21 世纪高等院校教材

GIS 空间分析原理与方法

刘湘南　黄　方
王　平　佟志军　编著

科学出版社
北　京

内 容 简 介

本书共分 9 章。第 1 章和第 2 章探讨了 GIS 环境下空间分析的基本框架和基础问题;第 3 章至第 7 章主要阐述了空间量测与计算、空间表达变换分析、空间几何关系分析、空间统计学分析,以及三维分析技术;第 8 章介绍了网格计算技术及其对 GIS 空间分析的影响;第 9 章结合地理空间数据的不确定性问题,探讨了智能计算技术的基本原理、方法体系及其应用于地理空间数据分析的实例。

本书可作为地理、遥感、地理信息系统专业本科生和研究生教材,也可供相关专业高等院校师生和科研技术人员参考。

图书在版编目(CIP)数据

GIS 空间分析原理与方法/刘湘南等编著. —北京:科学出版社,2005.7
(21 世纪高等院校教材)
ISBN 7-03-015499-1

Ⅰ.G… Ⅱ.刘… Ⅲ.地理信息系统-高等学校-教材 Ⅳ.P208

中国版本图书馆 CIP 数据核字(2005)第 045592 号

责任编辑:郭 森 杨 红/责任校对:朱光光
责任印制:张克忠/封面设计:陈 敬

科学出版社发行 各地新华书店经销

*

2005 年 7 月第 一 版 开本:B5(720×1000)
2006 年 5 月第二次印刷 印张:21 1/2
印数:3 501—5 000 字数:400 000

定价:32.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

前　言

地理信息系统(GIS)是地理空间数据处理、分析的重要手段和平台。从诞生至今的40多年时间里,地理信息系统技术的日益革新为众多应用领域创造了丰富的地理空间信息财富,使地理空间数据的存储、检索、制图和显示功能越来越完善。地理信息系统的核心功能是空间分析。空间分析使GIS超越一般空间数据库、信息系统和地图制图系统,不仅能进行海量空间数据管理、信息查询检索与量测,而且可通过图形操作与数学模拟运算分析出地理空间数据中隐藏的模式、关系和趋势,挖掘出对科学决策具有指导意义的信息,从而解决复杂的地学应用问题,进行地学综合研究。GIS的奠基人之一Goodchild M F曾指出"地理信息系统真正的功能在于其利用空间分析技术对空间数据的分析"。

虽然GIS已取得了巨大的发展,但目前多数地理信息系统的应用还局限于数据库型GIS的层面,多维信息空间分析能力的不足及空间分析的结果不实用越来越明显,无法真正满足全球变化和区域可持续发展研究对空间数据分析、预测预报、决策支持等多方面的应用要求。同时,由于空间分析内容十分繁杂,GIS学术界对空间分析的理解和认识存在较大的模糊性和差异,GIS空间分析一直滞后于空间数据结构、空间数据库、地图数字化和自动绘图技术等方面的研究,在理论和技术上尚没有根本的突破。本书在参阅了国内外大量相关文献的基础上,结合我们的工作实践,试图从新的角度理解空间分析的概念内涵,不拘泥于基本分析方法的概念、算法原理及其应用,选择性地引入了许多新的空间分析技术方法,如网格计算、智能计算等,力求比较系统地论述GIS空间分析的原理方法和技术及其发展前沿,为地理学及相关专业的本科生、研究生和广大读者提供全面、详细了解和掌握GIS空间分析理论与方法的途径,并为进一步完善GIS空间数据分析理论和方法体系奠定基础。

全书共分9章。第1章简要回顾了20世纪50年代以来地理空间数据处理与建模领域,如数量地理学、地理信息系统、地理计算、地理空间数据挖掘等重要的技术方法及其进展,探讨了GIS环境下空间分析的基本框架。第2章主要概括了空间分析的基础问题,包括地理空间的理解方式,地理空间坐标系统的建立方法,地理网格系统、地理空间数据的特征及基本的地理空间问题等。第3章介绍了从GIS中获取地理空间目标基本参数的方法,即空间量测与计算。第4章主要阐述了空间数据结构、空间参考系统、时空尺度和图形表达等不同GIS的空间表达方式,提出空间表达变换不仅是空间数据操作的一种手段,更是GIS空间分析的重

要方法的观点。第 5 章详细地介绍了空间几何关系分析方法,即如何从 GIS 目标之间的空间关系中获取派生信息和新知识的有关分析技术,主要包括邻近度分析、叠加分析、网络分析等。第 6 章讨论了空间统计学的基本原理方法及其在地理空间数据挖掘和分析中的应用。第 7 章介绍了三维地理空间数据分析方法,主要包括三维景观建模、三维景观分析与计算,以及三维可视化表达技术等。第 8 章主要阐述了新一代 Web 技术——网格计算技术的基本特点,网格计算技术对 GIS 空间分析的影响,以及网格 GIS 基本概念和关键技术等内容。第 9 章结合地理空间数据的不确定性问题,主要探讨了智能计算技术的基本原理和方法体系,介绍了智能计算方法应用于地理空间数据分析的若干实例。

本书在编写过程中参考和吸取了近年来国内外诸多专家和同行的研究成果,在此表示诚挚的感谢。王静、任春颖、史晓霞、罗智勇、关丽、廖晓玉、吴雨航、邹滨等研究生参与了部分章节的编写工作;于思扬、付博、黄猛、许淑娜、曾文华、李世熙等研究生协助查阅了大量参考资料;本书得到了东北师范大学教材建设项目的支持,在此一并表示衷心的感谢。

《GIS 空间分析原理与方法》一书虽然酝酿准备了近三年,但在实际撰写中,原有的一些想法和内容没有得到充分的表达和展现,由于作者水平有限,书中不足之处敬请专家和读者批评指正。

目　　录

第 1 章　地理空间数据分析与 GIS

地理信息系统技术的日益革新为众多应用领域创造了丰富的地理空间信息财富,使地理空间数据的存储、检索、制图和显示功能越来越完善,但同时越来越多的复杂应用问题也对 GIS 产生了更多新的要求。各种类型的 GIS 中存储了海量的地理空间数据,且数据还在以指数级方式不断增长,迫切需要高效、精确、科学地分析这些数据,以找出数据所蕴涵的寓意,进而了解事物的性质与规律,为科学决策提供必需的信息。所以,开发一些工具来进行一般性地理空间数据分析和复杂的地理空间对象模拟,以将数据"点石成金"是一项艰巨而又紧迫的任务。因此,GIS 领域由原来重点关注数据库创建和系统开发建设,逐渐转向重点关注空间分析和空间建模。事实上,GIS 本身就是空间数据分析技术的重要组成部分和有效依赖平台。GIS 的奠基人之一 Michael F G 曾指出:"地理信息系统真正的功能在于它利用空间分析技术对空间数据的分析"。空间分析使 GIS 超越一般空间数据库、信息系统和地图制图系统,成为不仅能进行海量空间数据管理、信息查询检索与量测,更可通过图形操作与数学模拟运算分析出地理空间数据中隐藏的模式、关系和趋势,挖掘出对科学决策具有指导意义的信息,从而解决复杂的地学应用问题,进行地学综合研究的技术系统。

然而,目前多数地理信息系统的应用还局限于数据库型 GIS 层面上,没有充分利用和开发 GIS 的空间分析功能,不能真正满足全球变化和区域可持续发展研究对空间分析、预测预报、决策支持等多方面的应用要求,GIS 空间分析功能偏弱已经严重地阻碍了其作为空间数据分析工具和决策支持系统的应用。因此,建立完善的空间数据分析理论和方法体系,集成先进的空间数据分析工具,增强 GIS 的空间分析能力,使数据库型 GIS 上升为分析型 GIS,是 GIS 技术与应用的发展目标和趋势。本章首先对 20 世纪 50 年代以来地理空间数据处理与建模领域重要的技术方法如数量地理学、地理信息系统和地理计算等进行简要的回顾,然后论述数据分析领域中迅速发展的新技术——数据挖掘,在此基础上,讨论 GIS 环境下空间分析的基本框架。

