

地球信息科学基础丛书

空间数据多尺度表达模型 及其可视化

李霖 吴凡 著



科学出版社

www.sciencep.com

内 容 简 介

本书着重对空间数据多尺度表达研究领域的理论、原理和方法等多个方面的问题展开了研究。从地球信息科学的发展阐明了该问题研究的科学意义,及其研究的必要性,从分析当今问题域的研究主题入手探明了其研究发展现状和存在问题。以此为基础,从对空间数据的尺度本质及其特征展开研究,提出广义尺度的概念和理论框架,进而提出尺度依赖的空间数据模型,到应用小波分析的理论与方法剖分地理空间为多尺度嵌套的序列空间,再到以线状数据、场数据等具体的空间数据类型来实现所提出的模型等内容为研究主线,建立了一系列的理论模型和实现方法。

本书可供地球信息科学、地学相关学科的科研人员、高校教师以及研究生等阅读参考。

图书在版编目(CIP)数据

空间数据多尺度表达模型及其可视化/李霖,吴凡著. —北京:科学出版社,2005

(地球信息科学基础丛书)

ISBN 7-03-014640-9

I. 空… II. ①李… ②吴… III. 地理信息系统-比例模型
IV. P208

中国版本图书馆CIP数据核字(2004)第134539号

责任编辑:彭胜潮 韩 鹏 董 巍/责任校对:李奕莹
责任印制:钱玉芬/封面设计:王 浩

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新 蕾 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2005年1月第一版 开本:787×1092 1/16

2005年1月第一次印刷 印张:8 1/2

印数:1~3 000 字数:184 000

定价:26.00元

(如有印装质量问题,我社负责调换〈新欣〉)

序 言

现代科学技术的进步为人类认识和理解自然现象提供了更加优越的方法论和越来越精细的观测平台。20世纪60年代数字计算机技术的发展,催生了信息科学——这个正改变人们日常生活和工作方式的学科。地球信息科学作为信息科学的一个重要组成部分,由于将空间(或地理)元素作为主要研究对象,进一步开拓了人类在空间认知、理解和把握地理现象规律性等方面的智慧之路。

地理信息从哲学的观点看是人们对地理(自然)现象的规律性认识,它既包含感性认识,也包含理性认识。由于地理现象的巨复性和目前人类认知水平的局限性,人们对地理现象的认知总是在一定抽象过程的基础上完成的,空间及时间的粒度以及空间规模是人类揭示地理现象规律性的关键因素。不同的粒度和尺度将揭示不同的规律,试想一下,在细粒度(高分辨率)和小尺度空间(大比例尺)下,Alfred Wegener(1880~1930)能发现大陆漂移学说吗?因此,对地理信息空间尺度的研究是深入探索地球信息科学本质特征的关键课题。空间数据的多尺度处理要解决由于用户对地理数据的不同应用需求和分析的需要,而导致需要对空间数据产生不同表示的问题。

李霖教授等自1994年在《测绘学报》发表颇有影响的论文“GIS中二维空间目标的非原子性和尺度性”以来,一直从事着地理信息空间尺度问题的研究;吴凡博士借鉴小波分析方法重点研究了地理实体几何数据的多尺度处理方法。在国家自然科学基金和国家测绘局科研基金的支持下,他们取得了一系列可喜成果,其研究成果和相关研究心得主要体现如下:

(1) 基于人类空间认知的特性,分析了空间对象的尺度概念以及表示特性,研究了空间数据的尺度语义特征;在数据库模式的基础上,提出了尺度依赖的空间数据模型,分析了此模型的重要性质,并探讨了基于此模型的空间数据维护更新模式。

(2) 根据空间对象的表示特点,分析了空间目标的非原子性(可分割性)和尺度性,通过形式化定义这些目标、目标之间的关系以及目标行为的尺度特性来研究了多尺度空间目标的空间和非空间属性随比例尺变化的一般规律,指出空间目标的尺度多态性可以用其特征的值来描述,值的变化在一定条件下会导致空间目标结构的变化。以集合论为基础,定义了空间目标的聚合算子,提出了多尺度空间目标的聚合模型。

(3) 分析研究了空间目标空间属性的基本形态特征及其尺度变化的形式

化描述方法,引入并初步探索了利用小波理论的多分辨率分析(Multiresolution analysis)方法来描述目标空间属性的尺度状态和变化特征,提出并实现了相应的多尺度处理与表达及其质量评估模型和方法。

(4) 基于小波分析原理,分析了自动获取不同尺度下空间数据几何特征的空间知识和最佳表示的机制,深入研究了空间线状地物和连续面状特征的尺度之间变换原理及方法,通过多个真实数据的实验,介绍了空间目标尺度转化的过程以及实际工作中的关键问题。

本书篇幅不长,但集中了作者的最新研究成果,从原理和实践的角度阐述了地理信息中尺度性质的本质特征和一般性转换规律。它的出版将促进地理信息认知、地理信息处理以及地理信息系统技术的发展。希望作者继续在这个研究方向上取得创新的成果。

中国科学院院士
中国工程院院士



2004年3月18日
于珞珈山(武汉大学)

前 言

空间数据是地理信息的数字载体,是客观地理世界的抽象表达。随着信息和知识时代的来临,人类对地理空间数据的需求结构向知识化方向发展,空间数据的原始积累也促使基于空间数据的信息服务成为主要需求,多尺度、多类型、多时态的空间数据已成为人类研究和解决人口、资源与环境等重大问题时所必需的重要信息资源,它既是国家空间数据基础设施和数字地球的重要内容,也是国家信息高速公路上的重要成员。

