

地球信息科学基础丛书

空间模式的测度及其应用

毛政元 李 霖 著

科学出版社

北 京

内 容 简 介

本书是地球信息科学基础丛书之一.书中将地理对象抽象为“空间模式”,将与“解决空间问题有关的定量研究”归纳为空间模式的表达、转换、测度和应用四个方面并在每一个方面展开系统研究.

本书共分七章.第一章阐述空间模式的定义及其与相关概念和相关研究领域的关系,论证空间模式的研究意义.第二章论述“空间模式表达”涉及的分析空间、数据模型、变量标度和存贮方式等主题,最后归纳出空间模式的分类体系.第三章系统研究点、区、面三类空间模式相互转换所涉及的各种方法.第四章引入空间随机过程和几何概率阐述了空间模式的测度原理,发展和完善了点、区、面三类空间模式及其相互关系的测度方法.第五章针对现有空间模式测度方法在研究集聚型点模式中的局限性,提出基于几何概率的集聚型点模式结构信息提取思想以及相应的指示函数和算法,推导指示函数的解析表达式,验证指示函数和算法的可行性.第六章设计了集聚型点模式结构信息提取算法,并以居民地坐标数据为例实现了该算法.第七章是全书的总结和归纳.

本书可供从事空间数据挖掘、地学和地球信息科学研究的科研人员以及高等院校有关专业的师生阅读参考.

图书在版编目(CIP)数据

空间模式的测度及其应用/毛政元 李霖著.—北京:科学出版社,2004
(地球信息科学基础丛书)

ISBN 7-03-013642-X

I.空… II.①毛… ②李… III.①地理信息系统②遥感技术—应用—测量③全球定位系统(GPS)—测量 IV.P2

中国版本图书馆 CIP 数据核字(2004)第 065732 号

责任编辑:彭胜潮 汪福舫/责任校对:鲁 素

责任印制:钱玉芬/封面设计:王 浩

科学出版社发行 各地新华书店经销

*

2004 年 9 月第 一 版 开本:787×1092 1/16

2004 年 9 月第一次印刷 印张:8 3/4

印数:1—2 500 字数:178 000

定价:26.00 元

(如有印装质量问题,我社负责调换〈路通〉)

序 言

空间信息与人类的生存和发展息息相关,它显式地或隐式地包含于空间数据之中.长期以来,为了得到并有效地管理空间信息,人类投入了大量的人力、物力和财力从事有关空间数据表达、获取、存贮、传输以及运算、分析等方面的试验和应用研究,这些努力导致了地理信息科学(Geographical Information Science)的兴起.毛政元博士和李霖博士合作完成的专著《空间模式的测度及其应用》正是这个领域的一部力作.

作为一个新兴的交叉学科,地理信息科学脱胎于一系列的相关学科,如何将各相关学科融合为一个有机的整体,使之成为一门内容体系完整又相对独立的学科一直是一个具有挑战性的难题,尽管这种倡导由来已久.本书对此进行了大胆、有益的探索.作者从分析空间信息的本体(地理对象)和载体(空间数据)以及两者之间的关系入手,然后引入分析空间的概念,接着导出空间模式的定义,再提出研究空间模式的内容体系,到最后逐一展开论述,思路自然而流畅.在这个过程中,作者实践和尝试了一种新的空间信息研究理念,并在构建空间模式的内容体系、完善空间模式的测度方法两个方面取得了重要的实质性进展.

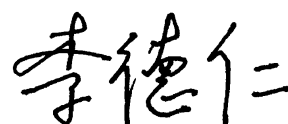
本书将研究空间模式的内容体系归纳为表达、转换、测度(包括关系测度)和应用四个方面,并就每一个方面展开系统的论述.作者首先根据空间数据的表达习惯建立起空间模式与地理对象的关系,从数据模型、变量标度和存贮方式等角度研究空间模式的现有表达规范,并提出基于空间认知的空间模式分类体系;然后详细地讨论了空间模式转换所涉及的各种方法;接着深入地探讨了空间模式测度的原理和方法,其中着重研究了基于几何概率测度点模式的新方法;最后设计了集聚型点模式结构信息提取算法,并以居民地坐标数据为例实现了该算法.全书思路新颖、结构严谨、语言精练,章节安排所体现的内在逻辑性增强了本书的可读性.

本书在理论上的主要贡献有二:其一,通过将地理空间中的地理对象抽象为分析空间中的空间模式,在现实的空间问题与抽象的空间模式测度之间建立起内在的对应关系,使得地理学有可能最终摆脱一一对应的经验研究;其二,成功地将几何概率引入地理信息科学之中,作为测度空间分布的客观标准,从而为空间相关、空间相似、空间扩散、空间竞争等各种与空间分布相关的专题研究提供了新的思路.本书在方法论上的主要贡献在于发展和完善了空间模式的测度方法.通过对数学分析方法的恰当运用,作者提出了基于几何概

率测度点模式的样方法和函数法,后者不仅可以成功地测度集聚型空间点集的结构信息,而且也能够有效地解决模式识别中的最佳分类数问题.

进入 21 世纪,空间数据获取技术飞速发展,计算机硬件性能指标记录不断刷新,网络技术日新月异,所有这一切昭示着一个新的时代已经来临.我们应该紧紧抓住技术进步带来的机遇,为促进地理信息科学的发展而努力.我很高兴读到这本专著,也希望有更多的学者关注这个主题,共同将这个领域的研究水平推向新的高度.

中国科学院院士
中国工程院院士



2004 年 6 月 28 日
于珞珈山(武汉大学)

前言

地理学源于人类对自身生存环境的关注,其发展历史悠久,但至今仍不完善.地理理论是一些基于定性分析的假说,一般缺乏与之对应的定量模型,不能精确地揭示空间过程的系统动力学机制,直接地导致地理学理论和方法的脱节,并间接地造成地理学界经常出现面对同样的地理事实众说纷纭的混乱局面.自 20 世纪 50 年代以来,越来越多的地理学者试图通过定量化摆脱这种困境(Haggett, 1966; Harvey, 1971; Bailey and Gatrell, 1995),他们的努力导致了定量地理学(Quantitative Geography)的兴起.

定量地理学在其发展过程中逐步渗透到地理学的三大研究主题(描述、解释和预测)之中,成为地理学最具活力的分支和方向.最先受到关注的是针对预测的空间模型研究.计量革命早期的一代地理学者在否定地理环境决定论、扬弃区域学派(以 R. Hartshorne 为代表的一种学派,认为地理学的任务在于辨明区域差异)的基础上,强调空间组织(即空间本身对地理事物的影响和决定作用)的观点,并以定量方法模拟空间现象及其变化与空间位置或空间实体的关系,空间模型的运用一时成为地理研究中的时尚.其中有成功的应用案例,但误用、甚至滥用的情况也不少.由于当时地理信息系统尚未出现,计算机能够处理的实际上只是空间数据中的属性数据部分,而几何数据要么不予考虑(这样空间数据已退化为一般的统计数据),要么仍采用传统的手工处理方式.许多人因此将空间模型的失败归咎于由技术条件限制造成的空间数据描述失真.与计量革命早期相比,今天的技术条件已经有了根本的改善,“3S”技术的逐步成熟和“数字地球”的全面实施正在为空间数据的获取、描述和处理提供全方位的定量化支持,然而,毋庸讳言,地理研究中成功的模型仍不多见.空间模型失败的另一个深层次原因逐渐凸现——这便是定量地理学本身的基础理论研究停滞不前所致.与对技术实现主题的高度重视(毋河海, 1991; 龚健雅, 1993; 毋河海, 龚健雅, 1997; 李霖, 1994, 1995, 2000)相比,地理信息科学对空间信息本身及其与空间过程的关联缺乏必要的关注,相关研究文献(李霖, 李德仁, 1994; 杜清运等, 2000)也少得多,定量地理学的发展主要地依赖于技术进步而不是其自身基础理论的完善.

与技术的巨大进步和技术为其提供的发展空间相比,定量地理学本身的进步缓慢,原因在于其研究对象和目标的定位模糊以及由此造成的大量重复研究.本书认为,定量地理学的研究对象不是具体的地理对象,而是映射于分析空间中的抽象地理对象.实际的地理对象是数不胜数、各式各样的,但却可以在抽象的层次上归纳为数量有限的若干类型;我们不必要也不可能对具体的地理对象进行一一对应的穷举研究,但却能针对抽象地理对象的每一类型进行合理的深入探讨.相应地,定量地理研究的目标不是为了得到针对某一个具体地理对象的研究结论,而是为了从总体上提供一种地理研究的方法论范式,指导设计具体地理问题的解决方案,并最终跨越存在于地理学各个分支学科之间的鸿沟,建立统一的地理学.地理学的研究对象属于复杂系统,其中机械运动、物理运动、化学运动、生物运动和社会运动相互交织渗透,在多种空间尺度和时间尺度上制约着其发展演变.目前人类的认识水平尚不能从整体上把握这类复杂系统,对此,地理学者的策略是将其划分为不

同的子系统并建立相应的分支学科.地理学各个分支学科的理论大多源自对邻近学科理论的借鉴和改良,地理学作为一个整体存在的价值在于从特定的角度或者在更高的层次上将这些从其他学科借鉴、改良过来的理论整合为一个有机的整体,而地理学发展过程中多次分合演变的历史说明了这种整合的不易和已有整合方式的不彻底性,定量地理学的任务在于为此提供方法论支持.