1.1　地理空间数据处理与建模

地理空间数据分析是地理学和地理信息科学领域的重要研究内容,它通过研究地理空间数据及其相应分析理论、方法和技术,探索、证明地理要素之间的关系,

揭示地理特征和过程的内在规律和机理,实现对地理空间信息的认知、解释、预测和调控。长久以来,人们一直不懈地致力于研究和探索高效的、适合于地理空间数据处理与分析的方法,从对地理现象及其空间关系的文字记载,到利用数学概念和方法进行解释性描述;从传统统计学方法和数学模型对地理现象和过程的模拟,到基于地理信息系统的多维地理空间数据表达与管理、地理过程的动态模拟、可视化分析和决策支持;从空间数据挖掘技术到高性能计算技术支撑下的地理计算方法,随着人们对信息需求水平的不断提高和科学技术的日益进步,地理空间数据分析的技术和方法得到不断完善和丰富。

1.1.1 数量地理学

数量化方法在感知、认识和解释现实世界的各种自然、人文、社会现象过程的相互关系中起着定性方法不能替代的作用。对于决策者而言,数量化方法是获取更为合乎理性、可信、有效决策信息的重要手段。它能够以多种方式,从多重侧面,详尽、准确地解释事物的状态特征和演变过程,合理推测未来发展规律,提供可供选择的多重决策。

马克思早就指出:“一种科学只有在成功地运用数学时,才算达到了真正完善的程度”。数量地理学(Quantitative Geography)又称计量地理学或地理数量方法,是应用数学思想方法和计算机技术进行地理学研究的科学。它试图以定量的精确判断来弥补定性文字描述的不足;以抽象的、反映本质的数学模型去刻画具体的、庞杂的各种地理现象;以对过程的模拟和预测来代替对现状的分析和说明;以合理的趋势推导和反馈机制分析来代替简单的因果关系分析。数量地理学提供了理性的复杂方法以传递有关诸如行为、决策的确定性程度、综合研究精度等有用的信息,与定性研究方法结合共同构筑了地理学研究方法的科学体系。数量地理学是对地理学传统研究方法的发展和变革,反映了地理学向定量化、科学化发展的趋势,使地理学由一门对地表事物进行解释性描述的学科,转变为一门进行确定性解释的科学。数量地理学是地理学领域中最先采用数学原理方法来探讨地理数据分析处理与建模的学科。

1. 数量地理学的产生与发展

地理学是一门研究地球表层自然要素与人文要素相互作用关系及其时空规律的科学。作为一门古老的空间科学,地理学与数学有着不解之缘。在古代,地理学与数学之源泉科学——几何学,几乎都是研究地表的科学,如运用几何学原理和方法测算河流长度、山体高度、土地面积等。古希腊学者、西方“地理学之父”艾拉托塞尼(Eratcsthenes)最早运用几何学原理和方法测算了地球的周长。在近代地理学时期,经济学中的区位论被移植到地理学中,开辟了地理学运用分析数学之先

河。20 世纪 20~ 30 年代,地理学研究中统计方法开始萌芽,主要采用一般的数理统计,进行地理要素的统计概括和相关关系探讨。前苏联地理学家马尔科夫指出:“更多的地理学家应当使主要的研究方向现代化,应当偏重于以基础科学、首先是精确性科学为基础的道路。”

现代地理学中的数量方法与理论模式的产生与形成,可以追溯到 20 世纪 50 年代末期开始的计量运动。计量运动主要由美国地理学家发起,早期集中在依阿华、威斯康星、普林斯顿和华盛顿等几所大学。不同学者所持观点不同,研究方向各异,由此形成了所谓的经济、统计、社会等学派。从世界范围看,计量运动的兴起首先要归功于加里森(Garrison W L)及其领导的华盛顿小组。加里森是第一个把地理学的理论和方法建立在定量基础上的倡导者和实践者,也是第一本《计量地理学》教材的作者。作为地理科学的方法论之一,数量地理学尽管历史不长,但发展速度很快,且时时充满着变革和创新。从 20 世纪 50 年代末开始,数量地理学先后经历了三个发展时期,各自呈现不同特征(图 1.1)。

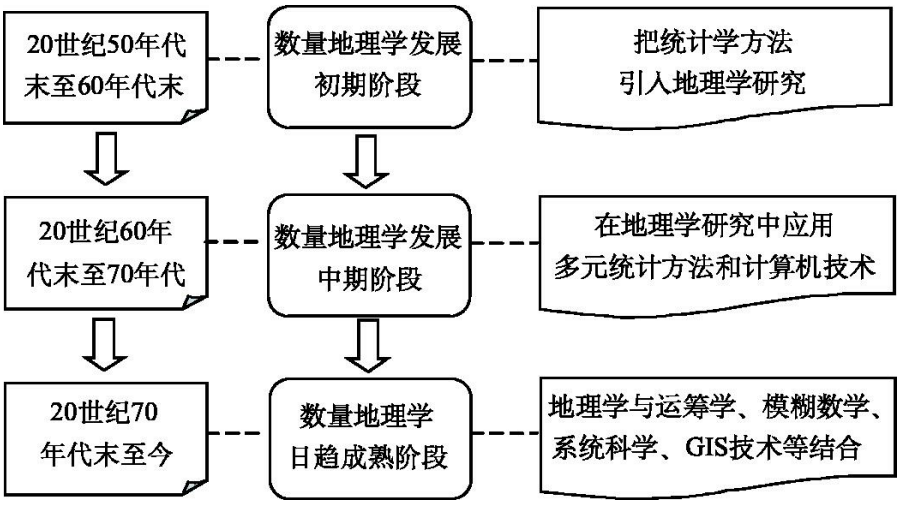


图 1.1　数量地理学的发展阶段

2. 传统地理学与数量地理学

数学方法是人们进行数字运算和求解的工具,能以严密的逻辑和简洁的形式描述复杂的问题,表达极为丰富的实质性思想。对于现代地理学而言,数学方法不仅是应用地理学研究中进行预测、决策、规划及优化设计的工具,也是理论地理学研究中进行逻辑推理和理论演绎的手段。世界上的任何事物都可以用数值来描述和度量,地理要素如区域范围、城市位置、道路长短、气温高低、雨量多少、山高水深、人口增减、物产丰欠等都可用数量来表示。各种地理要素的分布形态及其相互关系特征,亦可以用数学方法进行定量分析与研究。与地理学传统的思维模式相比,地理数量方法有着明显的优势(图 1.2)。

传统地理学分析方法所采用的推理方式以经验归纳型综合为主,以观察材料

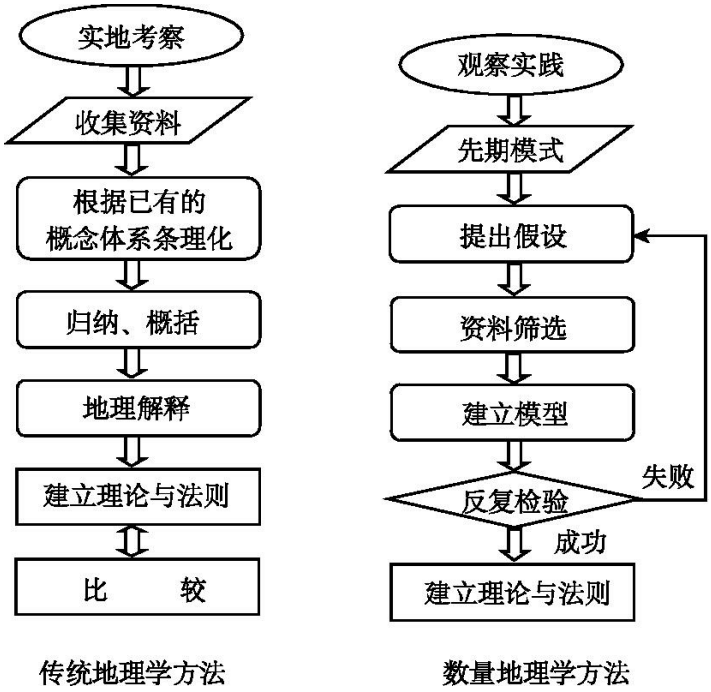


图 1.2 数量地理学与传统地理学研究方法比较

和事实为基础,由直接的类推得出现实世界的结论,这一方法难以回避特殊情况或解释者的主观好恶问题。而数量地理学以理论演绎为主,把感知到的地理事物通过假设予以条理化,继而经过模式化得出数据进行检验,在成功的情况下建立法则和理论,否则反馈回去重新制定假设。整个研究过程经历了提出假设、建立模式、检验假设和建立理论四个步骤,符合感性—理性—实践这一认识的过程规律。

