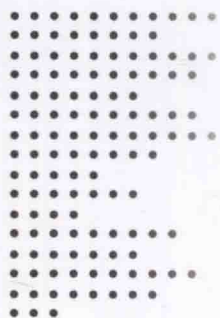




地球信息科学基础丛书

多尺度空间关系 理论与实践

◎ 杜世宏 著



科学出版社

国家自然科学基金项目(41171297) 联合资助
教育部“新世纪优秀人才支持计划”

地球信息科学基础丛书

多尺度空间关系理论与实践

杜世宏 著

科学出版社

北京

内 容 简 介

基于关系的空间信息处理是对基于几何的空间信息处理方式的一种有益补充,是地理信息科学、语言学、认知科学和人工智能等学科交叉的重要理论之一,在图像智能解译、空间数据挖掘、多源空间数据集成和空间场景知识表示等高级和智能化空间信息处理领域具有重要应用。本书简要介绍空间信息几何表现和关系表现与空间尺度的关系,重点阐述多尺度空间关系建模理论与应用,主要包括作者在拓扑关系多尺度建模理论、方向关系多尺度建模理论、定性位置多尺度建模理论与相似性问题,以及基于关系的多尺度数据分析的总体框架、多尺度数据查询、多尺度数据一致性检测应用等方面的研究成果。

本书可作为从事遥感与地理信息科学研究和应用的科技人员参考书,也可作为高校测绘、地理信息科学、遥感、计算机科学等相关专业的师生教学用书。

图书在版编目(CIP)数据

多尺度空间关系理论与实践 / 杜世宏著. —北京: 科学出版社, 2014.11
(地球信息科学基础丛书)

ISBN 978-7-03-042276-7

I. ①多… II. ①杜… III. ①空间信息技术—研究 IV. ①P208

中国版本图书馆 CIP 数据核字(2014)第 248216 号

责任编辑: 彭胜潮 苗李莉 李 静 / 责任校对: 朱光兰

责任印制: 徐晓晨 / 封面设计: 陈 敬

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>

北京厚诚则铭印刷科技有限公司 印刷

科学出版社发行 各地新华书店经销

*

2014 年 11 月第 一 版 开本: 787×1092 1/16

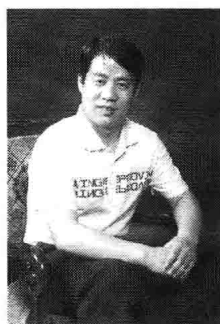
2014 年 11 月第一次印刷 印张: 18 7/8

字数: 436 000

定价: 89.00 元

(如有印装质量问题, 我社负责调换)

作者简介



杜世宏 北京大学遥感与地理信息系统研究所副教授、博士生导师。武汉大学资源与环境科学学院学士(1998)、测绘遥感信息工程国家重点实验室硕士(2001),中国科学院遥感应用研究所博士(2004),北京大学地球与空间科学学院博士后(2004~2006),英国剑桥大学地理系访问学者(2009~2010)。入选北京市科技新星计划(2009)和教育部新世纪优秀人才支持计划(2010)。主持国家自然科学基金面上项目、国家863专题项目、国家重大专项“高分辨率对地观测系统”子课题等项目多项。主要研究领域包括空间关系知识表达与推理、空间数据智能理解、时空统计分析等。已发表学术论文60余篇(SCI收录19),已出版专著3部,获得2008年度国家测绘科技进步二等奖。现担任中国地理信息产业协会理论与方法委员会委员、中国现场统计研究会空间统计学会理事、中文核心期刊《地理与地理信息科学》编委, *International Journal of Geographical Information Science*、*GeoInformatica*、*Transactions in GIS*、*Computers, Environment and Urban Systems*、*International Journal of Approximate Reasoning*、*GeoCarto International*、*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*、*Computers & Geosciences*、*PLOS ONE* 等国际期刊审稿人。

E-mail: dshgis@hotmail.com。

前 言

空间信息的多尺度表示与处理是当今空间信息科学领域理论与方法研究的重要前沿问题,主要解决由于不同部门用户对地理数据的不同应用需求和分析需要,而导致空间数据不同表示的问题。多尺度、多类型、多时态的空间数据是人类研究和解决人口、资源与环境等重大问题所必需的重要信息资源,是“数字奥运”“智慧城市”“智能交通”等应用的重要数据基础设施。

