

国家科学技术学术著作出版基金资助出版

空间数据挖掘理论与应用

李德仁 王树良 李德毅 著

科学出版社

北京

内 容 简 介

本书主要提出云模型、数据场、地学粗空间和空间数据挖掘视角等新技术,构建空间数据挖掘金字塔,研究空间数据挖掘的数据源,导出空间观测数据清理的“李德仁法”,研究基于空间统计学的图像数据挖掘,提出“数据场-云”聚类、基于数据场的模糊综合聚类和基于数学形态学的聚类知识挖掘算法,研究基于归纳学习的空间数据挖掘、基于概念格的遥感图像数据挖掘和地理信息系统(GIS)数据挖掘,结合滑坡监测、银行经营收益分析及选址评价、遥感图像土地利用分类、土地资源评价、火车运行安全检测等实例系统研究空间数据挖掘可操作性,并在此基础上自主研发了空间数据挖掘原型系统 GISDB Miner 和 RSImage Miner。

本书可供空间数据挖掘、计算机科学、地球空间信息科学、GIS、遥感(RS)、全球定位系统(GPS)、数据分析、人工智能、认知科学、空间资源规划、土地科学、灾害防治、管理科学与工程和决策支持等领域的研究人员和开发人员使用,亦可作为高等院校相关专业的本科生、研究生教学用书和参考用书。

图书在版编目(CIP)数据

空间数据挖掘理论与应用/李德仁,王树良,李德毅著. —北京:科学出版社,2006

ISBN 7-03-015123-2

I. 空… II. ①李…②王…③李… III. 数据利用-计算机应用-地理信息系统 IV. P208-39

中国版本图书馆 CIP 数据核字(2005)第 016603 号

责任编辑:朱海燕 韩 鹏 / 责任校对:朱光光

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>

印 刷

科学出版社发行 各地新华书店经销

*

2006年10月第 一 版 开本:787×1092 1/16

2006年10月第一次印刷 印张:38 1/4 插页:2

印数:1—3 000 字数:843 000

定价:78.00 元

(如有印装质量问题,我社负责调换〈科印〉)

作者简介



李德仁,1939年生,德国斯图加特大学博士,摄影测量与遥感学家,武汉大学教授,中国科学院院士,中国工程院院士,国际欧亚科学院院士,第九届全国政协委员,测绘遥感信息工程国家重点实验室学术委员会主任,全球对地观测卫星委员会中国副主席和一体化全球观测战略伙伴关系联合主席之一,国家有突出贡献的中青年专家,国家“973”专家顾问组成员,国务院学位委员会评议组成员,国家自然科学基金委员会学科评议组成员,武汉中国光谷首席科学家,中国测绘学会副理事长,国际摄影测量与遥感学会第Ⅲ委员会主席(1988~1992年)和第Ⅵ委员会主席(1992~1996年),原武汉测绘科技大学校长。1982年,提出的选权迭代法,被国际测量学界称之为“李德仁方法”;1985年,提出的可靠性理论,“科学地解决了测量学上一个百年来的问题”;20世纪90年代,提出地球空间信息科学的概念和理论体系;1994年,提出并奠定了空间数据挖掘的基础;21世纪以来,又提出了广义和狭义空间信息网格的概念,大力推进空间信息的社会服务,其成果相继获得原联邦德国“汉莎航空测量奖”、国家科技进步二等奖等10余项奖励。已发表论文450余篇,出版专著9部;已培养硕士生60多名,博士生80多名,4篇博士论文被评为全国优秀博士学位论文。



王树良,1975年生,武汉大学和香港理工大学博士,武汉大学教授,美国 Carnegie Mellon University 的 iCarnegie 教师,全国优秀博士学位论文获得者。担任 *International Journal of System Science*、*International Journal of Data Mining and Data Warehousing* 等国际刊物的客座编辑, *Data Mining and Knowledge Discovery*、*ISPRS Journal of Photogrammetric and Remote Sensing* 等杂志的审稿人和 *Advanced Data Mining and Applications* 学术委员会合作主席。由 Springer 等出版专著3部,已有多篇论文被 SCI 收录。主要研究领域为空间数据挖掘。



李德毅,1944年生,英国爱丁堡 Heriot-Watt 大学博士,计算机工程和人工智能学家,中国工程院院士,国际欧亚科学院院士,第十届全国政协委员,中国工程院信息与电子工程学部副主任,国家自然科学基金委员会信息学部主任,国家软件工程重点实验室学术委员会主任,中国电子学会副理事长,中国人工智能学会副理事长, *Knowledge and Information Systems*、*软件学报* 等学术期刊编委。已出版专著4部,发表论文120多篇,培养博士、硕士研究生40多名。

序 一

计算机技术的不断进步,使得人们从最初用它代替人工计算,扩展为用计算机系统来模仿人类的视觉、听觉及其思维等高级智能。人们曾经为建立计算机专家系统而努力工作,但不久就遇到了难以建立专家知识库的问题。随后,计算机专家们就转向机器学习,试图让计算机像人一样,有自学习功能。但这种努力也遇到问题,即让计算机从哪里学习呢?正在步履维艰之时,计算机数据库技术有了突飞猛进的发展,大量的数据通过文字、图表、影像、多媒体等技术被输入计算机中,数据库体之大可以达到上千个 TB 的量级。到了 20 世纪 90 年代,数据挖掘(或知识发现)作为数据库技术、人工智能技术、数据处理技术和可视化技术的集成而变得愈来愈引人注目。人们终于认识到很多知识原来就隐藏在大量的数据之中。从数据库中通过数据库管理系统和应用程序可以获得信息;而从数据库中通过一个知识发现的工具应当可以获得知识,这些知识可以自动构成计算机专家系统的知识库。

1993 年,李德毅院士把计算机科学界开展数据挖掘的研究动向告诉了其兄长李德仁院士。当时,李德仁院士认为空间数据库的数量、大小和复杂性也都在激增,特定问题或特定环境下的空间数据,似乎是一种原始的、混乱的、不成形的自然状态积累,但又是一种可以从中生长出秩序和规则的源泉,为什么不可以研究空间数据挖掘呢?1994 年,李德仁院士在加拿大第五届 GIS 年会上提出了“Knowledge Discovery from GIS”的理念,建议在智能化 GIS 中利用从数据库发现知识(KDD)的方法,透过表观上的千头万绪、杂乱无章,去挖掘蕴涵其中的规则性、有序性、相关性和离群性。