空间数据的多尺度表达是当今地理信息科学研究的前沿课题之一。它要解决由于用户对地理数据不同应用的需求,而导致需要对空间数据产生不同表示的问题。长期以来,国内外实际运行的地理信息系统,大多是对同一地区按不同比例尺要求重复采集和处理地理数据,重复建库。这不仅花费大量的人力、财力、物力,而且由于其巨大的数据量和冗余存储,以及同一空间对象的不同表示之间缺少相关和互动机制,造成空间数据的一致性难以维护,数据的及时更新也难以实现。因此,美国国家地理信息与分析中心(NCGIA)于1988年在其创新研究计划中开始提出研究空间数据的多重表示问题(Buttenfield and DeLotto, 1989; Buttenfield, 1993; NCGIA, 1993)。1992年SMALLWORLD的Richard G. Newell等人将空间数据的多尺度处理与表示列入GIS领域十个困难问题之一(Newell and Theriault, 1992)。1996年6月美国大学地理信息科学协会(UCGIS)也将该问题列为未来十年地理信息科学的十个优先研究领域之一(UCGIS, 1996)。1997年NCGIA的VARENIUS基金将“地理细节的形式化概念”列为高度优先的认知研究项目,以研究地理信息认知中的尺度、详细程度以及多尺度表示等多方面的问题(Mark et al., 1999; NCGIA, 1997)。研究支持相同现象的多重表示共存于同一数据库中的新的数据表达与管理方法,已列入由欧洲多家研究机构参与的、在欧共体资助下的MurMur(MULTiple Representations, MULTiple Resolutions)项目的研究计划中,并于2000年1月启动(Parent, Spaccapietra and Zimányi, 2000)。NCGIA近年来再次将空间数据的多种表达方式问题列为美国当代GIS研究的19个方向之一(张永忠译, 2001)。我国国家自然科学基金委员会也将地理空间尺度及基于尺度的智能化自动综合列为优先资助领域。在国家构建“数字中国”地理空间基础框架的总体战略中,同时也将空间数据的多尺度表达列入地理空间数据库关键技术研究的核心问题之一。国际摄影测量学会(ISPRS)数据综合与数据挖掘工作组联合国际制图协会(ICA)地图综合委员

会也于2002年7月在加拿大渥太华以“空间数据的多尺度表达”为主题召开学术会议,研讨与多尺度表达有关的基础理论与应用整个领域内的有关问题,会议共列出了12个研究主题进行了广泛讨论(ISPRS Commission IV and ICA,2002)。

本专著是在国家自然科学基金项目“多尺度空间目标聚合模型的研究”(No. 69773048)、国家测绘科技发展基金项目“基于小波理论的多尺度空间数据处理与表示研究”(No. 99013)、武汉测绘科技大学科技发展基金项目“基于小波理论的GIS多比例尺数据处理与表示研究”(No. 9809)以及国家基础测绘科技计划项目“国家基础地理信息数据跨比例尺处理试验”(No. 1469990324231)等与空间数据多尺度表达有关项目的共同资助下取得的研究成果之一。作者试图在广泛深入地分析国内外相应研究发展动态基础上,通过对相关问题的研究,建立新的空间数据多尺度表达模型,为最终实现无比例尺的地理信息系统奠定理论和技术基础。从思想方法上,本书的核心在于:

(1) 地理信息的自动综合是实现空间数据多尺度表达的核心技术,这一点已成为共识。但是长期以来,没有将其作为一种认识世界和模拟世界的思维方法和思维过程来对待。也就是说没有从建立地理信息系统空间数据表达模型这个根本问题上来看待综合问题,而仅仅是将它作为GIS或数字制图派生不同表示的一种算法或方法,从而使现有GIS与生俱来就没有尺度空间上的分析、处理和表达能力。这个问题的源头在于没有将尺度看作为与时间、空间一样的一度空间。因此研究尺度依赖的空间数据表达模型才是建立无级比例尺GIS,实现空间数据多尺度表达的基础和关键。这对于地理信息科学理论的进一步发展,研制具有原创性的新型地理信息系统平台有着极其重要的科学意义。

(2) 现有多级比例尺的GIS系统大多是基于静态的、低维的空间数据模型来建立其空间数据结构的,它既不支持实时的、动态的空间数据修改和空间关系维护,也不支持在数据表达详细程度即尺度上的变换。因此人们不得不从存储最详细的数据库开始,在需要时通过离线式综合逐级派生次详细的数据库。通常这种派生过程既不连续也没有逆过程。因此在派生比例尺跨度较大时,不仅计算量大,而且极难维护多尺度表示之间的一致性。而基于尺度依赖空间数据模型构建的GIS,其派生过程不仅可逆而且有完善重建的能力,即可完全回复最详细一级的数据库。因此其日常存储和维护的空间数据库不一定是最详细的,而可能是根据需要存储较概略或对某些应用是最概略的数据库,在必要时可通过动态的、增量的和连续的方式增减信息的详细程度来满足多种需求,因而可以是一个逐步求精的过程。由于在多尺度上的空间对象是

一个整体,内建了相关和互动机制,因此多尺度空间数据的自动更新和维护可以通过这个增量式的逐步求精过程方便地传递完成,从而确保了空间数据的现势性。这样的模拟框架其好处是显而易见的,它将空间数据的复杂性在尺度维上进行了分解,使得在空间推理的不同阶段可以集中推理的注意力,从而使问题求解系统需处理的信息量总在一个较低的数量级上。这对于海量空间数据的网络传输和发布,以及实现在线、实时、可交互地浏览多尺度的空间数据尤其具有重要的现实意义。

(3) 目前公认的 GIS 三要素是空间、时间和属性,将尺度作为与空间、时间一样的空间数据基本维的观点在国内外正在引起重视,尤其是在遥感科学应用领域。因此有必要按照空间结构和语义一致性的要求,将尺度与空间数据模型及其空间数据多尺度表达联系起来,提出一个有可操作性的尺度依赖空间数据表达模型,并研究其相应的一系列处理模型和处理方法,从而使基于该模型开发的 GIS 具有尺度维上的分析处理、更新维护和表达能力。这对于实现空间数据的智能化分析处理和集成有非常重要的学术价值,是建设多尺度地理空间数据库和实现数字地球、数字中国、数字省区的技术基础,对推动地理信息产业的进一步发展有着重要的实际意义。

值此成书之际,作者要感谢导师李德仁院士、祝国瑞教授、毋河海教授、郭仁忠教授所给予的指导和一如既往的支持。感谢王家耀院士多年来所给予的指导。

由于问题研究的复杂性和研究时间的限制,本专著的研究仍是初步的,在一些方面尚不成熟,旨在抛砖引玉,敬请提出宝贵意见,作者在此表示感谢。感谢曾经帮助和关心过我们的所有老师、同事、同学和朋友们!

