

GIS空间关联模式发现

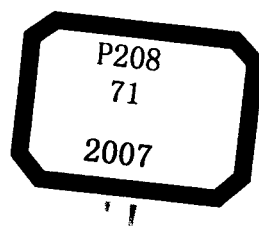
Mining Spatial Association Patterns
from GIS Database

● 马荣华 蒲英霞 马晓冬 著

2



科学出版社
www.sciencep.com



GIS 空间关联模式发现

Mining Spatial Association Patterns
from GIS Database

马荣华 蒲英霞 马晓冬 著

科学出版社

北 京

内 容 简 介

本书论述了 GIS 空间关联模式发现的理论和方法。全书共分 8 章, 主要内容包括: GIS 空间关联模式发现的认知基础和认知过程、空间关联位置模式发现的方法、空间关联结构模式的探索性发现和证实性发现的方法, 以及空间关联模式发现的实际应用。全书结构严谨, 理论、方法和应用结合紧密。

本书可以作为地理信息系统专业及相关专业研究生的选修教材或教学参考书; 同时可供地理学、计算机科学、测绘科学与技术等领域从事空间信息处理与分析的研究人员和工程技术人员阅读参考。

图书在版编目(CIP)数据

GIS 空间关联模式发现/马荣华, 蒲英霞, 马晓冬著. —北京: 科学出版社, 2007

ISBN 978-7-03-018458-0

I. G… II. ①马…②蒲…③马… III. 地理信息系统—研究
IV. P208

中国版本图书馆 CIP 数据核字(2007)第 007300 号

责任编辑: 彭胜潮 韩 鹏 / 责任校对: 刘亚琦

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

*

2007 年 1 月第 一 版 开本: 787×1092 1/16

2007 年 1 月第一次印刷 印张: 14

印数: 1—3 000 字数: 313 000

定价: 38.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

序 一

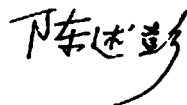
空间关联(spatial association)原是地理分布现象客观存在的本质特征。万千世界中的事物和现象,它们在地球上的存在,都是相互依存、相互制约、相互影响和相互作用的。美国科学院院士 A. Tobler(1970)非常强调它的重要性,称之为地理学的第一定律,指出地表所有事物和现象在空间上是关联的;距离越近,关联程度就越强;距离越远,关联程度就越弱。

人类对空间关联的认识最初是从地理空间的方位、距离、面积等一些几何要素着眼的,公元3世纪裴秀(224~271)的“制图六体”,就曾对空间关联性作过一番逻辑分析,作为地图编制的规范。比托莱密(Claudius Ptolemaeus, 90~168)把一大堆经纬度测量数据汇总在一起,就称为“地理学指南”,显然是一大进步。今天,人类文明进入到航天时代与信息社会,人类对地球和太空的观测能力空前强大,卫星遥感、全球定位系统、海陆定位观测台站网络等所获取的数据,有如浩瀚烟海,连篇累牍,不可胜计。分布式地理信息系统数据库应运而生,盛极一时,与新一代网络系统相融合,信息流通达全球,汇通天下!此时此刻,人们对地理关联性的认识就日新月异,深度、广度都拓展开来了,包括空间分布特征、空间序列模式、空间聚类规则、空间趋势分析等复杂性问题,都提到日程上来了。为了进一步理解和分析这些地球信息机理领域的新问题,空间关联性模式就成为当代地理学前沿首选的突破口。这是顺理成章的,符合科学发展观逻辑思维的必然性。

中国古代有一个“庖丁解牛”的故事,聪明的宰牛高手,批大郤,导大窾,顺应牛的生理结构,宰牛就毫不费力。今天,我们面对庞大无比的地理信息系统数据库群,要从海量数据中提取有用的信息,由此及彼,由表及里,从信息中去发现有助于决策的知识,以空间关联模式作为首选的切入点,也是非常高明的。

马荣华、蒲英霞、马晓冬博士的研究群体在这部科学专著中,对如何抓住地理关联模式这一核心技术,去开发知识发现的宝库,进行了系统、深入的分析,逻辑严谨,推理深入,行文流畅,读起来并不艰难。在这部专著中,他们把空间关联模式看作是一个广义的知识发现过程,剖析了从现实世界转化到概念世界和信息模拟(cybers)世界,然后又反馈到现实世界的全过程,促进地理科学认知的螺旋式升华。他们构建了从探索性空间数据分析(ESDA)到证实性空间数据分析(CSDA)的技术方法集,大大增强了利用地理空间关联模式