3. 数量地理学中地理数据分析模拟方法

数量地理学本质上是一门关于地理数据分析处理与建模的科学,其主要研究内容涉及地理要素的描述统计和数量分析技术,地理系统的分析方法,数学模型的构建和应用,数学模拟(仿真)技术,地理预测和决策的方法、程序、模型以及地理学理论表述的数学形式等,其研究方法归纳如下。

(1) 地理系统分析

地理系统分析是指扬弃地理事物繁琐的枝节,抽象出地理事物在结构与功能上的主线,揭示地理事物动态演变的方向与强度,预测其状态变化和稳定性程度,将复杂、高级的地理系统简化为次一级简单的系统,进而探讨地理要素之间的数量关系。一般是首先列出所研究等级系统的要素清单,根据地理系统的实际绘出各要素的联系框图,再以定量方法研究系统要素之间的关系。

(2) 随机数学方法

地理系统输入与输出之间一般具有随机性,数量地理学研究方法中随机数学占很大比重。地理系统是多级、多元系统,在进行系统分析时,分析一组或几组地

理要素之间的关系经常应用多元统计分析方法,如多元线性回归、逐步回归、主成分分析、因子分析等;地理系统是具有空间范围和地域界线的系统,确定界线、进行地理区域的划分等经常应用二级判别分析、多级判别、逐步判别等数学分类技术;在探讨地理系统结构、类型组合、空间关系时,常运用系统聚类分析方法;分析地理系统的空间特性时,常用趋势面分析方法;地理系统研究十分重视系统目标、系统结构的研究,为此运用运筹学和最优化方法以使地理系统达到符合一定目标的最佳状态。此外,模拟地理系统状态的转移规律时还包括马尔柯夫链、多元线性方程组、微分方程的应用等。

(3) 地理系统数学模拟

建立地理系统数学模型的过程称为地理系统的数学模拟(简称地理模型)。地理模型成为表达地理现象的状态,描述地理现象的过程,揭示地理现象的结构,说明地理现象的分级,认识该现象与其他地理现象之间联系的概念性和本质性的表征方式。地理系统数学模拟的一般过程是:①从实际的地理系统或其要素出发,对空间状态、空间成分、空间相互作用进行分析,建立地理系统或要素的数学模型。②经验检查,若与实际情况不符,则要重新分析,修改模型;若大致相符,则选择计算方法,进行程序设计、程序调试和上机运算,从而输出模型解。③分析模型解,若模型解出错,则修改模型;若模型解正确,则对成果进行地理解释,提出切实可行的方案。可见,地理系统数学模拟过程是反复修改数学模型、调试和修改程序的过程。

1.1.2　地理信息系统

英国著名地理学家 Johnston R J 在 1995 年曾指出,“计量革命的直接成果是导致了 GIS 革命的到来”。GIS 起源于 20 世纪 60 年代,是对地理空间数据进行采集、存储、表达、更新、检索、管理、综合分析与输出的计算机应用技术系统。GIS 是以应用为导向的空间信息技术,强调空间实体及其关系,注重空间分析与模拟,是重要的地理空间数据管理和分析工具。

1. GIS 是客观现实世界抽象化的数字模型

客观现实世界极其复杂,运用各种数据采集手段和量测工具,如野外调查、遥感技术等,获取有关客观世界的数据,把各种来源和类型的地理空间数据数字化,输入计算机,按一定的规则组织管理,构建客观现实世界的抽象化数字模型,即 GIS(图 1.3)。

存储于 GIS 中的地理空间数据不是客观世界的完全再现,而是在地理认知的基础上对真实世界进行抽象和概括而形成的数字模型,在一定比例尺下表达客观事物的分类、分级、空间过程和空间格局。GIS 应用成功与否不仅在于空间信息技

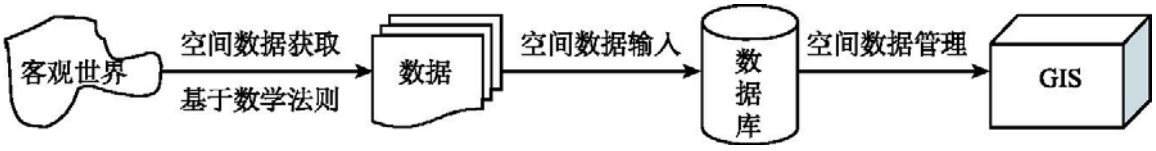


图 1.3 客观世界的抽象化过程

术的发达程度,更多地依赖于人类定义客观世界认知模型的恰当程度。在 GIS 中,对现实世界的理解是从数据、信息、知识到智慧逐渐深入的。

2. GIS 是地理空间数据管理、显示与制图的集成工具

地理信息系统不仅是客观世界抽象化的数字模型,同时还是一种对空间数据进行采集、存储、管理、显示与制图的计算机系统和集成工具,这是地理信息系统最主要的功能之一。GIS 处理的数据可以归纳为两大类:一类描述地理实体的空间位置和空间拓扑关系的图形图像信息;另一类描述地理实体的属性文字、数字信息等。通过数据的获取、管理、显示、分析与制图输出,保证了地理信息系统数据库中数据在内容与空间上的完整性、数值逻辑上的一致性与正确性。地理信息系统拥有所有大型数据库管理系统所具有的功能,如地学空间数据的采集、监测、编辑、存储与管理等,能够高效地组织海量数据,为解决空间复杂问题奠定基础。地理信息系统还为用户提供了许多用于显示地理空间数据的工具,其表达形式既可以是计算机屏幕显示,也可以是报告、表格、地图等硬拷贝方式。GIS 除了具有计算机辅助设计(CAD)、计算机辅助制图(CAC)等一般显示功能外,还具有多幅图层叠加、阴影透视、网状透视、用户格网、地图动画等高级显示功能。一个完备的地理信息系统应能提供一种良好的、交互式的制图环境,使地理信息系统的使用者能设计和印制出具有高品质的地图。

3. GIS 是地理空间数据分析模拟与可视化的技术平台

地理信息系统支持多种数学模型综合运用,可以建立一系列具有分析、模拟、仿真、预测、规划、决策、调控等多功能的模型系统。这种模型系统的运行既需要海量地理数据构成的地理数据库支持,也依赖强有力的计算方法与计算机程序,最终的研究结论则以可视化的地图、统计图或者三维图等形式输出。GIS 用户可以完成对空间数据的一系列处理、分析与建模任务,实现空间数据的可视化。

(1) 空间数据分析与建模

现实世界中的地理现象在 GIS 中都以数字形式表达,形成地理空间数据库。对数据库中的空间数据进行分析与建模以挖掘出有用的空间信息是 GIS 最具生命力的核心功能,也是 GIS 区别于其他计算机系统的主要标志之一。目前常用的

GIS 空间分析方法有缓冲区分析、叠加分析、网络分析、拓扑结构分析、三维分析等。对于复杂的地理空间问题可以为其建立空间分析模型,如数字地形模型(DTM)、空间统计分析模型、人工神经网络模型、粗集模型等。借助 GIS 进行地理模型分析是研究地球系统的重要途径,如综合评价模型、预测模型、规划模型、决策分析模型等应用分析模型在分析地理空间信息、探究地学研究对象的本质特征及其动态变化方面具有重要价值。

(2) 空间信息可视化

科学可视化技术贯穿 GIS 空间分析的始终,它将分析结果以易于理解的方式直观地表达出来,最大限度地利用信息,实现信息共享。从某种角度讲,GIS 可以称为“动态的地图”,它提供了比普通地图更为丰富和灵活的空间数据表现方式,如动态信息表达、虚拟现实等。地学专家对可视化在地学中的地位和作用已进行了深入探讨,提出了与可视化密切相关的地图可视化、地理可视化、GIS 可视化、探析地图学、地学多维图解、虚拟地理环境等概念,但不同的专家有不同的理解,对其相互关系的认识目前仍不明确。

1.1.3　地理计算

随着计算机技术、数学方法的不断进步,空间数据分析处理方法论也随之革新。20 世纪 90 年代,一门融合了计算机科学、地理学、地球信息科学、信息科学、数学和统计学理论与方法的地理计算学(Geocomputation)开始形成并逐渐发展起来,数量地理学进入全新的计算地理学(Geocomputational Geography)时代,地理空间数据分析与建模有了一个新的技术平台。