随后,李德仁院士和李德毅院士兄弟二人筹划共同指导空间数据挖掘的博士。由于该项研究的难度较大,直到 1995 年邸凯昌决定跟随他们做博士论文后,他们的上述想法才得以初步实现。邸凯昌博士在完成其博士论文后立即被美国俄亥俄州立大学邀请为博士后研究员,武汉大学出版社将其博士论文出版。1999 年,在邸凯昌博士答辩时,他们发现王树良对空间数据挖掘有浓厚的兴趣和一定的研究基础,于是又指导王树良在邸凯昌的基础上继续向前研究。后来,他们组建了自己的研究团体,目前已经发展为由数十名院士、教授、副教授、博士后、博士生和硕士生组成的老、中、青相结合、地球空间信息科学和计算机科学等学科交叉的梯队。兄弟二人敏捷的思维、渊博的学识、高屋建瓴的学术视野、实事求是的治学态度、执著的人生信念和宽以待人的品质,引导和激励着整个研究团体,在空间数据挖掘的道路上逐步前进。

在这部力作中,系统研究了空间数据挖掘的国内外进展,在概述了可用于

空间数据挖掘的理论方法的基础上,主要提出了云模型、数据场、地学粗空间和空间数据挖掘视角等新技术,构建了空间数据挖掘金字塔,全面分析了空间数据挖掘的数据源及其清理方式,导出了空间观测数据清理的选权迭代法,研究了基于空间统计学的图像数据挖掘,提出了“数据场-云”聚类算法、基于数据场的模糊综合聚类算法和基于数学形态学的聚类算法,研究了基于归纳学习的空间数据挖掘、基于概念格的遥感图像数据挖掘和 GIS 数据挖掘。为了展现空间数据挖掘的可操作性,具体研究了长江三峡宝塔滑坡监测、银行经营收益分析及选址评价、遥感图像土地利用分类、土地资源评价、火车运行安全检测等多个关系国计民生的空间数据挖掘实例。最后,自主研制了一个空间数据挖掘原型系统。

可见,作者们既有地球空间信息科学的背景,又具备计算机科学的知识,从事理论研究和应用开发多年,而且都有所建树,是地球空间信息科学和计算机科学在空间数据挖掘中真正有机的交叉、融合和升华。在空间数据挖掘中,同时可以满足地球空间信息科学的学者掌握计算机科学的要求,以及计算机科学的学者了解地球空间信息科学的需要。在研究空间数据挖掘时,他们紧密围绕空间数据的特点,理论切实联系实践,并始终以解决实际问题为要旨。

在科学界,如果说兄弟院士不多的话,那么不同学科的兄弟院士共同指导的博士就更少了;更进一步地,兄弟院士共同指导自己的博士,联名著述的就更少之又少了。李德仁院士和李德毅院士为其中之一,令人敬佩!他们第一次把自己最近十几年来在空间数据挖掘中的丰富研究成果汇集熔炼,凝聚了系列国家重大课题的创新研究及其横向应用成果,汇总了十几名教授、副教授、博士后、博士和硕士的智慧。其创新内容突出,利用自己的创新成果研究空间数据挖掘,而不是单纯利用已有的数据挖掘策略。绝大部分成果都已经被邀请在国际会议上做口头报告,并在交流时获得与会国际同行学者的赞同,部分成果已经被国际学术期刊发表,为 SCI、EI、ISTP 等国际权威检索机构收录;而且理论研究始终以解决实际问题为要旨,积极把理论成果应用在关系国计民生的实际问题中,服务社会,造福人民。如长江三峡滑坡监测。全书内容丰富、结构严谨、层次清晰、逻辑严密,成稿后多人阅读、多次征求意见,数易其稿,精益求精。这种一丝不苟的治学态度,值得称赞和学习。

随着信息时代、网络社会和知识经济的到来,空间数据挖掘开始引起越来越多学者的兴趣。目前,它已经成为国际研究的热门课题。祝作者的研究更上一层楼,做出更大的成就;愿空间数据挖掘的理论和应用研究,开放出更多、更鲜艳的花朵!



2005年1月28日

序 二

现代数据采集技术的发展,促使数据量迅速膨胀,也导致了人们利用数据的困难。数据挖掘应运而生,空间数据挖掘为其主要的研究内容。长期以来,人们常常使用数据挖掘的方法研究空间数据挖掘,而较少考虑空间数据的独有特征。实际上,相对一般的事务数据挖掘而言,空间数据挖掘更为复杂,挖掘的对象不仅包含位置数据和属性数据,还有实体间的空间关系,而且空间数据的结构也比较复杂,既有表格数据,也有矢量数据和栅格数据。

李德仁院士是地球空间信息科学的专家,他首次提出了“Knowledge Discovery from GIS”的理念。李德毅院士是计算机科学的专家,他也较早研究数据挖掘。为了把数据挖掘原理和地球空间信息科学有机整合,从本质上研究空间数据挖掘,他们兄弟二人先后共同指导了邸凯昌、王树良两位博士。两位院士高屋建瓴,不仅成功培养了邸凯昌和王树良,还培育了一个空间数据挖掘的研究团队,并共同指导他们对空间数据挖掘展开了卓有成效的开创性研究。而且,他们的研究团队成功申请并出色完成了国家自然科学基金等多项课题,并把研究成果成功地应用于滑坡灾害监控等关系国计民生的实际问题。他们的很多成果,在国际交流时获得同行学者的认可,引起国际学术界的关注。

更为可贵的是,李德仁院士和李德毅院士携王树良博士,集中自己研究团队多年来的空间数据挖掘智慧,汇集熔炼成为一本专著,并强调系统性、可读性和可操作性。本书的可读之处良多,既有云模型、数据场、地学粗空间、挖掘视角、挖掘金字塔和挖掘机理等创新技术,又有崭新的数据清洗法和聚类算法,还研究了遥感图像分类、宝塔滑坡监测、银行收益评价、土地资源评价、火车安全检测等多个实例等等。纸短笔陋,实难以尽述。

阅读本书,计算机科学的读者可以了解地球空间信息科学的知识,地球空间信息科学的读者可以学习计算机科学的知识,数据挖掘的读者可以发现空间数据挖掘的独特魅力,空间数据挖掘的读者可以找到自己的归宿。