进行知识发现的能力。他们选用了全局及局域空间关联性测度、空间异质方差模型、空间趋同模型等方法,对中国流动人口、江苏省城镇群格局,苏州、徐州等城市化过程和江苏省区域经济差异等诸多应用案例进行了实证分析,对包含、共生、相连以及相邻等一体化的空间关联位置模式进行了有益探索,帮助读者加深对“空间关联模式”的理解。与此同时,该专著也启发我们读者,从“空间关联模式”着手,还应该进一步关注自然演化动态过程中时间维度的研究与时空转换的可能性。这必将使地理信息系统的知识发现,前景更加广阔,对决策支持的贡献更加全面,对落实科学发展观,落实五个统筹,建设和谐社会,贡献将更加巨大。我们殷切期望马荣华等博士们的研究群体再接再厉,进一步加强对时空关联综合模式的探索,为知识发现做出突破性的贡献。



2006 年 6 月

序 二

地理学的科学化是现代地理学的主要特征。它包括解释途径的确定化、分析方法的模式化和研究领域的系统化等相互联系的方面*。由马荣华等著的《GIS 空间关联模式发现》一书再次诠释了现代地理学这一主要特征。

由自然-社会-经济构成的地理系统是一个复杂的空间系统,但是它们是可以被认知的。作者以地理空间认知为基础,以地理空间现象和空间过程的本质特征——空间关联为研究内容,通过空间关联模式发现的认知过程,构建了一个基于认知的双库协同的 GIS 空间关联模式发现的总体框架(SAP-GIS-DC)。这个框架从问题来源到问题归宿,不同于单纯的描述归纳方法,体现了以理论演绎为主的地理学研究方法。

对空间关联问题的研究提出了模式化的分析方法,包括空间关联位置模式和空间关联结构模式。前者主要关注 GIS 空间数据,后者主要关注 GIS 属性数据,但所使用的主要对象都是 GIS 数据库。这种模式化的分析方法可以通过计算机的形式化语言进行描述表达和运行求解,将有力促进地理学定量化和数字化方法的发展。

该书是在国家自然科学基金项目“GIS 数据挖掘的认知逻辑模型研究”等课题研究的基础上撰写而成。内容选题新颖,旨在关注基于地理学第一定律的空间关联,强调了地理空间认知理论及 GIS 数据组织的重要性,构建了 GIS 空间分析模型,为地理学从空间数据库获取地理信息提供了平台支持,并通过实例证实了该平台在流动人口空间分布、城镇群空间结构、城市化格局与过程、经济发展区域差异等方面研究所具有的重要理论和实践意义,是应用基础研究的一项创新性成果。谨此表示最诚挚的祝贺!

黄杏元

2006 年 6 月

* 杨吾扬. 1988. 地理学思想简史. 北京: 高等教育出版社

前言

地理研究的“数据贫乏”时代已经一去不复返了。作为现实世界重要认知结果的传统地图,虽然长期以来也一直是获取知识、认知世界的重要手段,但只是对空间对象(现象)的距离、方位等空间信息的一种简单表达,是对地理实体的自我表述。GIS的诞生,让现实世界有了更加适合的客观刻画和数字表达,地理空间数据有了表达和传播的合适媒介——GIS数据库,地理学家从而可以从不同视角在空间与时间尺度的变换中认识世界,探索规律。随着应用领域的不断扩展与深化,更值得我们关注的是,地理空间数据以指数方式不断增长的海量特征凸显,“数据丰富”的时代已经来临。

随着数据库技术的成熟和信息应用的普及,在海量空间数据背后隐藏的具有决策意义的信息越来越受到人们的重视。作为最重要的一类空间数据库,GIS数据库具有不同于其他数据库的独特性,如何从GIS数据库中发现更为隐含的知识,为决策提供服务,将是一个重要课题。在科学技术不断发展的今天,学科交叉与借鉴,分析方法的提高已使得地理学逐步由定性描述、概括总结为主的经验科学向以实验为主要手段、机理探讨为主要目的、数学模型表达为主要特色的现代科学迈进。然而,我国的地理学由于理论的归纳和创新缺乏科学的研究范式的支撑,特别是对地理学综合集成研究方法探索的不足,已经在某种程度上制约了地理学对前沿科学问题的研究。

地理信息科学的发展,使得空间关联模式发现的地位日益显现。目前,地理学的诸多研究领域均涉及空间关联模式问题,获取空间关联模式的工具和分析空间关联模式的方法日渐增多,真正基于学科前沿问题的研究方法和面向应用的分析工具尚不多见。国外相关方面的研究已在理论、方法和工具(软件)等方面有了很多积累,正在走向实用化。在国内,这方面的研究也取得了十分丰富的成果,但仍需深化,尤其在应用领域。本书并没有追求面面俱到,旨在关注基于地理学第一定律的空间关联,把GIS空间关联模式发现看作一个广义的发现过程,强调了地理空间认知理论以及GIS数据组织的重要性,构建了发现空间关联位置模式和空间关联结构模式的方法框架,证实了它们在诸多研究领域的适用性和实用性。