李霖 吴凡

2003年9月于武汉大学

教育部地理信息系统部级重点实验室

目 录

序 言	
前 言	
第 1 章 绪论	1
1.1 空间数据的多尺度表达问题	1
1.2 国内外相应研究的成果与动态	6
第 2 章 空间数据的尺度特征及其形式化描述	26
2.1 地理空间与地理信息空间	26
2.2 广义尺度下的地理空间数据及其基本特征	31
2.3 广义尺度模型	35
2.4 尺度特征的空间认知	40
2.5 智能化空间数据多尺度处理流程	46
第 3 章 尺度依赖的空间数据模型	47
3.1 尺度依赖空间数据模型的需求分析	47
3.2 尺度依赖空间数据模型	48
3.3 多尺度空间数据库的一致性约束模型	49
3.4 多尺度空间数据的传递更新模型	51
第 4 章 空间数据的多尺度聚合模型	54
4.1 基于空间数据库的聚合操作定义	54
4.2 聚合模型	56
4.3 地理对象的聚合	59
第 5 章 小波理论及其在空间数据多尺度表达中的应用原理	62
5.1 小波理论的基本原理	62
5.2 空间数据多尺度表达中应用小波分析的哲学思考	74
5.3 面向空间数据多尺度处理的小波函数选择	75
第 6 章 基于小波理论的空间数据多尺度表达模型	77
6.1 基于小波理论的空间数据多尺度表达模型	77
6.2 尺度依赖综合程度评估模型	79
6.3 多尺度一致性维护模型	80
6.4 基于小波分析的空间数据挖掘方法	81
6.5 多尺度处理模型应用流程	84
第 7 章 尺度依赖空间数据的表达及其可视化	85
7.1 线状特征空间数据的多尺度表达及其自适应可视化	85
7.2 场空间数据的多尺度表达及其自适应可视化	102
7.3 结束语	114
参考文献	115

第 1 章 绪 论

1.1 空间数据的多尺度表达问题

1.1.1 空间数据的多尺度表达与地理信息系统

空间信息，特别是具有地理参考特征的各种属性信息，约占信息总量的 75% 到 80% (McKee, 1996)，因此对地理空间信息的开发利用尤为重要。地理信息可以定义为地球表层特定地方的一组事实 (UCGIS, 1996)，它是有关位于地球表层附近的要素和现象的信息 (Goodchild et al., 1999)。地理信息不仅内容丰富，信息量大，而且结构复杂，利用的难度大，所以它的处理和应用已成为世界各国学术界和产业界在多学科交叉领域研究和开发的对象。

地理空间数据是地理信息的数字载体，是客观地理世界的抽象表达。它是地学家们为了直接或间接地描述具有空间特征的现象而选择的一些特性，通常是数字特性 (Dent, 1989)。由于地球表层的无限复杂性，人们不可能观察地理世界的所有细节，地理信息对地球表面的描述总是近似的 (UCGIS, 1996)，近似的程度如何反映了对地理现象及其过程的抽象程度或抽象尺度。因而尺度必定是所有地理信息的重要特性，它定义了人们观察地球的一种约束。只有经过合理的尺度抽象的地理信息才更具利用价值。不同的尺度不仅在所表达的信息密度上有很大差异，而且还会影响所表达的地理信息是否正确，因为不少地理现象和规律只在一定的尺度出现 (Cao and Lam, 1997; Xia and Clarke, 1997; 邬建国, 2000)。从哲学和认识论观点看，多尺度的、分层次的分析也是一种思想方法。在观察、理解和传播地理空间知识过程中，地理现象的表现不仅取决于其本身的特征，而且依赖于观察者所用的尺度和方向，因而通常会得到多方面的表现结果。由于多种地理现象和过程的尺度行为并非按比例线性或均匀变化 (UCGIS, 1996)，因而进行一系列的尺度和方向分析则能有效地反映出该现象的本质特征，从而在多尺度的分析框架中精确重建其空间意义 (空间含义)。因此，地理信息系统 (GIS) 不仅需要多种详细程度的空间数据支持，而且需要把这些多尺度表示的信息动态地联接起来，建立不同尺度之间的相关和互动机制，以进行有效的综合分析和辅助决策，从而构成了多尺度的 GIS (multi-scale GIS)，也称无比例尺 (Scale-free, Scaleless)、不依比例尺 (scale-independent)、自由变比 (on the fly) 的 GIS。所谓多尺度指示了数据库中包含多种详细程度的空间数据，并允许在可变分辨率 (可变尺度) 上表示、分析和处理空间数据。

到目前为止，我国已建成了覆盖全国范围的 1:100 万和 1:25 万基础地理信息系统，包括相应比例尺的地形数据库、数字高程模型库、地名数据库等，其中地形数据库包括水系、交通、境界、居民地、地形、植被等 14 层要素；组织生产了全国七大江河重点地区 1:1 万正射影像和 DEM 库；国家级数字地球空间数据框架和国家空间数据基础设施建设已初具规模，具有多尺度、海量数据的特点，它们在国家的经济建设和社

会发展以及重大自然灾害的速测速报和灾情评估中发挥了重要的作用。目前正在加紧建设 1:5 万的基础地理数据库（包括 4D 产品），各省、市、自治区也正在加紧建设各省区 1:1 万的基础地理信息系统和更大比例尺的城市基础地理数据库，以满足全社会的发展需要。