有幸先读本书,在赞赏作者的丰厚学术成果之余,也向大家鼎力推荐这本力作。

韓詠亭
Jian Han

2006年8月16日

序 三

计算机数据采集和存储技术的发展,使得数据库急剧膨胀,这造成了“数据过量而知识贫乏”的瓶颈。在这些数据中,大约 80% 与地理分布有关,例如空间数据。面对持续增长的海量空间数据,人们由于难以在没有知识辅助的情况下完全理解数据,因此开始关注数据挖掘中的空间数据挖掘。

除了数据挖掘的共性,空间数据挖掘还有自己的特性。它不仅包含位置和属性数据,还有实体间的空间关系,而且空间数据的结构也比较复杂,既有表格数据,也有矢量数据、栅格数据。可是,人们几乎是在用数据挖掘的方法研究空间数据挖掘,却很少顾及数据的空间特征。

中国科学院院士、中国工程院院士李德仁教授在 1994 年国际地理信息系统会议上提出从地理数据库发现知识,中国工程院院士李德毅教授是计算机科学领域的数据挖掘专家,他们兄弟二人共同指导王树良博士研究空间数据挖掘,其博士学位论文被评为全国优秀。而且,他们的研究团队成功申请并出色完成了国家自然科学基金、国家重大基础研究计划(973)、国家高技术研究发展计划(863)等 10 余项课题。同时,他们的理论技术也被成功地应用于国民经济建设中基于空间数据的决策支持。在研究空间数据挖掘的过程中,他们把计算机科学的数据挖掘原理和地球空间信息科学的空间特征紧密地结合在一起,得到我们国际同行的认可。

本书提出了云模型、数据场、地学粗空间和挖掘视角等新技术,构建了空间数据挖掘金字塔,给出了规则加例外的多层次挖掘机理,导出了空间观测数据清理的选权迭代法和 DHP 法,提出了“数据场-云”聚类算法、基于数据场的模糊综合聚类算法和基于数学形态学的聚类知识挖掘算法,研究了基于空间统计学、归纳学习和概念格的遥感图像数据挖掘,研究了长江三峡宝塔滑坡监测、银行经营收益分析及选址评价、遥感图像土地利用分类、土地资源评价、火车运行安全检测等多个关系国计民生的实例,自主研制了空间数据挖掘原型系统。

由此可见,《空间数据挖掘理论与应用》一书是集体智慧的结晶,作者在研究空间数据挖掘时,学科交融,理论联系实际,真正实现了学科的有机交叉。

感谢作者让我先睹为快!同时,我也诚恳地向大家推荐这本力作。

Lotfi A. Zadeh

2005 年 7 月 31 日

The technical progress in computerized data acquisition and storage results in the growth of vast databases. It has become an imminent bottleneck that data are excessive while knowledge is scarce. Among the excessive data, 80% are geo-referenced, i. e. spatial data. Faced with the large mounts of increasing spatial data, a terminal user has more difficulty in understanding them without the helpful knowledge from spatial databases. In order to overcome the bottleneck, spatial data mining was proposed under the umbrella of data mining. Now, a growing attention has been paid to it.

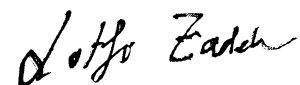
Besides the common public properties of data mining, spatial data mining has its own private characteristics. Spatial data includes not only positional data and attribute data, but also spatial relationships among spatial entities. Moreover, spatial data structure is more complex than the tables in ordinary relational databases. Besides tabular data, there are vector and raster graphic data in spatial database. However, people have almost studied spatial data mining in the context of data mining, paying inadequate attention to the spatial characteristics.

Professor Deren Li is an Academician of Chinese Academy of Sciences and an Academician of Chinese Academy of Engineering in Geo-informatics. He proposed the knowledge discovery from geographical databases on the international conference on GIS in 1994. Professor Deyi Li is an Academician of Chinese Academy of Engineering in computer science, especially to data mining. The brothers both supervised Dr. Shuliang Wang, whose thesis was honored one of the best 100 Ph.D. theses in China. Moreover, their group have successfully applied and finished more than 10 projects on spatial data mining, such as National Natural Science Fund of China (NSFC), National Key Fundamental Research Plan of China (973), National High Technique Research and Development Plan of China (863), China Postdoctoral Science Foundation and so on. At the same time, their theoretical and technical results have been applied to support and improve spatial data-referenced decision-making in the real world. During the process of their study, the fundamental theories on data mining in computer science are really combined with the spatial characteristics. Their work has further been accepted by the scholars in the world, of course, including myself.

In this monograph, there are several contributions. Some new techniques are proposed, i.e. cloud model, data fields, geo-rough space, mining view-angles, and a pyramid of spatial data mining. The discovery mechanism is believed to be a process of discovering a form of rules plus exceptions at hierarchal view-angles with various thresholds. Two algorithms are also put forward on spatial data cleaning, which are the iterative weight selection, and cloud model based Delphi hierarchy process. And three clustering techniques are further given, i.e. clustering discovery with cloud model and data fields, fuzzy clustering under data fields, and mathematical morphology based clustering. Remote-Sensed image databases are mined with spatial statistics, inductive learning, and conceptual lattice. The applicability and examples are further studied, such as monitoring landslides near Yangtze River, deformation recognition on train wheels, land use classification on remote-sensed image data, land resources evaluation, and bank place selection. Finally, a prototype system on spatial data mining is developed.

Seen from the abovementioned, the monograph of *Spatial data mining theories and applications* is the fruits of a collective work. The authors really study spatial data mining as an interdisciplinary subject, along with the applicability.

Thank the authors for inviting me to share their contributions earlier. And I am really pleased to recommend this monograph to you.