1. 地理计算的概念与内涵

20 世纪 90 年代中期,英国著名地理学家、里兹大学计算地理研究中心 Stan Openshaw 教授认为空间数据挖掘已成为数量地理学中一个重要分支,并以 Geocomputation 命名这个新的学科,Stan Openshaw 因此被称为“地理计算之父”。此后,许多学者纷纷从不同角度对地理计算的定义与内容框架进行设计,并论证其作为一个学科的必要性和合理性。

Openshaw(1999)认为地理计算本质上是继地理信息科学之后的革命。他在 2000 年又进一步深化对于地理计算的理解,认为地理计算是一种高性能计算,用以解决目前不能解决的、甚至未知的空间问题的科学。地理计算具有三方面特点:一是强调地理主题;二是对现存问题承认有新的或更好的解决办法,且可以解决以前不能解决的问题;三是地理计算需要独特的思考方式,由于以基于海量计算代替残缺的知识或理论,故能够增强机器的智能。

英国里兹大学著名地理学家 Rees 等提议将地理计算定义为:应用计算技术求

解地理问题的理论、方法和过程。从构词来看,Geocomputation 由前缀“Geo”和主词“computation”组合而成,前者指地理计算要做什么,后者则是指如何去做。Gahegan在 1999 年发表的论文中细致地谈到“……地理计算关注利用一系列方法的工具箱丰富地理模拟和分析大量高度复杂的、非确定性的问题……这是人类有意识地努力探索地理学与计算机科学之间的关联。这是一门真正的数量地理学技术,也是计算机科学家进行计算性应用的丰富源泉。”Conclelis(1998)采用相对简洁的定义:地理计算是应用数学计算方法与技术来描述空间特征、解释地理现象、解决地理问题。Openshaw 和 Abrahart(2000)认为:地理计算是一门新兴的交叉学科,它是在科学方法的整体范围内利用各种不同类型的地理数据发展相关的地理工具和模型。

2003 年 8 月,我国亚运村地理学术沙龙谈到“虚拟地理实验室”建设,认为地理计算既不是数量地理学,也不是 GIS,而是智能计算在地理学中的精确应用,是强大的高性能计算,其理论驱动是科学。地理计算能够有效地用于非线性复杂地理问题的模拟、计算与求解。

地理计算是利用不同类型的地理与环境数据,在计算科学方法的整个体系中发展相关的计算工具。它依赖于新计算技术、算法和范例,并且利用高性能计算(high-performance computing,HPC)和高效率计算机(HTC),包括空间数据分析、自动建模、模拟、时空动力学、可视化和虚拟现实。

地理计算试图回归计量革命时代的地理分析和建模,吸收了新的计算机科学成果,如高性能计算,模式识别、分类、预测与模型技术,知识挖掘,可视化等一系列计算方法和工具,建立地理模型并分析复杂的、具有不确定性的地理问题,从而丰富了地理学的研究。Geocomputation 不仅仅是计算机在地理信息领域中的应用,关键是可以辅助地理研究,从而获得基于数据驱动的地理信息管理和地理信息分析。

综上所述,地理计算这一学科的统一视角就是“计算”,它被认为是一系列有效的程序或算法(如神经网络、模糊逻辑、遗传算法等),当应用到地理问题时必然产生结果,不同算法之间由于基本假设的不同而产生结果的差异。地理计算本质上可认为是对地理学时间与空间问题所进行的基于计算机的定量化分析。

2. 地理计算模型与方法

地理计算的目标是将地理学领域的知识引入计算机工具,设计合适的地理数据挖掘和知识发现操作,研发时空尺度上的集群算法,获得超越目前软件、硬件能力的地理数据分析方法,用可视化和虚拟现实的手段实现地理问题的理解与交流。

地理计算学是数量地理学向深层次的拓展,强调数学模型与模拟实验并重的理念,凭借计算机工具对地理学问题进行定量或非定量分析的抽象概括和综合研

究,解决海量、复杂数据集或数据库分析的复杂空间问题。Geocomputation 包含丰富的模型和方法体系,不仅采纳了传统的数量地理学理论与模型,还涉及一系列新的理论技术方法:GIS 为之创建数据库;人工智能技术(artificial intelligence, AI)和智能计算技术(computational intelligence, CI)为之提供计算原理和计算工具;高性能计算服务系统为之提供动力。智能计算技术中的神经网络模型(neural network, NN)、模糊逻辑模型(fuzzy logic)、遗传算法模型(genetic algorithm, GA)、元胞自动机模型(cellular automata, CA)以及分形分析(fractal analysis)等不断被引入并成为地理计算的核心。高性能计算(HPC)是利用超级计算机对大容量资料、需要进行实时分析与控制的系统以及那些复杂而又不能用其他手段来处理的现实世界所实施的计算。地理研究的实践,更多的是充分利用 GIS 技术,结合 GPS 和 RS 技术,以向量或并行处理器为基础的超级计算机为工具,对海量数据资料所表征的地理学问题实施高性能计算,从而探索并构筑新的地理学理论与应用模型。

在目前 GIS 技术下,计算机表达地理空间基本上是静止的。地理计算研究的重要内容之一是如何建立一种模型将空间(地理目标)的结构元素与改变这种空间结构(人类活动及其影响)的过程相结合。这种模型将改变对于空间的静止描述观点,强调作为地理空间基本部分的动态组成,如使用元胞自动机技术模拟城市和区域增长等。

1.2　地理空间数据挖掘

人类在空间科学技术、遥感(RS)、地理信息系统(GIS)、全球定位系统(GPS)等领域取得了巨大成就,对地球系统的不同层面、不同现象的综合观测能力达到了空前的水平,获得了大量对地观测数据。同时,随着数据库技术的成熟和信息应用的普及,人类累积的数据量正在呈指数级增长,全世界每天存入的数据数量超过万兆字符。未来学家 John Naisbitt 惊呼:“人类正被数据淹没,却饥渴于信息”。面临浩如烟海的数据,人们呼唤从数据的汪洋大海中去芜存精、去伪存真,因此,“从数据库中发现知识”(KDD)及其核心技术——数据挖掘(data mining)应运而生。

1.2.1　地理空间数据挖掘概述

数据挖掘是一个由数据库、人工智能、数理统计和可视化等多学科与技术交叉、渗透、融合形成的交叉学科(邸凯昌,2000)。它试图综合应用上述领域技术,在庞大的数据库中探索事先并不知道,但潜在有用的、新的结构形态或者关系特征,即关于数据的高层次信息结构和知识。地理空间数据挖掘(geospatial data mining)是数据挖掘的一个研究分支,其实质是从地理空间数据库中挖掘时空系统中

潜在的、有价值的信息、规律和知识的过程,包括空间模式与特征、空间与非空间数据之间的概要关系等。由于空间数据具有海量、多维和自相关性等特征,使得地理空间数据挖掘更为复杂。

地理空间数据挖掘技术可以有效地解决一些地学问题。例如,地球系统的基本驱动力是什么?整个地球系统是如何变化的?如何能更好地预测地球系统未来的变化?某一种流行病的分布模式?流行病发展变化范围、趋势及速率等?其中许多分析都是基于空间位置关系的,因此地理空间数据挖掘技术最根本的是基于事物的空间特性(如拓扑、距离、方位等)。

近些年来,国内外开展了许多有关地理空间数据分析与挖掘方面的研究。加拿大 Simon Fraser 大学计算机科学系 Han Jiawei 教授领导的小组进行了基于关系数据库挖掘系统的研究,在 MapInfo 平台上开发了空间数据挖掘原型系统 GeoMiner,并设计了专门用于空间数据挖掘的语言 GMQL,实现了空间数据特征描述、空间比较、空间关联、空间聚类和空间分类等空间数据分析方法的集成。该系统具有空间数据库模型、空间数据立方体、空间 OLAP 等模块(图 1.4)。武汉大学李德仁院士等提出从 GIS 数据库可以挖掘出包括几何信息、空间关系、几何性质与属性关系以及面向对象知识等多种知识,认为空间数据分析与挖掘使 GIS 的有限数据变成无限的知识。图 1.5 为数据挖掘与知识发现的进化历程(陈述彭等,1996)。