构建多尺度的空间数据库框架，是数字地球和空间数据基础设施的战略重点之一（陈军，1999）；实现空间数据的多尺度处理和自适应地无级缩放是实现无比例尺 GIS 的关键技术（承继成等，1999、2000；杨崇俊，2000；王家耀，2001）。它们的明显特征是以基于多分辨率、多尺度的框架为基础的地理空间数据贯穿始终。多尺度空间数据处理方法也是超大型 GIS 的根本解决方案（张家庆，1994），是空间数据挖掘和空间数据仓库的核心技术（Han and Kambr，2000）。基于多尺度的分析处理，允许在速度和精度之间寻求折衷，从而可以在非常大的数据库中快速获取近似的、概括或概略的答案，因此多尺度 GIS 与单一尺度的 GIS 相比更能满足用户的需要。空间数据库（数字地图）是 GIS 的核心，是地理世界的一种有序的数字表示形式，基于这种表示可以完成有关现实世界的某些推理任务，这里面包含了很重要的智能性活动——地理信息的综合，其目的在于派生更抽象更高层的信息。它通过对空间数据的空间和非空间属性进行层次抽象与聚合来完成给定的一系列的地理推理任务，因而多尺度 GIS 应是基于空间数据的知识发掘的产物。为解决海量数据未能充分利用和社会对信息的需求之间的矛盾提供了坚实的基础，为社会各领域利用和共享空间信息提供了保障。

1.1.2 空间数据多尺度表达问题的提出

20 世纪 60 年代地理信息系统的出现和日益广泛的应用，大大提高了人类处理和分析大量有关地球资源、环境、社会与经济数据的能力。随着信息和知识时代的来临，人类对地理空间数据的需求结构向知识化方向发展，空间数据的原始积累也促使需求结构逐步向地理信息数据集成和信息资源知识产品开发转移，空间数据的信息服务将成为主要需求。需求结构的多元化趋势，带来了对多种尺度空间数据分析和显示的大量需求，并成为 GIS 实用化的核心问题之一。这主要表现在：用户对于解决不同的问题，对空间数据需求的详细程度是不一样的，GIS 如何提供给用户多尺度或多种详细程度的空间数据？GIS 以其丰富的地理信息内容作为数字地图制图生产的基础之一，如何满足输出不同详细程度和不同内容的地图产品？前一个问题对应着模型综合，它强调空间数据的模型抽象和深层次的地理空间知识，以达到压缩表示层上的细节并按特定的抽象程度和一致性的空间结构表达地理现实的目的。它具有广泛的适用性，也是多尺度地理空间分析的基本手段；后一个问题则对应着制图综合，以地图显示（包括电子地图媒体的和常规地图媒体的显示）和生产各种比例尺地图为目的。它强调基于各种媒介，在有限的表达空间内进行抽象并消除视觉冲突以可视化地表示地理信息，从而使 GIS 中地理空间信息的压缩和再现与尺度变化自适应。通过模型综合从已有数据库（数据集）派生新的综合的数据库（数据集），既要满足相应尺度的地理空间分析的需要，保持与抽象程度相适应的较高的信息密度和空间特征，同时也是派生视觉综合模型和地图生产的先决条件（Lagrange，1997）。

地理信息的自动综合是空间数据多尺度表达的核心内容（王家耀，2001），但这一

问题却一直是困扰地图学及 GIS 界实现空间数据自动处理与合理可视化的极具挑战性的国际性难题。实现自动综合必须具备两个基本条件：地理空间数据及其关系的表示（即数据模型）和自动综合操作算子（郭庆胜，1998）。由于自动综合的极度复杂性，至今仍然没有一整套现成的自动综合操作，这意味着不能保证对每一操作至少有一种算法（Lagrange, 1997）。由于不同尺度下的空间目标，其运动状态及其关系有所不同，维数也会发生变化，因此多尺度空间数据处理模型的建立异常复杂（Li, 1999；吴凡、李霖，2000），而且在不同尺度上仅用单一的自动综合方法（非尺度依赖）或算法来实现多尺度的表示，其通用性、可靠性、实用性较弱（Muller, 1993；李霖、李德仁，1994；吴凡、李霖，2000；王家耀，2001）。作者认为尺度依赖的综合方法还远不够丰富，大多数自动综合算法与尺度无直接或精确的关系（吴凡、祝国瑞，2001）。因而算法与尺度之间没有相互适应的关系，也就是说综合算法不能随尺度变化而对自身进行相应的调整。这常常会导致不同尺度空间目标之间的非一致性，尤其在交互式缩放（放大、缩小）时图形特征会出现不协调。此外，大多数算法不能预知和控制其综合效果，不能从整体上（全局上）评估结果的一致性（Lagrange, 1997）。综合程度如何至今仍未找到合理的、数量化的评估方法和标准（Joao, 1998；郭庆胜，1998），当前常用于评估的数量化方法，由于其没有反映尺度的规则和需求，因而提供不了多少实际的指南（Dutton, 2000）。而数量化的综合程度评估方法与标准正是实施计算智能技术（Computational Intelligence Techniques, CIT）的关键（Joao, 1998），因为它们能为智能化计算提供目标函数和约束。20 世纪 90 年代以来，地理信息的自动综合开始强调要面向地理要素，而不能仅仅是解决综合的纯几何问题，即要考虑每个要素与其所处背景的空间和语义关系，每一空间目标与其周围目标的相互依赖的关系。这就是所谓全局综合或上下文综合（Holistic Feature Generalization）问题，包括综合操作的形式化、综合过程的理解及其形式化等。然而由于问题的极度复杂和硬软条件的限制仍没有取得多少有效的进展（Lagrange, 1997；Ruas, Plazanet, 1996；Richardson, Mackaness, 1999）。近年来由于多媒体与网络技术的快速发展，为利用空间数据开发交互式地图提供了机会，从而对检索和显示地图的速度提出了更高要求。但在现有硬件条件下，当前大多数综合算法的性能离实时计算的要求差得太远（Lagrange, 1997；王家耀，2001）。这既有算法上的效率问题，也缘于当前的空间数据模型大多尚不能支持动态和增量地更新修改功能（Gold, 1995；Jones et al, 2000）。而多尺度的处理往往并不以地图生产作为惟一目标，有时只是为了实时观察地理要素的特征分布，原始数据保持不变，改变尺度可以得到相同数据不同的暂时表示或多种表示，以确定分析是否进行或判断分析的结果如何，所以产生的图形表示有些只是暂时的图像。因此相对常规制图而言，在数字环境下（GIS 和地图数据库）计算速度比高质量的地图输出效果更重要（Lagrange, 1997）。