July 31, 2005

前 言

空间数据挖掘旨在解决“空间数据海量而知识贫乏”的瓶颈问题。长期以来,人们基本上是套用数据挖掘的策略研究空间数据挖掘,没有充分顾及到空间数据和普通事务数据的不同特点。现有的空间数据挖掘书籍或涉及空间数据挖掘的某些章节的作者,其学科背景在很大程度上是来自于计算机领域。虽然计算机科学与地球空间信息科学在方法上有所渗透,但是两者的学科基础和思维方法还是非常不同的,这决定了空间数据挖掘和数据挖掘的差异。作者深感传统研究方法的局限性,并在自己的研究和应用过程中努力突破其束缚。

1994年,在加拿大渥太华举行的GIS国际学术会议上,李德仁首次提出了从GIS数据库中发现知识(KDG)的概念,认为它能够把GIS有限的数据库变成无限的知识,精练和更新GIS数据,使GIS成为智能化的信息系统,率先从GIS空间数据中发现了用于指导GIS空间分析的知识。后来,他组建了自己的研究团体,指导他们把KDG进一步发展为空间数据挖掘,系统研究或提出了可用的理论、技术和方法,并取得了可喜的创新性成果,从而敲响了武汉大学在地球空间信息学中进行空间数据挖掘的战鼓。

作者指导的研究团队,由数十名教授、副教授、博士后、博士生和硕士生组成老、中、青相结合的梯队,地球空间信息科学和计算机科学等学科交叉,优势互补。先后申请并完成了有关空间数据挖掘的国家自然科学基金优秀重点实验室项目、国家重大基础研究计划(973)、国家高技术研究发展计划(863)、国家自然科学基金重点项目、国家自然科学基金面上项目、香港特别行政区政府基金项目、国家教育部博士点基金项目、测绘遥感信息工程国家重点实验室基金项目、软件工程国家重点实验室基金项目、中国博士后基金项目 and 武汉大学科研基金项目等10余项课题。同时,理论研究成果也被成功地应用于国民经济建设,如长江三峡宝塔滑坡监测数据挖掘、火车车轮的变形识别等,都直接关系到人民的生命财产的安全。先后培养了3名博士后、6名博士和20名硕士。在国内外公开发表了相关的学术论文50余篇,其中,被SCI、EI、ISTP等三大检索机构收录30余篇。

为了真正解决“空间数据海量而知识贫乏”的瓶颈问题,我们把计算机界数据挖掘的基本理论与由图形数据库、影像数据库和属性数据库集成的空间数据库特征相结合,对空间数据挖掘的理论、方法及其应用进行了十多年的研

究和探讨,现将这些研究成果进行加工和系统化,汇集成为一本较为系统的、可读性强、具有可操作性、理论联系实践的书——《空间数据挖掘理论与应用》。而且,竭力把许多重要的但局部的结果统一到一个令人满意的空間数据挖掘框架内,试图使人们切实认识到空间数据挖掘的独特魅力和功能。

本书从综述空间数据挖掘的国内外研究进展入手,深入研究了空间数据挖掘的内涵、外延、特征、可用的理论技术等,构建了空间数据挖掘金字塔;全面总结了空间数据挖掘的数据源及其获取方式、数据结构和数据模型,讨论了空间数据库、空间数据仓库、国家空间数据基础设施、数字地球;探讨了空间数据清理,给出了清理空间观测数据误差的方法,基于云模型提出了选择空间数据的 DHP 法,从验后方差估计原理导出了空间观测数据清理的选权迭代法;研究了基于空间统计学的图像数据挖掘,并以本征随机过程成功应用在基于纹理的特征提取和图像检索中;在分析模糊集的不彻底性的基础上,根据粗集,在地球空间信息学内提出了由粗实体、粗关系和粗算子组成的地学粗空间;在研究空间数据挖掘中的不确定性的基础上,提出了集成随机性和模糊性的云模型,系统地阐述了云模型的产生背景、基本概念、基本云模型、云发生器、虚拟云、云变换和不确定推理等基本技术,并在此基础上取得了一定的见解或创新;从空间数据辐射的角度,建立了数据场的概念、场强函数、势、势场及其可视化方法,讨论了数据场的影响因素,及其对空间数据挖掘的作用;提出和应用了空间数据挖掘的视角技术,基于云模型和数据场,在发现状态空间中研究了宝塔滑坡监测的空间数据挖掘技术与方法,并挖掘得到了微观、中观和宏观三个不同认知层次的可视化空间知识,而且和宝塔滑坡区的自然现象非常吻合;研究了空间聚类知识挖掘,提出了“数据场-云”聚类算法、基于数据场的模糊综合聚类算法和基于数学形态学的聚类算法;从空间数据挖掘的角度,研究了从空间数据库发现知识用于遥感图像分类的问题。提出了一套基于归纳学习的遥感图像分类技术和流程,能较好地解决同谱异物、同物异谱等问题,显著提高了分类精度,并且能够根据发现的知识进一步细分类,扩展了遥感图像分类的能力;将归纳学习用于银行经营收益分析和选址评价,直接从银行数据和相关图层数据中挖掘出银行经营收益与多种地理因素关联的知识,提高了 GIS 空间数据分析和决策支持的智能化水平;分析了概念格理论的基本含义和算法,并研究了基于概念格的遥感图像数据挖掘,以及基于概念格的 GIS 数据挖掘;研究了空间数据挖掘软件的体系结构和开发策略,在总结几个典型的国际软件系统的基础上,自主研制了一个空间数据挖掘原型系统。最后,思考展望了空间数据挖掘的未来进一步发展。

本书的研究成果,先后获得了国家自然科学基金优秀重点实验室项目“利用空间数据挖掘进行新型遥感影像目标提取和自动分类”(40023004)、国家重

大基础研究计划(973)项目“数据开采和知识发现的理论与方法研究”(G19980305084)和“对地观测数据-空间信息-地学知识的转化机理”(2006CB701305)、国家高技术研究发展计划(863 计划)项目“多源空间数据挖掘技术”(2001AA135081)、国家自然科学基金重点项目“遥感、全球定位系统、地理信息系统集成的理论与关键技术”(49631050)、国家自然科学基金重点项目“基于 Internet 的管理信息系统研究”(70231010)、国家自然科学基金项目“用模糊数学综合处理观测误差的理论与应用”(49574201)、国家自然科学基金项目“信息扩散原理在测量数据处理中的应用研究”(49874002)、全国优秀博士学位论文基金项目“多源变粒度可视化空间数据挖掘及其实现”(2005057)、香港特别行政区政府基金项目“Development of Infrastructure for Cyber Hong Kong”(1.34.37.9709)、香港特别行政区政府基金项目“Advanced Research Centre for Spatial Information Technology”(3.34.37.ZB40)、国家教育部博士点基金项目“从 GIS 数据库中发现知识及其在遥感图像理解中的应用研究”(98049801)、测绘遥感信息工程国家重点实验室基金项目“空间数据挖掘的理论框架研究”(WKL(97)0302)和“基于形部概念分析的遥感图像数据挖掘研究”(03-0101)、武汉大学科研基金项目(216-276081)、中国博士后基金项目“基于数据挖掘视角的多层次管理”(2004035360)、软件工程国家重点实验室基金项目“变粒度空间数据挖掘”(SKLSE05-15)和 GIS 教育部重点实验室基金项目“基于云理论的可视化空间数据挖掘”(WD200603)等 10 余个科研项目的资助,作者对以上各方面的支持表示热忱的感谢!

