织的概念和方法,反映现实世界中空间实体(spatial entity)及其相互之间的联系,是描述 GIS 空间数据组织和进行空间数据库设计的理论基础[陈军,1995]。

本章将首先简要介绍 GIS 空间数据模型的重要性,然后分别论述 GIS 的空间概念数据模型、逻辑数据模型和物理数据模型,继而介绍 GIS 空间数据模型的国际学术前沿和本书的主要安排。

1.1 GIS 空间数据模型的重要性

1.1.1 GIS 空间数据模型的三个层次

传统的纸质模拟地图是根据地图模型(map model),按一定的数学法则、符号、制图综合原则和比例,将地球空间实体和现象的形状、大小、相互位置、基本属性等表示在二维平面上。在计算机化的 GIS 中表达和组织管理地球空间实体和现象的空间数据,比起传统的纸质模拟地图复杂得多。其是根据空间数据模型(spatial data model),要表达和处理空间实体及其相互间关系,还要考虑计算机存储和处理的性能,因而与地图模型有着本质的区别。空间数据模型类似于一种语言,是说明和描述空间数据、数据间关系、数据语义、数据一致性、数据操纵等的一种方法。

GIS 空间数据模型由概念数据模型、逻辑数据模型和物理数据模型三个有机联系的层次所组成(如图 1-1 所示),其中概念数据模型是关于实体及实体间联系的抽象概念集,逻辑数据模型是表达概念数据模型中数据实体(或记录)及其间关系,而物理数据模型则是描述数据在计算机中的物理组织、存取路径和数据库结构[Lee and Isdale, 1991; 陈军,1993]。

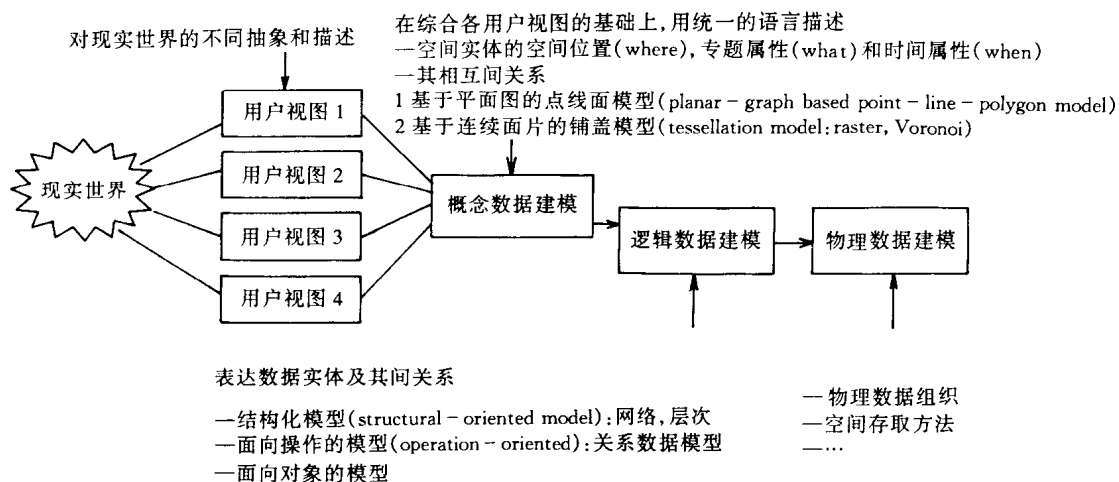


图 1-1 GIS 空间数据模型的三个层次

空间数据模型是空间数据库模式(data schema)设计的基础,空间数据库模式是关于空间数据库数据组织的描述。如果说空间数据模型相当于 C 语言(看成是抽象的程序设计语言),空间数据库模式则相当于 C 语言程序中的变量描述(程序设计语言中的数据说明语句)[Lee and Isdale,1991]。空间数据模型与空间数据库模式的层次对应关系如图 1-2 所示。其中概念数据模型对应于概念数据模式(扩展语意结构),常用的表示方法为实体—联系方法(Entity-Relationship Approach),即用 E-R 图来描述实体及实体间的联系。逻辑数据模型对应于逻辑数据模式,是对整个数据库中的全部数据进行定义和操纵的数据模型,包括网状模型、层次模型、关系模型及面向对象的模型等。物理数据模型对应于内模式(即内部语意结构)[孙道明等,1998]。

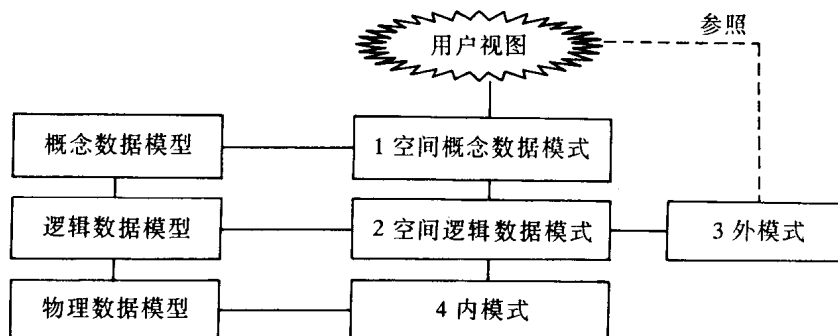


图 1-2 三层数据模型与四级空间数据模式对应关系

空间数据模型为空间数据库模式设计提供了目标类型、数据操作算子和完整性规则等语法规则。其中目标类型是指空间数据模型能够表达的实体类型及实体间联系,包括空间实体本身的几何和非几何特征及两者之间的直接或间接关系。在设计一个空间数据库时,其所表达的空间目标是由所选定的空间数据模型决定的。算子是指可用来对数据库目标对象进行的检索、更新等各种操作,包括其定义及操作符号、操作规则及语言定义。完整性规则给出了数据模型中数据及其联系所具有的制约和依存规则,用于说明空间实体的几何和非几何特性之间的相互制约机制及限定时间序列下的动态变化,以保证数据的正确、有效及相容性。

