

PEARSON  
Prentice  
Hall

# 空间数据库

Spatial Databases  
A Tour



(美) Shashi Shekhar 著  
Sanjay Chawla

谢昆青 马修军 杨冬青 等译

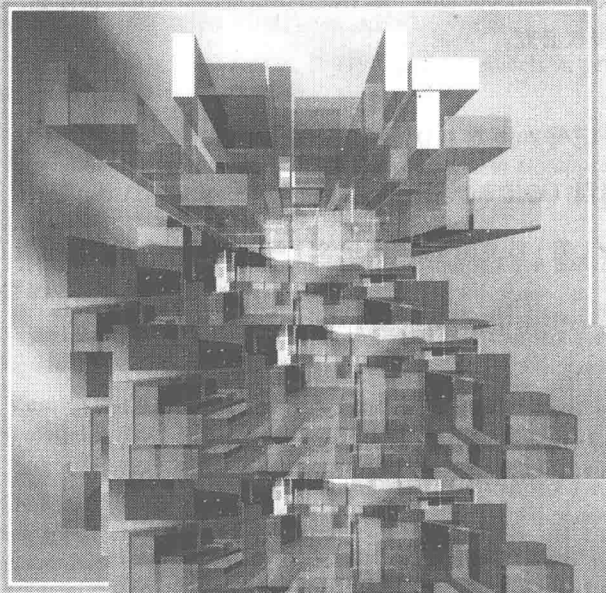


机械工业出版社  
China Machine Press

# 空间数据库

Spatial Databases

A Tour



(美) Shashi Shekhar 著  
Sanjay Chawla 著

谢昆青 马修军 杨冬青 等译



机械工业出版社  
China Machine Press

空间数据库是近年的热点研究领域，是一门前沿的交叉学科。本书全面介绍了空间数据库的概念、应用领域、查询语言、空间数据的索引和存储机制、空间查询处理和优化等内容，对空间数据挖掘和空间数据仓库也有精彩的论述。本书条理清晰，叙述严谨，实例丰富，曾得到业内权威人士的赞誉。本书的每章之后都附有习题，帮助读者检验学习效果。本书既适合作为计算机及相关专业的本科生、研究生的教材，也适合IT业的研究人员、技术人员阅读。对于想了解空间数据库的初学者来说，本书也是一本极有价值的参考书。

Simplified Chinese edition copyright © 2004 by Pearson Education Asia Limited and China Machine Press.

Original English language title: *Spatial Databases: A Tour* (0-13-017480-7), first edition by Shashi Shekhar and Sanjay Chawla, Copyright © 2003.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall, Inc.

本书封面贴有Pearson Education（培生教育出版集团）激光防伪标签，无标签者不得销售。版权所有，侵权必究。

**本书版权登记号：图字：01-2003-4986**

**图书在版编目（CIP）数据**

空间数据库 / (美) 沙克哈 (Shekhar, S.) 等著；谢昆青等译. —北京：机械工业出版社，2004.1

书名原文：Spatial Databases: A Tour

ISBN 7-111-13221-1

I. 空… II. ①沙… ②谢… III. 数据库系统 IV. TP311.13

中国版本图书馆CIP数据核字（2003）第100371号

机械工业出版社（北京市西城区百万庄大街22号 邮政编码 100037）

责任编辑：朱 劼

北京昌平奔腾印刷厂印刷·新华书店北京发行所发行

2004年12月第1版第2次印刷

787mm × 1092mm 1/16 · 20印张

印数：4 001-5 000册

定价：39.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换  
本社购书热线：（010）68326294

# 译者序

空间信息是指与位置（特别是地理位置）有关的信息，它在信息中占有相当大的比例（有人统计可以达到80%）。然而，空间信息又有其特殊的一面，它具有诸如数据量巨大、结构复杂多样、操作是计算密集型的、具有自相关性等特性。随着IT技术的迅速发展，以GIS为代表的空间信息技术在各领域得到了应用，同时遥感等空间信息获取技术不断进步，现代社会对位置服务和分析决策的需要也日益迫切，因此深入研究和掌握空间信息技术的理论与方法的重要性也日益凸显出来。