空间信息表示主要有基于几何和基于关系两种方式。一方面,传统的几何表现以笛卡儿/地理坐标系为参考,以几何坐标和属性为处理对象,侧重于表达地理要素的几何位置、几何形态和几何分布,采用高效的几何数据结构进行存储和管理,为空间信息的高效组织、管理、查询、几何分析和可视化提供服务。这种信息处理方式易于计算机操作,但与人类对空间信息的理解方式有一定的差异。另一方面,关系表现以人类对地理世界的认知和理解为基础,对各个地理要素的空间关系进行抽象和理解,并用一些关系符号进行表达,服务于空间知识表达、推理和理解,与人类对空间信息的表述和理解一致。几何表现和关系表现都是对地理世界的客观表示,但由于出发点不同,二者差异明显。几何表现重在对空间信息的精确表达、管理和几何分析,因此具有严格的坐标系统。而关系表现重在对关系概念的认知和表达,由定性符号系统及相关运算组成,因此基于关系的空间信息处理是对基于几何的空间信息处理的一种有益补充。在地理信息科学领域,几何表现方式经过长期发展,在理论、方法、技术和软件系统等方面都比较成熟,达到了实用化的程度;而关系表现尚处于理论发展阶段,未达到实用化阶段。由于关系表现是从几何表现转换而来的,几何表现的多尺度特征导致关系表现也是多尺度的。空间关系的多尺度表达无论在理论、方法和应用上都不成熟,急需开展相关研究,弥补该领域的不足。

基于这样的认识,本书在简要阐述和概括空间信息表现方式与空间尺度关系的基础上,重点阐述了作者近年来在多尺度空间关系方面的研究成果,主要包括多尺度拓扑关系建模、多尺度方向关系建模、多尺度定性位置建模和基于关系的多尺度数据分析等内容,并从理论、方法、技术和实验等方面进行了论述。

全书共有5章。第1章详细阐述空间信息的两种表现方式:几何表现和关系表现,以及它们与空间尺度的联系;第2章主要论述拓扑关系的多尺度建模理论和方法,概括总结拓扑关系的尺度问题及研究进展;第3章主要论述方向关系的多尺度建模理论,概括总结方向关系的尺度问题及研究进展;第4章全面阐述空间定性位置的多尺度建模理论和方法,包括基于拓扑的定性位置、基于方向的定性位置和多尺度位置相似性度量等内容;第5章全面总结基于关系的多尺度数据分析的总体框架及关键技术问题,包括多尺度数据匹配、多尺度关系构建、多尺度定性位置构建、多尺度数据一致性检测、多尺度数据查询,以及多尺度数据分析软件系统和分析实验等工作。

本书全部内容由杜世宏撰写、校正和定稿。在写作过程中,研究生张方利、郭舟、

王婉祎、锥立群、赵文智、张修远和熊紫倩帮助整理资料、绘制插图和校正书稿，在此表示感谢。

本书在研究和出版过程中得到了国家自然科学基金项目“多尺度空间关系建模与异常检测方法研究”（编号：41171297）和教育部“新世纪优秀人才支持计划”（编号：NCET-10-0189）的联合资助。在研究过程中，得到了龚健雅院士、王桥研究员、秦其明教授、曾琪明教授、邬伦教授、方裕教授、李琦教授、李培军教授、程承旗教授、毛善君教授、闫国年教授、李清泉教授、朱庆教授、杜清运教授、艾廷华教授、张晓东教授、陆峰研究员、吴立新教授、童小华教授、齐清文研究员的关心和指导，在此表示感谢。

由于时间仓促，作者水平有限，相关理论和技术还在不断发展和更新中，书中难免存在不妥之处，恳请读者批评指正。

杜世宏

2014年8月于北京大学

目 录

前言

第 1 章 空间信息表现与空间尺度	1
1.1 空间信息表现方式	1
1.2 几何表现与空间尺度	4
1.3 关系表现的基本问题	11
1.4 关系表现与空间尺度	16
1.5 本书内容组织	22
参考文献	22
第 2 章 拓扑关系多尺度建模	27
2.1 拓扑关系与尺度变换	27
2.2 维数退化与拓扑关系尺度建模	32
2.3 区域合并与拓扑关系尺度建模	39
2.4 属性归纳与拓扑关系尺度建模	41
2.5 形状化简与拓扑关系尺度建模	70
2.6 拓扑不变量尺度建模	92
2.7 本章小结	100
参考文献	101
本章数学符号	105
第 3 章 方向关系多尺度建模	106
3.1 方向关系与尺度变换	106
3.2 区域合并与方向关系尺度建模	110
3.3 维数退化与方向关系尺度建模	124
3.4 本章小结	136
参考文献	137
本章数学符号	139

第 4 章 空间定性位置多尺度建模	140
4.1 空间参照框架与空间位置	140
4.2 基于拓扑的多尺度位置模型	145
4.3 基于方向的多尺度位置模型	165
4.4 基于方向的通用位置模型	188
4.5 多尺度位置相似度量	208
4.6 本章小结	228
参考文献	228
本章数学符号	233
第 5 章 多尺度关系应用实践	235
5.1 多尺度关系分析总体框架	235
5.2 多尺度数据匹配	242
5.3 多尺度网络数据初匹配	248
5.4 多尺度网络数据优化匹配	255
5.5 多尺度离散数据匹配	260
5.6 多尺度数据查询	265
5.7 软件系统	267
5.8 多尺度数据分析实验	272
5.9 本章小结	292
参考文献	292