1.1.2 GIS 空间数据模型的重要作用

综观国内外 GIS 研究和应用的发展状况,无论是针对具体应用目标进行 GIS 应用系统设计,还是研制 GIS 基础软件平台,均是以空间数据模型理论为基础的。

就 GIS 基础软件平台而言,ARC/INFO 等是在地理-关系模型(geo-relational model)基础上发展起

来的;SYSTEM 9 是在面向对象空间数据模型基础上发展起来的。总的说来,可将 GIS 空间数据管理基础软件平台的设计思想分为以下三种情况[Vijlbrief and Oosterom,1992];

1) 混合结构体系(Hybrid design):早期设计研制的 GIS 空间数据管理系统采用的是混合结构体系,即分别采用常规的数据库管理系统管理属性数据和用专门的空间数据管理系统管理定位数据。若用关系型数据库管理系统管理属性数据的话,则成为地理-关系(Geo-relational)结构,如图 1-3(a)所示。当前国际上流行的 ARC/INFO、MGE、GENEMAP 和 SICARD 等就是采用这种设计思想。例如,美国 ESRI 公司的 ARC/INFO 用 INFO(或 ORACLE、INGRES)管理属性数据,用 ARC 管理和处理空间数据。值得说明的是,人们以往主要是将通用的关系型数据库管理系统(General purpose relational DBMS)用于处理表格型数据,是因为其在处理空间数据方面有一些局限性,包括:(1) 由于不同空间目标的坐标串长度往往不统一,空间数据的记录是变长的,而通用的关系型数据库管理系统适于处理定长数据记录。此外,需要扩充通用关系型数据库管理系统的功能,方可处理与空间坐标数据交织在一起较为复杂的拓扑关系和属性数据;(2) 空间数据操作涉及到诸如相邻、连通、包含、叠加等空间概念,通用关系型数据库管理系统的查询语言不具备这方面的功能;(3) 通用关系型数据库管理系统一般不具备空间数据处理要求的高级图形功能;(4) GIS 数据之间的高度相关性要求一个高性能的安全系统(security system),以保证空间数据多文件的一致性。

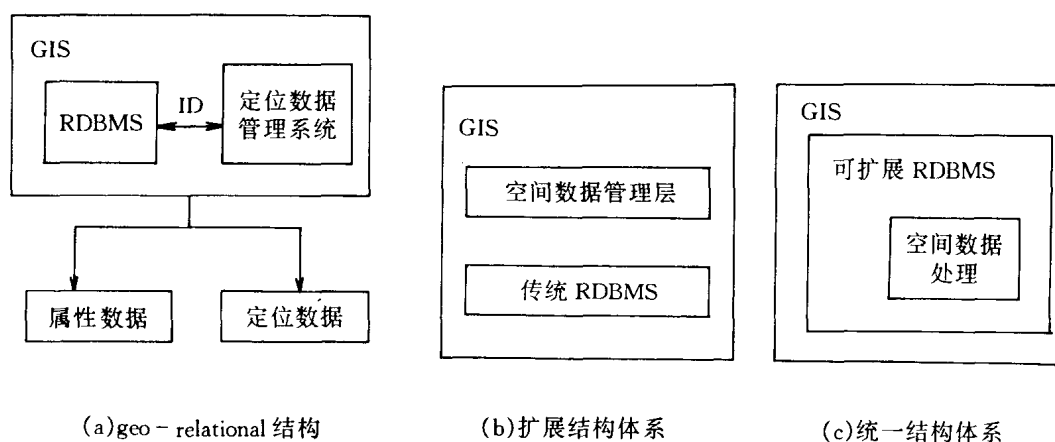


图 1-3 空间数据库的结构体系

2) 扩展结构体系(extended design):即采用同一 DBMS 存储空间数据和属性数据。System 9 和 Small World 都是采用这种方法。其做法是将空间实体划分成若干部分,用独立的关系表格存储,检索需要进行关系的“并”运算,为此还要在标准的关系型 DBMS(如 Oracle, InGress 等)上增加空间数据管理层,如图 1-3(b)所示,以将空间查询转换成标准的 SQL 查询。其优点是省去了空间数据库和属性数据库之间的繁琐联结,提高了系统的效率。

3) 统一结构体系(integrated design): GEO⁺⁺、TIGRIS 等采用的是这种设计思想,该结构不再是基于传统的 DBMS,而是在开放型 DBMS 基础上扩充空间数据表达的功能,如图 1-3(c)所示。近年来,Oracle 等商业化 RDBMS 开始支持空间数据的定义和空间查询,但有待于进一步研究和发展[Kim 等, 1993]。

对于 GIS 应用系统设计来说,空间数据库管理系统是其最基础的部分,直接关系到整个信息系统的运行质量和成败。这是因为它不仅影响着分析或显示模块对系统内数据的有效利用和存取效率,而且也影响着用户对数据库的概念。为了有效地进行空间数据库设计,需要针对具体的应用问题,选用合适的空间数据模型,分别进行概念、逻辑和物理设计。在概念模型层上,用户根据其问题的性质,将把现实世界中的客观对象抽象为一种信息结构,在此基础上对现实世界进行数字化表达、处理和管理。在逻辑模型层,系统设计员通过对各种用户观点的综合分析,指明数据实体及其间的关系、记录及其间的关系,但不涉及过于详细的实现技术;在最低一层(物理模型)中,程序设计者根据系统设计,综合地运用各种计算机技术,设计