2000年5月,我们参加了由南京大学主办的“遥感科学进展及其在地球科学中的应用”高级研讨班(第二期),美国密执根大学中国研究数据中心鲍曙明博士主讲“空间统计与建模技术在环境研究中的应用”,让我们受益匪浅。当时国内正兴起空间数据挖掘和知识发现研究,李德仁院士领导的研究团队发表了诸多相关文章。我们在思考:如何把空间统计学应用到GIS数据分析以及空间数据挖掘和知识发现的过程中?这种空间数据挖掘和知识发现的研究方法如何在地理学中得到应用,进而为地理学从经验科学向试验科学的迈进提供方法基础?随后,我们的博士论文均选择了与此有关的研究方向。幸运的是,博士毕业后,我们的研究课题都与此有关,或者直接对相关问题进行研究,或者

可以借鉴其中的研究方法。2004 年 10 月的一次学术沙龙上,我们萌生了一个念头:在相关课题研究的基础上,写一本有关空间关联模式发现的书。同年年底,黄杏元教授审阅了本书的写作提纲,并提出了诸多宝贵的修改意见。以此为基础,2005 年 7 月本书初稿基本完成。2005 年 11 月,借中国地理信息系统协会 GIS 理论与方法专业委员会在南京召开之机,我们请闻国年教授和史文中教授审阅了本书修改后的提纲。在两位教授提出的诸多宝贵意见的基础上,于 2005 年 12 月完成本书第二稿。之后,我们有幸参加了由南京大学主办的“中国城市化研究海内外研究组专题讨论会”以及顾朝林教授主持的国家自然科学基金重点项目(40435013)的研究工作,并将部分研究成果充实到本书中,于 2006 年 3 月完成了本书第三稿。之后不断修改完善,2006 年 6 月经研究组成员共同努力,最终成稿。本书写作过程中,自始至终得到了黄杏元教授的悉心指导,李满春教授、杨桂山研究员和朱传耿教授对本书提出了许多有益的建议,德高望重的陈述彭院士在百忙之中欣然为本书作序。

本书的研究工作得到了**国家自然科学基金(40301038)**的资助;另外,**国家自然科学基金重点项目(40435013)**也给予了资助,让 GIS 空间关联模式的发现有了鲜活的实证,为 GIS 与人文地理学的结合提供了典型案例。在这些科研项目的研究工作中,研究组成员做了大量的创新性研究工作,部分研究成果被收集到本书中。朱传耿教授研究了中国流动人口的空间分布结构模式,何增有博士研究了空间离群点以及频繁邻近类别集的挖掘算法,祈毅博士程序实现了位置关联规则的挖掘算法。

全书共分 8 章。第 1 章简要指出了研究背景,阐释了相关概念,介绍了主要研究内容;第 2 章分析了 GIS 空间关联模式发现的地理空间认知基础;第 3 章阐述了 GIS 空间关联模式发现的认知过程,构建了 GIS 空间关联模式发现的基本框架;第 4 章提出了空间关联位置模式发现的方法,重建了相关问题的定义和算法;第 5 章和第 6 章分别探讨了空间关联结构模式发现的探索性分析方法和证实性分析方法;第 7 章给出了空间关联结构模式发现方法在流动人口的空间分布、城镇群体的空间结构、城市化的格局与过程、经济发展的区域差异等方面的应用;第 8 章进行了全书总结。

在此,谨向为我们提供资助的单位、参与研究的全体成员以及提出宝贵意见的各位有识之士表示诚挚的谢意!向给予我们支持、理解和帮助的同行人、学者表示诚挚的谢意!另外,还有许多人为了本书的顺利出版提供了帮助,在此一并致谢。本书引用了许多学者的研究成果和学术思想,虽然已有标注和说明,但在主要参考文献中未能一一列出,敬请谅解!