可以说,《空间数据挖掘理论与应用》一书是三位作者与整个研究团队集体智慧的结晶,是整个研究团队辛勤劳动的成果。其中,本书第一作者和第三作者,对全书进行了缜密的构思与组织,并共同指导第二作者执笔完成了此书。而且,本书的英文版即将脱稿,将应邀由世界著名的出版公司出版。

作者衷心感谢中国科学院地学部资深院士陈述彭先生多年来对后辈们的关爱和扶持,并亲自为本书作序。作者也衷心感谢模糊数学创始人、美国科学院院士 Lotfi A. Zadeh 教授以及国际数据挖掘权威、数据挖掘经典《Data Mining: Concepts and Techniques》著者韩家炜教授的帮助,他们在阅毕本书英文版初稿后,不仅给予了具体指导意见,而且欣然为之作序。

还要感谢科学出版社,特别是朱海燕女士、韩鹏先生、姚岁寒先生的大力支持。他们的艰辛劳动,促成了本书的顺利出版。

本书的完成,王新洲教授、张良培教授、秦昆博士、马洪超教授、周焰教授、巫兆聪博士、邸凯昌博士、王焯硕士、吕辉军硕士、淦文燕博士、曾旭平博士等也起到了重要的作用;王任享院士、王守觉院士、Michael Frank Goodchild 教授、Benjamin Zhan 教授、Mircea Neogita 教授、Vladimir Gorodetsky 教授、史

文中教授、陈国青教授、林宗坚教授、陈鹰教授、周成虎教授、胡占义教授、龚健雅教授、曹建农博士、游扬声博士、邹逸江博士等给予了很多宝贵建议；国家自然科学基金委员会、国家科学技术部、武汉大学测绘遥感信息工程国家重点实验室、武汉大学软件工程国家重点实验室、武汉大学国际软件学院、武汉大学遥感信息工程学院、清华大学经济管理学院等提供了很多帮助。在此，一并表示衷心的感谢！

感谢所有曾经、正在或将要鼓励、爱护和帮助我们研究空间数据挖掘的单位和个人。

我们深知，本书所反映的研究工作虽然取得了一定进展，但是对于整个空间数据挖掘领域来说，我们的成果只是“沧海一粟”。尽管我们数易其稿，字斟句酌，成稿后又请不同学科的多位学者阅读，多次征求意见，集思广益，可是，由于研究深度和水平所限，本书只能是抛砖引玉，书中难免存在疏漏和不足之处，敬请广大读者批评和指正。

李德仁 王树良 李德毅

2005年1月8日

目 录

序一

序二

序三

前言

第 1 章 绪论	1
1.1 空间数据挖掘的由来	1
1.1.1 过量的空间数据	1
1.1.2 数据带来的灾难	3
1.1.3 数据利用的尝试	5
1.1.4 空间数据挖掘的提出	6
1.1.5 空间数据挖掘的任务	7
1.2 空间数据挖掘国内外研究进展	8
1.2.1 学术活动	8
1.2.2 理论方法	9
1.2.3 应用成果	10
1.3 空间数据挖掘的难点	12
1.3.1 海量的空间数据	12
1.3.2 高维的空间数据	13
1.3.3 有污染的空间数据	13
1.3.4 不确定的空间数据	14
1.3.5 空间数据挖掘的角度	15
1.3.6 发现知识的表示	15
1.4 本书的内容和组织结构	16
参考文献	17
第 2 章 空间数据挖掘基础	19
2.1 基本概念	19
2.1.1 基本特性	19
2.1.2 不同角度的理解	20
2.1.3 空间数据挖掘金字塔	21
2.2 空间数据挖掘视角	22
2.2.1 人类思维	22
2.2.2 概念空间	23
2.2.3 特征空间	24

2.2.4	发现状态空间	24
2.2.5	数据挖掘机理	26
2.3	从空间数据到空间知识	26
2.3.1	数字和空间数值	27
2.3.2	空间数据和空间概念	27
2.3.3	空间信息和空间知识	28
2.4	空间粒度和空间尺度	30
2.4.1	空间粒度	30
2.4.2	空间尺度	31
2.4.3	泛概念树	31
2.5	空间数据挖掘可发现的知识类型	32
2.5.1	普遍几何知识和面向对象知识	33
2.5.2	空间特征规则和区分规则	33
2.5.3	空间分类规则和回归规则	34
2.5.4	空间聚类规则和关联规则	35
2.5.5	空间依赖规则和预测规则	36
2.5.6	空间序列规则和空间例外	36
2.6	空间知识的表达	37
2.6.1	自然语言	37
2.6.2	云模型的贡献	37
2.6.3	空间知识的测度	38
2.6.4	空间规则+空间例外	39
2.7	空间在线数据挖掘	40
2.7.1	网络资源	40
2.7.2	主要内容	40
2.8	空间数据挖掘与相关学科的关系	41
2.8.1	数据挖掘	41
2.8.2	机器学习	42
2.8.3	人工智能	43
2.8.4	模式识别	43
2.8.5	推理方法	44
2.8.6	地学数据分析	44
2.8.7	空间数据库系统	44
2.9	小结	45
	参考文献	46
第3章	空间数据挖掘的数据源	48
3.1	空间数据的内容和特性	48
3.1.1	空间数据的内容	48
3.1.2	空间数据的特性	51

3.1.3	空间数据的种类	52
3.2	空间数据获取	53
3.2.1	点方式获取空间数据	54
3.2.2	面方式获取空间数据	61
3.2.3	移动方式获取空间数据	73
3.2.4	卫星应用技术的作用	86
3.2.5	基础空间数据集中的主要问题	86
3.3	空间数据结构	88
3.3.1	矢量数据结构	88
3.3.2	栅格数据结构	89
3.3.3	矢量和栅格的比较	90
3.3.4	矢栅一体化数据结构	91
3.3.5	超图数据结构	98
3.3.6	纯关系数据结构	98
3.4	空间数据模型	99
3.4.1	层次模型	100
3.4.2	网络模型	101
3.4.3	关系模型	102
3.4.4	面向对象数据模型	103
3.4.5	基于 DEM 的可视化模型	111
3.5	空间数据库	114
3.5.1	空间数据库的分类	114
3.5.2	空间数据索引机制	116
3.5.3	空间数据的无缝组织	118
3.5.4	空间数据融合	122
3.5.5	空间数据库技术的不足	123
3.6	空间数据仓库	126
3.6.1	数据仓库	126
3.6.2	空间数据立方体	126
3.6.3	空间数据仓库和数据挖掘	128
3.7	小结	129
	参考文献	129
第 4 章	国家空间数据基础设施	131
4.1	美国的国家空间数据基础设施	131
4.1.1	美国总统 12906 号行政令	131
4.1.2	美国 FGDC	132
4.1.3	美国国家空间数据交换网站	133
4.1.4	美国 FGDC 空间数据转换标准	135
4.1.5	美国国家数字地球数据框架	136