第 1 章 空间信息表现与空间尺度

空间信息多尺度表示与处理是当今地理信息科学领域理论与方法研究的重要前沿,主要解决由于不同部门用户对地理数据的不同应用需求和分析需要,而导致空间数据的不同表示问题。多尺度、多类型、多时态的空间数据是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,是“数字奥运”“智慧城市”“智能交通”等应用的重要数据基础设施。

空间信息服务已成为空间信息技术社会化、实用化的核心问题,其典型代表是以 Google Earth 为代表的多尺度空间数据网络服务、多尺度电子地图自适应可视化等平台。空间信息服务的主要问题有:①不同行业领域对空间数据详细程度的需求不一样,如何为用户提供多种尺度或多种详细程度的空间数据?②如何统一管理多种尺度或多种详细程度的数据?③如何保证不同部门或行业生产的多尺度空间数据的一致性?④如何在多个尺度上查询和分析不同详细程度的空间数据?第一个问题是多尺度空间数据的生产问题,第二个问题是多尺度数据建模和管理的问题;第三个问题是多尺度数据的一致性问题;第四个问题是多尺度数据查询、信息提取和匹配分析问题。这四个问题构成了多尺度数据的生产、管理、分析和应用等各个阶段。到目前为止,针对前两个问题已有较多研究和相关应用软件,本书主要解决第三个和第四个问题。

1.1 空间信息表现方式

空间信息表现方式是空间信息科学的基础,主要涉及如何把地理世界中各种地理要素的几何、属性,以及空间关系等重要信息经过认知、加工和抽象后,科学地加以组织和表达,以利于空间信息分析处理。常用表现方式有两类:面向计算机的几何表现及面向认知的关系表现。

1.1.1 表现方式

如图 1.1 所示,空间信息可从两个方面进行表示。一方面,为有利于计算机精确表现地理世界,必须先定义坐标系统,然后基于坐标系表达地理要素的几何位置,采用高效的几何数据结构进行存储和管理。这种表现方式侧重地理要素的几何位置、几何形态和几何分布的表示,因此称为几何表现方式。另一方面,人类基于自己对地理世界的认知和理解,对地理要素的空间关系进行认知和抽象,用一些关系符号表达,并进行交流和推理。这种表现方式侧重地理要素关系表达,并对每种关系赋予一个词汇或符号表示,因此称为关系表现方式。

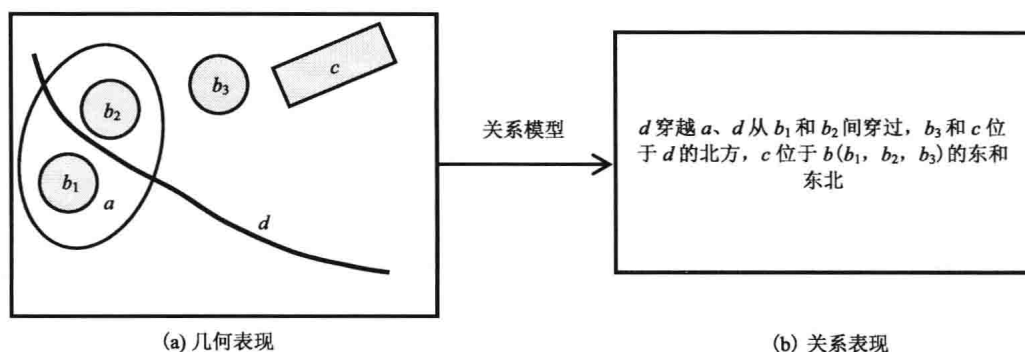


图 1.1 空间信息表现方式

如图 1.1 所示,几何表现把地理场景抽象为由面对象 a 、 b (由 b_1 、 b_2 和 b_3 组成)和 c ,以及线对象 d 组成的对象集合,并完整记录了每个对象的几何坐标位置。几何分析可从对象几何位置进一步进行面积、长度量算,以及分布模式推导等。关系表现则把对象间的空间关系表示为一些自然语句的集合,方便人类理解,如“ d 穿越 a 、 d 从 b_1 和 b_2 间穿过, b_3 和 c 位于 d 的北方, c 位于 $b(b_1, b_2, b_3)$ 的东和东北”等。这里,“穿越”“穿过”“北方”和“东北”等就是空间关系符号表示。