本书深入浅出地介绍了空间数据库的相关内容和知识。作者以清晰、有条理和严谨的方式，系统地阐述空间数据库的表达、查询和分析等诸多方面，举例通俗恰当，各章都附有针对性练习。本书对空间数据挖掘和空间数据仓库等当前热点研究的精彩描述，得到了Benjamin Wah教授、Michael F. Goodchild教授以及微软的Jim Gray等业内权威人士的高度评价。空间数据库方面的专著可谓凤毛麟角，许多著作只是涉及其中的部分内容，而本书却系统全面地论述这个主题，并采用业界标准，兼顾前沿研究、商业化趋势以及初学者实践几个方面，是一本不可多得的教材。

本书的内容安排如下：

- 第1章介绍空间数据库学科的应用领域和背景知识。
- 第2章介绍空间数据模型。
- 第3章介绍传统查询语言的空间扩展，并对扩展SQL的空间特性进行讨论。
- 第4章论述空间数据存储和索引机制以及压缩方法，探讨提高查询处理性能的途径。
- 第5章介绍空间查询处理和优化的方法，包括过滤-精炼策略。
- 第6章是空间数据库技术在空间网络方面的应用，包括网络数据模型和查询语言。
- 第7章介绍空间数据挖掘，包括空间依赖概念及其建模方法。
- 第8章讨论栅格数据库、基于内容的检索方法以及空间数据仓库的发展趋势。

本书由谢昆青、马修军和杨冬青组织并参加翻译和审校。参加翻译工作的还有陈冠华、吴杰、张远志、张敏、雷小锋、马秀莉、张惠彬、应莺、张月祥、孙岩峰、徐丹、刘晨、韩亮、蔡颖琨和袁杰。

由于翻译水平有限，书中难免有疏漏和错误之处，欢迎读者批评指正。

译者

# 序 1

早在三万五千年前，克鲁马努猎人在法国拉斯科附近的洞穴岩壁上绘制行走轨迹图形和有意义的刻符来描绘迁徙路线。从那时起，人们就使用带有地理信息的图形了。今天，地理信息系统（GIS）得到广泛的应用：从跟踪北美驯鹿和北极熊的迁徙路线，确定石油开采对野生动物的影响，协助农民最低限度地施用杀虫剂，帮助公司供应部门的经理预测配送仓库的最佳设立位置，直到建立降水和航空照片信息与湿地在每年特定季节变干之间的联系。

从最严格的意义上讲，地理信息系统是一个计算机系统，用于汇集、存储、操纵和显示与位置相关的数据。然而，现代地理信息系统通常要从多个不同来源接收各种形式的数据，以便处理查询和帮助分析信息。从广义上讲，地理信息系统不仅能将地理信息转换和存储为数字形式进行分析，也必须能在空间数据库中进行收集、变换、聚集、索引、链接和挖掘。现代地理信息系统能够集成用其他方法很难关联起来的信息，同时可以结合地图化的变量来构建和分析新变量。

Shekhar和Chawla正是从这种高度完成了一项展示地理信息处理的原理和趋势的非凡工作。本书是讲述概念与方法的力作，循序渐进地讲解模型、语言和算法，层次清晰，结构合理。作者不仅解释概念，而且使用大量例子加以说明。本书强调了将空间数据集成到传统数据库中的很多重要主题，从深奥的空间建模本体论问题到重要的文件管理问题无不涉及。每一章后面都有许多引人思索的习题，帮助读者更好地理解书中的概念和算法。本书最后所展示的空间数据挖掘和空间数据库未来发展趋势尤为精彩，有助于读者了解新兴的研究领域。

本书适合作为地理信息系统方面交叉学科的课程教材，也可以作为该领域从业人员的参考书。即使没有受过正规数据库方面的培训，读者也可以轻松地理解和应用从书中学到的概念和算法。其他学科背景的人也能从本书所讲述的技术中获益，使这些技术得以广泛地应用于政府、商业和工业等部门。