4.1.6	美国国家空间数据基础设施战略	145
4.1.7	开放式地理信息系统协会	151
4.1.8	美国地质调查局及其地理信息产品	155
4.2	其他国家和地区性空间数据基础设施	159
4.2.1	英国皇家测量局的地理空间数据系统	160
4.2.2	德国官方地形和制图信息系统	162
4.2.3	加拿大国家地形数据库	164
4.2.4	澳大利亚土地和地理信息系统	166
4.2.5	日本地理信息系统	171
4.2.6	亚太地区空间数据基础设施	175
4.2.7	欧洲空间数据基础设施	183
4.3	全球空间数据基础设施	190
4.3.1	产生背景	191
4.3.2	全球空间数据基础设施的组成	193
4.3.3	全球空间数据基础设施的主要参与者	194
4.3.4	全球空间数据基础设施的实施观点	196
4.4	数字地球	199
4.4.1	基本概念	199
4.4.2	数字地球的技术基础	200
4.4.3	数字地球的应用	202
4.5	中国国家空间数据基础设施	203
4.5.1	建设必要性	203
4.5.2	建设可能性	205
4.5.3	建议内容	207
4.5.4	国家测绘局对实施 CNSDI 的设想	208
4.5.5	中国地球空间数据框架的设计思想与技术路线	213
4.5.6	中国空间数据交换格式的设计思想与原则	225
4.6	小结	228
	参考文献	228
第 5 章	空间数据清理	230
5.1	空间数据清理的必要性	230
5.1.1	有污染的空间数据	230
5.1.2	空间观测数据的数学模型及其误差	231
5.1.3	空间数据误差处理的发展阶段	234
5.1.4	空间数据清理的滞后	236
5.2	空间数据清理的概念	237
5.2.1	基本概念	237
5.2.2	基本内容	237
5.2.3	特点	238

5.3	空间数据清理的基本技术	238
5.3.1	不完整的空间数据	238
5.3.2	不准确的空間数据	239
5.3.3	重复记录的空间数据	240
5.3.4	不一致的空间数据	240
5.4	空间图形图像数据的清理	241
5.4.1	辐射形变的校正	241
5.4.2	几何形变的改正	243
5.4.3	一幅图像清理实例	243
5.5	空间观测数据的清理	244
5.5.1	系统误差处理方法	244
5.5.2	偶然误差处理模型	246
5.5.3	粗差处理模型	250
5.6	基于 DHP 法的空间数据选择	256
5.6.1	德尔菲法和层次分析法	256
5.6.2	DHP 法基本原理	256
5.6.3	DHP 法的专家选择	260
5.6.4	DHP 法的指标规范化	261
5.7	空间数据的定性定量转换	262
5.8	小结	262
	参考文献	263
第 6 章	空间数据挖掘可用的理论方法	264
6.1	确定集合理论	264
6.1.1	概率论	264
6.1.2	证据理论	265
6.1.3	空间统计学	266
6.1.4	规则归纳	266
6.1.5	聚类分析	268
6.1.6	空间分析	268
6.2	扩展集合论方法	269
6.2.1	模糊集	269
6.2.2	云模型	270
6.2.3	粗集	270
6.2.4	地学粗空间	272
6.3	仿生学方法	273
6.3.1	神经网络	273
6.3.2	遗传算法	274
6.4	可视化	275
6.5	决策树	276

6.6	理论方法讨论	276
6.6.1	对比分析	277
6.6.2	选择原则	277
6.7	小结	278
	参考文献.....	278
第7章	图像纹理的空间统计分析理论.....	282
7.1	研究进展	282
7.2	随机场理论	283
7.2.1	随机过程的基本概念	283
7.2.2	随机过程的基本类型	285
7.2.3	Markov 随机场与 Gibbs 场	288
7.2.4	Markov 随机场在纹理表示中的应用	291
7.3	本征随机过程	292
7.3.1	变差函数	292
7.3.2	变差函数的物理性质	293
7.3.3	变差函数的数学性质	295
7.3.4	正则化和块金效应	298
7.3.5	变差函数的理论模型	300
7.4	在空间数据挖掘中的应用前景	303
7.5	小结	304
	参考文献.....	304
第8章	地学粗空间.....	306
8.1	空间实体的描述近似性	306
8.2	地学粗空间的概念	307
8.2.1	数学基础	307
8.2.2	符号系统	308
8.2.3	粗元	310
8.2.4	测度	311
8.2.5	粗多维空间	312
8.3	粗实体	312
8.4	粗关系	314
8.5	粗算子	317
8.6	基于向量的属性简化	318
8.6.1	向量的核	319
8.6.2	粗集的核	319
8.7	在地球空间信息学中的应用	320
8.7.1	空间保真性	320
8.7.2	属性不确定性	321
8.7.3	对已有技术的数学解释	322

8.7.4	矢量数据和栅格数据	322
8.7.5	在空间数据挖掘中的作用	324
8.8	小结	324
	参考文献	325
第9章	云模型	327
9.1	空间数据挖掘需要云模型	328
9.1.1	自然语言的不确定性	328
9.1.2	概率论和数理统计的方法	328
9.1.3	模糊集的不彻底性	330
9.1.4	粗集的笼统性	331
9.1.5	GIS模型的不足	332
9.1.6	灵敏度分析	334
9.1.7	随机性或模糊性	334
9.1.8	云模型的方法	335
9.1.9	云模型的研究进展	335
9.2	云的定义和特性	337
9.2.1	云的基本定义	337
9.2.2	云的数字特征	338
9.2.3	云的“3En”规则	339
9.2.4	云的可视化	340
9.2.5	云的数学外延	341
9.3	基本云模型	342
9.3.1	正态云模型	342
9.3.2	衍生云模型	342
9.4	云发生器及其误差	344
9.4.1	正向云发生器	344
9.4.2	逆向云发生器	345
9.4.3	条件云发生器	349
9.4.4	不确定性推理器	351
9.4.5	云发生器的误差	352
9.5	虚拟云	354
9.5.1	浮动云和综合云	354
9.5.2	分解云和几何云	356
9.6	云变换	357
9.6.1	峰值法云变换	357
9.6.2	原子云模型集的归整	359
9.7	基于云模型的不确定推理	359
9.7.1	单规则推理	359
9.7.2	多规则推理	361