几何表现和关系表现都是对地理世界的客观表示,但由于出发点不同,因此差异明显。几何表现重在空间信息的精确表达、管理和几何分析,因此必须严格定义空间坐标系,并基于坐标系进行描述和分析,但与人类的认知不符。关系表现重在关系概念的认知和表达,因此由定性符号系统及相关运算组成。关系表现不是一种精确表达方式,但人类有足够的进行推理、理解和交流关系符号所表达的空间信息。

尽管几何表现和关系表现间差异明显,但它们都是重要的空间信息表现和处理方式,互为补充。高层次的空间分析和理解不仅需要几何信息,更需要关系知识。例如,制图综合中,形状化简前后的数据需要保持拓扑一致性。因此,需要在几何表现和关系表现间建立联系。空间关系模型能根据对象类型,把几何数据转化为关系表示。不同的关系模型可以把相同的几何数据转换为不同的关系符号,并进行关系推理运算。

1.1.2 几何表现

为表达地理世界,几何表现首先定义了空间坐标系,然后把地理对象表示为空间坐标系中点的集合。由于每个点都具有坐标,因此可测量面积、长度等定量几何信息。对于多源数据,只要具有相同的坐标系,就可进行集成、融合、匹配、管理和分析。

(1)空间坐标系。GIS 坐标系是笛卡儿坐标系,由一个原点和垂直的坐标轴组成。GIS 中常用坐标系主要包括平面直角坐标、地理坐标系等。地球椭球面上任一点的位置,可由该点的纬度(B)和经度(L)确定,即地面点的地理坐标值,由经线和纬线构成两组互相正交的曲线坐标网称为地理坐标网。由经纬度构成的地理坐标系又称为地理坐标系。根据地图投影方法,可把地理坐标系转为平面直角坐标系。

(2)空间数据模型。简单地讲,空间对象为点的集合。但由于地理要素和地理应用

的复杂性,不同要素、不同应用需要不同的模型表达对象。通常可分为矢量模型和栅格模型。矢量模型是一种基于对象的表达方式,完整准确表达了每个对象的几何位置、形状和形态特征。栅格模型是一种基于像元的表达方式,用具有一定大小的像元近似表示空间对象,每个像元可具有属性。影像是一种常用的栅格数据,每个像元用光谱表达地物的信息。由于影像没有显式表达每个对象及其属性,因此需要对影像数据进行分析理解,转化为基于对象的表示。

(3)空间数据管理。几何表现将空间场景表示为对象(点、线、面)的集合,并采用空间索引技术(格网索引、四叉树索引、R 树等)对空间对象进行高效组织,以帮助管理和快速查询。在此基础上,对每个对象进行符号显示和制图,或者进行三维场景虚拟显示。无论是空间数据索引或可视化,都是在空间坐标系支持下,以空间对象的几何坐标、位置、形态为基础的。

(4)空间几何分析。几何表现把空间对象表示为点、线、面以及三维实体等基本结构。基于这些对象结构,可分析对象的基本信息,如空间位置(基于空间坐标系传递空间对象的定位信息)、空间分布(同类空间对象的群体分布信息,包括分布、趋势、对比等)、空间形态(空间对象的几何形态)和空间距离(空间对象的接近程度)。把空间对象的几何数据和属性数据结合,可进行几何分析和量算,如缓冲区分析、叠加分析、网络分析等,再从空间数据中提取有用信息。

1.1.3 关 系 表 现

关系表现首先基于空间关系模型,把几何表现中的坐标数据转化到概念空间中的关系语言,然后基于关系语言进行关系知识的表达、管理、推理及场景分析。

(1)空间关系定位。计算机是基于笛卡儿坐标系进行对象定位的,而人类是基于关系进行对象定位的。例如,维基百科把太行山位置描述为“跨越北京市、河北省、山西省、河南省4省(市);北起北京西山,南达豫北黄河北崖,西接山西高原,东临华北平原”。基于关系的定位把目标对象(太行山)和参照对象(用地名表示)通过空间关系关联来进行定位。这种定位方式显然没有笛卡儿坐标系准确,但人们更习惯用基于关系的位置表达。

(2)空间关系模型。人类对关系的表达和抽象是自动进行的,为使计算机也具备空间关系处理能力,必须在几何表现和关系表现间进行转换。空间关系模型是一种数学模型,根据空间对象的几何特征,从对象的几何表现获得它们的关系符号描述。其中,空间对象间存在多种类型空间关系,包括:拓扑关系、方向关系、邻近关系、距离关系等。针对不同类型关系,有不同的空间关系描述模型,如拓扑关系的九交模型和方向关系的方向关系矩阵模型等。

(3)空间关系管理。关系表现将空间场景仍然看做对象集合,用一个关系网络来表达。在关系网络中,对象是没有大小和长度的,只是表示为网络的节点;而对象关系表示为节点间边的属性。因此,关系网络可管理和表达空间场景中所有对象的关系知识。基于关系网络,可进行空间推理、场景一致性和相似性分析等工作。