本书将抽象的地理对象称为空间模式,并从空间认知的角度将空间模式分为变量模式和实体模式,实体模式进一步细分为点、线、区、体 4 种类型.空间模式测度是有关空间模式研究的核心内容,Clark 和 Evans(1954)最早提出了测度点模式的 R 统计量,可视为该领域的开创性贡献.按照本书关于空间模式的定义,迄今为止空间模式测度方面已取得的代表性研究成果包括点模式测度的样方法和函数法、区模式测度的指标法和两个区模式关系测度的匹配法,可以归入该领域的相关成果还包括有关图像处理、DTM 和图论的一些内容.在现有工作的基础上,本书提出了空间模式的研究体系,将其归纳为表达、转换、测度(包括关系测度)和应用 4 个方面;通过引入几何概率为改善空间模式的测度效果、提取空间模式的结构信息提供了理论依据;推导出正方形区域内基于几何概率的 P 函数全部定义域上的解析表达式,设计出基于 P 函数测度点模式的 H 函数和测度区模式整体结构的二维有序聚类算法,设计和实现了集聚型点模式结构信息提取算法.

感谢李德仁院士在百忙中抽空为本书作序,感谢陈述彭院士和王钦敏院士对作者的关心与支持,感谢张超教授和闫国年教授提出的宝贵意见.

书中错误之处,恳请读者指正.

摘要

空间模式是介于空间原型和空间模型之间的研究领域,其目标是解释空间现象和空间过程、探索地理学的方法论范式.空间模式测度是研究空间模式的核心内容,是许多专题研究工作者感兴趣的内容,不同分析空间中的空间模式测度可能对应空间分布、空间相关、空间扩散、对象相似性描述、类的可分性等理论问题以及资源分配、区位分析、区域规划、交通布局等应用问题,是一个有重大理论意义和应用价值的研究课题.现有关于空间模式的研究至少存在以下 3 个方面的不足:其一,内容零散,不成系统;其二,忽略了空间模式的结构信息,测度指标只能说明一种平均统计状况,不能全面反映空间模式的特征;其三,缺少相关算法的研究,与应用问题衔接困难.本书围绕空间模式这个主题,全面研究了空间模式的表达、转换、测度、应用 4 个方面的内容,试图通过建立空间模式的研究体系、引入几何概率、设计相关算法促进空间模式研究领域的发展和完善.本书的主要论点是:空间模式是联系空间原型和空间模型的纽带,是地理学中承上启下的重要研究层次,是建立空间模型的基础.尽管有关空间模式的研究在某些具体方面取得了一些进展,但其研究体系的整体框架尚未建立;空间模式的研究体系可以概括为表达、转换、测度(包括关系测度)和应用 4 个方面,其中表达和转换是测度的基础,应用是测度的目的;引入几何概率可以改善空间模式的测度效果,为提取空间模式的结构信息提供理论依据;与空间模式测度相关的算法研究是空间模式研究领域不可分割的重要内容.全书共分 7 章,各章内容如下.

第一章 绪论.本章通过引入“地理对象”的概念将“空间模式”定义为“地理对象的抽象”,并从空间认知的角度划分地理对象和空间模式的子类.提出空间原型、空间模式和空间模型 3 大研究层次,论证空间模式在空间原型和空间模型之间的纽带和桥梁作用,从理论和方法两个角度阐明研究空间模式的意义.分析“空间模式”和“空间模型”两个概念之间的区别、“空间模式”和“模式”两个概念之间的联系.探讨空间模式与其 3 个相关研究领域——模式识别、空间数据挖掘、地理信息系统的关系.其中模式识别研究基于概率论、模糊数学、人工智能和知识工程的对象分类判识问题,是空间模式研究领域的理论基础;空间数据挖掘研究如何运用确定型空间数据分析方法、探索型空间数据分析方法、软计算方法、可视化方法从一个已经建立的空间数据库中发现知识,是空间模式研究领域的理论基础;地理信息系统研究基于计算机技术的空间信息处理问题,是空间模式研究领域的技术基础.

第二章 空间模式的表达.本章通过引入“分析空间”(由几何维、属性维和时间维任意组合而成的 N 维抽象空间)的概念进一步明确空间模式的含义,指出空间模式是地理对象映射在分析空间中的像,空间模式的数据集是此像的记录,空间模式的可视化是此像的可视表达.针对概念层次上的空间数据模型(对象模型、场模型)和逻辑层次上的空间数据模型(矢量模型、栅格模型)论述空间模式数据集的具体形式,说明空间模式与其数据集、可视化三者之间的关系.阐述空间数据 4 种标度(名称标度、顺序标度、区间标度、比率

标度)的性质、相互关系和可视化方案,探讨多维标度统一问题.介绍空间数据 5 种存贮类型(对象、事件、空间连续数据、区数据、流数据)的含义和相互关系.根据空间模式的数据集形式、地理变量的标度和地理实体的几何维归纳出空间模式的分类体系,分析空间变量模式和空间实体模式的本质特征与记录方式.

第三章 空间模式的转换.本章论述点、区、面 3 类空间模式相互转换所涉及的各种方法(邻近多边形、密度计算、分布中心计算、地理编码、特征点提取、聚类分析、插值分析和抽样).提出基于距离、维数和发生元性质的广义邻近多边形分类方案,分析 Delaunay 三角网的性质,说明生成 Delaunay 三角网的分治算法、逐渐插入算法和三角网生长算法的步骤及其各种实现方案的效率,阐述狭义邻近多边形基于 Delaunay 三角网的间接生成算法和基于矢量、栅格的直接生成算法.阐述局部平均法、局部加权平均法、中心估计法和邻近多边形法 4 种密度插值算法的思路、实现步骤、适用场合,比较四者的优劣,针对密度计算中存在的边缘效应提出修正算法.论述分布中心的数学含义,设计基于三角形几何中心计算公式的任意多边形分布中心的矢量求解算法、基于逐级细分递推和积分思想的任意不规则区域分布中心的矢量求解算法以及任意区域分布中心的栅格求解算法.分析地理编码对应的问题原型,从自然和社会原因两方面论证不均匀空间分布存在的普遍性以及对应地理编码的重要意义,设计不均匀空间分布的地理编码算法.介绍面模式形态特征点的提取标准.指出作为归类依据的相似(关)系数和距离的各种形式,阐述其对应的含义,根据理论内涵和算法实现总结聚类方法的分类体系,比较各种聚类方法的优劣.提出插值算法的分类方案,阐述其中提及的各种插值方法的算法思想、实现方案和应用场合,论述插值效果较好的克里金插值法的理论依据,解释 3 种抽样方案的含义.

第四章 空间模式的测度.本章从方法和内容两个方面提出空间模式测度的分类体系,阐述基于概率论的空间模式测度原理,通过引入几何概率完善测度空间模式的理论基础,提出测度矩形和椭圆区域内线段长度和方向分布的 P 函数.分析测度点模式的各类方法的特点,比较 F 函数、 G 函数、 K 函数和 L 函数测度效果的差异,介绍普通样方法和可变样方法的取样方案、原理、步骤,将样方法扩展到各种分析空间类型,设计基于几何概率的样方法并以正方形区域为例说明其具体细节.讨论针对区模式的空间邻近性测度指标和空间滞后邻近性测度指标,论述空间自相关的含义,介绍测度区模式空间自相关水平的 Z 指标(二值邻接关系法)、 I 指标、 C 指标,设计测度区模式整体结构的二维有序聚类算法.论述以值结构和值分布结构测度面模式的思想,提出针对面模式测度的值区间频率函数、趋势面分析、特征点和特征线提取方法.将两个空间模式的关系作为空间模式测度的一种类型——关系测度,提出通过转换数据集的形式测度两个异质模式的思路,阐述两个同质模式关系测度的方法.

第五章 集聚型点模式结构分析.本章在分析集聚型点模式集聚结构多样性的基础上,针对现有空间模式测度方法在研究集聚型点模式中的局限性,提出以基于几何概率的指示函数图形峰值个数和组合知识为理论依据的点模式结构信息提取思想.推导集聚方向数、集聚距离数与集聚型点集的集聚中心数和集聚方向性的定量关系.推导基于几何概率测度正方形区域内集聚型点模式的 D 函数、 R 函数、 P 函数在其全部定义域上的解析表达式,分析 D 函数和 R 函数的图形特征,设计出基于 P 函数测度点模式的 H 函数.设计并实现集聚型点模式结构信息提取算法,并以单中心、双中心和多中心集聚型点集验证

该算法的可行性.分析 H 函数对于随机和均匀型点集的指示作用.

第六章 空间模式的应用研究.本章阐述空间模式测度方法和相关算法在分析居民地空间分布规律中的应用——集聚中心确定和提取.以集聚型点模式结构信息指示函数和结构信息提取算法为基础,通过使用聚类分析与 Voronoi 图设计并实现点集的集聚中心提取算法.通过对比点集的集聚中心与其几何中心、最佳区位点的差异,明确点集的集聚中心所代表的理论意义——点集的“形心”.分析点集的集聚中心在地图注记、空间分布、空间相关、空间扩散等研究中的重要作用,指出散列式居民地集聚中心在解决行政区划调整、乡村政权驻地选址、乡村公共设施建设、区域规划和交通布局等空间决策问题中的应用价值.

第七章 结论.本章总结了全书的内容,归纳了本书在研究空间模式测度方面的主要工作和得到的结论.