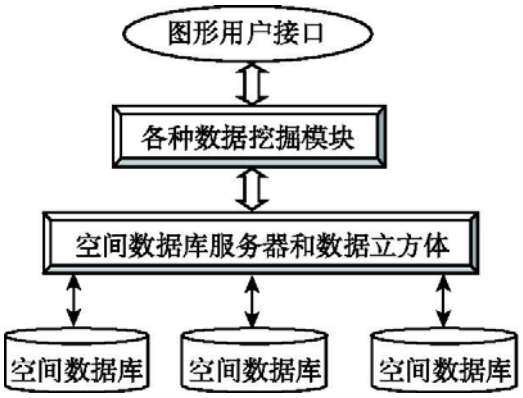


图 1.4 GeoMiner 系统结构

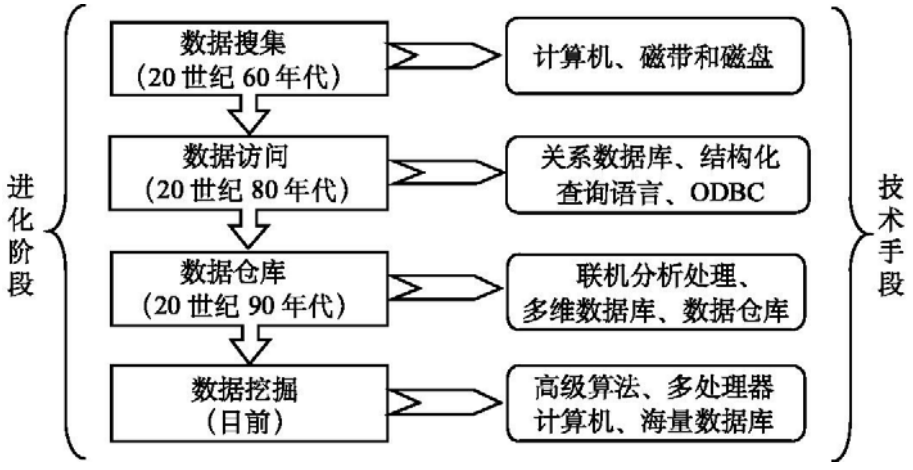


图 1.5 数据挖掘与知识发现的进化历程

地理空间数据挖掘包含从地理空间数据库中发现有用却尚未发现的模式的一系列技术。传统观点认为数据挖掘技术植根于计算科学和数学,不需要也不得益

于数据立方体。这种观点今天看来并不正确,数据挖掘成功的关键之一就是先通过访问正确、完整和集成的数据库,才能进行深层次的分析,寻求有意义的信息。而这些正是数据立方体所能提供的,数据立方体不仅是集成数据的一种方式,其联机分析功能——OLAP 还为数据挖掘提供了一个极佳的操作平台。实现空间数据挖掘与数据立方体有效的联结,将给空间数据挖掘带来各种便利操作和新的功能。

按照不同的挖掘任务,地理空间数据挖掘可以分为预测模型发现、聚类、关联规则发现、序列模式发现、依赖关系发现、异常值分析和趋势发现等。由于空间数据库包含了大量的拓扑/距离信息,需要按照复杂的多维空间索引结构组织数据。在访问这些数据时,需要采用空间推理、地理计算和空间知识的表示技术。地理空间数据挖掘系统包括三大支柱模块:地理空间数据立方体、联机分析处理(OLAP)模块和空间数据挖掘模块。

地理空间数据挖掘的体系结构如图 1.6 所示,由以下四部分组成:①图形用户界面(交互式挖掘);②挖掘模块集合;③数据库和知识库(空间、非空间数据库和相关概念);④空间数据库服务器(如 ESRI/Oracle SDE,ArcGIS 以及其他空间数据库引擎)。

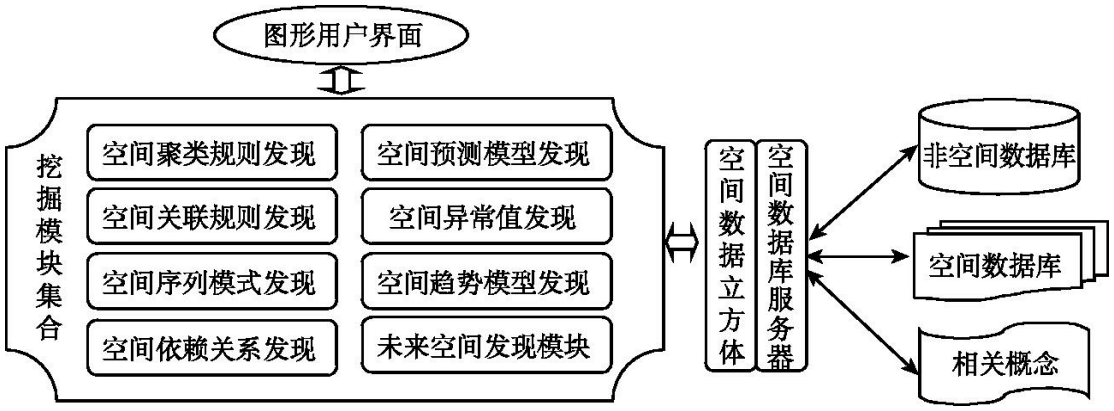


图 1.6　地理空间数据挖掘体系结构

1.2.2　地理空间数据立方体

地理空间数据立方体(geospatial data cube)是一个面向对象的、集成的、以时间为变量的、持续采集空间与非空间数据的多维数据集合,组织和汇总成一个由一组维度和度量值定义的多维结构,用以支持地理空间数据挖掘技术和决策支持过程。地理空间数据立方体绝非仅在数据库上加一层空间外衣,而是真正地以空间数据库为基础,进行复杂的空间分析,反映不同时空尺度下的动态变化趋势,为决策者提供及时、准确的信息。地理空间数据立方体中的数据是经过选择、整理、集成等处理的,为空间数据挖掘提供了良好的数据基础,因而在地理空间数据立方体中进行数据挖掘比在原始数据库中更加有效。

数据立方体法的基本思想是把那些经常被查询到的求和、求平均值、求最大最

小值等成本较高的计算进行具体化,并将这些具体化的视图存储到数据立方体中,便于知识发现。

所谓“立方体”并非指数据仅包含 3 个维度,事实上一个数据立方体可以包含 128 个维度。数据立方体在处理时预先计算好一些汇总数据,称为聚合。聚合提供了一种便于使用、快捷且响应时间一致的数据查询机制。数据立方体在逻辑上一般由一个事实数据表和多个维度表构成一种星形构架(图 1.7),其核心是事实数据表。事实数据表是数据立方体中度量值的源,维度表是数据立方体中维度的源。

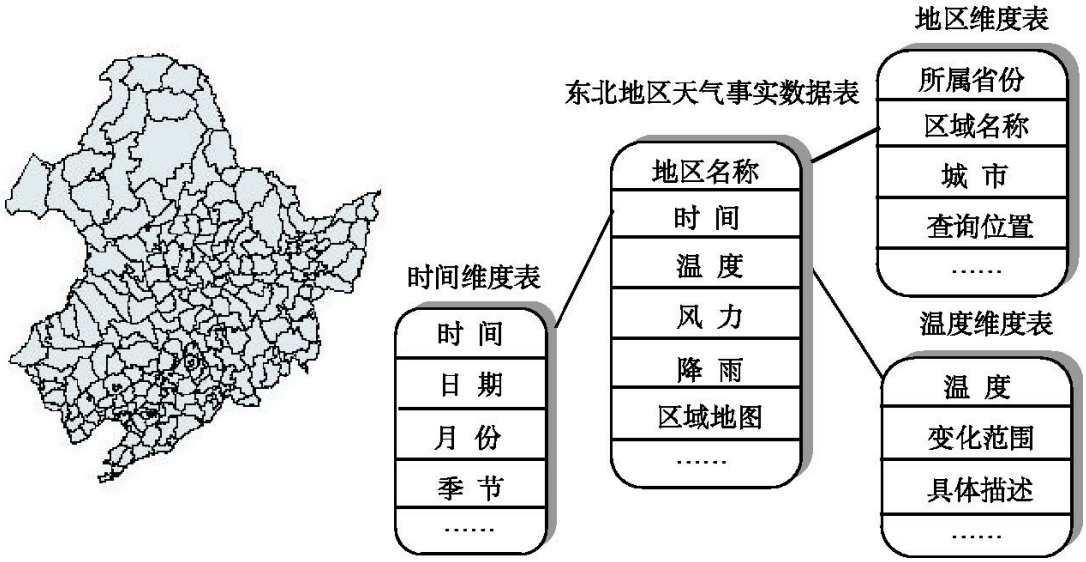


图 1.7 东北地区天气数据立方体星形构架

地理空间数据立方体涉及的概念如下所述。

(1) 维度

维度是数据立方体的一种结构特性,是描述事实数据表中数据级别的有组织的层次结构。这些级别通常描述相似成员的集合,用户根据它们进行分析。例如,某个地理维度可能包括国家、省以及城市等级别。在地理空间数据立方体中有三种维度类型:

1) 非空间维度,包含非空间信息,如城市名称、城市人口数、气温、湿度等。

2) 空间-非空间维度,该维度的初始数据是空间维度,其解释数据变为非空间维度。例如,作为空间维度的城市分布是中国地图的一部分,假设该城市分布被表达为“长江以北”,尽管“长江以北”是一个空间概念,但它从表达上是一个字符型,属于非空间维度。

3) 空间-空间维度,初始数据和解释数据均为空间维度。例如,等温区维度包含空间数据,其解释数据为 0~ 5℃、5~ 10℃区域的空间维度数据。

(2) 度量值

度量值是在数据立方体内基于该数据立方体的事实数据表中某列的一组值,它们通常是数字。度量值是进行聚合和分析的主要数值。空间数据立方体的度量值有两种类型:

1) 数值度量,仅包含数字数据。例如,已知一个区域的人均月收入,便能计算出总体收入(年、国家等)。

2) 空间度量,包含空间目标的指示性聚集信息。例如,相同的温度和风力范围的区域可以被合成为一个单元。

(3) 成员属性

成员属性是维度表的一个可选特性,为最终用户提供成员的其他信息,仅从属于级别。成员属性在级别中创建,该级别应包含应用该成员属性的那些成员。

1.2.3　联机分析处理技术

1. OLAP 概念

联机分析处理(on-line analytical processing, OLAP)的概念最初是由关系数据库之父 Codd E F 于 1993 年提出的。Codd 认为联机事务处理已不能满足终端用户对数据库查询分析的需要,SQL 对大型数据库的简单查询也无法满足用户分析的需求,因此提出了多维数据库和多维分析的概念,即 OLAP。OLAP 是共享多维信息的、针对特定问题的联机数据访问和分析的软件技术,具有汇总、合并、聚集以及从不同角度观察消息的能力。它可以跨越空间数据库模式的多个版本,处理来自不同组织的信息和由多个数据存储集成的信息。联机分析处理对空间数据立方体进行的多维数据分析主要有切块、切片、旋转、钻取等分析动作,目的是进行跨维、跨层次的计算与建模。在多维空间数据结构中,按二维进行切片,按某一维进行切块,对片、块或整个多维数据库在维数不变的前提下通过改变维的层次或位置,进行数据钻取和旋转。

利用 OLAP 对空间数据立方体进行多维分析的一般过程是:先按某一维切块得到关注的内容,然后钻取空间数据到达适当的综合层次,再通过旋转动作更换空间数据观察角度,选取重要的空间数据进行切片分析。每个环节可能有一定的重复,但是经过如此切片、切块、旋转、钻取可以形成对空间数据新的观察角度和综合层次,可能提取出有价值的空间信息,得到潜在知识。

2. OLAP 与地理空间数据立方体

OLAP 和地理空间数据立方体密不可分,但两者概念内涵不同。如前所述,地理空间数据立方体中的数据不能用于联机事物处理系统(OLTP),而 OLAP 技术

则可利用数据立方体中的数据进行联机分析,将复杂的分析查询结果快速地返回用户。OLAP 利用多维数据集和数据聚集技术对数据仓库中的数据进行组织和汇总,用联机分析和可视化工具对这些数据迅速进行评价。从图 1.8 中可以发现,OLAP 用多维结构表示空间数据立方体中的数据,能有效地满足用户复杂查询的要求。因此,空间数据立方体的结构将直接影响立方体的设计和建立,进而影响 OLAP 的工作效率。

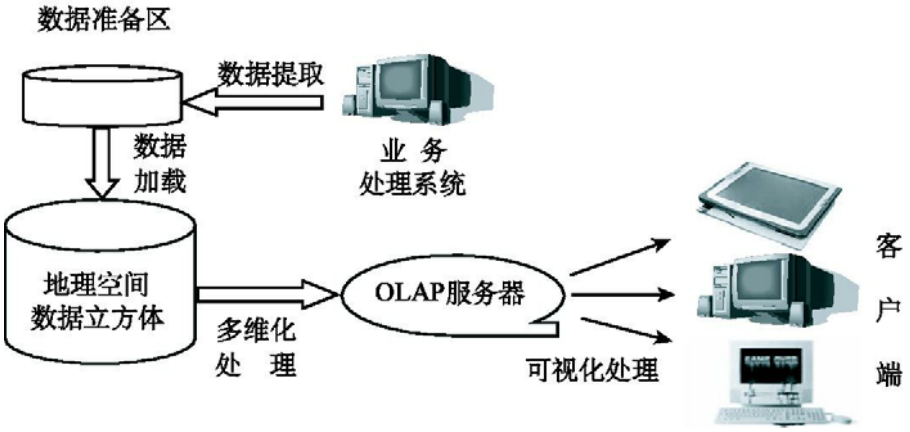


图 1.8 地理空间数据立方体与 OLAP 的关系

1.2.4 地理空间数据挖掘典型方法

1. 地理空间统计方法

地理空间统计是指分析地理空间数据的统计方法,主要是基于空间中邻近的要素通常比相距较远的要素具有较高的相似性这一原理。它是通过空间位置建立数据间的统计关系,其应用范围极广,包括地质、大气、水文、生态、天文、遥测、地震、环境监测、流行病及影像处理等。事实上,除极少数情况外,真实世界的空间数据大多无法仅基于物理化学机制用简单的公式来描述。为解决数据中所隐含的空间不确定因素,地理空间统计模型尝试从凌乱的地理空间数据中,用统计方法发掘地理空间变化规律。

地理空间统计分析与传统统计分析主要有两大差异:①空间数据间并非独立,而是在 D 维空间中具有某种空间相关性,且在不同的空间分辨率下呈现不同的相关程度;②大多数空间问题仅有一组(不规则分布空间中)观测值,而无重复观测的资料。因此,真正地了解与描述空间现象是极为复杂的任务。传统的统计分析技术,特别是基于独立样本的统计方法,并不适于分析处理空间数据。而地理空间统计分析与时间序列分析最大的差异在于空间中并无过去、未来的次序,因而不易透过某种因果关系的描述来建构空间模型。

目前地理空间统计模型大致可分为三类:地统计(geostatistics)、格网空间模

型(spatial lattice model)和空间点分布形态(spatial point pattern)(表 1.1)。地统计是以区域化变量理论为基础,以变差函数为主要工具,研究空间分布上既具有随机性又具有结构性的自然现象的科学。它可以根据离散数据生成连续表面,通过空间自相关进行空间预测。格网空间模型用以描述分布于有限(或无穷离散)空间点(或区域)上数据的空间关系。例如,在流行病学中通过地理区域(如县市、乡镇)的发病人数数据研究疾病发生率与地理位置的关系,在影像处理中利用扭曲或带有噪声的数字影像(如医学或卫星影像)数据,重建背后的真实影像等。在自然科学研究中,许多资料是由点(或小区域)所构成的集合,比如地震发生地点分布、树木在森林中的分布、某种鸟类鸟巢的分布、生物组织中细胞核的分布、太空中星球的分布等,称之为空间点分布形态,其中点的位置为事件。由于形成机制不同,空间点分布形态具有随机、丛聚或规则等不同类型。基于空间点分布形态的研究,可以找寻丛聚所在,并了解其形成的原因及其可能的影响。空间点分布形态通常由一个 D 维的空间点过程描述。此类模型的随机机制在于位置本身,其中最基本的空间点过程为均匀泊松点过程,通常用于定义所谓完全空间随机的点分布形态,并与丛聚或规则的分布区别开来。

表 1.1　数据类型与统计模型

		点处理	基于格网的统计	地统计
栅格			•	•
矢量	点	•	•	•
	线			•
	面	•	•	
图表				

空间数据统计分析是分析空间数据广泛使用的一种方法,能够很好地处理数字数据,提出空间现象的现实模型。然而,需要指出的是统计分析方法往往假设在空间中分布的数据具有统计独立性,而在现实中,空间物体相关性很大。此外,绝大多数统计模型需要在有丰富领域知识和掌握统计专门技术的专家的协助下才能实现。而且,统计模型不能很好地处理字符值、不完整或非确定性数据。