作为认知主体的人类不仅具有不同的文化背景和专长，而且从不同的角度来认知客观世界。而对于不同的应用目的，通常也需要相同地理现象的不同表示。因而空间数据的多尺度处理与表示是符合人类推理习惯的一种自然表达方式。作为一种分析和思想方法，它不仅仅是为了满足使 GIS 能自动化地、合理地、自适应地提供多种详细程度的空间数据，实现空间数据的多尺度可视化，同时它也是一种更自然的空间认知和空间分

析方法。特别在地理学中，常常需要利用多个尺度以提高研究效率并确定其结论正确与否 (Herot et al., 1980; Minsky, 1985)。例如在城市中旅行一般采用 1:1 万到 1:3 万比例尺的地图比较合适，这个尺度范围的地图提供了比较详细的信息；对于城市之间的旅行则应使用 1:5 万到 1:20 万的地图；国家之间的旅行应使用 1:30 万到 1:100 万的地图，这样可以确定总的方向和方位概念并满足一览性的要求。Mark 曾指出，制图综合的重要性不仅在于图形显示，而且也为了进行有效而适当的空间分析 (Mark, 1991)。人类对客观地理世界的认知是分层次的，所进行的地理空间分析有很强的区域性。例如研究一个国家范围内各城市制造业从业人员的分布，通常将城市看作点，因此城市内部从业区域的差别就成为次要。但如果研究这种现象的内部分布时，研究人员就应把注意力放在城市内部从业区域的分布上。全球环境变化的研究也是如此，其核心问题是探讨土地利用变化和气候变化对生态系统的影响及其反馈机制以及人们在未来气候 (环境) 变化下所要采取的适应性管理对策。与传统的生态模拟相比，有关全球变化对生态系统影响的模型模拟都是以大尺度 (全球或区域尺度) 的空间格局及其动态变化作为主要的研究对象。然而大尺度的生物地理模型都应从模拟小尺度的生态过程开始，并最终校正到大尺度的地理格局上 (傅伯杰等, 2001)。因此，其研究必须在大 (区域)、中 (景观)、小 (斑块) 多个尺度上进行，如何建立能够抓住所有尺度上的重要过程的简单的模型结构是问题关键所在，这就要求对各种尺度效应应有所了解。

GIS 是基于空间推理实现辅助决策的，多尺度、分层次的空间表示有助于在空间推理的不同阶段集中推理的注意力。应用空间数据的多尺度处理方法，人们可以以不同的抽象程度描述现实世界，使每一类用户的注意力可以集中在其感兴趣的基本信息上，而这个基本信息是通过智能化的综合手段从原始空间数据库中优化出来的。用简单的方法和足够好的近似去回答局限于相应层次的问题是十分经济的，只有在必要时才需深入到更详细的层次。当多尺度空间数据的表示设计得合理时，可以做到在不同的推理阶段，问题求解系统需处理的信息量在同一数量级。当在顶层作推理时，尽管空间数据涉及的地理范围可能很大，但由于大量细化信息的忽略，问题求解系统处理的对象并不多，所需处理的信息量也较低；而问题求解系统深入到较详细的层次时，地理范围已缩小到感兴趣的地方，所以虽然信息细化了，但由于范围小，要处理的信息量仍然不大。也就是说，由于在增加细节的同时缩小了范围，使问题求解系统能有效地处理整个空间推理的复杂性。正如 Muller 所说的那样，人类的推理是以一种有序的方式对思维对象进行各种层次的抽象，以便使自己既看清了细节，又不被枝节问题扰乱了主干，因为“超过一定的详细程度，一个人能看到的越多，他对所看到东西能描述的就越少” (Muller, 1993)。人们运用不同比例尺、不同专题的地图来分析并表达客观地理世界就是一个很好的例证。同样，模拟人脑思维活动的智能化推理工具使用自适应的分层推理方法，以及使用四叉树和影像金字塔结构也是这种思维方式的例证。由此可见，地理分析不仅需要不同尺度的空间数据，而且需要利用多尺度的分析手段，这既是空间分析本身的需要也是计算效率的要求。需要指出的是，用四叉树那样的递归区域分解表示是不同于多尺度表示的。主要的区别在于多尺度表示方法在不同尺度 (层次) 上表示抽象程度不同的信息，而四叉树是对空间的划分，而不是对现象的划分，它的任何层次均表示抽象程度相同的信息。换言之，多尺度表示依靠的是信息的层次抽象，而四叉树表示则依靠空间