ABSTRACT

Spatial pattern is a transitional research field between spatial prototype and spatial model, the aim of which is to explain spatial phenomena and spatial processes, and form the geographical methodological paradigm. The Measurement of Spatial Patterns (MSP), an interesting topic of many thematic studies, is the kernel in the research area of spatial patterns. MSP of different analytical space can correspond to theoretical problems such as spatial distribution, spatial correlation, spatial diffusion, similarity description of objects, classifiable classes, ... etc. and practical problems such as resource allocation, location analysis, regional planning, transportation distribution and so on. Therefore, MSP is a scientific topic with important theoretical significance and pragmatic value. There are at least three disadvantages in the current situation of studying spatial patterns. First, the related content is not systematical. Second, the structure information of spatial patterns is disregarded and thus the measuring indices, an average statistical indication, cannot reflect their overall characteristic. Third, the relevant algorithms are not included. Centering on the theme of spatial pattern this book systematically studies its representation, transformation, measurement, and application; attempts to promote the research area by establishing the system of studying spatial patterns, introducing geometrical probability, and designing relevant algorithms. The following are its main points. Spatial pattern, the link between spatial prototype and spatial model, an important bridging research level in geography, is the foundation of spatial modeling. Although there are some achievements in studying spatial patterns, its entire framework has not been built. The system of studying spatial patterns can be concluded as four components: representation, transformation, measurement (including relationship measurement), and application among which application is the purpose of measurement which is underlain by representation and transformation. Introducing geometric probability can improve the measuring effect of spatial patterns, and provide theoretical basis for extracting their structure information. The algorithm studying is an indispensable part related to MSP. The thesis consists of 7 chapters and the content of each chapter is as follows.

Chapter 1 introduction. In this chapter, the concept of spatial pattern is defined as “the abstraction of geographical objects” after explaining the concept of geographical objects. The geographical objects and spatial patterns are classified from the perspective of spatial cognition. The three research levels—spatial prototype, spatial pattern, and spatial model are put forward. The bridging effect of spatial pattern between spatial prototype and spatial model is demonstrated. The differences between spatial pattern and spatial model, the links between spatial pattern and pattern are analyzed. The relationship between spatial pattern and its three relevant fields—pattern recognition, spatial data mining, and GIS is discussed.

Pattern recognition—a field studying classification and recognition of objects based on probability theory, fuzzy mathematics, artificial intelligence, and knowledge engineering—is the theoretical basis of spatial pattern. Spatial data mining—a field discovering knowledge from an established spatial database by utilizing confirmatory spatial data analysis, exploratory spatial data analysis, optimizing-computation, and visualization—is the methodological basis of spatial pattern. GIS—a field studying the problem of processing spatial information based on computer technology—is the technological basis of spatial pattern.

Chapter 2 the representation of spatial patterns. At the beginning of this chapter, the connotation of spatial patterns is further clarified via introducing the concept of analytical space (an abstract space with N dimensions constructed by arbitrary combination of geometric dimension, attribute dimension, and temporal dimension). Spatial patterns are images of geographical objects mapped into the analytical space, the datasets of spatial patterns are the record of the images, visualization of spatial patterns is the visual representation of the images. Then concrete forms of spatial patterns' datasets are expounded at the concept level of spatial data model (object model and field model) and the logical level of spatial data model (vector model and raster model). The relationship among spatial patterns and their datasets, their visualization is explained. After that, the property, visualization schemes, and inter-relationship of 4 spatial data measurement categories—nominal, ordinal, interval, ratio—are clarified. The integration of multi-dimension measurement is probed. The connotation and inter-relationship of spatial data's 5 memory categories (objects, events, spatial continuous data, area data, spatial interactive data) are stated. Finally the classification scheme of spatial patterns is induced according to their dataset forms, the measurement of geographical variable, and the geometric dimensions of geographical entities. The essential traits and recording ways of spatial variable patterns and spatial entity patterns are analyzed.

Chapter 3 transformation of spatial patterns. This chapter concentrates on all kinds of methods involving transformation among point pattern, area pattern, and surface pattern, including proximal polygon, density computation, distribution centroid computation, geocoding, characteristic point extraction, clustering, interpolation and sampling. First, it raises the classification scheme of general proximal polygon based on distance definition, dimension number, and generator property; elaborates the property of Delaunay Triangulation; explains the steps of algorithms to construct Delaunay triangulation (Divide-and-Conquer Algorithm, Incremental Insertion Algorithm, Triangulation Growth Algorithm) and the efficiency of their implementation; expounds the indirect algorithm based on Delaunay Triangulation and the direct algorithm based on vector and raster to generate ordinary proximal polygons. Second, it states the thoughts, implementation steps, and right utilizable situations of 4 density computation algorithms (local average, local weighted average, kernel estimation, and proximal polygon), and compares their advantages and disadvantages; puts forth the algorithms to correct edge effect in density computation. Third, it illustrates the mathematical connotation of distribution centroid; designs the vector algorithms for computing distribution

centroid of any polygon based on the equation of triangular centroid, and any irregular area based on gradual hierarchical division, recurrence, and calculus. The raster algorithm to calculate distribution centroid of any region is also given. Fourth, it analyzes the corresponding problem prototype of geo-coding; demonstrates the existing universality of uneven spatial distribution and the corresponding significance from nature and society; designs the geo-coding algorithm for uneven spatial distribution. Fifth, it states the extracting standard of specific morphological points in a surface pattern. Sixth, it lists various forms of correlation coefficient and distance, which are utilized as the basis of classification; elucidates the corresponding meaning; concludes the classification scheme of clustering in accordance with theoretical connotation and algorithm implementation; and compares the advantages and disadvantages of various clustering methods. Seventh, it advances the classification scheme of interpolation algorithms; expounds their ideas, implementation plans, and utilizable situations; demonstrates the theoretical foundation of Kriging that is one of the best interpolation methods. The meaning of three sampling ways is also explained.

Chapter 4 the measurement of spatial patterns. This chapter focuses on the methodological system of MSP. Based on probability theory it elucidates the principle of MSP, which can be improved by introducing geometric probability; creates P function to measure the distribution of direction and length of line segment in a rectangle or an ellipse; analyzes the traits of all kinds of methods to measure point patterns; compares the differences of measuring effect among F, G, K and L function; states the sampling scheme, principle and steps of ordinary and variable resolution quadrat; expands quadrat to all the analytical space types; designs the quadrat in the light of geometric probability and gives the related details by taking a square as the case. It also discusses the spatial proximal indices and the spatial delayed proximal indices for area patterns; explains the meaning of spatial autocorrelation and its measurement such as index Z (Joins count statistic), I, and C; designs the 2-dimension orderly clustering algorithm for measuring the whole structure of area patterns. In this chapter, the idea to measure surface patterns with value structure and value distribution structure is demonstrated. The value-interval frequency function, the trend surface analysis, and the methods of extracting specific morphological points and lines are put forth for measuring surface patterns. The relationship of 2 spatial patterns is treated as a kind of MSP—relation measurement. The conception of measuring relationship between 2 heterogeneous spatial patterns through transforming the involved spatial patterns' datasets is offered. The relevant methods of measuring relationship between 2 homogeneous spatial patterns are elaborated.

Chapter 5 the structure analysis of clustered point patterns. This chapter mainly addresses the importance and methods of extracting the structure information. On the basis of analyzing cluster structure diversity of clustered point patterns, being directed against the limitation of presently available MSP methods in studying the clustered point patterns, it raises the idea of extracting structure information from point patterns which is theoretically

underlain by combination knowledge and the peak value number in the graph of indicative functions based on geometric probability. It infers the quantitative relationship among cluster direction number, cluster distance number, and cluster center number. It also infers the possible analytical form of D, R, and P function of measuring the clustered point patterns in a square area based on geometric probability, analyzes the graphic trait of D and R function, designs H function to measure point patterns based on P function. The algorithm to extract the structure information from point patterns is designed and implemented, and its feasibility is testified by utilizing point sets with 1, 2, and 3 cluster centers respectively. The indicative ability of H function to random and even point sets is analyzed.

Chapter 6 the application of spatial patterns. This chapter elucidates the application of MSP methods and relevant algorithms in analyzing spatial combination law of residence—the determination and extraction of cluster centers. It designs and implements the algorithm to extract cluster centers of a point set through clustering and Voronoi diagram in the light of structure information indicative functions and the structure information extracting algorithm of clustered point patterns. Through contrasting the differences of the cluster center with the corresponding geometric center and the optimizing location point of the same point set, it makes clear the theoretical meaning that the cluster center of a point set represents “shape-center”. It also analyzes the important function of cluster center of a point set in the research such as map annotation, spatial distribution, spatial correlation, spatial diffusion, and points out the practical values of cluster centers in scattered residences with respect to spatial decision such as administrative areas adjustment, location allocation of rural local governments, rural public facilities construction, regional planning, transportation distributing and so on.

Chapter 7 conclusion. In this chapter, the entire book is summarized. The main content and the conclusions in studying MSP are induced.