和实现数据的实际存储结构以及在存储介质上进行检索存储操作的各种算法。

1.2 GIS 空间概念数据模型

由于职业、专业、志向、社会地位等的不同,人们往往在所关心的问题、研究的对象、期望的结果等方面存在着差异,因而对于现实世界有着不同的看法、观念和兴趣,对现实世界的描述和抽象亦是不同的,形成了不同的用户视图。换句话说,对于不同的应用目标,就有其特定的用户视图。GIS 空间数据模型的概念模型是考虑用户需求的共性,用统一的语言描述和综合、集成各用户视图。其基本任务是,确定所感兴趣的现像和基本特性,描述实体间的相互联系,从而确定空间数据库的信息内容。目前广为采用的是基于平面图的点、线、面数据模型(planar-graph based data model)和基于连续铺盖(tessellation)的栅格数据模型(raster data model)。

1.2.1 平面图数据模型

平面图数据模型是把现实世界的空间实体抽象地看作是由平面上的点、线、面空间目标(spatial objects)组成的。点状目标(point object)的几何特征主要由空间位置表征,线状目标(line object)的几何特征由空间位置、长度、弯曲度、方位等表征,而面状目标(area object)的几何特征还包括范围、周长、重叠等。这些点、线、面空间目标之间存在着一些空间关系,如点、线、面目标与其组成弧段、结点、坐标之间的相交(junction)、连接(connections)、连通(connectivity)和包容(containment)等拓扑空间关系[Laurini and Thompson, 1992]。一般说来,人眼易于看出地图上的点、线、面之间的空间拓扑关系,但计算机难以直接识别。因此,基于平面图的点、线、面数据模型的一个核心问题是:描述和表达点、线、面空间目标及其相互间的拓扑空间关系。

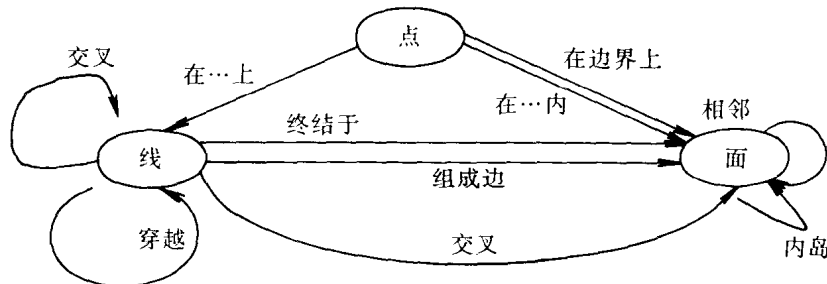


图 1-4 点、线、面目标间的部分空间关系

二维空间目标及目标间的空间关系可以用二维欧几里德空间上的拓扑理论和图论来解释与描述[Corbett, 1979; White, 1984]。设有一个二维平面图(2D planar graph) G ,

$$G = (V, R)$$

其中, $V = \{x | x \in \text{空间目标}\}$ 表示图中所有结点的有穷非空集合;

$R = \{VR\}, VR = \{\langle x, y \rangle | P(x, y) \wedge (x, y \in V)\}$, 表示图中两个结点之间有向边的集合。

二维欧氏空间中任意一个空间实体均可看作图 G 的一个子图 G' , 可表示为:

$$G' = (V', R'), \text{ 其中 } V' \subseteq V, R' \subseteq R.$$

子图 G' 对应于二维拓扑空间上的一个 n -单纯复形(n -complex, $0 \leq n \leq 2$)。点状、线状和面状空间目标分别对应于 0, 1, 2 单纯复形。任意一个 n -单纯复形($0 \leq n \leq 2$)由一系列 k -单纯形组成($k \leq n$)。其中 0-单纯形和 1-单纯形对应于图 G' 中的结点与边, 2-单纯形对应于图 G 的一个空间剖分, 若干个相邻接的 2-单纯形组成与平面图 G' 相关的面状目标。

用二维单纯复形描述二维平面图的一个主要目的是能够通过单纯形间的拓扑性质有效地表述空间目标间的拓扑关系。例如, 每个 1-单纯形为两个 0-单纯形所边界; 每个 1-单纯形为两个 2-单纯形所共享; 任

意两个 1-单纯形不能相交,即不存在一个交点,它不是一个 0-单纯形等;每个 2-单纯形为三个 1-单纯形所边界;2-单纯形由一系列 0-单纯形及 1-单纯形构成的环所包围且满足欧拉公式。这些为构造二维拓扑空间数据模型提供了理论基础。值得说明的是,空间目标间的拓扑信息有着重要的用途,包括反映空间目标之间的关系、区分和存储公共边、判断相邻性和流通性等,这是 GIS 区别于 CAD 等的一个根本之处[Laurini and Thompson, 1992]。

除了空间目标的位置、形状、大小等几何图形信息及相互间关系外,在空间数据模型中还应表达每个空间目标的属性信息,如建筑物的高度、建筑面积、使用面积等。对同一个空间目标而言,其几何图形信息和属性信息之间是通过目标标识号(object ID)相联接的,如图 1-5 所示。此外,每个空间目标还有着时间特征。

在基于平面图的空间数据模型中,点、线、面目标的空间位置是用采样点的空间坐标 x, y, z 来表达空间目标边界的。这种矢量数据结构的突出优点是容易定义目标和便于操作,能方便地表达地物之间的拓扑空间关系,图形精度高,数据存贮量小。其缺点是数据结构相对来说比较复杂,且难于处理叠置(overlap)操作,并缺乏与遥感及数字地面模型直接结合的能力。