2. 地理空间聚类方法

地理空间数据聚类是按照某种距离度量准则,在大型、多维数据集中标识出聚类或稠密分布的区域(图 1.9),从而发现数据集的整体空间分布模式。该方法把空间数据库中的对象分为有意义的子类,使同一子类内部的成员有尽可能多的相同属性,而不同的子类之间差异较大。比如,空间聚类方法可以将距离很近的、散

布的居民点聚类成居民区,也可将精准农业中的作物产量图聚类成高、中、低产区。事实上,聚类分析技术把大型数据库分为多个较小的部分,采用“分而治之”的策略使用户可以更好地分析空间数据,更容易把握大局。地理空间聚类是空间数据挖掘中的主要方法之一,对于处理海量数据、提取大型空间数据库中的有用信息和知识具有十分重要的意义。在实施其他空间数据挖掘任务之前,应用空间聚类方法也可大大提高挖掘精度与效率。

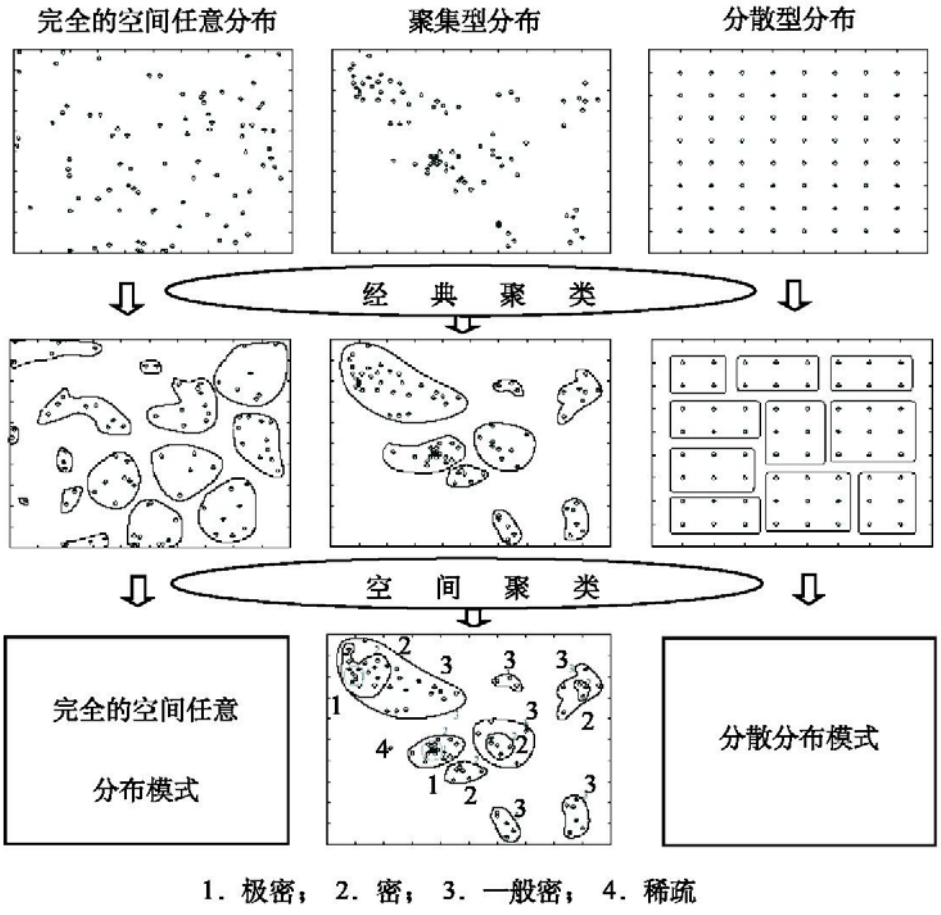


图 1.9 地理空间聚类

目前,地理空间聚类方法主要有四类:分割法、层次法、基于密度的方法及基于网格的方法。而经典聚类法包括 K-mean、K-meriod、ISODATA 等。近年来,围绕 DMKD 领域发展了 CLARANS(Ng R,1994)、DBSCAN(Ester M,1996)、Murray(Murray A J,1998)等算法。Kohonen 自组织特征映射网络、竞争学习网络等自组织神经网络方法,在空间聚类应用中亦有较好的效果。

3. 地理空间关联分析

空间数据库存储了大量与空间有关的数据,与关系数据库存在很大区别。空间数据表现了地理空间实体的位置、大小、形状、方向及几何拓扑关系。地理空间关联分析利用空间关联规则提取算法发现空间数据库中空间目标间的关联程度,

是空间数据库知识发现研究中的一个重要课题。GIS 数据库是典型的空间数据库,从 GIS 数据库中挖掘空间关联规则是理解 GIS 模型和将 GIS 数据转化成知识的一种有效方法。

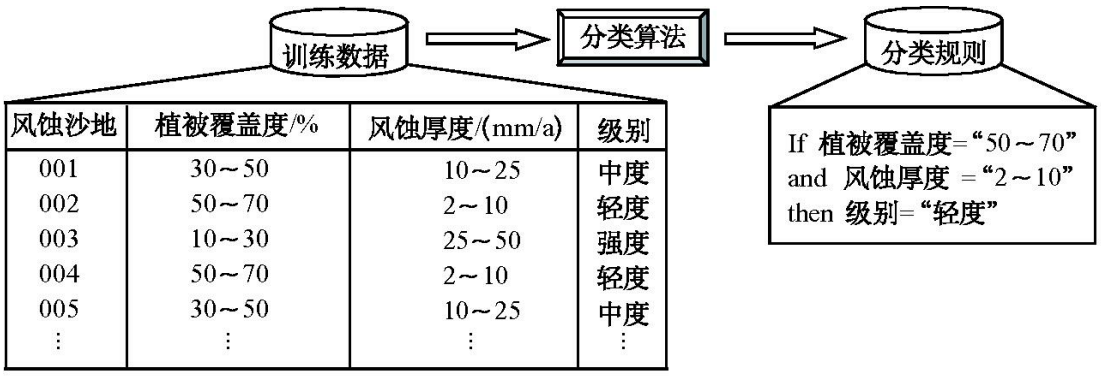
地理空间关联分析的核心内容是挖掘空间关联规则。空间关联规则是指空间目标间相邻(如村落与道路相邻)、相连(如火车站与铁路相连)、共生(如蒙古包与草场的关系)、包含(如区域中包含城市)等空间相关关系。具体而言,空间关联规则中包含各种不同的空间谓词,它们不但可以表示空间对象之间的拓扑关系(如相交、不相交、相邻等),还可以表示空间方位、排列次序(如东、西、南、北等)以及距离信息(如靠近、远离等)。空间关联规则指明了空间谓词与非空间谓词间存在的关联性。例如,通过挖掘 GIS 数据库,可能发现“靠近海滩的房屋”有 90%“价格很贵”,“加油站”有 75%“靠近高速公路”等。空间关联规则提取算法并不唯一,较常用的是利用 MBR 技术、R + 树及其他快速方法进行空间分析,并采用概念层次树对拓扑关系进行概化形成拓扑关系数据表,从而提取关联规则。

4. 地理空间分类与预测分析

地理空间分类与预测是根据已知的分类模型把数据库中的数据映射到给定类别中,进行数据趋势预测分析的方法。

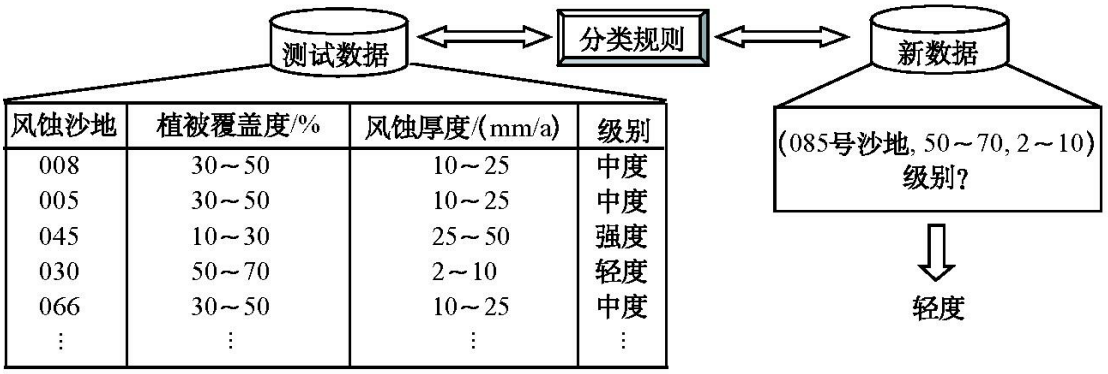
分类是将数据库中的对象根据一定的意义划分为若干个子集。它和聚类算法的差别在于:聚类算法是根据一定要求将对象聚为一个集合,最后得到的分布模式是聚类之前未确知的;分类算法则是根据已知分布模式的属性要求,将数据库对象归入相应的分类中。在机器学习中,数据分类一般称为监督学习,而数据聚类则称为非监督学习。分类目的是通过学习确定一个分类模型(或分类器),该模型能把数据库中的数据项映射到给定类别中。地理空间数据分类包括两个步骤(图 1.10)。第一步通过分析数据库中各数据行的内容建立一个分类模型(分类器),描述已知数据集类别或概念。第二步利用所获得的分类模型(分类器)进行分类操作。首先对模型分类的准确率进行评估,若分类准确率可以被用户接受,则利用该分类器对目标数据库进行分类。分类器的典型构造方法有决策树法、贝叶斯法、神经网络法、近邻学习或基于案例学习等。