目 录

序 言

前 言

摘 要

ABSTRACT

第一章 绪论	1
1.1 空间模式的含义	1
1.2 空间模式的相关概念	2
1.3 空间模式的相关研究领域	3
1.3.1 模式识别	3
1.3.2 空间数据挖掘	3
1.3.3 地理信息系统	6
1.4 空间模式的研究意义	8
1.5 本书的研究目标与主要内容	9
1.6 本章小结	10
第二章 空间模式的表达	11
2.1 地理空间与分析空间	11
2.2 数据集与可视化	12
2.3 空间数据的标度	13
2.4 空间数据的存贮类型	14
2.4.1 对象	15
2.4.2 事件	15
2.4.3 空间连续数据	15
2.4.4 区数据	15
2.4.5 流数据	15
2.5 空间模式的类型	16
2.6 本章小结	19
第三章 空间模式的转换	21
3.1 邻近多边形	21
3.1.1 邻近多边形的类型	22
3.1.2 生成邻近多边形的算法	22
3.2 密度计算	25
3.2.1 局部平均法	25
3.2.2 局部加权平均法	26
3.2.3 中心估计法	27

3.2.4 邻近多边形法	28
3.3 分布中心计算	28
3.3.1 矢量算法	29
3.3.2 栅格算法	30
3.4 地理编码	31
3.5 特征点提取	33
3.6 聚类分析	33
3.6.1 聚类的判据——归类指数	34
3.6.2 聚类的方法	37
3.7 插值分析与抽样	41
3.7.1 插值的类型	41
3.7.2 整体插值	42
3.7.3 规则格网局部插值	43
3.7.4 局部函数法	43
3.7.5 局部加权平均法	43
3.7.6 克里金法	45
3.7.7 抽样	47
3.8 本章小结	48
第四章 空间模式的测度	49
4.1 测度的原理	50
4.1.1 空间随机过程	50
4.1.2 几何概率	51
4.1.3 矩形和椭圆区域内的 P 函数	53
4.2 点模式的测度	56
4.2.1 函数法	56
4.2.2 样方法	57
4.3 区模式的测度	62
4.3.1 空间邻近性测度	62
4.3.2 空间自相关测度	63
4.3.3 组合结构测度	66
4.4 面模式的测度	67
4.5 空间模式的关系测度	70
4.5.1 点模式与点模式	70
4.5.2 区模式与区模式	73
4.5.3 面模式与面模式	74
4.6 本章小结	75
第五章 集聚型点模式结构分析	76
5.1 理论基础	76
5.2 指示函数	78

5.2.1	$D(\theta)$ 函数	78
5.2.2	$R(r)$ 函数	80
5.2.3	$P(\theta, \Delta\theta, r, \Delta r)$ 函数	80
5.3	函数 P 的推导	81
5.3.1	分母的计算	81
5.3.2	区间 $[0, a]$ 上分子的计算	83
5.3.3	区间 $[a, \sqrt{2}a]$ 上分子的计算	84
5.3.4	分子计算总结	90
5.4	函数 H 的检验	90
5.4.1	单中心集聚	90
5.4.2	双中心集聚	91
5.4.3	多中心集聚	92
5.4.4	随机和均匀点集	94
5.5	本章小结	95
第六章	空间模式的应用研究	97
6.1	数据处理	97
6.1.1	数据来源	97
6.1.2	研究目的	97
6.1.3	过程与方法	98
6.2	结论分析	104
6.3	集聚中心的几何意义	106
6.4	本章小结	107
第七章	结论	108
参考文献		111
附录		118

第一章 绪 论

1.1 空间模式的含义

地理学研究人类赖以生存和发展的空间,它是一个以三相界面(大气圈与岩石圈、水圈的交接面)为核心、有一定厚度的地球圈层.关于这个圈层的边界有多种理解(林超、杨吾扬,1990;牛文元,1992),但无论采用何种理解,其垂直厚度相对于它的水平延展和内外环境均显得微不足道,故习惯上又将它近似地当作一个面处理.在地理文献中,该圈层曾经被称为地理环境(E.雷克吕)、地理壳(圈)、景观壳(圈)、地理面和地球表层巨系统(陈传康,1990;Сочава,1978;牛文元,1992;钱学森,1994).由于其规模宏大、要素众多、结构复杂,即使是在科学技术高度发达的今天,要从整体上把握和研究它仍然十分困难,因此在绝大部分情况下,地理学者研究的只是它的子系统.传统上子系统的划分可按空间范围或组成要素进行,形成区域地理学和部门地理学两大流派.这种划分存在多种空间尺度和要素等级层次,空间尺度和要素等级层次之间又存在各种各样的组合方案,每一个特定的方案都可以得到与之相对应的确定的子系统,它们正是地理学的研究对象,简称地理对象.

从空间认知的角度分析(Mark *et al.*,1999; Mennis *et al.*,2000),地理对象包括两种基本的类型,即地理实体与地理变量.前者表现为具有确定二维边界的区域,通常假设它的内部具有一致的或相对一致的属性,从而作为一个整体来描述和研究;后者是反映地理环境性质特征的各种要素,通过建立要素特征值与空间位置的关系对其进行描述和研究.两者的区别不是绝对的,实际上,地理变量取值相对一致的区域可以当作地理实体处理,而一组地理实体也可以看成一个多维地理变量.

空间模式是个被广泛使用的概念,本书的定义是“空间模式是地理对象的抽象”.所谓抽象,便是忽略地理对象所代表的具体专题含义,抽取其几何位置、拓扑关系、时间和属性的测度值.空间模式中的地理对象可以是一个(或一组)地理变量,也可以是一组(或一个)地理实体.一个和一组的分别是相对的,因为一组地理变量可以视为一个多维地理变量,一组地理实体可以视为一个更高层次的复杂实体.

由于空间问题具有多学科内涵,“空间模式”一词也经常出现在图像处理(Parlow,1989)、人工智能(Domingo *et al.*,2002)、环境学(Jacquez *et al.*,2002)、特别是生态学(Boerner *et al.*,1996;Walters *et al.*,1999;Malanson,2002)等领域的文献中,用于指代有某种几何含义的对象或排列分布格局.此种用法与本书的定义并无矛盾,但其外延较小,可视为本书所定义的“空间模式”的子概念.

空间模式按地理对象性质的不同分成实体模式和变量模式,前者按描述实体几何形态所需要的维数又分为点实体、线实体、区实体和体实体;后者依变量的标度类型分成名称、顺序、区间、比率4种类型.在上述空间模式的各种子类中,最早被提出和研究最深入的当属点模式(Point Patterns).1954年,Clark和Evans提出了测度点模式的最近距离统

计量 R (按本书的归类体系属于指标法), 可视为研究空间模式的开创性贡献, 1960 年 Dacey 将该成果引入地理学中 (Brimicombe & Tsui, 2000). 此后, 有关空间模式的研究不断深入, 指标法逐步完善, 样方法 (Greig-Smith, 1964; Harvey, 1966)、函数法 (Bailey and Gatrell, 1995)、可变样方法 (Brimicombe & Tsui, 2000) 等测度空间模式的方法相继问世.

1.2 空间模式的相关概念

与空间模式对应的英文词是 Spatial Pattern. 此处 Spatial 是个特化了的空間概念, 这种特化体现在以下两个方面: 其一, 范围的特化; 这里的空間指的是地理空間, 即 1.1 节所说的“以三相界面为核心、有一定厚度的地球圈层”. 其二, 尺度的特化; 这里的空間既不是微观空間, 也不是宇观空間, 而是指宏观空間, 其最大尺度小于地球的范围, 最小尺度大于人类正常自然视觉的空間分辨率. “Pattern”一词在英文中内涵深刻、外延广泛, 在不同的使用场合, 其可以用作“模范、典型、式样、样品、模子、模型、图案”等含义. 本书采用“模式”作为其对应的中文词是由于“Pattern Recognition”的中文译名“模式识别”在国内学术界已经约定俗成. 在“模式识别”的有关文献中, “模式”被定义为“对象的科学抽象”或“对象的描述”. “抽象”或“描述”的过程是通过量测得到表征对象特征的一组数据, 并将其记录下来; 其结果可能是矢量、符号串、图或数学关系; 目的在于运用定量手段有效地识别或区分对象 (孙即祥等, 2002). “空間模式”中“模式”的含义与“模式识别”中“模式”的含义是一致的, 即“空間模式”是一类特殊的“模式”, 其特殊性体现在与之相对应的对象具有空間内涵, 即对象显式地或隐含地与一组几何数据相关联.

与空間模式紧密相关的另一个词是空間模型 (Spatial Model). “模型” (Model) 一词在英文中含义很多, 此处它是原型 (Prototype) 的对偶概念. 原型是指现实世界中存在的实际对象, 而模型是原型的替代物 (姜启源, 1993), 通常模型会简化原型的一些细节并保持某些方面的相似性. 空間模式与空間模型的差别在于: 空間模式是空間原型 (地理对象) 的抽象, 空間模式与空間原型是一种映射关系; 而空間模型是空間原型在抽象基础上的简化, 空間模型与空間原型只是一种相似关系. 空間模式包含空間原型在约定含义下的全部信息, 是合理构建空間模型的科学依据, 故对应同一空間原型的空間模式和空間模型在时间顺序上存在先后关系. 在一些文献中 (姜启源, 1993; 寿纪麟, 1993; 何宗宜, 屠龙海, 1994; 李霖, 1997), 与模型对应的原型是人们解决问题的方法或过程, 相应地, 空間模型 (也称地理模型) 则被用于指代研究和解决空間问题程序的抽象和简化 (张超, 沈建法, 1993; 徐建华, 1994; 刘昌明等, 2001). 空間模式没有与此相对应的含义或用法.