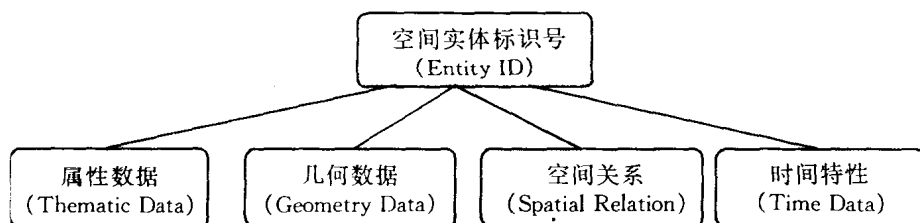


图 1-5 空间实体的数据组成

根据对坐标数据的组织与存贮方式的不同,矢量数据结构可分成面条结构(spaghetti data structure)和拓扑结构(topological data structure)两种形式[Laurini and Thompson, 1992]。在面条结构中,地图用一系列的坐标串来表示,点对应于一个 x, y 坐标,线对应于一个 x, y 坐标序列,面对应于一个起点与终点坐标相同的多边形。虽然这种结构记录了空间实体的形状信息,但空间实体间拓扑关系信息却必须通过在数据文件中搜索所有实体的信息并经过大量计算才能得出,因此,该方法难以有效地进行空间分析,但它比较适于诸如地图制图等不考虑空间关系的系统。

拓扑结构是将实体间的某些拓扑空间关系直接进行存贮,以提高空间分析的效率。该方法的优点是不通过坐标数据即可完成某些空间分析功能,从而可以大大提高系统的时间效率。但是,该方法在对一个新的图形构筑拓扑空间关系表或进行系统更新时,所需的时间代价较大。一个具有代表性的拓扑数据结构由结点与弧段、弧段与面之间的相互关系,构造结点和弧段的拓扑关系表。通过这些关系表,可以很方便地判断线状实体的连通关系(connectivity)和面状实体的邻接关系(contiguity)。该结构在空间数据的组织、拓扑空间关系的表达、数据模型的拓扑一致性检验及图形恢复等方面均具有较强的能力,因而成为一些基于拓扑结构的 GIS 软件(如 Arc/Info, Genamap, TIGER 等)在构造数据模型时所广泛采用的数据结构。

值得指出的是,同一类型的若干空间目标组成复合空间目标,例如由诸多河流组成的水系。而不同类型的若干空间目标组成复杂空间目标,如由建筑物、院墙和附属物组成的宅院等。

1.2.2 基于连续面片的铺盖模型

基于连续铺盖的栅格数据模型是将连续空间离散化,即用二维铺盖或面片(tessellation)覆盖整个连续空间;铺盖可分为规则的和不规则的,后者可当作拓扑多边形处理,如社会经济分区、城市街区;铺盖的特征参数有尺寸、形状、方位和间距[Laurini and Thompson, 1992]。对同一现象,也可能有若干不同尺度、不同聚分性(aggregation or subdivisions)的铺盖。

1. 规则型铺盖

在边数从 3 到 N 的铺盖中,方格(栅格)、三角形和六角形可完整地铺满一个平面,是空间数据处理中

常用的建模方法。其中栅格是用规则栅格覆盖连续空间,栅格格网的空间位置由行、列数确定,相邻单元彼此邻接、连通和不交叉,其间的空间关系是隐含表达的。这种栅格数据模型的突出优点是它可以直接利用遥感、数字摄影测量、扫描等方式获取的栅格形式的数据,且结构简单,比较容易实现地图或图像的叠加组合及进行各种空间分析。其缺点是不能精确地记录点和线状目标,难以建立地物间的拓扑空间关系。

随着格网尺寸(精度)的增加,逐个格网记录导致数据量不断增大,从而对设备的存贮量要求过高及查询速度较慢。为此,人们研究发展了行程编码(Run-Length Encoding,简称 RLC)、目标行程编码(Object RLC,简称 ORLC)等压缩编码方法。RLC 是考虑到栅格数据中相邻网元往往是相同的,逐行将相邻等值网元合并,记录其值和行程的起点与终点,或记录其值和行程的长度。ORLC 是对一连通的面状目标,将其边界网元先按行顺序(y),再按列坐标(x)排列为

$$\begin{aligned} & y_1, (x_{11}, l_{11}, x_{12}, l_{12}, \dots) \\ & \dots \\ & y_i, (x_{i1}, l_{i1}, x_{i2}, l_{i2}, \dots) \\ & \dots \\ & y_n, (x_{n1}, l_{n1}, x_{n2}, l_{n2}, \dots) \end{aligned}$$

其中 y_i 是第 i 行的行坐标, x_{i1} 是该行第 1 行程段的起始列坐标, l_{i1} 是该行程段的长度(网元数)。

2. 四叉树(quadtrees)

四叉树是一种具有可变空间分辨率的空间数据模型,适于多层次空间表达。其是将空间按四个象限进行递归分割,直到分割为均质块为止。换言之,四叉树是将 $2n \times 2n$ 像元阵列的区域,逐步分解为包含单一类型的区域,最小的方型区域为一个像元[Samet, 1984]。

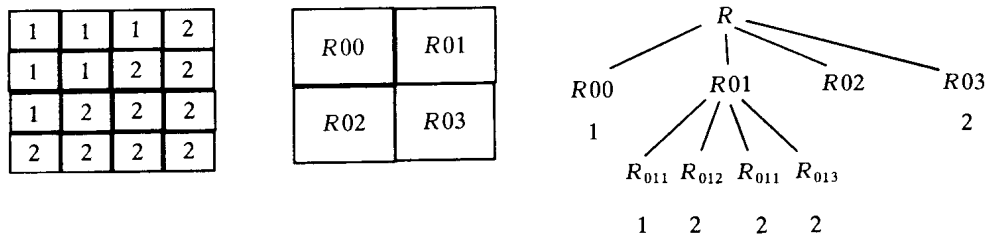


图 1-6 四叉树的递归分割

3. 不规则三角形

不规则三角形是用一些不相交的三角形面来表示空间目标(如地形表面)。每个三角形面有三个顶点,分别对应于 x , y , z 坐标。通过顶点、三角形边与面之间的相互关系,即可反映地形表面空间目标之间的拓扑空间关系[Saalfeld, 1987; Smaet, 1990]。与基于栅格结构的数字高程模型(DEM)相比,不规则三角形是用角点表示重要转折点,用边表示线状要素,因而更容易表达一些复杂的地形情况,且所需的数据量相对来说要小得多。