本书是空间数据库领域的第一本著作，作为本书的读者，我确信，从这个令人兴奋和有重要意义的领域中所学到的一切将使你获益匪浅。

Benjamin Wah

伊利诺斯大学电子与计算机工程系教授

IEEE计算机协会主席



## 序 2

空间信息，也就是在某个空间框架（例如地球表面）中对象的位置信息，长期以来被视为特殊的计算问题。早在1972年，空间数据处理（spatial data handling）这一术语就开始使用，它是指一群研究人员的研究活动。这些研究人员承诺共同开发电子化的数据处理方法，用以提高诸如地图编撰、地图测量和空间数据分析等领域的生产率。在高层结构中经常会用到空间信息。尽管在20世纪60年代就开始出现的每一个经典数据库管理模型都对空间应用领域有所考虑，但不管是关系模型还是面向对象的建模都不能完全适合这一领域。关系模型能够较好地处理拓扑关系，但对表示横跨空间区域的复杂层次关系却无能为力；而面向对象模型能够处理拓扑和层次关系，但难以处理空间中重要的连续性现象。

阐述空间数据库这一内容纷繁的领域的书籍可谓凤毛麟角，因此本书一定会大受欢迎。它涵盖了整个空间数据库领域，以清晰的概念、有条理的结构和严谨的表述，介绍了从表现、查询到分析的各个方面。数据挖掘一章尤其受人欢迎，它不仅讲述传统的空间数据分析方法，而且介绍近几年来发展的许多新技术。这些技术利用当今的高速计算能力，能够在超大型空间数据库中自动搜索异常情况和模式。本书是为计算机系的学生编写的，但同样适用于那些具有其他学科背景又希望学到比一般地理信息系统教材更严谨和更基础的方法的学生。

空间数据库的重要性日益增加，其中一个原因在于其应用范围已超出传统GIS领域。位置和时间是鉴别和刻画信息的强有力方法，因为许多数据集都具有空间和时间的“印记”。地图和地球照片显然如此，而许多报告、书籍、照片以及其他类型的信息亦不例外。因此，位置信息成为在分布的信息源（如因特网）中搜索相关信息的强大基础。人们逐渐认识到，空间（和时间）提供了集成信息的重要方法，这些信息已经远远超出了传统的空间数据库和GIS领域。本书在最后一章探讨了其中一些问题，并对许多人深信空间数据库的重要地位在未来几年会快速提升的原因提出了自己的见解。

Michael F. Goodchild

美国国家地理信息及分析中心，加州大学地理系

# 前 言

近年来,许多计算机应用领域通过扩充数据库管理系统的功能来支持与空间相关的数据。空间数据库管理系统(spatial database management system, SDBMS)研究是找到有效处理空间数据的模型和算法的重要步骤。

经过20多年的发展,空间数据库已成为一个热点研究领域,其研究成果(如空间多维索引)开始应用于许多不同领域。正是已有应用的需求推动了空间数据库管理系统的研究,这些应用包括地理信息系统(geographical information system, GIS)和计算机辅助设计(computer-aided design, CAD),以及诸如多媒体信息系统、数据仓库、美国国家航空航天局(national aeronautics and space administration, NASA)的地球观测系统等潜在应用。这些空间应用拥有上百万的使用者。

商业数据库的主要厂商已推出专门处理空间数据的产品,其中包括ESRI开发的空間数据引擎(spatial data engine, SDE),以及Intergraph、Autodesk、Oracle、IBM和Informix等公司在对象-关系数据库服务器上开发的空間数据插件,研究的原型系统有Postgres、Geo2和Paradise。这些系统都提供一组空間数据类型(如点、线和多边形)和一组空間操作功能(求交(intersection)、闭合(enclosure)和距离(distance))。开放地理信息系统(Open Geographic Information System, OGIS)协会制定出一套空間数据类型和空間操作的现行标准,使得空間类型和操作可以像SQL3那样成为对象-关系查询语言中的一部分。为了增强性能,这些系统还为空間存取方法、空間范围查询以及空間连接提供了多维空間索引和算法。