2001 年, 陈述彭等根据“图谱”在生物、化学等领域的成功应用提出“地学信息图谱”的概念, 并认为“图是可感知的外部特征表现, 谱则反映内在规律; 图表现静态现象特征, 而谱则表现动态, 包括方向性、过程性; 图是可感知的现象, 谱则是推演得出的规律; 图表现空間单元特征, 谱则表示事件发展之起点与过程; 图用于二维表达非线性源, 谱则用一维表达规律, 实现多维分析; 图描述复杂系统的初始与边界条件, 而谱为运动变化规律描述的主导方程”. 简言之, 图指空間序列, 谱指时间序列. “图谱主要运用图形语言, 进行时间与空間的综合表达与分析. 其中地学图谱则是应用于地学分析的系列多维图解” (陈述彭等, 2001). 并依据这样的定义, 将“地学信息图谱”按研究对象、表达内容特征和功能进

一步细化为若干子类.“地学信息图谱”又被简称为“地学图谱”或“图谱”。

虽然“图谱”与“空间模式”都是针对“空间原型”提出的概念,但两者的差别是显而易见的.首先,“图谱”侧重于表现“空间原型”的时空特性,属性维从属于时空维,分析空间基本上是固定的;而“空间模式”则将对象的属性作为与空间维和时间维同等重要的独立的维,分析空间的维数根据需要确定.其次,在研究的思路上,“图谱”强调“形”、“数”、“理”

的高度统一,而“空间模式”将数值计算和可视化视为两种并行的数据处理手段,数值计算和可视化既可配合使用,也可单独使用,但通常以数值计算为主,可视化为辅。

本节中各概念的关系如图 1.1,其中虚线表示层次关系,实线表示包含关系。

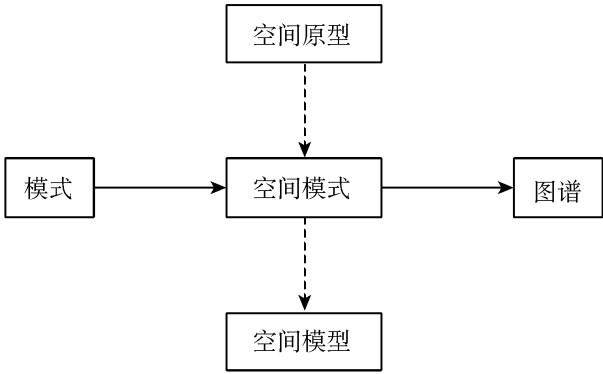


图 1.1 空间模式与其相关概念的关系

1.3 空间模式的相关研究领域

1.3.1 模式识别

模式识别(Pattern Recognition)是人工智能的一个重要分支,它是根据研究对象的特征或属性,利用以计算机为中心的机器系统运用一定的分析算法认定它的类别.模式识别的理论和已成功地应用于诸多领域,比如:信件分检、指纹识别、遥感图像机器判读、数据自动输入等.一个完整的模式识别过程包括特征提取、特征选择、学习和训练、分类识别 4 个环节.目前主流的模式识别技术是:统计模式识别、句法模式识别、模糊数学方法、人工神经网络法和人工智能法.其中,统计模式识别通过直接或隐含地利用各类的概率密度函数、后验概率等进行分类识别,包括聚类分析法、判别类域代数界面法、统计决策法和最近邻法.句法模式识别将对象分解为若干基元,用基元以及它们的结构关系来描述对象,然后运用形式语言理论进行句法分析而决定其类别,该类技术主要用于复杂对象的识别.模糊数学方法运用模糊数学的理论解决模式识别问题,适用于对象本身或识别结果有模糊性的场合.人工神经网络法通过由人工神经元相互联结而构成的非线性系统进行分类识别.人工智能法则运用人工智能中有关学习、知识表示、推理等技术进行模式识别(孙即祥等,2002).此 5 类技术中,统计模式识别的理论基础最为坚实,对研究空间模式的指导和借鉴意义最大。

1.3.2 空间数据挖掘

数据获取曾经是地理研究中难度最大、耗时最长、需要人力物力最多的环节,数据不足的问题曾经长期困扰古代和近代的地理学家,使得相关研究进展缓慢.遥感技术、空间定位技术的兴起和逐步完善使空间数据获取的难度迅速降低而获得的数据精度不断提高,社会统计数据的积累正在解决另一类数据不足的问题.计算机的发明,特别是数据库技术的推广应用使得各种数据能够以标准的格式存贮于电子媒介上供人们使用;地理信息系统技术使空间数据中的几何数据和属性数据能够根据需要在各种尺度上进行匹配、

融合、定量计算和可视化；而网络技术的应用则使得数据的异地共享变得空前便利。今天，空间数据获取、空间数据库的建立和维护已经从地理研究的环节中分离出来，成为一个标准化的工程问题。地理研究正面临数据过剩的问题，如何在一个已经建立的空间数据库的海量数据中提取我们所需要的信息、特别是隐含信息成为地理研究的新课题。数据过剩是当代许多研究领域中有共性的问题，数据库中的信息提取技术随之受到广泛的关注，一个新的研究领域——“数据挖掘”从而诞生。

数据挖掘(Data Mining)的主要任务可以概括地分为两类，即描述性数据挖掘(Descriptive Data Mining)和预测性数据挖掘(Predictive Data Mining)。前者以简明和概括的方式指出数据集令人感兴趣的总体特征；后者则通过构造一些模型，以已知的数据集推断模型的参数，并试图预测新生数据的性质；其中前者又可进一步细分为类特性描述、关联分析、分类、集聚检测以及时序分析(Han, 1996; 1999)。空间数据具有维数高、类型多、结构复杂和数据量大等特征，其对应的数据挖掘技术难度大、牵涉面广，具有特殊性(Knorr and Ng, 1996; Kang *et al.*, 1997; Sander *et al.*, 1998; Lin *et al.*, 1999; Wang *et al.*, 1999; Zhou *et al.*, 1999)，空间数据挖掘(Spatial Data Mining)因之逐渐从数据挖掘领域中分离出来。从 20 世纪 90 年代开始，有关空间数据挖掘的研究不断深入(Lu *et al.*, 1993; Ng and Han, 1994; 1996; Han *et al.*, 1997; 1998; Ester *et al.*, 1995; 1996; 2000)，逐步发展成为数据挖掘领域中一个相对独立的分支。空间数据挖掘主要采用 4 大类方法，即确定型空间数据分析(Confirmatory Spatial Data Analysis)方法、探索型空间数据分析(Exploratory Spatial Data Analysis)方法、软计算方法和可视化方法。

1.3.2.1 确定型空间数据分析方法

地理研究中应用定量方法的历史最早可以追索到 19 世纪早期，但形成一大流派则发生在计量革命以后。20 世纪 50 年代源于美国、继而波及欧洲和整个世界的计量革命对地理学的发展产生了广泛而深远的影响。计量革命初期的一大特征是强调数理统计方法在地理学中的应用(Taaffe, 1979; Billinge, 1984; Barnes, 2001)，并且用一个新名词“空间统计学”(Spatial Statistics)或“空间统计分析”(Spatial Statistical Analysis)来命名这个新的领域，确定型空间数据分析便是该领域的主要内容。

确定型空间数据分析方法以经典的概率统计理论为基础。对于给定的观测数据，一般是首先提出一种统计假设，然后在这一假设条件下，导出它的一些理论和方法，作出统计推断。参数估计和假设检验是统计推断中的两大类问题，参数估计是在总体的分布函数的形式为已知但它的的一个或多个参数为未知的情况下，根据总体的样本值估计总体所服从的分布函数中的未定参数。参数估计可以采用点估计或区间估计两种方法完成，点估计又包括矩估计法和极大似然估计法。假设检验是在总体分布完全未知或只知其形式、不知其参数的情况下，为了推断总体的某些性质，针对已提出的关于总体的某些假设作出接受或拒绝的判断。假设检验包括 u 检验、 t 检验、 χ^2 检验、 F 检验以及秩和检验等(盛骤等, 1989)。

此类分析方法的使用有个重要的前提，便是问题中所考虑的总体要服从某个预定的统计模型(比如独立和正态性条件)。然而实际上许多数据资料，由于种种原因可能不符合现有的某个统计模型的假设或者根本不知道它应该符合哪一种统计模型，空间数据更是

如此.这意味着此类方法用于处理空间数据存在理论上的缺陷.

1.3.2.2 探索型空间数据分析方法

1970年,Gould指出在地理研究中应用传统数理统计分析方法存在以下问题:①回归模型的线性形式;②数据样本的随机特性;③随机变量和误差项的正态分布;④随机变量和误差项的独立性假设(Gould,1970;Brunsdon,2001).同年,Tobler提出“空间自相关”(Spatial Autocorrelation)的著名假说,即认为空间数据之间存在固有的相关性,而且采集数据的样点距离越近,相关性也就越大(Tobler,1970).该假说在西方地理学界立即得到了广泛的认同,被称为地理学的第一定律.1973年Cliff和Ord发表专著论述空间自相关的度量方法(Cliff and Ord,1973).空间自相关的存在使得地理变量不符合相互独立的条件,进一步明确了确定型空间数据分析方法应用于地理研究中的缺陷.

与此同时,多元数据分析方法(纯代数处理)作为应用数学中一个重要的领域得到迅速发展.这类方法的基本思路是:一般不对数据作模型假设,而直接从实际数据出发,考察和分析大量数据的结构和特征,从中提取出主要而准确的信息,必要时进一步归纳出反映数据的较切实的模型为统计推断作准备.这种数据分析的观点已为学术界所接受(Johnston,1978;胡定国,张润楚,1990;项静恬,史久恩,1997).这类方法包括回归分析、相关分析、主成分分析、因子分析、典型分析、相应分析、判别分析、聚类分析和周期分析等.这类方法中的一部分是通过将相应的多元统计分析方法赋予新的内涵发展而来的,如回归分析、主成分分析等.在传统多元统计分析中,回归分析是采用极大似然估计法对正态总体(随机样本)的一种参数估计方法,主成分分析是确定一个多元正态分布等密度椭圆面主轴的方法(由样本来估计).在多元数据分析中回归分析的含义变为采用最小二乘法的数据拟合,主成分分析的含义则是简化原始数据的降维处理.这类由多元统计分析方法演变而来的多元数据分析方法无疑具有较好的数学基础,另外也有一些多元数据分析方法(如聚类分析等)其数学理论依据还不是很充分,因而带有更多实用的色彩.多元数据分析方法在生物、医学、地质、工程、气象等领域已得到了广泛的应用,在地理研究中,又将其称为探索型空间数据分析方法(Exploratory Spatial Data Analysis)并成为空间数据分析中的主流.