(4)空间关系分析。空间关系是定性概念,因此不能进行量化分析。距离是分析的基础,关系间的距离是通过概念间的定性距离来度量的。不同类型对象(如线/线、线/面、面/面)、不同类型关系(如拓扑关系、方向关系)可用不同概念邻居图来表达任意两个关系的距离。基于两个关系的距离,可定义和度量两个关系网络的相似性,从而进行场景相似性分析。

表 1.1 总结和比较了几何表现和关系表现的特点。

表 1.1 几何表现和关系表现

	几何表现	关系表现
定位方式	笛卡儿坐标系	基于关系的定位
定位性质	绝对定位	相对定位
定位精度	量化、精确	定性、近似
对象表现	几何表现(点、线、面等)	关系知识表达、推理
空间分析	几何计算、欧式距离等	概念距离、关系距离、场景相似
信息度量	定量处理(面积、长度)	定性概念表达

1.2 几何表现与空间尺度

地理世界中的各种地理现象和过程都可经过人类认知、抽象和建模,表达为空间几何数据和属性数据。地理信息系统主要用于存储、检索、操作、查询和分析空间几何数据(如点、线、面、空间划分和网络等),以及属性数据,为空间问题或地理问题解决和决策提供服务。几何表现提供了基本的数据类型来抽象和表达地理现象的几何结构、关系、属性,以及相关操作。因此,基本数据类型的定义对空间数据表达和管理非常重要。

但由于地理世界的复杂性、人类认知能力的有限性,以及分析应用需求的多样性,同一地理现象需要在不同尺度上进行抽象和表达。度量或观测的空间尺度在很大程度上决定了哪些现象和过程要表达、如何表达,以及哪些现象要舍弃的问题。因此, GIS 或空间数据库包含了地理现象在一定和有限尺度上的表达。为在多个尺度上创建、管理、制图和发布地理现象的几何表达,这就需要管理空间对象的多尺度表现。

1.2.1 几何表现模型

从对象建模的角度看,主要有两种模型:基于对象的(entity-oriented / feature-based view)和基于场的模型(space-oriented / field-based view)。基于对象的模型主要建模和表达位置或范围明确的地理要素,它们在空间上占据一定范围,并且是离散的(如道路、建筑物等)。基于对象的模型主要由三个要素组成:①空间位置;②非空间属性(如名称、类型等);③运算(如对于道路而言,其操作运算有确定道路中心线、确定道路长度等)。空间对象一般指地理对象的空间属性,在 GIS 和空间数据库中也称为空间数据类型。

空间对象有单个对象类型和集合类型之分。

(1) 简单点对象(图 1.2(a)), 用坐标系中的一个点表示, 主要建模和表达对象在空间的位置, 而不表达对象的范围(如水井、电线杆、城市等)。

(2) 简单线对象(图 1.2(b)), 用一系列坐标串表示, 不能自相交, 且只有 2 个端点。主要表达对象在空间上的连通性, 或者对象在空间上的运动(如河流、道路、电力线、路径等)。

(3) 简单面对象(图 1.2(c)), 指由首尾相同的一系列坐标串围成的封闭区域, 主要表达对象的空间范围(如湖泊、城市、林地、农地等)。简单面对象是连通的, 且不能包含岛屿。

(4) 复杂点对象(图 1.2(d)), 由多个点对象组成, 表示一些复杂的对象。

(5) 复杂线对象(图 1.2(e)), 用一系列坐标串表示, 但允许自相交, 端点个数可多于 2 个。

(6) 复杂面对象(图 1.2(f)), 由多个简单面对象组成, 因此可以不连通, 或包含岛屿。

(7) 空间划分集合类型(图 1.2(g)), 主要指把平面空间划分为分离的面对象, 每个面对象都有一个属性, 且可表示为简单或复杂的几何类型。一般而言, 划分类型中的每个元素不允许出现带岛屿的面对象。例如, 每个国家的行政区域划分、土地利用分类图等。