4. Voronoi 图(Voronoi diagram)

Voronoi 图是 Delauny 三角网的对偶,是计算几何中的一种基本几何数据结构[Aurenhammer, 1991]。若将 N 个空间点做为 N 个生长点,按距每一点最近原则将空间剖分为若干个面块,使每一个面块只包含一个生长点,由此将整个连续空间划分为 Voronoi 铺盖(Voronoi tessellation) (a fundamental geometric data structure)[Okabe et al., 1992]。由于在 Voronoi 图中每一空间目标唯一地被一个 Voronoi 区包含,其兼具了矢量数据模型下图形与空间实体一一对应和栅格数据模型对空间连续铺盖的双重特性,因而在气象、地学、测绘、计算机科学、考古、化学、生态学等领域得到了较为广泛的应用[Gold, 1992]。其在地学领域的应用包括空间内插、空间关系表达和空间分析(如聚类、骨架化、路径)等,在计算机科学方面的应用包括文件搜索(associative file searching)、集群分析(cluster analysis)、存取访问(scheduling record accesses)、碰撞检测(collision detection)等[Gold and Edwards, 1992]。

1.2.3 两种基本空间概念数据模型的讨论

实体—关系法(entity-relationship approach)是建立空间概念数据模型时采用的一种主要方法[Caron and Bedard, 1993; Webster, 1994]。其将现实世界看作是由具有相似性质的实体或实体类(entity type)组成的,单个实体或实体类之间存在着多种关系,这种关系可以是一对一、一对多的,也可能是多对多的。如道路这一实体穿过城市,对同一城市而言一般地有 N 条道路穿过,表现为 ER 模型中的 $1:N$ 关系。此外,实体和关系都有属性。对于一个街区来说,其中一街道有若干路段,一个路段有若干组边组成,而每一地块的边界可能有 M 个边组成。空间实体的点、线、面之间也存在着一些关系,可以用 ER 方法表示。如果将某些实体进一步分类的话,则得到 ECR 模型(entity-category-relationship model)[Armstrong and Densham, 1990; 唐治锋,陈军,1995]。

值得指出的是,前述的两种空间概念数据模型在描述、表达现实世界空间实体及其相互关系和进行动态模拟方面,均表现出一定的局限性,存在着一些有待解决的基本问题[Lee,1990; Dutton,1991; 陈军,1993; 陈军,蒋捷,2000],其中包括:

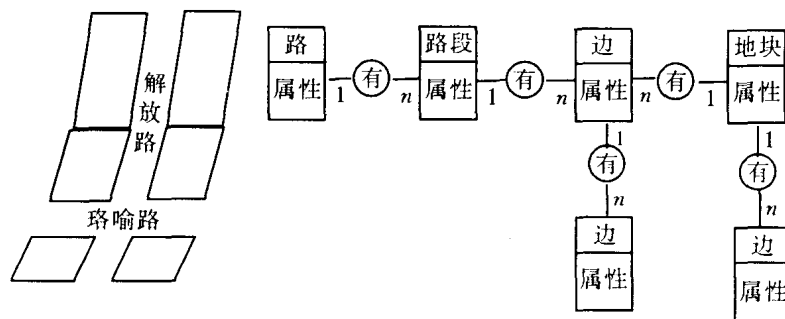


图 1-7 空间实体间关系的 ER 表达

1) 面向实体法和铺盖法是表达连续空间的两种基本方法,各有其特点。如规则铺盖用于表征连续空间,而不是空间实体的特征或位置,且不直接考虑拓扑空间关系。出于精度的考虑,一般不用规则铺盖表达曲状要素和不规则多边形。而面向空间实体的点、线、面模型适于表达曲状要素和不规则多边形。在实际应用中,往往需要将两者结合起来。例如,在城乡规划管理的过程中,由于规划管理的对象、内容、阶段的不同,往往涉及到多种不同形式的数据资料。从数据处理和管理的角度来看,其不仅需要统一地存贮管理多种数据结构的空間数据,而且要求综合地运用矢量、栅格等多种空间数据处理办法,因此单一数据结构的 GIS 工具往往难以满足空间数据存贮管理、查询检索、分析评价、输出表达等方面的要求。为了解决这一问题,人们试图设计出能有效表达矢量和栅格两种数据的混合数据结构(hybrid data structure),曾提出了 Chain Coding, Vaster, pseudo-vector 等模型,它们在表示不同类型的空间数据(线、面和影像)时各有所长,但难以顾及各方面的要求。因此,人们一方面继续探索新的混合数据结构[Edwards,1991],另一方面采用集成式方法(integrated approach),在同一空间数据库管理系统中支持或管理矢量和栅格两种数据结构,允许数据库的每一目标采用适宜的矢量或栅格数据结构,通过“目标”实现统一的数据管理[Derenyi, 1991; 唐治锋,陈军,1993]。

2) 主要适于描述和表达二维空间实体,难以有效地表达现实世界中的三维空间实体及其相互关系。一般说来,利用这两种概念数据模型描述和表达三维空间实体(如城市建筑物)时,是将其投影到二维平面上去,将其第三维的信息作为属性处理,这在许多场合下导致空间目标划分与表达的困难[Sarkozy, Ferenc, 1994; 陈军,1995; 孙敏,陈军,张学庄,2000]。