1.3.2.3 软计算方法

软计算方法是一类通用的优化算法.优化技术是一种以数学为基础,结合各学科专业知识,优化求解各类实际问题的应用技术.传统的最优化方法不仅要导出实际问题目标函数的确定解析形式,而且要求目标函数是凸的、高阶连续可微、可行域是凸集.这些弱点使其面对非线性复杂系统,特别是包含不确定信息的非线性复杂系统时显得无能为力.随着计算机技术的推广应用、特别是人工智能研究的逐步深入,针对不确定非线性复杂系统的最优化研究渐渐发展为一个新的研究领域——软计算方法.软计算方法包括模糊计算方法、进化计算方法、人工神经网络计算方法以及它们之间的结合.

模糊计算方法源于L. A. Zadeh博士提出的“模糊集”(Fuzzy)概念.模糊性是客观事物的根本属性之一,是一种比随机性更深刻的不确定性.模糊计算方法通过使用模糊集合、隶属度函数等概念和模糊推理等手段实现以精确的方法解决模糊问题.进化计算(Ev-

olution Computation)方法主要包括遗传算法、模拟退火算法和禁忌搜索算法.其中遗传算法(Genetic Algorithms)是一种基于生物自然选择和基因遗传学原理的优化搜索方法,通过循环使用复制、交叉和变异三个步骤模拟生物在自然界经遗传、变异和选择的进化过程搜寻问题的最优解;模拟退火(Simulated Annealing)算法通过模拟物理退火过程寻找大规模组合优化问题的最佳解;禁忌搜索(Taboo Search)算法模拟人类具有记忆功能的寻优特征,通过局部邻域搜索机制和相应的禁忌准则来避免迂回搜索,并通过破禁水平释放一些被禁忌的优良状态,保证多样化的有效探索,最终实现全局优化.人工神经网络(Artificial Neural Network)计算方法通过使用以工程技术手段模拟人脑神经网络结构与特性得到的人工神经网络,经过学习和训练形成自适应、自组织和自学习能力,进而模拟人脑的自然智能,获得问题的最优解(张颖,刘艳秋,2002).

软计算方法的主要优点是不要求目标函数和约束条件的连续性与凸性,甚至可以没有解析式;对计算中遇到的数据不确定性有很强的适应力且计算速度快.缺点是理论上不完善,不能确保解的最优性,但其作为导出最佳解的“启发式方法”仍然被广泛地应用于空间数据挖掘之中(Parashar *et al.*, 1997; Whigham, 2000; Dai and Khorram, 1999; Maier and Dandy, 2000; Dawson and Wilby, 2001; Duczmal and Assunção, 2002).

1.3.2.4 可视化方法

可视化的概念源于对不可视原型(对象或过程)的可视表达研究.由于任何原型在计算机中对应的都是一个数据集,可视化的概念在研究过程中进一步深化为相对数据集的可视表达,而数据集对应的既可以是不可视原型,也可以是可视原型.可视化科学涉及计算机图形学、地图学、心理学和知识工程,相应地,其应用于空间数据表达会遇到来自技术手段、感知能力、计算方法和地理主题方面的限制, M. Gahegan(1999)曾经将其归纳为响应速度、综合效果、映射关系和标示评判.其中,响应速度是指影响大数据量交互显示和操作的技术障碍;综合效果是指同时显示多个数据层(主题或变量)即多维空间数据可视化遇到的问题;映射关系是指在空间数据与视觉域以及可视化环境提供的巨大可能性之间建立对应关系的复杂性(制图综合便是其中的问题之一);标示评判包括让用户了解可视化场境的标识和可视化解决问题的有效性两个方面.

在空间数据挖掘中,可视化主要应用于两个环节:其一,用于表达空间数据挖掘的结果;其二,用于获得空间数据的直观形象,以便利用人类的自然视觉直观地感知空间数据的特征及其中隐含的规律,为进一步处理提供启发式支持(Gahegan, 2000).与前3类方法不同,可视化不做定量分析,取而代之的是提供一个直观形象,其效果取决于个体的感知能力,主观色彩浓厚,因而难于评判.

1.3.3 地理信息系统

20世纪对地理学发展产生重大影响的另一事件是空间信息处理技术,特别是地理信息系统(Geographical Information System,简称GIS)的崛起.地理信息系统源于机助制图(Computer Aided Cartography,简称CAC),CAC最初的应用领域是工程图、尤其是机械制造的设计图纸.1960年代初 Roger Tomlinson 在加拿大首次提出应用计算机处理地理数据的设想,并且用“Geographic Information System”这个术语来表示这个新的应用领域

(Tomlinson,1998).技术和市场强有力的推动使地理信息系统经历了从少到多(数量上的扩展)、从简单到复杂(功能的完善)、从集中到分散(网络化)以及平台的更替(微机化)等一系列重大变革,并且衍生出多种含义.

最初,GIS作为“Geographic Information System”的首字母缩写词出现在各种文献中,其代表的是一类“应用系统”(硬件平台+软件平台+空间数据).自20世纪80年代开始,微型计算机性价比不断提高,成为这类“应用系统”的主流硬件平台,而软件平台的开发则逐渐商业化,这些开发GIS软件平台的商业公司在广告宣传中往往省去“软件平台”,径直地将自己的软件产品称为“GIS”.此后,“GIS软件平台”的开发应用在涉及空间数据管理的应用领域获得了巨大的成功,成为IT产业中发展最快的分支之一,“地理信息系统产业”的概念出现.随着地理信息系统产业的不断扩展,对相关人才的同步需求使得地理类院系将越来越多的地理信息系统课程包括在其学位教育的课程中,与空间信息处理有关的一般问题也受到了学术界越来越多的关注,成为一个充满活力的交叉学科领域,1991年Rhind等将该领域命名为“地理信息科学”(Geographical Information Science,简称为GIScience或GISci)(Rhind *et al.*,1991).Goodchild认为其中包含“数据收集和度量,数据获取,空间统计分析,数据建模和空间数据理论,数据结构、算法和过程,显示,分析工具,制度、管理和伦理问题”8个分支(Goodchild,1992).1995年,Goodchild进一步论述道:“地理信息系统在消除航空摄影测量、遥感、大地测量学、地图学、测量和地理学之间的传统隔阂方面做出了很大的贡献,……我认为将其作为‘地理信息科学(Geographical Information Science)’的缩写词,也就是将其看成一门学科,而不是局限于今天的地理信息系统产品所能提供的解决方案更有意义”(Goodchild,1995).1997年,地理信息系统领域的国际权威刊物“International Journal of Geographical Information Systems”更名为“International Journal of Geographical Information Science”,GIScience作为一个学科的地位正式确立.“GIScience”有时又简写为“GISci”或“GIS”,而地理信息系统对应的英文词则演变为“Geographical Information Systems”,简写为“GISs”.

Geomatics和GeoComputation是GIScience的近义词.“Geomatics”的概念源于法国,法国科学家Bernart Dubuisson于1975年在正式的科学文献中使用“Geomatique”描述空间信息获取与处理这个新兴的领域.此后“Geomatique”被引入加拿大并演变成“Geomatics”.1996年国际标准化组织(ISO)给出了它的定义:“Geomatics是一个使用系统的方法整合各种作为涉及空间信息的生产和管理的科学、行政、法律和技术操作的一部分,用于获取和管理空间数据的手段的活动领域.这些活动包括但不限于地图学、控制测量、数字制图、大地测量、地理信息系统、水文、土地信息管理、土地测量、采矿测量、航空摄影测量和遥感”,李德仁建议将其译为“地球空间信息学”(李德仁,1998).

“GeoComputation”(中文文献中译为“地理计算”或“地学计算”)的概念由英国地理学家Stan Openshaw教授在20世纪90年代中期提出(毛政元,李霖,2003),他给出的定义是:“GeoComputation意在利用各种不同类型的地理数据发展相关的科学计算方法”(Openshaw and Abrahart,2000).为了区别于其相关领域,Openshaw(2000)进一步明确了它的两个侧重点,即地理主题和高质量的运算.从定义中不难看出两者的区别:即Geomatics涉及的是地球空间信息的获取与管理,侧重技术与应用层面;而GeoComputation的主题是地理信息的分析处理,侧重认知与研究层面.

作为一个学科, GIScience 的任务和目的是为地理信息系统的发展和完善提供理论和方法支持, 它研究地理信息系统尚未解决或即将遇到的各种问题, 提供前瞻性的指导和解决方案. GIScience 涵盖了 Geomatics 和 GeoComputation 的内容, 其现阶段的研究热点包括数据自动输入、空间数据模型、空间分析、空间信息可视化 (Muller *et al.*, 1995)、Web-GIS、易用界面和系统集成 (毛政元, 李霖, 2002; Mao and Li, 2002) 等方面. 其中以空间分析与空间模式的关系最为紧密.

在英文文献中, “空间分析 (Spatial Analysis)” 一般用做 “空间统计分析 (Spatial Statistical Analysis)” 或 “空间数据分析 (Spatial Data Analysis)” 的缩略词 (Anselin, 2000; Fotheringham, 2000; Frank, 2000), 有时也作为 “定量地理学” (Quantitative Geography) 的同义词 (Bailey and Gatrell, 1995; Brown, 2000; Longley, 2000). 此处, “空间分析” 对应的是与地理信息系统软件平台中的一组空间数据处理方法 (杜道生, 1995) 和运用这组方法处理空间数据的过程相关联的研究内容. 一些文献上建议将此种含义的 “空间分析” 称为 “空间数据操作 (Spatial Data Manipulation)” (Unwin, 1996; Unwin and O’Sullivan, 2002).