(8) 空间网络集合类型(图 1.2(h)), 主要指由离散对象及其网络连接关系组成的集合。例如, 道路网络、水系网络等。

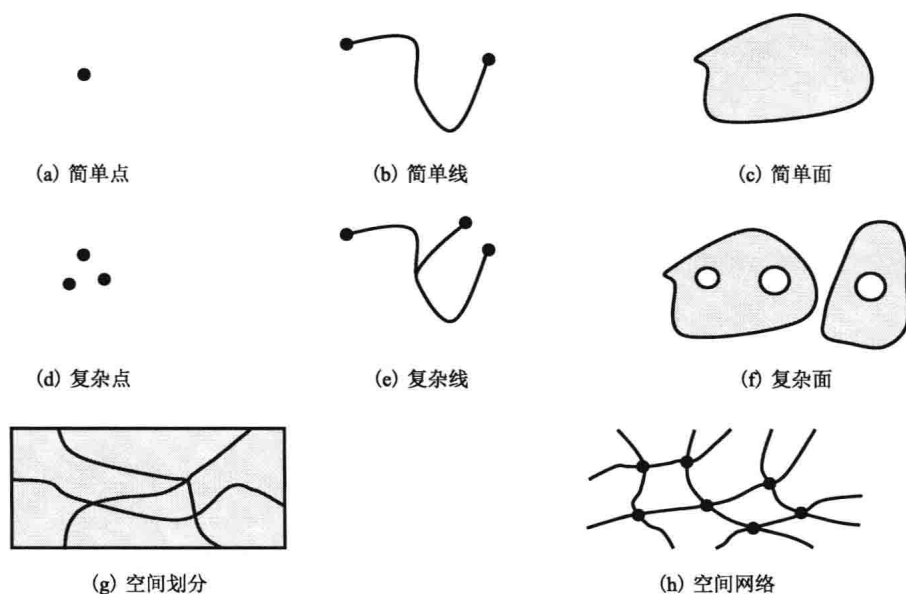


图 1.2 空间数据类型

通常基于对象的模型用矢量数据结构实现, 而基于场的模型主要建模和表达在空间上连续分布的地理要素(如温度场、气压场), 通常用栅格数据结构实现。

1.2.2 几何表现与空间尺度

多尺度数据建模和管理是多尺度应用的核心,也是当前 GIS 领域研究的主要前沿问题。美国国家地理信息与分析中心(NCGIA)从 1988 年就开始关注并支持研究空间数据的多尺度表达问题(Buttenfield, 1995)。UCGIS(1996)认为未来的 GIS 是尺度依赖的,但对用户来讲是尺度敏感的。为对同一数据库中相同地理对象的多尺度表达进行管理,欧洲多家研究机构共同发起,由欧共体资助 MurMur(Multiple Representations, Multiple Resolutions)项目,并从 2001 年开始实施(Spaccapietra et al., 1999)。国际摄影测量学会(ISPRS)数据综合与数据挖掘工作组和国际制图协会(ICA)地图综合委员会也于 2002 年 7 月在加拿大渥太华以“空间数据多尺度表示”为主题召开学术会议,研讨与多尺度表示有关的基础理论与应用问题。但 GIS 的尺度问题,尤其是多尺度数据查询与分析问题,远未达到人们所期望的程度(李霖和吴凡, 2005; 王艳慧等, 2006)。

1. 尺度概念

在 GIS、地理学和环境科学领域,尺度有多种含义(Goodchild, 2001)。例如,对于制图学家来讲,尺度是指地图上单位长度与其代表现实世界实际长度的比率;对于环境学家来讲,尺度是指空间范围或分辨率。因此,大尺度在不同领域含义是不同的。例如,对于制图学家,大尺度是指详细细节;而对于环境学家,大尺度是指粗略细节。

地理学领域有关尺度的具体含义解释如下。

(1)空间粒度或分辨率(spatial grain / resolution),指像素的大小或最小的可分辨单元。这种情况下,只记录了单元上的总体信息,而忽略了单元内的信息。例如,ETM 影像的空间分辨率是 30m, IKONOS 全色波段分辨率是 1m。

(2)空间范围(spatial extent),指研究区域的大小或最大的可分辨单元。例如,地理学研究中的全球尺度、区域尺度、城市尺度,以及城市内的小尺度和微观尺度等。这里的尺度主要指研究范围,并且当研究范围不同时,地理现象的结构组成和地理规律都会发生变化(鲁学军等, 2004)。例如,大尺度研究主要关注地带性规律,中尺度研究主要关注地理景观类型单元组成要素的作用关系,而小尺度或基本尺度研究关注景观单元内各基本功能单元的要素组成及要素关系。从空间范围的角度看,尺度越大,数据越粗略;尺度越小,数据越详细。

(3)制图尺度(cartographic scale),指地图上单位长度与其代表的现实世界实际长度的比率。对于纸质地图,其精度是有限制的。例如,由于线宽和符号大小的限制,纸质地图上很难显示小于 0.1mm 的地物,这意味着 1 : 50000 比例尺地图的精度为 5m。然而,对于数字地图,则没有这样的限制。从制图角度看,尺度越大,数据越详细;尺度越小,数据越粗略。

(4)时间尺度(temporal scale),指获取数据的时间频率。例如,SPOT 卫星的重访周期为 26 天, WorldView 卫星的平均重访周期 1.7 天, IKONOS 的平均重访周期为 3 天。