把空間数据集成到传统数据库中意味着要在不同层次上解决许多重要问题,这些问题的范围广泛,从关于空間建模的深奥本体论问题(例如,“它应该是基于场的还是基于对象的”,很像物理学中的波粒二象性)到文件管理这类平凡但重要的问题。这些不同课题使空間数据库管理系统研究实际变为多学科的问题。

我们用一个国家数据集的例子来说明空間数据库的特别需求。一个国家至少有一个非空間的数据(国名)和一个空間数据(国界)。国名的存储或表示不会产生任何问题,但国界的存储或表示就不那么简单了。假定用一个直线段的集合表示国界,这时会要求数据库系统能支持空間数据类型“线”、“点”和“面”,以便对“国家”这个对象进行空間查询。操作和组合这些新的数据类型需要遵从某些固定的规则,于是空間代数便应运而生了。由于空間数据具有可视性和数据量庞大的性质,所以必须扩展数据库系统以提供可视化查询处理和特殊的空間索引。数据库的其他重要问题(如并发控制、批量加载、存储和安全机制等)也都必须重新加以考虑和调整,以便构建高效的空間数据库管理系统。

本书以明尼苏达大学的科学数据库(Csci 8705)这门研究生课程的讲义为蓝本。计算机科学及其他系的研究人员与学生都认为该课程非常有用并能应用到他们的工作中。尽管这样的科目在学生中得到很好的反响并引起他们极高的兴趣,但市面上却找不到一本教材能满足听课

者对多学科的需要。最近, [Scholl等, 2001]写的一本书主要介绍了与查询语言和访问方法有关的传统主题, 并没有涉及像空间网络(比如, 道路地图)和空间模式的数据挖掘这类流行的主题。Adam和Gangopadhyay在1997年编纂了一本GIS领域中与数据库相关的问题的专著, 但是没有提到业界的最新技术。另一本关注GIS的书[Worboys著, 1995]也只有两章讨论了数据库问题。这类书多数没有采用像OGIS这样的行业标准, 也缺乏适当的辅导, 例如在每一章节后面提供习题和讨论问题, 帮助学生理解主要概念。毫无疑问, 从事数据库、并行计算、多媒体信息、土木机械工程和林业方面的学术研究的人员迫切需要一本详尽叙述空间数据库的教材。一项调查表明, 业界的专业技术人员, 包括GIS和CAD/CAM软件的开发者, 也希望了解有关空间数据库的知识。

在开始撰写本书之前, 我们完成了一份“空间数据库: 成就和研究需求”(Spatial Databases: Accomplishments and Research Needs)的调研报告, 发表在IEEE Transactions on Knowledge and Data Engineering (1999年1月)上。我们发现计算机科学方面的文献多侧重于研究某些特定问题(例如空间索引、空间连接算法), 而很少涉及其他许多重要问题(例如空间数据的概念建模)。在本书有关这些问题的章节中, 我们参考了来自GIS业界和非计算机科学专业的GIS研究人员的很多思想和观念。

从1998年的假期开始, 我们着手编写本书。通过充实原有的课程讲义, 完成了许多章的初稿。本书的初稿在明尼苏达大学的数据库课程中使用。随后, 利用审校人、同事和学生的反馈, 对本书进行了修订。

本书的特点如下:

- 本书旨在提供空间数据库管理系统的全面概述。它不仅涵盖传统主题(如查询语言、索引和查询处理), 还包括许多目前流行的问题(如空间网络和空间数据挖掘)。
- 每章都提供一组习题, 读者可以测试自己对核心概念的理解, 将这些概念应用于新的领域, 甚至举一反三地应用该章内容。本书的网站还安排附加的教学辅导(例如实验、课程讲义等)。
- 尽量将视野超越GIS。空间数据库管理系统技术在许多领域得到应用, 包括多媒体信息、CAD/CAM、天文学、气象学、分子生物学和计算力学。
- 每章都尝试给出对象-关系数据库的框架, 这是商业数据库应用的发展趋势。在适当的情况下, 该框架允许空间数据库重用关系数据库的功能, 同时根据需要来扩展关系数据库的功能。
- 采用符合标准(例如OGIS)的空间数据类型和操作来说明常见的空间数据库查询。这些遵从标准的类型和操作可以兼容像SQL-3这样的对象关系查询语言。
- 本书内容自成一体, 读者无需具备GIS或数据库的知识。
- 完全覆盖空间网络的建模、查询和存储方法等各个方面。
  - 详尽讨论有关空间数据挖掘的问题。
  - 既有前沿研究的讨论, 也有商业趋势的讨论。
  - 包含很多易于理解的常识性应用领域的例子。



## 本书组织方式

本书共有8章，每一章介绍空间数据库的一个重要部分。第1章介绍空间数据库的基本概念。第2章重点介绍空间数据模型，引入场和对象的二分概念，以及它们在数据库设计中的应用。第3章探讨如何扩展传统查询语言来支持空间数据库，广泛讨论各种扩展SQL空间能力的提议。第4章描述空间数据存储和索引方案。由于空间数据库处理的是海量数据，所以提供合理的存储、压缩和索引方法来提高查询处理的性能，对于数据库管理系统来说至关重要。第5章从查询语言和索引入手，继续讨论查询处理和优化。此时读者会发现，要放弃或从根本上修改传统数据库的许多标准技术才能使之应用于空间问题。该章还引入空间查询处理中的过滤-精炼策略。第6章说明如何将空间数据库技术应用于空间网络，该章还将介绍网络数据模型和查询语言。第7章全面涵盖新兴的空间数据挖掘领域。在这一章我们为读者揭示出空间数据集普遍存在的空间依赖性，以及如何对其建模并整合到数据挖掘过程中。第8章讨论空间数据库的发展趋势。

## 致谢

在本书的编写过程中，我们得到了许多人的帮助，在此向他们深表谢意。没有明尼苏达大学计算机科学系的Vipin Kumar教授和ESRI的主席Jack Dangermond博士的鼓励，本书是不会面世的。本书从ESRI的研究人员和产品中获益颇多。我们也要感谢Oracle公司的Siva Ravada博士和Illustra公司的Robert Uleman博士帮助我们理解他们的空间插件。感谢Alan Apt及其Prentice Hall出版社的优秀员工，他们在本书撰写过程中不断给予我们帮助和鼓励。一些未留姓名的评论者也对本书提出了宝贵的意见，这里一并表示感谢。

我们对空间数据库的研究还得到了很多组织的支持，这些组织包括联合国开发计划署 (United Nations Development Programme, UNDP)、美国国家科学基金会 (National Science Foundation, NSF)、美国国家航空航天局 (National Aeronautics and Space Agency, NASA)、美国陆军研究实验室 (Army Research Laboratories)、美国农业部 (U.S. Department of Agriculture)、美国联邦公路管理局 (Federal Highway Administration, FHA)、美国运输部 (U.S. Department of Transportation)、明尼苏达州运输部 (Minnesota Department of Transportation)、城市与区域事务中心 (Center for Urban and Regional Affairs) 以及Computing Devices International。其中许多研究项目总结为调研成果文献，并据此开发出用于空间数据库相关问题的技术。

特别要感谢明尼苏达大学计算机科学系空间数据库研究组的成员。他们从各个方面为本书的出版做出了贡献，包括查找文献、设计样例和图表、对各种方法提出见解，以及设计恰当的问题表述和有新意的解决方法。我们非常感谢Vatsavai Ranga Raju，他仔细地审阅并多次修改了本书的早期版本。还要感谢学习课程Csci 8701和Csci 8705的同学试用本书并为本书的修订提出有益的建议。

多年来与其他许多人士的探讨也使我们获益良多。他们是Marvin Bauer、Yvan Bedard、Paul Bolstad、Nick Bourbakis、Thomas Burk、John Carlis、Jai Chakrapani、Vladimir