的划分实现信息量的简单分割。从另一方面看，多尺度表示和四叉树表示可以综合应用，例如多尺度表示中的每一层（每一尺度）均可表示为四叉树。

空间数据的多重表示 (Multiple Representation) 是指“随着在计算机内存储、分析和描述的地理实体的分辨率 (尺度) 的不同，所产生和维护的同一地理实体在几何、拓扑结构和属性方面的不同数字表达形式” (NCGIA, 1993)。就是说，具备“多重表示”机制的 GIS 能够以不同分辨率 (尺度) 表达其数据库中相同的地理场、特征或对象。这需要在不同的空间或时间尺度上综合数据的能力 (Buttenfield and DeLotto, 1989; Joao, 1998)。然而由于当前 GIS 尚缺乏完善而又实用的自动综合方法可用，尤其是实时计算和实时自动综合的方法、模型，因此远不能满足空间数据多尺度表达的要求。当前 GIS 数据库为了满足人们浏览和利用空间数据集的不同需求，不得不存储多种来源、多种版本、多种空间分辨率或比例尺、多种详细程度的空间数据，构成多尺度的异构空间数据库 (Heterogeneous Spatial Databases) (Egenhofer, 1998)，使同一空间实体的多种表示共存于同一数据库或信息系统中，从而会产生大量的数据冗余、数据矛盾及与其相关的弊端。尤其困难的是，由于同一空间实体在多个尺度上的不同表示之间毫无联系，使这种多重表示的 GIS 在数据更新时，需要逐级尺度分别、重复地进行，不仅耗费巨大、重复投资，而且难以保证同一实体的多种表示之间的一致性，造成 GIS 数据库难以维护，数据的及时更新难以实现，从而导致其数据库数据现势性差，难以实用，不能满足用户对数据现势性的要求。这在大区域、大型异构 GIS 的更新中表现的特别明显 (张家庆, 1994; Egenhofer, 1998)。多尺度数据库的一个关键是维护空间目标之间的一致性。多重表示中的不一致性往往会导致严重的后果。因为高层决策者掌握的信息与具体实施者所掌握的详细信息不一致，当高层决策意见从上层传达到下层时，就会导致错误和矛盾，反之亦然。长期以来，由于 GIS 空间数据模型的设计没有对尺度及其尺度效应给予应有的重视，从而先天性不足，使现有 GIS 没有尺度维的处理和表达能力 (Li, 1999)，缺少维护地理目标多重表示一致性的方法 (João, 1998)。我们面临的挑战是找到灵活和恰当的数据模型，它能满足多尺度地理信息的综合、分析、维护和多种表示的要求。目前对于尺度问题的简单理解常常不适合于有效解决方法的设计，因此有必要澄清和建立一些概念并讨论其相关的性质，以进一步地研究地理对象的空间形态和过程随尺度变化的规律，从而设计有效、合理的尺度依赖模型，这是建立多尺度空间数据处理模型和表示方法的基础。

“尺度”是与地理信息细节相关的最基本的但也是难以理解、易混淆和过载的概念之一，它有多种含义，如绝对大小、相对大小、分辨率、颗粒度、详细程度等 (NCGIA, 1997)。尺度也是认知地理对象、地理空间和地理现象的基础 (Montello, 1993; UCGIS, 1996)，是地理信息和地图综合的主要动力。在与地理信息相关的学科中，尺度问题是虽经努力但仍未解决的主要问题 (Li, 1999; Bruegger, 1995)。尺度既不是新问题，也不仅仅为地理信息学家所关注。长期以来，人们已经知道尺度变化会影响信息被观察、表达、分析和传输的详细程度。改变数据的尺度，而不首先了解改变尺度将产生的结果，会使有关现象的过程或形态的表达得不到预想的结果。研究表明，多种地理现象和过程具有明显的尺度依赖特征，其尺度行为并非按比例线性或均匀变化 (UCGIS, 1996、1998; Goodchild and Quattrochi, 1997)。例如改变高程的采样间隔

或分辨率,会影响坡度、汇水区等地形属性的分析;全球环境变化研究其复杂程度随尺度变化而变化,但的确不是作线性或均匀变化(UCGIS, 1998)。数字环境下,地理空间数据作为地理信息的载体和客观地理世界的抽象表达,试图以离散方式描述、模拟自然界的连续分布现象,势必要受采样分辨率(也是一种尺度)的约束(Bruegger, 1994)。地球的大小和测绘技术的精度限定了地理空间分辨率的合理取值范围,它可以大到约 10^7m 的分辨率(将地球作为一个整体),到通常可用于对地观测和制图的最精细分辨率 1m ,或大多数实地测量的最小分辨率如 1mm (Goodchild et al., 1999)。在如此丰富的分辨率框架中,用离散方式模拟自然界的连续分布现象时,每一空间数据对象(或对象集)所代表的充其量只是真实地理实体全部信息的一个子集。这个子集的真实度是随着为了在计算机中存储、分析和描述而对该地物进行数据编码时的分辨率的不同而变化的。在最坏的情形下,该数据子集不但在地物信息详细程度上远不如人意,甚至会出现对地物歪曲表达的情况。这种有关地理实体的外貌和隐含结构真实程度的尺度依赖特征,限制了GIS基于空间数据进行地理空间分析和可视化以及地图制图对各种信息层次的要求。因此,Cola认为,尺度是在空间数据分析处理中到处都要碰到的问题,应将它作为与空间、时间和主题一样的基本因素(Cola, 1997)。Quattrochi和Goodchild也指出,尺度毫无疑问是任何研究最基本的问题之一(Goodchild and Quattrochi, 1997)。

近些年来,由于地理信息采集技术,尤其是遥感、全球定位系统和国际互联网的飞速发展,空间数据的集聚能力和数据量迅速增加,数据内容的变化周期正在大大缩短。数据量之大,变化之快已远远超出人类的手工处理或常规的数据库更新和处理技术的能力。空间数据库的现势性已难以保证。如何实现智能化的快速自动更新已迫在眉睫并成为地图学与GIS界研究的热点问题之一。由于空间数据的挖掘和整合问题日益突出,集成、尺度、过程模拟和易用性已成为地理信息科学面临的主要研究问题(Mark, 1999)。多尺度的空间数据分析、处理与表达和地理信息的自动综合在当今的地理信息技术中正显得越来越重要和必不可少。多尺度空间数据处理与表示的理论与方法是建立多尺度GIS或无级比例尺GIS的核心,是多尺度空间数据自动更新的基础。在一系列国际GIS与地图学大会上,多尺度GIS的有关问题均被列为中心议题(Devogele et al., 1996; João, 1998; Jones, 1991; Kidner, 1996; Kilpellinen, 1995; NCGIA, 1997; Timpf and Frank, 1995a, 1995b; UCGIS 1996, 1998; ICC' 2001; GEOINFORMATICS&DMGIS' 2001),但GIS中尺度问题的解决仍远未达到人们所期望的程度,已成为GIS实用化的核心问题之一,也是建立无比例尺GIS和实施数字地球计划的关键技术之一。