空间数据挖掘和知识发现方面的研究成果已有很多,我们仅结合自己的研究心得,关注从 GIS 数据库中发现空间关联模式,试图为空间数据挖掘和知识发现添枝加叶。由于作者知识所限,学识疏浅,加之时间仓促,难免存在不当之处,恳请读者批评指正。

目 录

序 一

序 二

前 言

第 1 章 绪 论	(1)
1.1 研究背景	(1)
1.2 基本概念	(3)
1.3 研究内容	(6)
主要参考文献	(8)
第 2 章 空间关联模式发现的认知基础	(9)
2.1 认知、空间认知与地理空间认知	(9)
2.2 地理空间认知的心理学基础	(13)
2.3 地理空间认知的心理学模型	(15)
2.4 地理空间认知假设	(18)
2.5 地理空间认知理论	(19)
2.6 地理时空认知	(23)
2.7 小结	(27)
主要参考文献	(28)
第 3 章 空间关联模式发现的认知过程	(30)
3.1 从现实世界到概念世界	(31)
3.2 从概念世界到 GIS 世界	(34)
3.3 从 GIS 世界到空间关联模式发现	(60)
3.4 空间关联模式发现的基本框架	(62)
3.5 小结	(63)
主要参考文献	(65)
第 4 章 空间关联位置模式发现	(68)
4.1 位置关联规则发现	(68)
4.2 空间离群点挖掘	(86)
4.3 频繁邻近类别集挖掘	(91)
4.4 小结	(94)
主要参考文献	(95)
第 5 章 空间关联结构模式的探索性发现	(97)

5.1	探索性空间数据分析	(98)
5.2	半变异函数	(99)
5.3	空间自相关	(101)
5.4	空间马尔可夫链	(111)
5.5	模式可视化	(113)
5.6	小结	(118)
	主要参考文献	(119)
第 6 章	空间关联结构模式的证实性发现	(120)
6.1	证实性空间数据分析	(121)
6.2	空间关联模型	(121)
6.3	空间关联模型的估计	(124)
6.4	空间关联效应的检验	(127)
6.5	空间关联模型的设定	(132)
6.6	空间关联模型的验证	(135)
6.7	小结	(137)
	主要参考文献	(137)
第 7 章	空间关联模式发现的应用	(139)
7.1	流动人口的空间结构模式:分异与区划	(139)
7.2	城镇群体空间结构模式:集聚与扩散	(148)
7.3	城市化时空演进结构模式:格局与过程	(159)
7.4	区域经济发展空间结构模式:趋同与分异	(179)
7.5	小结	(204)
	主要参考文献	(205)
第 8 章	总结与展望	(208)

第 1 章 绪 论

1.1 研究背景

1. 地理数据不断丰富需要空间分析技术进步

地理学对空间现象的描述和空间规律的揭示始终建立在空间分析的基础上。随着地理学的不断发展,以及信息科学的日益进步,空间分析由对地理实体空间特征的概括式、图解式的描述逐渐发展为对空间数据库中的信息提取、空间关系的表达、空间模式的发现和空间机理的揭示等。早期的地理空间分析是基于地图的,地理学者往往利用地图——这种地理实体“自我表述”的工具来得到空间要素间的距离、方位和面积等信息,并将这些信息作为其他较高层次信息(如社会、经济、文化和军事等)分析的空间基础。即使是 20 世纪 50 年代的地理学计量革命,也只是将数理统计的方法引入具有空间出处的数据分析中,并将统计结果利用地图空间化或专题化。所以,基于地图的空间分析在此阶段的主要作用是揭示地理对象的空间异质特征。由于缺乏有效的空间数据采集、储存以及传输技术的支撑,传统的地图所能承载和表达的空间信息非常有限,更无法在多空间尺度和时间幅度上来进行空间分析,从而使得空间数据成为制约地理空间分析的重要“瓶颈”。20 世纪 60 年代以来,随着地理信息系统和遥感等地理信息科学技术的飞速发展,使得地理空间数据的储存、检索、制图和显示功能越来越完善,以 GIS 数据库为主体的空间数据库得到极大发展,数据量极大丰富且数据还以指数级方式不断增长,这使得地理学面临着数据丰富而分析不足的尴尬局面;迫切需要高效、精确、科学地分析这些数据,地理空间分析的平台由地图转向空间数据库。而 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 Goodchild 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”。所以,基于 GIS 数据库的空间分析不仅是对地理数据的空间分布、空间形态和空间关系等进行精确地分析;更需要从地理数据的空间关系中挖掘新的空间信息、发现有用的空间模式、揭示内在的空间机理,并进行空间知识的归纳、空间规则的推理和空间过程的模拟等。

2. 空间数据挖掘和知识发现推动地理认知深化

随着数据库技术的成熟和信息管理系统应用的普及,在海量空间数据背后隐藏的具有决策意义的信息日益受到人们的重视。人们开始在数据库中探索事先并不知道的,但潜在有用的、新的结构形态或关系特征,即关于数据的高层次的信息结构和知识。20 世纪 90 年代,“从数据库中发现知识”(KDD)及其核心技术——数据挖掘(DM)应运而生。作为地理信息的载体,地理空间数据具有不同于一般数据的特质,即数据的空间关联性、