Cherkassky, Douglas Chub, William Craig, Max Donath, Phil Emmerman, Max Egenhofer, Michael Goodchild, Ralf Hartmut Gueting, Oliver Gunther, John Gurney, Jia-Wei Han, Ravi Janardan, George Karypis, Hans-Peter Kriegel, Robert McMaster, Robert Pierre, Shamkant Navathe, Raymond Ng, Hanan Samet, Paul Schrater, Jaideep Srivastava, Benjamin Wah, Kyu-Young Whang 和 Micheal Worboys。

# 目 录

- 1-1 GIS这一缩写在过去20年中的演化。在20世纪80年代，GIS表示地理信息系统，在20世纪90年代更流行的提法是地理信息科学（geographic information science），而现在GIS则朝着地理信息服务（geographic information service）的方向发展
- 1-2 由Landsat卫星提供的带空间信息叠加图层的明尼苏达州Ramsey郡的地图
- 1-3 边界ID为1050的人口普查区
- 1-4 在关系数据库中，容纳多线数据类型所需的4个带重叠属性的表
- 1-5 数据库的演化[Khoshafian and Baker,1998]
- 1-6 三层体系结构
- 1-7 用来说明连接与空间连接间的区别的两个关系
- 1-8 用于减少计算时间的过滤-精炼策略
- 1-9 确定相交矩形对。a) 两个矩形集合： $R$ 和 $S$ 。b) 标出左下角和右上角坐标的矩形 $T$ 。c) 排序后的矩形集合。注意，平面扫描算法的过滤特性在这个例子中，有12个可能的矩形对将被连接。过滤阶段将可能的数目降低到5。然后用准确几何测试来验证这5对对象是否满足查询谓词[Beckmann et al.,1990]
- 1-10 a) 程序员的观点。b) DBMS设计者的观点
- 1-11 对多维数据排序的两种方法。a) 行排序。b) Z排序。如果按照数字的升序来画线，Z模式就会很明显
- 1-12 a) 二叉树。b) B树
- 1-13 为了处理空间对象将B树扩展为R树
  
- 2-1 对象-场的二分法。a) 一幅显示3种林分（松树、冷杉、橡树）的地图。b) 对象的观点：将地图表示为3个对象的集合，每个对象有唯一的标识符、主要的树种和一块区域。区域的边界（一个多边形）由坐标指定。c) 场的观点，这时区域中的每个点被映射为主要树种对应的值
- 2-2 OGIS提出的关于空间几何体的基本构件（采用UML概念表示）
- 2-3 九交模型[Egenhofer et al., 1989]
- 2-4 州立公园例子的ER图
- 2-5 州立公园例子的关系模式
- 2-6 point、line、polygon和elevation的关系模式

- 2-7 州立公园例子的带象形符号的ER图
- 2-8 State-Park例子的UML类图
  
- 3-1 World数据库的ER图
- 3-2 河流的Buffer区域以及在该区域内外的点
- 3-3 示例对象[Clementini and Felice,1995]
- 3-4 州立公园数据库的ER图
  
- 4-1 将记录从Country、City和River表映射到磁盘页
- 4-2 City表的散列文件组织方式
- 4-3 City表的有序文件组织方式
- 4-4 生成一条Z曲线[Asano et al.,1997]
- 4-5 生成一条Hilbert曲线[Asano et al.,1997]
- 4-6 计算z值的例子
- 4-7 查找z值的记录
- 4-8 Hilbert曲线转换的例子
- 4-9 聚类的图示: a) Z曲线的两个聚类。b) Hilbert曲线的两个聚类
- 4-10 City表中的二级索引
- 4-11 City表中的主索引
- 4-12 点(A、B、C、D)的固定网格结构
- 4-13 一个二维网格目录和数据页面
- 4-14 线性比例的网格文件
- 4-15 一个空间对象集合
- 4-16 R树的层次结构
- 4-17 R+树的层次结构
- 4-18 R+树内部结点的矩形
- 4-19 分解对象的不规则多边形
- 4-20 一个R链接树的一部分[Kornacker and Banks,1995]
- 4-21 在两个关系上建立一个连接索引
- 4-22 等值连接的元组级邻接矩阵和空间连接的比较
- 4-23 从一个连接索引建立一个PCG
- 4-24 练习题1的图示
  