1.2 国内外相应研究的成果与动态

NCGIA为空间数据的多尺度表示问题定义了五个研究领域(NCGIA, 1993):数据模型、多重表示之间的联接、所实现视图的维护、空间模拟、综合问题等。同时还应研究由多个机构提出的有关对多尺度数据的要求和需求的问题。NCGIA还为多重表示数据库定义了研究内容和要求(Abler, 1987);形式化描述不同细节程度(Level Of

Detail——LOD) 上的目标; 形式化如何联接不同 LOD, 以使一个层次上的变化传递到其他层次上; 允许自动演绎多重表示。对于地图制图综合, 其综合算法应该附加约束, 保持目标的总体结构; 保持不同 LOD 层次的一致性。UCGIS (1998) 认为未来的 GIS 应该是尺度依赖的, 对用户来讲是尺度敏感的, 并提供尺度管理工具。为此应该研究: 开发能了解并描述尺度行为的空间数据模型; 先进的、高质量的对尺度影响的理解能力; 具有描述数据尺度的新颖方法; 自动改变尺度的智能化方法等。ISPRS 数据综合与数据挖掘工作组和 ICA 地图综合委员会为 2002 年召开的学术会议列出了: 多尺度/多重表示数据库; 数据库/模型综合; 基于多尺度表示, 从图像和矢量数据集中提取目标; 利用层次、多尺度结构进行图像分析 (解译、分类); 图像与 GIS 数据的集成与匹配; 三维可视化与综合; 几何与语义数据的匹配/一致; 多尺度数据库中传递更新与版本控制机制; 基于空间位置的服务和小目标的地图显示; 与国际互联网、互操作性和空间数据基础设施等有关的多尺度方面; 矢量域提取目标的算法在图像域的应用; 质量评估技术等 12 个研究主题。

从理论与技术方法研究的角度, 归纳为以下 4 个方面进行分析。

1.2.1 变尺度空间数据的存取与多重表示

多重表示是地理信息系统中一个重要的研究主题 (Buttenfield and Delotto, 1989)。人们在认识周围世界时, 在其思想中自然地使用多种甚至是相矛盾的表达 (Tversky, 1993)。其原因在于作为认知主体的人类不仅具有不同的文化背景和专长, 而且从不同角度来认知客观世界, 因而在 GIS 数据库中储存现实世界实体的多种表示是十分自然的。GIS 数据库必须能在不同等级表示实体并支持不同分辨率级别的修改。按每一分辨率级别描述实体, 必须能形式化描述它们之间的关系, 以致一个实体的变化能传播到其他实体, 从而能自动演绎出其他分辨率级别, 并建立其互动机制。多重表示问题就在数据提取及生产过程中产生。这要求地理信息系统必须具有新的模拟和推理方法才能满足多方面的需求。

从逻辑上有三种方法可为用户提供存取变尺度的空间数据以实现多重表示的能力。

第一种方法是在数据库中储存来自同一个现实世界实体的多种表示, 即建立对应于多种比例尺的多个数据库 (图 1-1)。其核心是根据 GIS 中对地理信息的不同信息需求, 划分出一系列比例尺空间数据, 来模拟空间变化的连续性。数据库通过采集不同比例尺的数据来建立, 图形显示时通过控制当时屏幕比例尺的变化, 轮换调入和释放相应尺度的数据实现。当前大多数 GIS 数据库存储的就是具有固定比例尺的多个版本的数据, 如我国建立的 1:50 000、1:250 000、1:500 000 和 1:1 000 000 国家基础地理信息系统。法国国家地理院 (IGN) 制图署也采取这种方法开发了三个主要数据库: 1:25 000 的 BD Topo、1:50 000 的 BD Carto 和 1:1 000 000 的 BD。在一些商用 GIS 软件和计算机图形学中常采取这一方案, 如著名的 ARC/INFO (ArcView)、MAPINFO 等 GIS 软件。该方法在一定程度上缓解了系统对不同详细程度数据的迫切需求与自动综合相对落后的矛盾, 对早期 GIS 的发展和应用起了很大的推动作用。其问题是在存取时该拾取哪一表示和该如何在数据库中适当地储存它们才能正确地完成任务。这种方法在不同尺度上, 其表示同一地

理实体的空间对象之间没有任何联系，尺度之间没有平稳过渡关系，也不考虑多尺度空间对象的一致性和可维护性。因此多尺度上图形层次明显不连续和不协调，数据更新困难，耗费巨大。

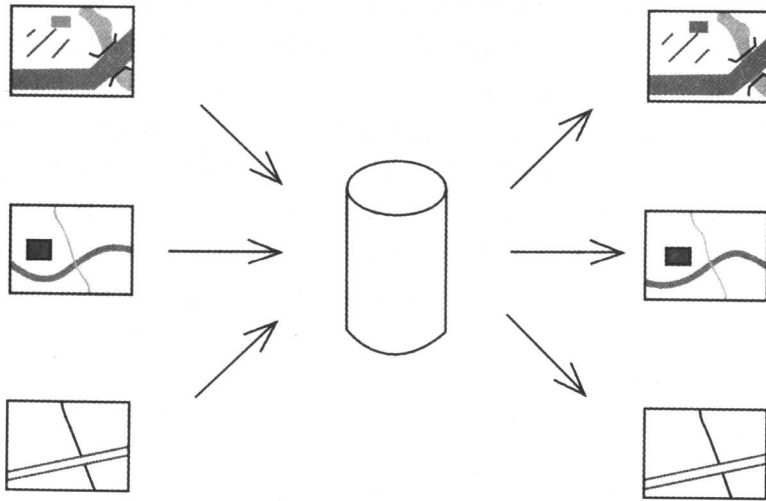


图 1-1 多尺度多版本建立多重表示模型

第二种方法是多重表示被储存在数据库中的一些“关键尺度”上。其数据库包括了多个层次的、详细程度逐渐增加的不同尺度的空间数据，高层某一对象可能由多个低层对象来表达，从而构成层次树数据结构，它是多尺度数据库中实现多重表示的先决条件(图 1-2)。这种方法仰赖层次数据结构的实现，即所有空间对象按其尺度来排序。通过这种多分辨率的层次数据结构表达空间数据，可以实现存储和检索不同详细程度的内容。空间数据集可以随尺度变化而增加或减少，但只能在该数据集允许的尺度范围内变化。关键尺度的确定取决于利用该数据集完成什么样的任务。Van Oosterom (1995) 利用反作用树 (Reactive-Tree) 技术，在开放式的面向对象数据库管理系统 Postgres