3) 这两种概念数据模型侧重于空间点、线、面目标自身的描述,没有顾及到空间点、线、面目标之间的一些重要空间关系(如相交 intersect、穿越 cross 等)和空间实体之间从属、组合关系的描述与表达,因而在一定程度上不能有效地表达现实世界中的空间现象和支持空间分析[Kainz, Egenhofer and Greasley, 1993; Egenhofer and Mark, 1995; 陈军,赵仁亮,1999]。

4) 点、线、面数据模型和栅格数据模型最早是针对宏观(或小比例尺)应用发展起来的,因而在将其用

于表达城市空间实体时,出现了一些新的问题[Lee,1990]。例如,城市主要道路要用双线表示,因此原有线状目标的定义必须予以修正或扩充;再如,城市主要道路在小比例尺数据库中为单线目标,在大比例尺数据库中为双线目标,如何实现两者之间的有机联系,对于进行宏观与微观相结合的城市空间分析十分重要。

5) 从本质上说,前述的空间数据模型主要是表达某一时刻(snapshot)的空间实体及其相互之间的联系,没有考虑表示空间实体的时空变化,本质上是非时态的。这种非时态 GIS 没有记录其历史状态和变化过程,因而难以回答诸如何时何地发生了什么变化、变化的类型和速率、变化的周期等问题。实际上,人们要求 GIS 能够方便地提供空间实体分布和状态的最新信息,能够重构空间变化(或演进)的过程,预测和模拟未来的空间分布和状态[Gail Langran, 1992; Armenakis, 1992; 蒋捷, 陈军, 2000]。

6) 在点、线、面数据模型中,点、线、面空间目标之间的邻接(adjacency)是隐含表达的,在寻找某一空间目标周围的其它一些空间目标时往往要花费较长的检索与处理时间,这在一定程度上影响了其在空间现象动态分析和空间过程动态模拟中的应用[Gold, C. M., 1989]。

7) 前面关于空间数据模型的讨论仅限于对集中式空间数据组织与管理的考虑。如果一些相关联的空间数据库分布在不同地域上,或用户分散在一定的地域范围内,往往需要给用户一个总的、聚合的、唯一的空间数据集合及其统一管理方法,实现分布式空间数据管理。在一些情况下,这些相关联的空间数据库可能是非均质的,如是基于不同的数据模型或设计方法,这就要求将诸多数据库模式集成起来(schema integration)。

1.3 GIS 空间逻辑数据模型

逻辑数据模型是根据前述概念数据模型确定的空间数据库信息内容(空间实体及相互关系),具体地表达数据项、记录等之间的关系,因而可以有若干不同的实现方法。一般说来,可将空间逻辑数据模型分为结构化模型(structural model)和面向操作的模型(operation-oriented model)两大类[Armstrong and Densham, 1990]。

1.3.1 结构化模型

结构化模型是用显式表达数据实体之间关系的树形结构。其中的层次数据模型是按树形结构组织数据记录,以反映数据之间的隶属或层次关系。网络数据模型是层次数据模型的一种广义形式,是若干层次结构的并。

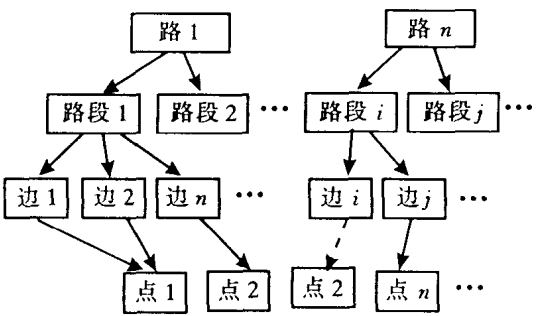


图 1-8 按树形结构组织数据记录

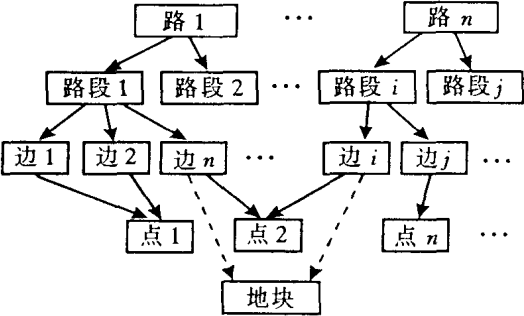


图 1-9 按网状组织数据记录

一般说来,结构化模型的优点是能反映现实生活中极为常见的多对多的联系,直接地反映现实世界中空间实体之间的联系。其缺点是复杂,若要检索信息,往往要回溯整个结构。

1.3.2 关系数据模型

关系数据模型是面向操作的逻辑数据模型,它是用二维表格表达数据实体之间关系,用关系操作提取

或查询数据实体之间的关系,因此称之为面向操作的逻辑数据模型。早在 1970 年美国统计局就设计和采用了能显式地表达拓扑关系的线段结构(line segment structure);HARVARD 大学研制的 ODYSSEY 系统采用了一种链模型(chain model),用线段表、结点表、多边形表和点表显示地表达线的中间点和端点。目前 Arc/Info 等商业化 GIS 是采用这种方法,用关系表显式表达和记录节点、弧段与面块之间的邻接关系、连通关系、组成关系等拓扑关系。

在图 1-10 中,1,2,3,4,5 为结点;*a,b,c,d* 为线段(弧段);*A,B,C,D* 为面(多边形)。

其相邻性是由边—面表或连通矩阵、面表加以表示的,如图 1-11 所示。其中对一个面而言,在关系表中列出其组成边(左、右边),对一条边则在关系表中列出其左、右面和起止结点及中间点。