- 5-1 多步处理 [Brinkhoff et al., 1994]

- 5-2 查询优化器的模式
  - 5-3 查询树
  - 5-4 下推：选择操作
  - 5-5 选择下推并不总是有效
  - 5-6 执行策略：查询计算计划
  - 5-7 执行策略：查询树
  - 5-8 两种分解方式
  - 5-9 两个分布关系：FARM关系有1000个元组，DISEASE\_MAP关系有100个元组
  - 5-10 Web GIS体系结构
  - 5-11 并行体系结构选项
  - 5-12 不同数据分配方法示例
  - 5-13 地形可视化系统的组件
  - 5-14 示例多边形地图和范围查询
  - 5-15 并行模式的不同模块
  - 5-16 处理器间多边形/外包框分割的可选方案
- 
- 6-1 空间网络的两个例子
  - 6-2 图的三种不同表示
  - 6-3 河流网例子的图模型
  - 6-4 关系 $R$ 和它的传递闭包 $X$
  - 6-5 SQL的CONNECT子句操作
  - 6-6 BFS和DFS算法的结果（源结点为1）
  - 6-7 寻径的例子（一）
  - 6-8 寻径的例子（二）
  - 6-9 图以及它的非规范化的表
  - 6-10 明尼阿波利斯市的主要公路
  - 6-11 明尼阿波利斯市的主要公路的CCAM分页中的分割边
  - 6-12 一个示例网络的聚集和保存（key表示空间顺序）
- 
- 7-1 数据挖掘过程。数据挖掘过程需要领域专家与数据挖掘分析员进行密切交互。挖掘过程的输出是一组假设（模式），这些结果可以用统计工具进行严格验证，可用GIS可视化地表现出来。最后，分析员可以解释这些模式，制定并推荐合适的方案
  - 7-2 一个数据挖掘算法的搜索结果。a) 一个可能的模式，总共有 $2^{16}$ 种可能模式。b) 如果限



定每个 $2 \times 2$ 的块只能归入一个类，则可能的模式总数就减少到 $2^4$ 个。在其他一些信息的基础上，数据挖掘算法可以快速发现“最理想的”模式

- 7-3 Darr湿地，1995年。a) 学习数据集：沼泽地的几何形状和红翅黑鹇巢穴的位置。b) 植被韧性在沼泽地上的空间分布。c) 水深的空间分布。d) 到开阔水体距离的空间分布
- 7-4 满足经典回归的随机分布假设的空间分布
- 7-5 a) 空间网格。b) 邻接矩阵。c) 行规范化的邻接矩阵
- 7-6 Moran's  $I$  系数。图像b和c的像素值集合相等，但它们拥有不同的Moran's  $I$  系数
- 7-7 两类预测问题的四种可能输出结果
- 7-8 ROC曲线。a) 利用1995年Darr湿地数据建立的经典模型与SAR模型的ROC曲线的比较。b) 利用1995年Stubble湿地数据构建的两种模型的比较
- 7-9 空间数据情况下ROC曲线的问题。a) 鸟巢的实际位置。b) 实际有鸟巢的像素。c) 通过模型预测到的位置。d) 用另一种模型预测到的位置。预测d在空间上比c更精确，经典的分类准确性度量无法捕获这种差别
- 7-10 隐含盐和辣椒空间模式的空间数据集
- 7-11 事务数据库、频繁项集和高置信度规则的例子
- 7-12 同位模式举例
- 7-13 空间聚类的两种解释。如果目标是确定主导周围环境（所谓影响力）的位置，则聚类是S1和S2；如果目标是确定均一属性值的区域，则聚类是A1和A2
- 7-14 a)  $4 \times 4$ 灰度图像。b) 使用EM算法产生的图像的标号。c) 使用邻域EM算法对同样的图像产生的标号。注意，空间平滑是通过修改目标函数来实现的
- 7-15 使用邻域EM算法。a) 由于预期的聚类没有考虑任何空间信息，所以结果不佳。b) 考虑空间信息（ $\beta = 1.0$ ）使结果有很大改善。c) 过于强调空间信息（ $\beta = 2.0$ ）再次导致糟糕的结果
- 7-16 孤立点检测的数据集
- 7-17 检测空间孤立点的变差云图和Moran散点图
- 7-18 检测空间孤立点的散点图和空间统计量 $Z_{(i)}$
- 7-19 交通传感器装置构成的网络
- 7-20 空间和时间邻域
- 7-21 交通流量数据中的空间孤立点
  