严格地说, 空间数据操作的理论基础是集合论和计算几何, 而空间数据分析的理论基础是概率论和数理统计. 但随着地理信息系统与其他空间信息处理平台集成 (Goodchild, 2000; Boots, 2000; Marble, 2000; Zhang and Griffith, 2000; Ungerer *et al.*, 2002) 的不断深入, 空间数据操作与空间数据分析将逐步融为一体 (Fischer and Unwin, 1996).

空间模式与各相关研究领域关系如图 1.2 所示.

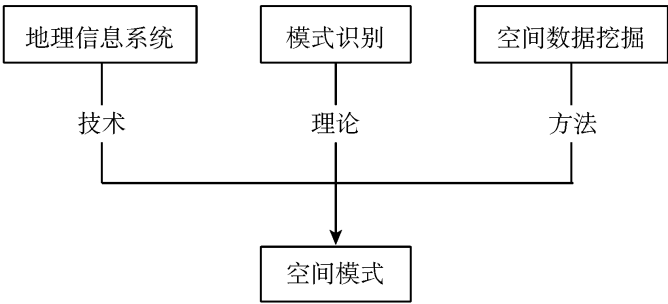


图 1.2 空间模式与各相关研究领域的关系

1.4 空间模式的研究意义

地理学在其漫长的发展过程中形成了定性和定量两大研究流派; 将两者截然分开或完全对立起来是不明智的 (Philip, 1998), 因为定性是定量的基础, 而定量是定性的深化. 描述、解释、预测是认识地理对象的 3 个层次, 毫无疑问, 预测是地理研究的最高层次和最终目标, 描述和解释是为了能够最终实现预测的目的. 预测的前提是要合理地构建地理对象的定量模型, 定量研究因之成为当代地理学的基本方向.

地理对象的特殊性决定了其建模方法的特殊性. 地理学的整体研究对象 (地理环境) 在空间上虽然是个有限的范围, 但地理研究的多尺度和多角度却能从中提取出几乎无限的局部研究对象, 加上其本身要素的多样性、分布的不均一性、变化的不确定性和联系的广泛性使其成为一个异常复杂的结构. 一代又一代的地理学家曾经提出各种各样的理论试图解释这个复杂的结构 (James and Martin, 1989; 刘盛佳, 1990; Johnston, 1991; 2000),

但实践证明这些理论只适用于局部对象,且一般不能导出定量模型.由于地理理论尚不能全面揭示地理对象的系统动力学机制,地理构模常常采用黑箱法.黑箱法的思路是将地理对象(过程)视为一个黑箱,寻找一个与其在输入输出定量关系上有相似性的替代物,这个替代物便是地理对象的模型.

地理研究的实践中曾经诞生过成千上万的模型(Chorley and Haggett, 1967; Haggett and Chorley, 1970; Wrigley and Bennett, 1981),但能够经受住时间考验流传至今的却不多(Fotheringham and Wegener, 2000).失败的原因大多是因为在构模过程中对地理对象做了过分的简化,或者使用了错误的假设,故防止过分简化和错误假设是构模成功的必要条件,前者主要取决于客观条件,后者主要取决于主观努力.在定量地理学发展的早期,计算机硬件性能低下,地理信息系统处在萌芽状态,遥感技术尚不发达,空间数据的获取、处理与分析十分不便.技术条件的限制和地理对象的复杂性决定了这一时期的地理学者在采用黑箱法构模的过程中不得不大量地使用简化进行归纳(Miller, 2000),频繁地使用假设进行演绎,构模失败主客观原因兼而有之.今天,计算机技术、遥感和空间定位技术已经高度发达,新一代地理信息系统中空间数据描述的精度和处理的速度已经能够满足地理研究的要求,空间数据获取、处理与分析的技术障碍基本消失,地理构模中过分简化的问题接近解决,构模失败主要应归咎于主观原因,即错误假设的问题.防止错误假设的关键是要在抽象的层次上更加深入地了解地理对象(空间原型)及其相互关系,空间模式正是这样一个研究层次.

空间模式是空间原型和空间模型之间的纽带和桥梁,空间模式为以定量方式解决空间问题提供了一种整体性的思路.通过恰当的抽象,空间模式将空间问题归纳为有限的类型,为避免不必要的重复、促进定量地理研究向纵深发展奠定了理论基础.此外,研究空间模式的意义还表现在以下两方面:其一,通过与可视化技术相结合,空间模式可以同时从定量分析和直观形象推理两个视角对同一地理对象进行探索,充分利用人类视觉感官与理性思维的潜能.其二,空间模式能够从不同的尺度和层次上描述地理对象之间的亲疏关系,解决类的可分性问题.类的可分性问题又称分叉律,是对象在不同尺度和层次上的相似性和差异性的归纳和总结,包括单级分叉和多级分叉,而后者又包括等分和不等分、对称和不对称等内容.

1.5 本书的研究目标与主要内容

本书试图实现两大目标:其一,建立空间模式的研究体系;其二,发展和完善空间模式的测度方法.此两大目标也是本书的两条主线,书中将空间模式的研究体系归纳为表达、转换、测度(包括关系测度)和应用4个方面,扩展了测度空间模式的3大类方法(指标法、函数法和样方法)的内容并提出结构测度的思想和结构信息提取算法.

全书共分七章.第一章绪论从地理对象的定义引出“空间模式”的概念,阐述空间模式与其相关概念的区别、空间模式与其相关研究领域的关系,论证空间模式的研究意义.第二章空间模式的表达通过引入“分析空间”的概念进一步明确空间模式的定义,研究“空间模式表达”涉及的数据模型、变量标度和存贮方式等主题,最后归纳出其分类体系.第三章空间模式的转换提出空间模式转换所涉及的各种方法,针对每一类方法分析其技术细节,

比较其优劣.第四章空间模式的测度从方法和内容两个方面提出空间模式测度的研究体系,阐述基于概率论的空间模式测度原理,分析空间模式现有测度方法的优劣并提出若干改进.第五章集聚型点模式结构分析针对现有空间模式测度方法在研究集聚型点模式中的局限性,提出基于几何概率和组合的结构信息提取思想以及相应的指示函数和算法,推导指示函数的解析表达式,验证指示函数和算法的可行性.第六章空间模式的应用设计了集聚型点模式结构信息提取算法,并以居民地坐标数据为例实现了该算法.第七章结论是全书的总结和归纳.

1.6 本章小结

本章主要内容如下:

- (1) 基于“地理对象”引出“空间模式”的定义,从空间认知角度分析其子类;
- (2) 比较空间原型、空间模式和空间模型三大研究层次,阐述“空间模式”和“空间模型”的区别、“空间模式”和“模式”的联系;
- (3) 论述空间模式与其三个相关研究领域(模式识别、空间数据挖掘、地理信息系统)的关系;
- (4) 从理论和方法两个角度阐明研究空间模式的意义;
- (5) 提出本书的研究目标,介绍本书的主要内容.

第二章 空间模式的表达

2.1 地理空间与分析空间

地理对象分布于一个有一定厚度的地球圈层中,以下将这个圈层称为地理空间.地理空间的核心是大气圈与岩石圈或水圈的交接面,又称地球表面.由于地球的物质分布不均,地球表面是个不规则的椭球面,地理空间是个不规则的空心椭球体.为了定位的方便,地球表面通常被投影到一个二维欧氏平面上,相应地,地理空间被变换为一个包含这二维投影面的三维欧氏空间(郭仁忠,1997).当分析地理问题时,地理学家通常在以上的二维欧氏空间或三维欧氏空间表示地理对象的几何分布,并用二维实数空间 R^2 或三维实数空间 R^3 来定量描述地理对象的几何特征和拓扑关系.地理对象的几何特征和拓扑关系并不是地理研究的全部内容,大量的地理问题都直接或间接地与地理对象的专题属性有关,而上述的二维和三维欧氏空间属于几何空间,不能表示地理对象的专题属性.为了表示地理对象的专题属性,引入属性空间的概念.属性空间由任意多个属性维构成,每一个属性维刻画一个属性的变化.当属性数据采用区间标度或比率标度时, n 维属性空间仍然可以用 n 维实数空间 R^n 来定量描述;当属性数据采用名称标度或等级标度时,属性空间只能进行定性或半定量描述.地理对象的时序变化是地理研究的另一个重要主题,可以用时间维来表示这种变化.时间维是时空整体中不可忽略的部分,它的表示是近来空间数据模型的重要研究方向之一,研究成果很多,但迄今为止商用地理信息系统基础平台尚无成熟的实现方案.本书将时间维作为一个独立的维处理,用一维实数空间 R^1 来定量描述.以下进一步明确分析空间的概念,分析空间是用于描述地理对象的 N 维抽象空间,它可以是几何维、属性维和时间维的任意组合.此处描述包含定性、定量和可视化 3 个层次,描述的内容可以是地理对象的几何特征、属性特征和两者的时序变化的任何子集.

表 2.1 分析空间的类型

序号	时间维	属性维	几何维	序号	时间维	属性维	几何维
1	0	0	0	13	1	0	0
2	0	0	1	14	1	0	1
3	0	0	2	15	1	0	2
4	0	0	3	16	1	0	3
5	0	1	0	17	1	1	0
6	0	1	1	18	1	1	1
7	0	1	2	19	1	1	2
8	0	1	3	20	1	1	3
9	0	n	0	21	1	n	0
10	0	n	1	22	1	n	1
11	0	n	2	23	1	n	2
12	0	n	3	24	1	n	3

分析空间中几何维的维数可以是 0,1,2,3,属性维的维数可以是 0,1, n ,时间维的维数可以是 0 或者 1,这样总共得到 $24(4 \times 3 \times 2)$ 种组合(表 2.1).其中组合 1 无意义;组合 2,5,13 为一维空间,应用场合很少;而组合 4,8,12,16,20,24 需要三维空间数据模型支持,目前虽有少数地理信息系统基础平台支持三维空间数据模型,但应用的难度较大.在剩下的 14 种组合中,3,7,9,14,17 五种组合最为常用.