连通性是由结点—链或连通矩阵、结点表表示,结点—链表包含每一条链的起始结点,连通矩阵和结点表是直接或间接地给出每一结点的连通结点,如图 1-12 所示。

对于面元来说,其包含关系的表达可用表格表示组成边及其坐标,用特殊码表示岛的存在,或表示嵌套(包含)关系的存在(在下表中用-号表示),如图 1-13 所示。

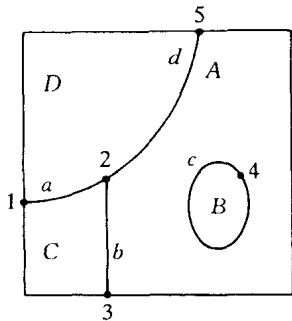


图 1-10 空间数据的拓扑关系

面域	A	B	C
A	—	1	1
B	1	—	0
C	1	0	—

图 1-11 部分面域邻接关系的表达

结点	1	2	3
1	—	1	0
2	1	—	1
3	0	1	—

图 1-12 结点间连通关系的表达

面域	弧段
A	<i>b d - c</i>
B	<i>c</i>
C	<i>a b</i>
D	<i>a d</i>

图 1-13 包含关系的表达,其中-表示岛的存在

为了实现概念数据模型向逻辑数据模型的转换,可将 ER 模型中的实体(如图 1-7 中的路、路段、路段组成边、地块、单位等)和它们之间的关系(如路-路段、路段-边、地块-边、边-点、单位-地块等)用关系框架表示,如图 1-14 所示。在此过程中,视需要可进行关系规范化,为进一步的物理实现提供逻辑框架。

就总体而言,面向操作的逻辑数据模型具有灵活简单的特点,但表示复杂关系时比其它数据模型困难;当数据构成多层联系时,存储空间利用效率较低。

路	路段 1	...	路段 n	属性
---	------	-----	--------	----

路段	边 1	...	边 n	属性
----	-----	-----	-------	----

边	起点	终点	中间点	左地块	右地块
---	----	----	-----	-----	-----

地块	左边 1	...	右边 1	...	属性
----	------	-----	------	-----	----

单位	地块 1	地块 2	...	...	属性
----	------	------	-----	-----	----

图 1-14 用二维表格表达数据实体及其间关系

1.3.3 扩展的网络模型

当前的一种发展趋势是将前两者的优点集中起来,形成新的或改进的逻辑数据模型,如扩展的网络模型(extended network model)[Armstrong and Densham, 1990]。例如,图 1-15 是用 ER 方法表示的某一建设项目管理信息系统 KCPIS 的概念数据模型,其将空间实体抽象为由平面上的点、线、面目标。其中结点(node)由坐标对(X,Y)构成,标识点串的起或止;弧段(chain)由简单点串构成,且保持等质(同质)性,或由两个结点和一系列坐标组成,可以从属于两个以上的空间目标;点状目标是结点的退化;线状目标由 $N \geq 1$ 条链构成;面状目标由一个点状目标(内点)和 $N \geq 1$ 条链构成(边界)。在上述 E-C-R 模型中,我们称弧段和结点为基本的数据库目标,而点、线、面为基本的空间目标。空间目标与其属性的组合反映了现实的空间实体。另外,空间实体类是根据主题属性对空间实体进行的分类。实体类定义了某类空间实体的属性结构,如果不同实体类之间存在部分相同的属性结构,我们可以在它们之上再构造一个超类,这个超类也具有一个超类属性表,并可被类所继承。这样就形成了一个类层次结构[唐治锋,陈军,杜道生,1995]。

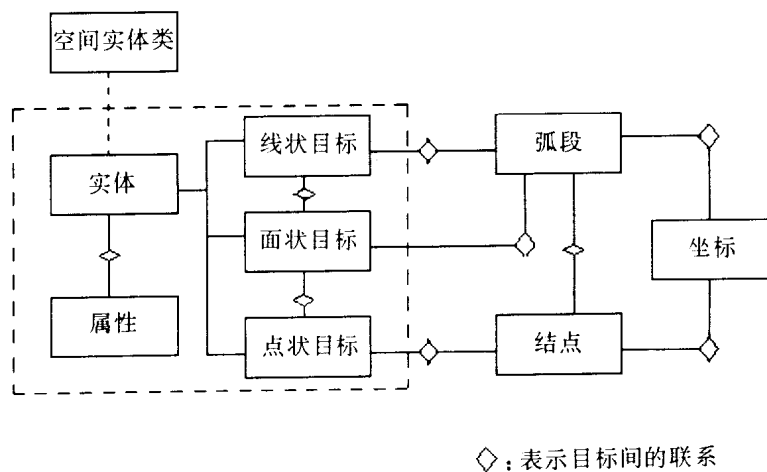


图 1-15 KCPIS 的 E-C-R 模型

KCPIS 系统涉及到两种类型的空间数据,即数字地图和具有地理位置的项目信息。其中数字地图中包括了如行政区划、水系流域、交通干线、地形地貌等各种地理层。同时系统还管理着大量的社会经济数据,这类数据往往附属于行政区划单元。因此对图 1-15 所示的数据模型进行扩充,即得到如图 1-16 所示的概念数据模型。其中面状目标被国家、省、市和县所代替,而建设项目数据由于具有地理位置被当作点状目标存于系统中