预测是利用历史数据记录自动推导出对给定数据的推广描述,实现对未来数据的趋势分析。分类和回归都可用于预测,空间回归规则与空间分类规则相似,也是一种分类器,其差别在于空间分类规则的预测值是离散的,空间回归规则的预测值是连续的。二者常表现为决策树形式,根据数据值从树根开始搜索,沿着满足条件的分支往上走,走到树叶就能确定类别。空间分类或回归的规则是普及知识,实质是对给定数据对象集的抽象和概括,可用宏元组表示。



风蚀沙地	植被覆盖度/%	风蚀厚度/(mm/a)	级别
001	30～50	10～25	中度
002	50～70	2～10	轻度
003	10～30	25～50	强度
004	50～70	2～10	轻度
005	30～50	10～25	中度
⋮	⋮	⋮	⋮

(a) 学习过程



风蚀沙地	植被覆盖度/%	风蚀厚度/(mm/a)	级别
008	30～50	10～25	中度
005	30～50	10～25	中度
045	10～30	25～50	强度
030	50～70	2～10	轻度
066	30～50	10～25	中度
⋮	⋮	⋮	⋮

(b) 分类过程

图 1.10　数据分类过程

5. 异常值分析

若一个数据库包含的数据目标与通常的行为或数据模型不一致,则这些数据目标被称为异常值。绝大多数数据挖掘方法把异常值作为噪音或例外数据,然而,在很多情况下这将会导致重要隐含信息的丢失。从另一角度讲,异常值是内在数据可变性的必然结果。例如,与我国其他城市的商业产值相比,我国经济中心——上海市的商业产值很自然地成为一个异常值出现。在一些应用,如赝品检测、定制买卖、数值分析等任务中,异常值分析有很重要的价值。一个人认为的噪音可能是其他人所需的重要信息,稀有事件往往比规律发生的事件更能说明问题。因此,异常值检测与分析也是一项重要的数据挖掘技术。基于计算机的异常值分析方法主要有三种:基于统计的异常值分析;基于距离的异常值探测;基于偏差的异常值探测。

聚类分析方法将异常值视为噪声,事实上可以将异常值探测作为聚类分析的副产品。另外,由于人眼能够迅速、有效地观察出异常数据,利用数据可视化方法探测异常值可以说是一个明智之举。需要指出的是,人眼只擅长于数字数据或二维、三维数据,在探测多类属性数据或高维数据时,数据可视化分析方法没有优势。

1.3　GIS 环境下的空间分析

1.3.1　空间分析概念

1. 空间分析的定义

空间分析(spatial analysis,SA)是地理学的精髓,是为解答地理空间问题而进行的数据分析与挖掘。目前,比较典型的空间分析定义有如下几种:

空间分析是对数据的空间信息、属性信息或二者共同信息的统计描述或说明(Goodchild,1987)。空间分析是对于地理空间现象的定量研究,其常规能力是操纵空间数据成为不同的形式,并且提取其潜在信息(Openshaw,1997;Baily,1995)。空间分析是基于地理对象空间布局的地理数据分析技术(Robert Haining,1990)。空间查询和空间分析是指从 GIS 目标之间的空间关系中获取派生的信息和新的知识(李德仁,1993)。空间分析是指为制定规划和决策,应用逻辑或数学模型分析空间数据或空间观测值(Landis J,1995)。空间分析是基于地理对象的位置和形态特征的空间数据分析技术,其目的在于提取和传输空间信息(郭仁忠,1996)。GIS 空间分析是从一个或多个空间数据图层获取信息的过程(Demers,1997)。

空间分析是集空间数据分析和空间模拟于一体的技术,通过地理计算和空间表达挖掘潜在空间信息,以解决实际问题。空间分析的本质特征包括:

1) 探测空间数据中的模式;

2) 研究空间数据间的关系并建立相应的空间数据模型;

3) 提高适合于所有观察模式处理过程的理解;

4) 改进发生地理空间事件的预测能力和控制能力。

2. 空间分析的研究对象

空间分析主要通过对空间数据和空间模型的联合分析来挖掘空间目标的潜在信息。

空间目标是空间分析的具体研究对象。空间目标具有空间位置、分布、形态、空间关系(距离、方位、拓扑、相关场)等基本特征。其中,空间关系是指地理实体之间存在的与空间特性有关的关系,是数据组织、查询、分析和推理的基础。不同类型的空间目标具有不同的形态结构描述,对形态结构的分析称为形态分析。例如,可以将地理空间目标划分为点、线、面和体四大类要素,面具有面积、周长、形状等形态结构,线具有长度、方向等形态结构。考虑到空间目标兼有几何数据和属性数据的描述,因此必须联合几何数据和属性数据进行分析。

空间数据分析实际上是对空间数据一系列的运算和查询。不同的应用具有不同的运算和不同的查询内容、方式、过程。应用模型是在对具体对象与过程进行大量专业研究的基础上总结出来的客观规律的抽象,将它们归结成一系列典型的运算与查询命令,可以解决某一类专业的空间分析任务。

3. 空间分析的研究目标

空间分析是指用于分析地理事件的一系列技术,分析结果依赖于事件的空间分布,面向最终用户,其主要目标如下。

1) 认知。有效获取空间数据,并对其进行科学的组织描述,利用数据再现事物本身,例如绘制风险图。

2) 解释。理解和解释地理空间数据的背景过程,认识事件的本质规律,例如住房价格中的地理邻居效应。

3) 预报。在了解、掌握事件发生现状与规律的前提下,运用有关预测模型对未来的状况做出预测,例如传染病的爆发。

4) 调控。对地理空间发生的事件进行调控,例如合理分配资源。

总之,空间分析的根本目标是建立有效的空间数据模型来表达地理实体的时空特性,发展面向应用的时空分析模拟方法,以数字化方式动态地、全局地描述地理实体和地理现象的空间分布关系,从而反映地理实体的内在规律和变化趋势。GIS 空间分析实际是一种对 GIS 海量地球空间数据的增值操作。

1.3.2 空间分析的萌芽与发展

空间分析在地理学研究中有着悠久的传统与历史。从某种意义上讲,空间分析孕育了地理学。在古代,人类出于生存和发展的需要,要学会分析周围地理事物的空间关系,因而始终在进行着各种类型的空间分析。

作为地理学的第二语言,地图的出现使人类的空间分析能力大大增强。从 1863 年 Lalanne L 提出六边形轨道模式到 1963 年 Tobler W R 提出图像转换方法,前 GIS 时期的地图学家对地理空间数据“自我表述”方法极为感兴趣。为使地图有助于空间分析,地理学家试图寻找一种能以形象方式描述数据空间分布的方法,这就是早期的空间统计方法。地图研究者一方面研究空间数据表达及空间数据归纳,一方面借助统计学等数学手段,探索从地图中提取尽可能多信息的方法。长期以来,人们在地图上量测各种地理要素间的距离、方位、面积,或者利用地图进行信息叠加与合成,也基于地图进行较高层次的信息分析,例如社会、经济、文化和军事等领域的各种区域性决策。

随着地图学理论与应用的广泛深入,物理、数学概念与方法的不断引入以及地学各分支学科的发展,传统的空间分析能力大大加强,人们对地图表达空间信息的