2.2 数据集与可视化

空间模式是地理对象的抽象.实际的地理对象分布于地理空间中,但因人类感官的局限性和地理对象的复杂性,直接在地理空间中研究地理对象不仅费时费力,而且难以奏效,甚至事与愿违,得到错误的结论.故地理对象的研究除了极个别的情况外,总是在分析空间进行.将地理对象从地理空间转换到分析空间涉及到空间数据建模和空间数据获取.

空间数据模型包括三个层次,即概念模型、逻辑模型和物理模型.在概念模型层次上要解决的问题是如何认知地理空间,包括场模型(Field Model)和对象模型(Object Model)两种解决方案.场模型将地理空间理解为连续的、具有属性描述的空间位置的集合.对象模型将地理空间理解为具有属性特征的空间对象的集合.在逻辑模型层次上要解决的问题是如何记录空间位置数据,包括矢量模型(Vector Model)和栅格模型(Raster Model)两种解决方案.矢量模型用坐标表示空间位置,栅格模型用栅格阵列表示空间位置.在物理模型层次上要解决的问题是如何在计算机存贮介质上存取空间数据,空间数据在该层次上的问题与其他类型的数据没有差别,属计算机科学研究范畴.

空间数据获取在空间数据建模的基础上进行,实践中通常根据所研究的地理对象本身的特点和数据处理设备选用相应的空间数据模型.空间数据获取有多种途径,其中实地测绘、遥感、统计、定点观测和野外考察是获取原始空间数据的主要方式,上述 5 种方式所获得的数据在内容或形式上是有区别的,实践中常常配合使用,以便全面地认识地理空间,取得地理对象的完整信息.由于遥感所产生的良好的综合效益,它已成为一种重要的数据获取方式,并有逐步取代其他数据获取方式的趋势.数据获取也可以通过利用历史资料或者共享的方式实现.

将地理对象从地理空间转换到分析空间所涉及的每一个步骤都有可能产生误差,如何度量和校正这些误差是一个非常重要的研究领域(史文中,1998).本书不打算研究这个主题,而将空间模式的研究建立在地理对象从地理空间到分析空间的转换已经完成的基础上.并假设分析空间中的数据正确、合理地反映了地理空间中的地理对象,地理空间中的地理对象和分析空间中的空间模式通过特定的空间数据模型和空间数据获取方式建立了一一对应的关系.基于这种假设,空间模式的定义可以进一步明确为——空间模式是地理对象的抽象,是地理对象在分析空间中的像.

当取得了构成空间模式的地理对象的空间数据时,实际上也就是记录了该空间模式.以下我们称代表一个空间模式的相关地理对象的空间数据为“空间模式的数据集合”,简称为“空间模式的数据集”或“数据集”.当空间模式中的地理对象为地理变量时,一般采用场模型组织其对应的空间数据,数据集的形式可以理解为: {位置 1, 属性; 位置 2, 属

性;…;位置 n ,属性}.当空间模式中的地理对象为地理实体时,一般采用对象模型组织其对应的空间数据,数据集的形式可以理解为: $\{\text{地物 } 1(\text{位置}+\text{属性}),\text{地物 } 2,\dots,\text{地物 } n\}$.数据集中的位置数据既可以用矢量模型记录,也可以用栅格模型记录.对象模型将地物抽象为点、线、面、体 4 种类型,点没有空间延展(零维),用一个坐标点(矢量格式)或一个栅格(栅格格式)记录;后 3 种地物类型分别有 1 维、2 维、3 维空间延展,分别用坐标串、相互关联的坐标集(矢量格式)或栅格串、相互关联的栅格集(栅格格式)表示.

采用场模型记录的数据集(空间模式中的地理对象为地理变量),其中的位置数据不能省略.采用对象模型记录的数据集(空间模式中的地理对象为地理实体),根据具体的应用场合,可省略或简化其中的位置数据.在不能省略或简化数据集(空间模式中的地理对象为地理实体)中的位置数据时,分析空间的几何维应该大于或等于地理实体的几何维.

地理对象客观地存在于地理空间之中,空间模式是地理对象从地理空间映射到分析空间得到的像,当空间模式的维数在可视化技术表达能力允许的范围之内时,借助于可视化技术可得到像(空间模式)的直观形象,而数据集则是像的记录.空间信息具有多维结构,空间模式的数据集和可视化也必然是多维的.对数据集而言,维数在理论上是没有限制的,而可视化却存在明显的维数限制.可视化使用的电子设备的显示屏为平面结构,可以很方便地显示二维信息;根据透视法则可以显示三维信息;第四维可用颜色表示;四维以外的信息则必须借助于符号的大小、形状、图案纹理来表示.如何突破维数的限制有效地表示多维空间信息是可视化研究领域的一大难题,理论上数据集是空间模式的记录,两者存在一一对应的关系;但高维的空间模式没有对应的可视化表示,因此空间模式及其数据集与其可视化表达不是一一对应的关系.

2.3 空间数据的标度

标度是规范变量取值的标准,Stevens(1959)将标度归纳为 4 种类型(表 2.2),标度的分类仅仅只是解决了关于标度的一个最基本的问题,标度的可视化、标度的转换和多维标度的统一处理等也是有关标度研究不可忽略的重要内容.空间数据标度的可视化表示(即制图符号)是地图学的经典研究内容之一,有许多现成的表示方案,图 2.1 是其中之一,注意图中采用了与表 2.2 不同的译法,序列、间隔、比例分别与 Ordinal,Interval,Ratio 对应.

表 2.2 空间数据的标度

标度	定义	性质	运算	实例
名称标度	是对象的命名或标识	定性	逻辑运算(等于)	地名、土地类型名
顺序标度	是同类对象按某种标准的排序或分级	半定量	逻辑运算(大于、等于)	土壤肥力等级;城市综合实力排序
区间标度	无自然原点的定量标度	定量	逻辑运算(大于、等于); 数值运算(减;加和除 (二区间))	温度、时空坐标
比率标度	有自然原点的定量标度	定量	逻辑运算(大于、等于); 数值运算(加、减、除)	人口、距离

标度的转换是指不同类型的标度能否替代以及如何替代的问题.定性标度具有简明和概括的优点,而定量标度则具有精确和易于运算的优点.由于每一种标度都有各自的优

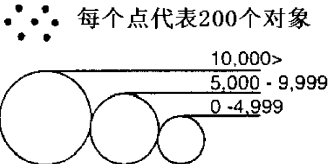
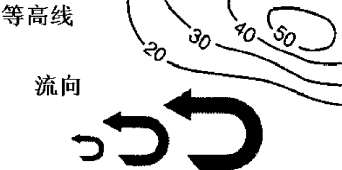
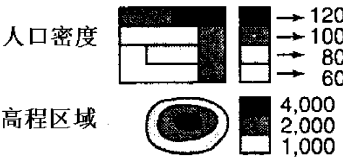
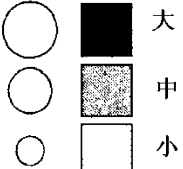
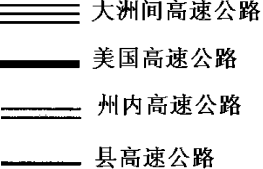
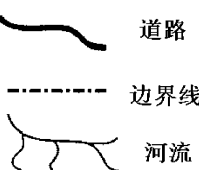
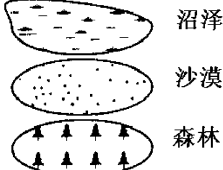
项目	点	线	面
间隔与比例	 每个点代表200个对象 10,000> 5,000 - 9,999 0 - 4,999	 等高线 流向	 人口密度 高程区域
序列	 大 中 小	 大洲间高速公路 美国高速公路 州内高速公路 县高速公路	 商业区 主要的 次要的 影响区
名称	 城镇 矿区 基准点	 道路 边界线 河流	 沼泽 沙漠 森林

图 2.1 各种数据标度的可视化(邬伦等,2001)

势,故标度的转换是个有重大应用价值的研究课题.由定量标度转为定性标度通常比较容易实现,区间标度和比率标度之间的换算以及顺序标度向名称标度的转换也不难;地理学中不乏成功的转换实例,其中的主要问题是合理性和准确性.由定性标度转为定量标度以及名称标度向顺序标度的转换技术难度较大,在很多时候这种转换是没有意义的,在有意义的场合这种转换需要大量的额外信息才能完成.

多维标度的统一包括两个方面的问题:其一,对具有多维结构的地理数据,在其每一维信息相互独立的情况下,如何保证它们能够得到相同程度的重视;其二,具有多维结构的地理数据,在其中的某几维或所有维存在关联的情况下,如何消除这种关联并用尽可能少的维数来表示.第一个问题又称量纲无关性处理,即消除同一指标因采用不同量纲所造成的值放大或值缩小,可以采用极差标准化、标准差标准化、正态化等方式处理.第二个问题又称降维处理,当关联为线性时,可采用主成分分析;当关联为非线性时,目前尚无良策.

2.4 空间数据的存贮类型

空间数据可以按各种标准分类,比如按其中所包含的意义,空间数据可分成几何数据、拓扑数据和属性数据 3 种类型或者按标度将空间数据分成 4 种类型,当然也可以按获取空间数据的时空尺度、质量精度、主题内容等标准对空间数据进行分类,以下的分类主要基于其在空间数据库中的存贮方式及对应的处理方式(Bailey and Gatrell, 1995).

2.4.1 对象

对象(Objects)是空间实体的数学抽象和图形描述,对应空间模式中的实体模式.按其