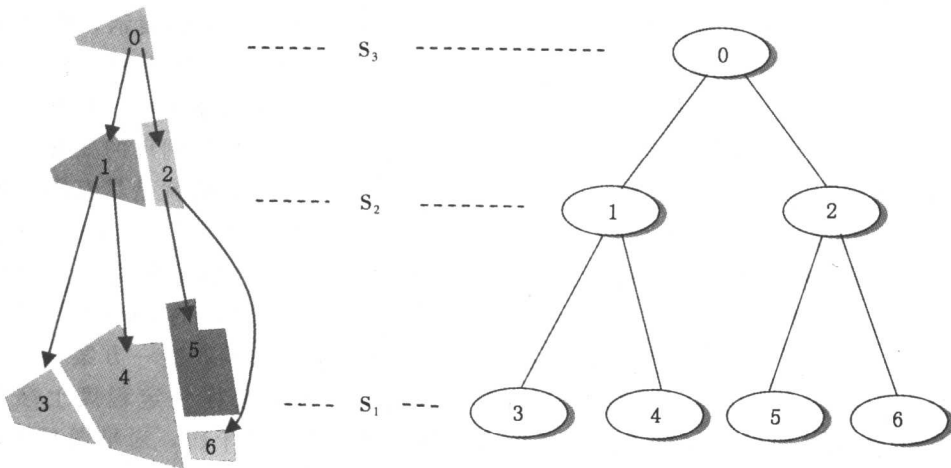


图 1-2 多尺度对象的识别和联接及其层次树结构 (有向非循环图)

中建立了一个交互式的多尺度 GIS——GEO++。Timpf 在其博士论文研究中基于图论也建立了这样的实验系统 DAG (Timpf, 1998)。这是目前认为比较现实可行的方法。但其数据库中各尺度 (层次) 的空间对象不能自动地彼此相联, 而这种联接对于识别在各层次上表达同一地理实体的空间对象是十分关键的。目前基于人工预置的识别方法缺乏动态可维护性, 因而层次树难以自动建立和维护。而且随着版本数量的增加, 其生成层次并建立联接的过程会越来越复杂, 效率也会越来越低。该方法努力的方向是开发可能的、更健壮的、智能化的、支持多重表示的数据结构, 包括面向对象的方法和语义数据模拟, 以及基于面向对象的方法建立智能识别处理程序并为每一要素附加丰富的语义数据。

第三种方法是所谓基于主导数据库的方法, 即维护一个单一的、有较高详细程度的主导数据库, 任何其他较概略的表示能按需要从该数据库自动演绎、派生 (图 1-3)。这是一种比较理想的然而也是最难以实现的方法, 是地图学家努力的方向 (Timpf, 1998; 王家耀, 2001)。这种方法是应用综合运算来产生数据库的多重版本, 因而其关键之处是需要自动综合的能力, 常规的人工地图制图综合是这一方法的模拟实现。这种 GIS 数据库能在不同等级上表示实体, 并能确定实体间的联接关系, 支持对数据的不同详细级别的修改。当一个实体被修改时, 其变化能通过联接关系传播到其他实体, 从而自动演绎出其他分辨率级别的数据, 实现多重表示。从存储的数据库对象派生所期望对象的变换, 必须予以准确地描述并形式化才有可能实现 (Bruegger and Frank, 1989)。一些商用 GIS 软件如 Smallworld GIS, Interlgraph DynaGen 等已开发了部分这种功能, 但都未从空间数据处理模型上作整体考虑, 目前仍需要进一步研究。仅从存储结构上对多尺度图形数据进行区分和管理, 以及与其相应的从地理要素的几何特征上实现其自动综合, 至今还没有取得突破性的进展。不仅如此, 要从根本上解决空间数据多尺度问题, 还必须研究属性的多尺度问题。因为空间目标不仅有图形形式, 还有一般性的说明信息, 即语义部分。只有对空间目标的图形部分和语义部分都进行多尺度的处理和表示, 才能真正完成从同一基础数据集导出不同尺度数据集的任务。此外还必须自动建立不同尺度之间的相关和互动机制, 才能满足基于多尺度的有效的综合分析和辅助决策

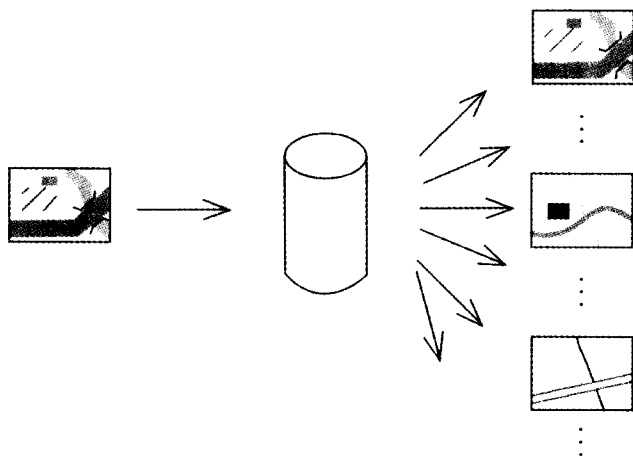


图 1-3 基于主导数据库建立多重表示模型