空间和时间上的多尺度性、属性数据的多维结构、数据表达的不确定性等等,使得基于地理空间数据库的数据挖掘和知识发现特别需要综合知识的支持和结合。空间数据挖掘和知识发现(spatial data mining and knowledge discovery,SDMKD)作为数据挖掘的一个新的研究分支,是指从空间数据库中提取隐含的、用户感兴趣的空间和非空间的模式和具有普遍特征的知识的過程。SDMKD挖掘的空间知识主要包括空间的关联、特征、分类和聚类等规则(李德仁等,2002)。GIS数据库是空间数据库发展的主体,作为一类最重要的空间数据库,从GIS数据库中提取的各种知识包括空间分布、空间分类、空间关联、空间分异、空间趋势和空间演变等。其中地理数据的空间关联是从GIS数据库中发现知识的重要基础,也是从空间数据库中进行数据挖掘和知识发现的重要条件。但目前从GIS数据库中进行数据挖掘和知识发现的过程,大多数由关系数据库和事务数据库中挖掘关联、分类、聚类、区分和特征等规则的相应算法在空间数据库中应用和扩展得到(Koperski and Han,1995;Koperski et al.,1996),尚未能首先基于地理数据的空间关联特征,也缺乏完善的理论指导(邸凯昌,2001)。

3. 地理数据的空间关联要求数据组织变革

空间数据挖掘和知识发现的对象是空间数据库。由于空间数据具有空间拓扑结构和位置关系,并且相互之间存在空间关联特征,所以较之关系型数据库和事务型数据库,空间数据库在数据组织结构、数据访问方式和数据存储机制等方面有诸多不同之处,且要复杂得多。这一点尤为明显地体现在空间数据的组织方面。作为概念层次的空间数据模型,空间数据组织反映了现实世界中地理实体及其相互联系,并在一定程度上映射了人们对地理现象空间认知的结果。在传统的空间数据库中,空间数据与属性数据是分别存储的,即空间数据(图形数据)以文件格式存放,非空间数据(属性数据)则存放在关系数据库中,形成“文件+关系数据库”的二元管理模式。由于空间数据库包含大量的空间关系信息,需要按照复杂的多维空间索引结构组织数据,二元管理模式在数据查询、索引和数据共享等方面都存在不少缺陷。目前也有空间数据管理平台具有了将空间数据与属性数据一体化储存管理的能力(如ArcGIS的Geodatabase),但其中空间数据分层组织的方式并不是真正的对真实地理空间的描述或表达。我们周围的感知世界并不是完全按照点、线或面进行抽象和组织的。人们认知的地理世界是一组具有高度空间关联结构的物质实体,也就是说,这些实体拥有一组允许人们在空间相关性基础上进行分类的共同属性。对地理实体属性和关系共性的认识是人们认知的起点(Usery,1993)。可见,人们对现实世界的空间认知基于地理特征,并强烈依赖于地理实体的空间关联。这种认知方式的存在要求变革当前的分层组织空间数据的方式,着重依据地理实体的空间特征及关系(形态、位置、邻接、结构等)来组织空间数据。

4. 地理学的发展需要从模式发现到机理分析的研究范式

作为揭示空间分异规律和区域之间联系的一门学科,地理学以“格局-结构-过程-机理”的研究思路贯穿始终(冷疏影等,2001)。基于空间(模式)视角对地理对象的结

构特征进行分析是地理研究的重要切入点。而揭示地理过程及其微观机理已成为地理学研究的重要目标。空间信息技术的发展极大地提高了获取地理信息的能力,为量化的区域研究提供了数据保证(陈述彭, 1999);同时,信息科学为地理学提供了更加高端的实验手段,使得地理学开展更高分辨率的区域研究成为可能。总之,在科学技术不断发展的今天,学科交叉与借鉴,分析方法的提高已使得地理学逐步由定性描述、概括总结为主的经验科学向以实验为主要手段,机理探讨为主要目的,数学模型表达为主要特色的现代科学迈进。然而,我国的地理学由于理论的归纳和创新缺乏科学的研究范式的支撑,特别是对地理学综合集成研究方法探索的不足,已经在某种程度上影响了地理学对前沿科学问题的研究(宋长青、冷疏影, 2005)。地理学的不断发展,在方法论上需要一体化的、综合的、集成的从空间模式发现到空间机理分析的研究范式。