- 8-1 一个连续函数及其栅格表示
- 8-2 局部操作的示例：阈值化

- 8-3 单元格的邻域和一个聚焦操作例子：聚焦求和
- 8-4 区域操作示例：区域求和
- 8-5 全局操作示例
- 8-6 剪裁操作：保留维数
- 8-7 切片操作：降低维数
- 8-8 存储矩阵的不同策略
- 8-9 拓扑相邻图[Ang et al., 1998]。两个关系的距离是图中它们之间的最短路径
- 8-10 方位邻近图[Ang et al., 1998]。两个关系的距离是图中它们之间的最短实线路径
- 8-11 视觉角度：将distance作为最小区分空间的谓词
- 8-12 图像和它的ARG[Petrakis and Faloutsos, 1997]。ARG被映射为一个N维的特征点
- 8-13 基于内容检索的通用方法
- 8-14 分布聚集函数的计算
- 8-15 代数聚集函数的计算
- 8-16 GIS聚集函数的几何并的例子
- 8-17 零维、一维、二维和三维数据立方体
- 8-18 数据立方体的一个例子
- 8-19 使用group-by的一个例子

# 表 目 录

- 1-1 常用GIS分析操作列表[Albrecht, 1998]
- 1-2 不同的空间类型及示例操作
  
- 2-1 拓扑和非拓扑操作举例
- 2-2 动态空间操作的典型示例[Worboys, 1995]
- 2-3 ER与UMLCD中的概念
  
- 3-1 World数据库中的表
- 3-2 关系代数中两个基本运算(选择和投影)的运算结果
- 3-3 关系R和S笛卡儿积运算
- 3-4 集合运算的结果
- 3-5 条件连接运算的步骤
- 3-6 SQL语言定义的Country表和River表的模式
- 3-7 选择、投影、选择并投影操作
- 3-8 查询例子的结果
- 3-9 SQL的OGIS标准定义的一些操作[OGIS,1999]
- 3-10 基本表
- 3-11 查询7的结果
- 3-12 Country表的创建语句
- 3-13 州立公园数据库的表
  
- 4-1 传统DBMS、应用程序和SDBMS在CPU和I/O相对代价方面的特征
- 4-2 Western Digital Caviar AC36400磁盘驱动器的物理和性能参数
  
- 6-1 BART系统的Stop表和DirectedRoute表
- 6-2 BART系统的RouteStop表
- 6-3 河流网的River关系和FallsInto关系
  
- 7-1 用于预测红翅黑鹇巢穴位置的习性变量(注意,这里有6个独立变量和1个依赖变量,

而且，依赖变量是二值的)

- 7-2 预测红翅黑鹇巢穴位置的解释性变量的Moran's  $I$ 系数
- 7-3 16 × 16灰度图像
- 7-4 三个规则的支持度和置信度
- 7-5 从1995年Darr湿地数据中发现的空间关联规则的例子
- 7-6 聚类中局部搜索的四种选项
- 7-7 设施数据库
  
- 8-1 聚集操作
- 8-2 地图再分类的SQL查询
- 8-3 GROUP-BY查询列表
- 8-4 在“Region”维上对值“America”切片
- 8-5 在图8-19中，表SALES-L2的“Year”维上对值“1994”切块并在“Region”维上对值“America”切块