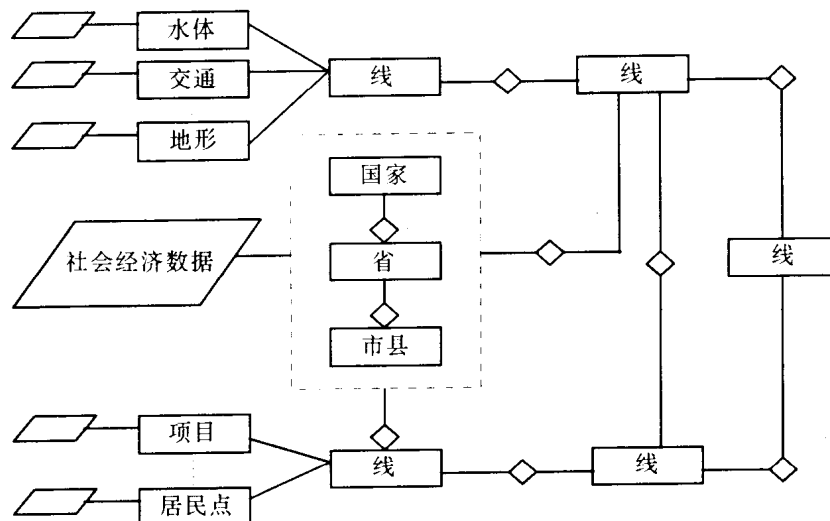


图 1-16 KCPIS 的概念数据模型

根据图 1-16 所示的概念模型,同时兼顾到空间目标及空间关系在数据库中的表达,我们进一步设计出反映数据记录以及记录之间关系的逻辑数据模型(如图 1-17)。这个扩展的网络数据模型的特点是将关系数据模型的灵活性和结构化网络模型有机地结合起来,并在原网络数据模型的基础上增加了系统集(system set),坐标、结点、链、点、线、面均被定义为系统集元素,这样在设计系统运行操作过程时,可直接对各种类型的目标进行查询检索而无需逐层搜索。在此模型中,所有记录都具有 ID,这是目标标识码,用于唯一地标识确定一个系统目标。每个空间目标记录中的 FC 域叫特征码(feature code),用于表达空间实体的分类信息(实体类)。

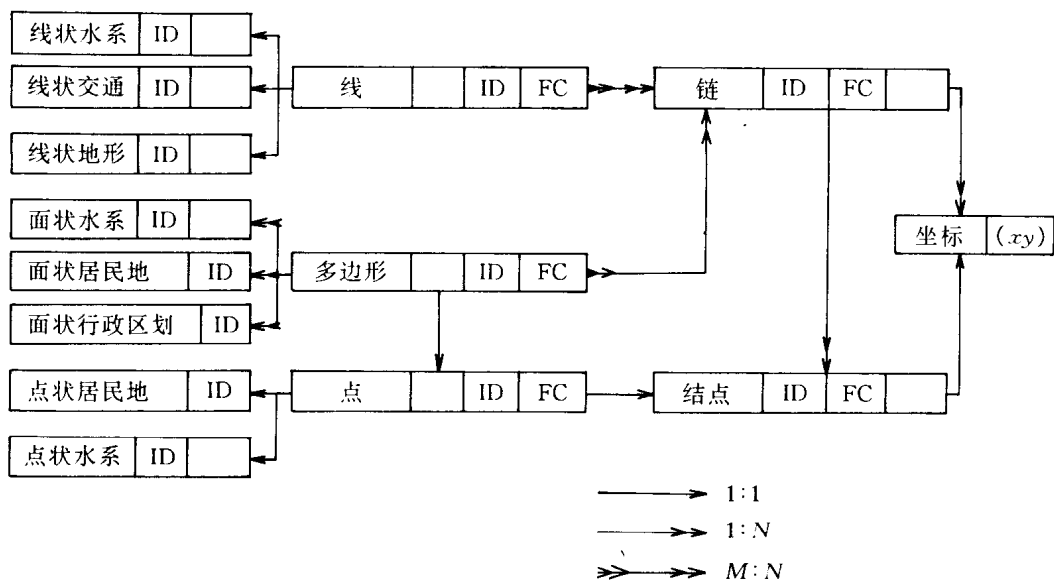


图 1-17 KCPIS 的扩展网络模型

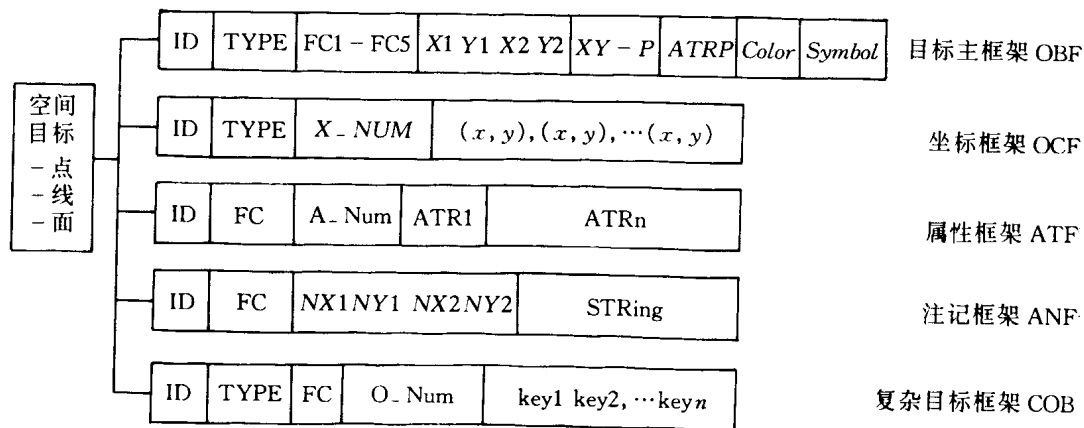


图 1-18 五个基本框架

根据该扩展网络模型,进一步抽象出四个关系框架:目标索引框架、坐标框架、注记框架、属性框架,作为 KCPIS 的空间目标统一表达框架,用于实现空间和非空间数据的统一组织和管理。其中主索引框架是把所有目标都具有的部分提取出来,包括范围($X1Y1X2Y2$)、指针($XY-P$)、分类码(FC ,如道路,河流,省界等)、目标类型($type$,如点/线/面/复杂目标)。如果 FC 为 1 的话,则为 Layer 模型; FC 为多个的话,为特征模型。

1.4 GIS 空间物理数据模型

逻辑数据模型并不涉及最低层的物理实现细节。但计算机处理的是二进制数据,必须将逻辑数据模型转换为物理数据模型,即要设计空间数据的物理组织、空间存取方法、数据库总体存储结构等[Kriegel,