1.2 基本概念

1. 空间认知

空间认知是人们对出现于周围环境中的各种事物或现象的存在、变化的方式以及它们相对位置的认知过程和认知能力,是对事物和现象的发生、影响、因果进行分析研究的基础。地理学家对空间认知的理解建立在对地理空间理解的基础上。地理空间是具有地理定位的几何空间,是联系特定对象集合关系的规范框架。地理空间认知广泛涉及地理知觉、地理表象、地理概念化等,进而成为空间关联、空间推理、空间信息获取与储存、空间知识发现等的重要基础。所以,地理空间认知是 GIS 数据表达与组织的桥梁和纽带。研究地理空间认知对 GIS 以及从 GIS 数据库中发现知识具有重大作用,主要涉及到 GIS 的三个主要功能:地理参考数据的存储、表达与空间数据分析,涉及的问题主要有:怎样对地理空间进行概念化? 怎样对地理空间进行推理? 怎样用地图语言表达空间信息? 怎样设计用户界面以促进空间数据模式的有效且精确地解译以及 GIS 操作? 怎样训练人们使用 GIS? 为什么不同的用户群对地理信息的感知和解译有差别? 使用 GIS 研究人们认知的前景怎么样?

2. 空间异质

现实世界存在的每一种事物和现象都包含着自身特殊的矛盾,这构成一事物和现象区别于其他事物和现象的特殊属性。从矛盾的角度看,这种属性就是差异性。在地球表面上,每一个空间位置上的事物和现象都具有区别于其他位置上的事物和现象的特点,这种空间差异性即空间异质性(spatial heterogeneity)(Anselin, 1988)。地理学是着重描述和解释地球表面地区差异的一门学科(哈维, 1996),揭示人-地系统相互作用的空间分异规律及其动力学机制一直是地理学的主要研究内容(蔡运龙等, 2004; 冷疏影、宋长青, 2005)。

空间异质与地表事物和现象的空间位置,如东、西、南、北、中等直接相关。地表事物和现象在空间上是异质的,一方面在于各种事物和现象本身在空间上缺乏平稳的结构;另一方面也在于空间单元本身远不是均质的,在面积、大小、形状等方面往往存在很大的差别。从统计学角度来看,空间异质性是指研究对象在空间上非平稳,这与经典统计学所要求的所有样本观测值都来自于同一个总体的假设相违背(Anselin, 1988)。空间异质性可以用下式表达:

$$y_i = f_i(x_i, \beta_i, \epsilon_i) \quad (1-1)$$

式中, i 代表空间观测单元; f_i 表示因变量 y_i 与自变量 x_i 、参数向量 β_i 和误差项 ϵ_i 之间的函数关系。

空间异质性意味着发现整体的或总的空间模式并不是空间数据挖掘的最终目的,必须识别空间模式的局部差异,揭示事物和现象的空间分异特征。

3. 空间关联

世界是普遍联系的。现实世界的事物和现象之间既相互区别,又相互联系。空间关联(spatial association)是事物和现象在空间上的相互依赖、相互制约、相互影响和相互作用,是事物和现象本身所固有的属性,是地理空间现象和空间过程的本质特征。“近朱者赤,近墨者黑”。地球表面上的事物和现象,如地表温度、土壤湿度、植被特征等,以空间分布的相似性反映了这种千丝万缕的联系。Tobler(1970)进一步将空间关联的普遍性上升为地理学第一定律(First Law of Geography),即“地表所有事物和现象在空间上都是关联的;距离越近,关联程度就越强;距离越远,关联程度就越弱”。

空间关联是自然界存在秩序、格局和多样性的根本原因之一(Goodchild, 1986)。造成空间关联的主要原因在于,地表事物和现象之间的位置、距离等空间要素具有非常重要的作用。由于位置的毗邻,造成了地表事物和现象之间的空间相互作用以及空间溢出效应,形成一定的空间组织或空间结构。该种形式的空间关联称为实质性的空间依赖(substantive spatial dependence)。这表明,一个位置上的事物和现象可由空间系统中其他位置上的事物和现象决定或部分决定,可用下面的空间过程加以表达(Anselin, 1988):

$$y_i = f(y_1, y_2, \dots, y_N) \quad (1-2)$$

式中, y_i 表示变量 Y 在第 i 个空间单元上的观测值, $i \in S$, S 是所有空间单元的集合。

空间关联或空间依赖性的产生还有可能来自于各种测量因素(Anselin, 1988; Anselin and Getis, 1992)。例如,研究对象本身的空间范围与观测的空间单元之间不能完全匹配,以及空间单元的任意合并等,都会造成某种程度的测量误差,而该误差有可能通过空间单元的边界发生溢出,从而造成冗余形式的空间依赖(nuisance spatial dependence)。