9.8 小结	362
参考文献	363
第 10 章 数据场	364
10.1 空间数据辐射	364
10.1.1 非完备的观测数据	364
10.1.2 数据辐射	365
10.1.3 辐射介质变化的影响	366
10.1.4 数据辐射和物理辐射	366
10.1.5 数据辐射和最小二乘配置的区别	366
10.2 数据场的概念和性质	367
10.2.1 概念	367
10.2.2 独立性和就近性	368
10.2.3 遍历性和叠加性	369
10.2.4 衰减性和各向同性	370
10.3 数据场的场强函数	370
10.3.1 衍生场强函数	371
10.3.2 场强函数	372
10.3.3 场强函数和协方差函数的区别	373
10.4 数据场的势	373
10.4.1 势函数	374
10.4.2 等势线	374
10.4.3 势场和势心	374
10.4.4 自然拓扑类	375
10.5 数据场的影响因素	376
10.5.1 数据辐射因子	376
10.5.2 数据辐射亮度	379
10.5.3 数据数量	380
10.5.4 综合作用	381
10.5.5 势间距和笛卡儿网格点密度	382
10.6 数据场的可视化	383
10.7 小结	384
参考文献	385
第 11 章 基于概念格的空间数据挖掘	386
11.1 概念的形成	386
11.1.1 特征表说和原型说	386
11.1.2 概念的内涵及其外延	386
11.1.3 从表象到概念	387
11.1.4 分层结构	388
11.1.5 概念聚类	388

11.1.6	数据挖掘的概念形成	389
11.2	关联规则挖掘算法	389
11.2.1	关联规则挖掘过程	389
11.2.2	Apriori 算法	390
11.2.3	优缺点分析	392
11.3	概念格理论研究	393
11.3.1	格论	393
11.3.2	概念格	394
11.3.3	Hasse 图	395
11.3.4	概念子格	395
11.3.5	单值属性的形式背景	396
11.3.6	多值属性的形式背景	396
11.3.7	基于概念格的知识表达与处理	399
11.4	概念格的构建和 Hasse 图的绘制	401
11.4.1	传统的批处理算法和增量算法	401
11.4.2	增量式概念格的构建算法	402
11.4.3	增量式概念格的快速构建算法	407
11.5	关联规则的生成	416
11.5.1	冗余规则和非冗余规则	416
11.5.2	频繁封闭项集	417
11.5.3	频繁封闭项集的产生子集	417
11.5.4	非冗余规则的生成	418
11.5.5	规则的直接提取法	419
11.6	算法时间复杂度	420
11.6.1	概念格与 Apriori 算法对比	420
11.6.2	概念格的化简	422
11.7	基于概念格的分类和聚类算法研究	422
11.7.1	分类分析	422
11.7.2	聚类分析	423
11.8	小结	423
	参考文献	423
第 12 章	宝塔滑坡的监测数据挖掘	425
12.1	宝塔滑坡	425
12.1.1	滑坡及滑坡灾害	425
12.1.2	滑坡监测方法	426
12.1.3	宝塔滑坡及其监测	428
12.1.4	宝塔滑坡监测的数据	429
12.1.5	宝塔滑坡的监测精度	430
12.2	滑坡监测数据挖掘的可行性	431

12.2.1	滑坡监测数据分析的不足	431
12.2.2	位移伪分布的缺陷	431
12.2.3	数据场的可用性	432
12.2.4	云模型的适宜性	433
12.2.5	发现状态空间的可操作性	434
12.3	宝塔滑坡形变监测数据挖掘的视角	434
12.3.1	视角类型	435
12.3.2	视角关系	435
12.3.3	基本视角	436
12.3.4	基本组合视角	436
12.3.5	宝塔滑坡监测数据的挖掘视角	438
12.4	同点异时同向的视角挖掘	438
12.4.1	X方向的数字特征	439
12.4.2	数字特征的定性诠释	441
12.4.3	Y、H方向的数字特征	442
12.4.4	数字特征可视化	446
12.5	异点同时同向的视角挖掘	448
12.5.1	滑坡变形概率分布密度辐射估计	448
12.5.2	数字特征	448
12.6	异点异时同向的视角挖掘	455
12.6.1	不同断面的数字特征值	455
12.6.2	滑坡的数字特征值	458
12.7	基于数据场的例外挖掘	460
12.7.1	不同方向上的例外	460
12.7.2	整体例外	461
12.7.3	规则+例外	461
12.8	宝塔滑坡形变监测的知识及讨论	465
12.8.1	发现的知识	465
12.8.2	挖掘机理	467
12.8.3	知识检验	468
12.8.4	方法讨论	468
12.9	小结	470
	参考文献	470
第 13 章	基于归纳学习和粗集的空间数据挖掘	471
13.1	基于归纳学习的空间数据挖掘	471
13.1.1	在空间数据库中的实施途径	471
13.1.2	遥感图像土地利用分类	472
13.1.3	银行经营收益分析及选址评价	478
13.1.4	空间关联知识挖掘	483