(5)生态尺度(ecological scale),指生态现象或过程发生的本征尺度。

(6)分析尺度(analytical scale),指地理或空间分析模型中的尺度。例如,空间点模式在不同尺度上往往表现出不同的规律, K 函数可以分析点模式在不同距离尺度上的规律(Dixon, 2002);不同邻域或距离尺度对空间自相关/交叉相关分析结果有重要影响,因此基于多种权重矩阵的分析,有助于揭示自相关在不同尺度上的规律(冯昕等,2011)。

上述几个尺度概念中,空间分辨率、制图尺度、时间尺度是关于数据的尺度,空间范围和分析尺度是关于数据分析和建模方法的尺度,而生态尺度是关于地理现象、过程和格局的本征尺度。本书主要涉及空间信息的几何表现,因此书中所涉及的尺度为制图尺度,尤其是数字环境下的制图尺度。因此,大尺度或详细尺度意味着更多的空间细节,而小尺度或粗略尺度意味着更少的细节。

空间信息需要多尺度表示,这是由多种因素决定的:①地理现象是复杂多变的,而人类认知和计算机处理能力是有限的,因此空间信息总是在特定尺度上对地理世界的抽象表示;②由于空间信息主要由纸质地图数字化、遥感影像解译或其他手段采集而来,而纸质地图的比例尺、遥感影像的分辨率,以及数据采集尺度/密度,这些都属于某种形式的尺度,因此导致获取的空间信息必然是多尺度的;③地理现象本身是有尺度的,为分析地理现象,必须采用与地理现象的尺度相匹配的数据,尺度过大或过小的数据都不利于地理现象的分析,因此导致空间信息的多尺度性。总而言之,由于原始数据的空间分辨率(遥感影像)或制图尺度(纸质地图)、地理现象的过程尺度,以及地理分析需求,需要对同一地区、相同地理现象进行多尺度表现。

2. 尺度管理

由于GIS对地理世界采取了多尺度表示,因此如何管理多尺度数据,为地理分析提供服务非常关键。从空间信息几何表现的角度来看,同一地理现象在不同尺度上的几何表现,应该具有某种关联性,即不同尺度表现间应该满足一些约束准则(具体包括拓扑、度量,以及视觉感知上的约束关系)。

目前主要有三种多尺度数据表示与管理方法(王涛和毋河海,2003;李霖和吴凡,2005;Tryfona and Egenhofer, 1997)。

(1)同一地区的多尺度数据按照多重副本独立存储。这种方法不需要对几何数据模型和数据库管理重新设计,只需按照已有几何数据模型组织数据,并对多尺度数据分库存储、管理,因此方法简单且便于实现。获取、查询和显示多尺度数据时,按照用户需求或事先设定的原则,查询和调入某种尺度的数据。但由于各个尺度版本的数据是独立存储和管理的,因此,相同地理实体的多尺度几何表示之间缺乏逻辑联系。为满足多尺度数据分析,应建立不同尺度数据间的联系,需用多尺度数据匹配方法自动实现(Saalfeld, 1998; Cobb et al., 1998; Walter and Fritsch, 1999; 童小华等, 2007)。由于不同尺度数据在几何位置、空间关系、获取时间、几何表现上均存在差异,匹配结果往往是不确定的。由于多尺度数据分版本管理,因此在建立不同版本逻辑联系时,会遇到困难。另外,如何保证不同尺度数据间的一致性是个难题。

(2)数据库中存储最详细的单一尺度数据作为基准,然后利用制图综合方法,从基准数据中自动推导多尺度数据。这种方法只维护、管理和更新一个尺度的数据,却可实现多尺度应用。优点在于可推导出用户需要的各种粗略尺度数据,不存在多版本、多库重复存储的问题,数据管理简单、无需维护不同尺度数据间的一致性。但这种模式受限于制图综合自动化的程度。

(3)基于层次结构的多尺度数据模型。基于单一详细尺度数据进行自动综合获取粗略尺度数据的方式受制图综合技术限制,尚无法自动实现;而多尺度数据分库存储的方法,无法建立不同尺度数据间的联系。因此,一个可行的方案是事先生成多尺度数据,基于层次结构索引技术,组织和存储同一地理实体不同尺度表现间的联系。这样既确保了多尺度数据的可获得性,同时也保证了多尺度数据间的关联性。

经过国内外长期努力和发展,制图综合方法取得了重要进展。例如,分形制图综合(王桥,1998)、基于特征的综合(齐清文和刘岳,1998)、渐进式综合(郭庆胜和任晓燕,2003)、基于三角网的综合(艾廷华,2000)、基于遗传算法的综合(王家耀和邓红艳,2005;武芳等,2008)、专题属性综合(杜道生等,2003)等。但距全自动化还相距甚远(艾廷华和成建国,2005),因此导致多尺度数据仍然分库存储、显式存储和管理同一地理实体在多个尺度下不同表现(艾廷华和成建国,2005;尹章才等,2007;Stefanakis,2003;Neun and Steiniger,2005)。同时,由于维护多尺度数据间的综合过程和联系的代价太大,因而多尺度数据库中并没有存储多尺度数据间的联系,仍然需要在分析时自动建立多尺度对象的匹配对应关系。