从统计学的角度来看,空间关联是指空间数据之间缺乏独立性,违背了经典统计学

的样本独立不相关假设。空间关联主要指空间自相关(Anselin, 1988)。空间自相关与时间序列分析中的时间自相关概念相类似,但两者具有本质的区别。空间自相关是多维的,而时间自相关是一维的。用来分析时间自相关的时间序列分析方法不能直接用来处理空间数据的自相关特征。

只要数据是以空间单元组织的,就有存在空间关联的可能性。空间关联意味着在空间数据挖掘过程中有必要发现数据之间的空间关联关系,揭示形成不同空间结构模式或空间分异的内在关联机制。

4. 空间模式

地表事物和现象在地理空间上并不是独立、随机分布的,而是相互联系、相互作用,呈现一定的空间分布格局。简而言之,空间模式(spatial pattern)就是地表事物和现象的空间分布格局、特征及其演变等。在不同研究领域,空间模式的表现形式不同,如景观生态学中的缀块-廊道-基底模式(邬建国, 2000)、区域经济学中的核心-边缘结构(崔功豪等, 1999)等。但无论如何,时间上和空间上的关联是各种不同的空间模式之所以产生、变化和发展的根本原因之一(Cliff and Ord, 1981)。图 1.1 展现了空间模式在时空维上的演变过程。

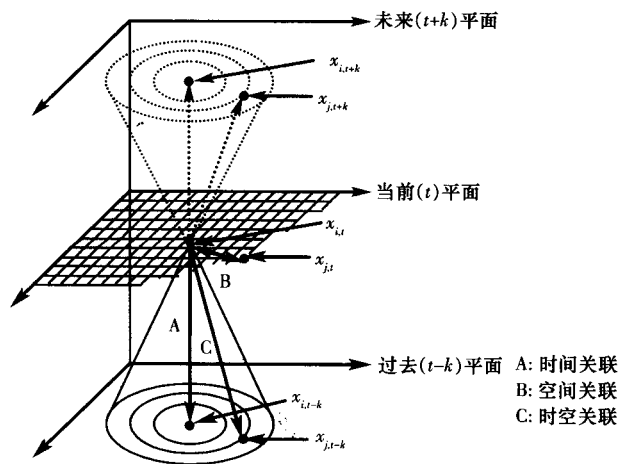


图 1.1 区域单元在时空维上的关联模式

(据 Cliff and Ord, 1981)

“空间模式”是个被广泛使用的概念。从不同的角度,按照不同的标准,可以得到不同空间模式,如点模式、线模式和面模式等。从成因上来看,空间模式大致包括以下几种类型:①时间关联模式(temporal association pattern)。时间不同、空间位置相同的事物和现象之间的关联模式(图 1.1A)。②空间关联模式(spatial association pattern)。时间相同、空间位置不同的事物和现象之间的关联模式(图 1.1B)。③时空关联模式(spatial-temporal association pattern)。时间不同、空间位置不同的事物和现象之间的关联模式(图 1.1C)。以地区经济增长为例,当一个地区的招商活动增

加时,该地区经济得到持续发展,属时间关联模式;当该地区的招商引资同时引起周边地区投资活动增加,属空间关联模式;当该地区的招商引资引起周边地区经济增长,属时空关联模式。

空间关联模式在本质上更加强调了不同空间位置上的事物和现象之间的关联所造成的空间模式,它具体包括侧重空间数据的空间关联位置模式和侧重属性数据的空间关联结构模式两种。通过观测到的空间关联模式,可以建立与空间过程有关的假设,并进一步指出空间模式形成的动力机制。

1.3 研究内容

以地理空间认知为基础,GIS 空间关联模式发现(简称“空间关联模式发现”,下同)的广义过程为主线,关注地理空间现象和空间过程的本质特征——空间关联在 GIS 数据中的隐含表现:空间关联模式,探讨从 GIS 数据库中挖掘空间关联位置模式和空间关联结构模式的方法集以及这些方法的实际应用。除了本章外,其他章节的构成如图 1.2,研究内容主要包括以下五个方面。

(1) 空间关联模式发现的认知基础

辨析认知、空间认知和地理空间认知之间的相互区别和内在联系,探讨地理空间认知的心理学原理,形成有关地理空间认知假设,总结有关地理空间认知理论,以此为基础,理解地理时空,探讨 GIS 时空表达,为 GIS 数据组织和空间关联模式发现奠定理论基础。

(2) 空间关联模式发现的认知过程

把空间关联模式发现看作一个“现实世界—概念世界—GIS 世界—空间关联模式发现—现实世界”的广义过程,分别探讨其中的涵义和内容,深入理解空间关联模式发现所涉及的问题和过程,为空间关联模式发现提供理论依据和方法基础。