1.2.3 几何表现与制图综合

制图综合主要对详细尺度的空间数据进行空间综合(如选取、合并、聚合、形状化简等)和专题属性综合,自动生产粗略尺度的空间数据,是多尺度空间数据生产的核心。McMaster 和 Shea(1992)概括总结了12种综合算子:化简(simplification)、光滑(smoothing)、聚合(aggregation)、融合(amalgamation)、合并(merging)、退化(collapse)、优化(refinement)、典型化(typification)、夸大(exaggeration)、增强(enhancement)、移位(displacement)和归纳(classification)。由于篇幅限制,主要介绍本书涉及的几个算子。

1. 选取运算

选取运算保留主要的要素实体,而舍弃次要的要素实体,以保证与成图比例尺相适应的要素数,满足用途要求的详细性,并反映地理实体的分布特点和密度对比。

选取包括两个方面的内容:①要素类选取,即选取规定的要素类,舍去不重要的要素类,结果是要素种类减少,如1:10000比例尺数据中的“土地庙”,在1:50000比例尺中被删除;②空间对象的选取,即在保留的要素中,进一步保留主要的、重要的对象,而舍去次要的对象,其结果是对象个数减少,如1:10000数据综合为1:50000时,长度小于1cm(地图距离)的河流可被舍弃。

地物要素选取的基本原则:选取的优先次序应遵循从高级到低级,从主要到次要、

从大到小、从整体到局部的原则。

地物要素选取的基本方法有如下3种。

(1) 资格法, 即按规定的质量标志或数量标志进行取舍。凡达到指标就选, 达不到指标就舍去。质量标志包括等级、品种、性质等, 数量标志包括线状对象的长度、面状对象的面积等。资格法的优点是目标明确, 操作简单易行。缺点是一个标志不能全面衡量对象的重要性。例如, 小湖泊在不同地区, 其重要程度不同。资格也无法预计选取后的地图容量, 很难掌握选取后各部分图面载负量。

(2) 定额法, 即规定单位面积内应选取对象的数量, 如规定单位面积内居民地个数、高程点个数等。定额法的优点是保证地图具有丰富的内容而又不失去易读性, 缺点是不能保证在不同地区具有相同质量指标。为弥补这种缺点, 可调整不同区域内选取的质量标准, 以及与相邻区域分布密度的逐渐过渡。

(3) 组合选取法, 在实际生产作业中, 往往是把基本的选取方法组合起来进行, 常用组合方法有按定额指标与分界指标的组合选取法、按定额指标与地物等级的组合选取法。

图 1.3 中, 在空间尺度粗化过程中, 过短的河流和过小的居民地都被舍弃了, 而只保留了较长的河流和较大的居民地。

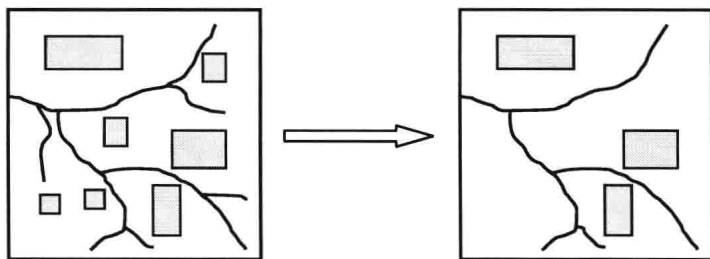


图 1.3 要素选取

2. 合并运算

聚合、融合和合并这三个算子的相似之处在于它们把多个相似的对象组合在一起, 并合并成一个对象; 不同之处在于它们所作用的对象维数不同。聚合运算主要搜索一群非常接近的点, 并把它们表示为一个连续的面。融合运算作用于面对象, 把多个面合并为一个面。主要有两种融合运算: 拓扑邻近和视觉邻近面对象融合。合并算子则把多个平行或接近的线合并为一条线。

图 1.4(a) 中, 详细尺度上两组点群内的点距离接近, 形成聚团; 当空间尺度缩小时, 点群内点的距离过小, 不能区分, 而两组点群间可以区分, 因此粗略尺度上两组点群分别被聚合为面对象。图 1.4(b) 中, 两个拓扑邻近且属性值相似的面被融合为一个面。图 1.4(c) 中, 三个视觉邻近的面被融合为一个面。图 1.4(d) 中, 当尺度减少一倍后, 线间距缩小, 不能很好区分, 因此将接近的两条线合并为一条。

尽管聚合、融合和合并这三个综合算子处理不同的对象, 但它们经常被统一称为合并算子, 因为它们都是把多个对象合并为一个。