(3) 空间关联模式发现的总体框架

在相关认知理论和认知过程的基础上,构建空间关联模式发现的总体框架,为空间关联模式发现提供技术准备和支撑。

(4) 空间关联位置模式发现的方法

总结前人研究,建立发现空间关联位置模式(位置关联规则)的方法和步骤,编程实现其中的示例,同时对在空间关联位置模式发现过程中的相关问题进行重新定义,构建新的算法。

(5) 空间关联结构模式发现的方法

建立发现空间关联结构模式的方法集,给出这些方法在相关问题研究中的实证,说明这些方法在实际应用中的作用。

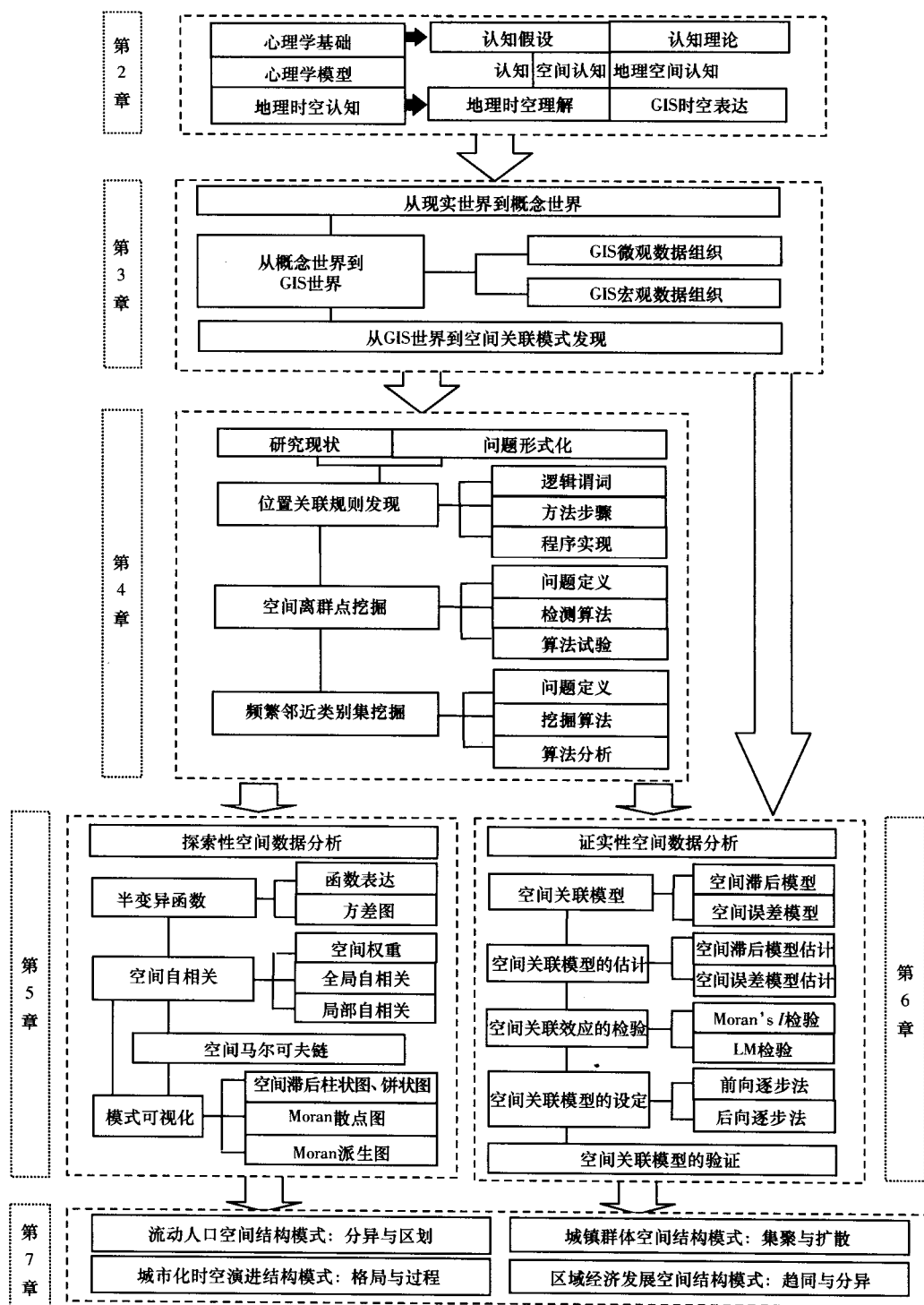


图 1.2 本书基本构成和研究内容