13.2	基于粗集的空间数据挖掘	487
13.2.1	基于地学粗空间的河流遥感影像分类	487
13.2.2	基于粗神经网络的遥感影像描述	488
13.2.3	基于粗集的遥感影像预处理方法	489
13.2.4	基于粗集理论的遥感影像分类知识获取	490
13.2.5	粗集在 GIS 属性分析中的应用	491
13.2.6	中国主要城市气温数据挖掘	492
13.2.7	中国大陆农业统计数据挖掘	494
13.3	粗集的潜在应用	499
13.4	小结	499
	参考文献	500
第 14 章	空间聚类知识挖掘	501
14.1	聚类挖掘算法回顾	501
14.1.1	分割算法	501
14.1.2	概念聚类	501
14.1.3	等间隔法	502
14.1.4	层次算法	502
14.1.5	其他聚类	503
14.2	“数据场-云”聚类	503
14.2.1	聚类算法	503
14.2.2	削除势心	504
14.2.3	二维数据集聚类	504
14.2.4	方法分析	506
14.3	基于数据场的模糊综合聚类	506
14.3.1	模糊综合评判和模糊聚类分析的不足	506
14.3.2	模糊综合聚类原理	507
14.3.3	模糊聚类置信水平	509
14.3.4	极大剩余法	509
14.3.5	平均绝对值距离法	510
14.3.6	土地评价	510
14.3.7	方法讨论	515
14.4	基于数学形态学的聚类	515
14.4.1	聚类算法	515
14.4.2	在矢量型空间数据库中的实现	519
14.4.3	发现偏差值和空洞	520
14.4.4	空间数据聚类试验及分析	520
14.5	小结	522
	参考文献	522

第 15 章 基于空间统计学的图像挖掘	524
15.1 基于纹理的图像检索	524
15.1.1 火车车轮检测	525
15.1.2 基本算法	526
15.1.3 人脸识别	528
15.2 图像模板窗口的确定	530
15.2.1 图像数据块	531
15.2.2 算法原理	532
15.3 遥感图像检索的特征提取	533
15.3.1 图像检索特征	533
15.3.2 灰度直方图特征	535
15.3.3 基于小波变换的边缘特征提取	537
15.3.4 火车车轮试验	540
15.3.5 基于形状的灰度直方图相似性比较	543
15.4 小结	545
参考文献	545
第 16 章 空间数据挖掘系统研发	546
16.1 系统结构和开发策略	546
16.1.1 系统功能及分类	546
16.1.2 系统开发策略	546
16.2 主要空间数据挖掘系统	547
16.2.1 GeoMiner	547
16.2.2 MultiMediaMiner	549
16.2.3 其他空间数据挖掘系统	551
16.3 GISDBMiner 和 RSImageMiner 的研制	551
16.3.1 GISDBMiner	552
16.3.2 RSImageMiner	554
16.4 小结	564
参考文献	564
第 17 章 结论和展望	566
17.1 创新和结论	566
17.2 应用前景	568

彩色图版

SPATIAL DATA MINING THEORIES AND APPLICATIONS

CONTENTS

Foreword I	
Foreword II	
Foreword III	
Preface	
Chapter 1 Introduction	1
1.1 Spatial data mining origin	1
1.1.1 Superfluous spatial data	1
1.1.2 Hazards from spatial data	3
1.1.3 Attempts to utilize data	5
1.1.4 Proposal of spatial data mining	6
1.1.5 Tasks of spatial data mining	7
1.2 The state of the art of spatial data mining	8
1.2.1 Academic research	8
1.2.2 Theories and techniques	9
1.2.3 Application results	10
1.3 Bottleneck of spatial data mining	12
1.3.1 Excessive spatial data	12
1.3.2 High dimensional spatial data	13
1.3.3 Polluted spatial data	13
1.3.4 Uncertain spatial data	14
1.3.5 Viewpoints of spatial data mining	15
1.3.6 Discovered knowledge representation	15
1.4 Monograph contents and structures	16
References	17
Chapter 2 Principles of spatial data mining	19
2.1 Concepts of spatial data mining	19
2.1.1 Characteristics of spatial data mining	19
2.1.2 Understanding spatial data mining from different views	20

2.1.3	Pyramid of spatial data mining	21
2.2	View-angles of spatial data mining	22
2.2.1	Human thought	22
2.2.2	Conceptual space	23
2.2.3	Characteristic space	24
2.2.4	Discovery state space	24
2.2.5	Discovery mechanism	26
2.3	From spatial data to spatial knowledge	26
2.3.1	Number and spatial numerical value	27
2.3.2	Spatial data and spatial concept	27
2.3.3	Spatial information and spatial knowledge	28
2.4	Spatial granularity and spatial scale	30
2.4.1	Spatial granularity	30
2.4.2	Spatial scale	31
2.4.3	Pan-conceptual tree	31
2.5	Spatial knowledge to be discovered	32
2.5.1	General geometry rules and object-oriented rules	33
2.5.2	Spatial characteristics rules and discriminate rules	33
2.5.3	Spatial classification rules and regressive rules	34
2.5.4	Spatial clustering rules and association rules	35
2.5.5	Spatial dependent rules and predictable rules	36
2.5.6	Spatial serial rules and exceptions	36
2.6	Spatial knowledge representation	37
2.6.1	Natural language	37
2.6.2	Contribution from cloud model	37
2.6.3	Spatial knowledge measurement	38
2.6.4	Spatial rules plus exceptions	39
2.7	Spatial online data mining	40
2.7.1	Network resources	40
2.7.2	Essential issues	40
2.8	Relationships between spatial data mining and related subjects	41
2.8.1	Data mining	41
2.8.2	Machine learning	42
2.8.3	Artificial intelligence	43
2.8.4	Pattern recognition	43
2.8.5	Generic reasoning	44
2.8.6	Geo-analysis	44
2.8.7	Spatial database system	44
2.9	Chapter conclusions	45

References	46
Chapter 3 Data source of spatial data mining	48
3.1 Contents and characteristics of spatial data	48
3.1.1 Contents of spatial data	48
3.1.2 Characteristics of spatial data	51
3.1.3 Diversity of spatial data	52
3.2 Spatial data acquisition	53
3.2.1 Position method to acquire spatial data	54
3.2.2 Polygon method to acquire spatial data	61
3.2.3 Mobile method to acquire spatial data	73
3.2.4 Satellite techniques to acquire spatial data	86
3.2.5 Problems to acquire basic spatial data	86
3.3 Spatial data structure	88
3.3.1 Vector structure	88
3.3.2 Grid structure	89
3.3.3 Comparison between vector structure and grid structure	90
3.3.4 Vector-grid structure	91
3.3.5 Super-map structure	98
3.3.6 Pure relational structure	98
3.4 Spatial data model	99
3.4.1 Hierarchical model	100
3.4.2 Network model	101
3.4.3 Relational model	102
3.4.4 Object-oriented model	103
3.4.5 DEM-based visual model	111
3.5 Spatial databases	114
3.5.1 Spatial database classification	114
3.5.2 Spatial database index	116
3.5.3 Seamless organization of spatial data	118
3.5.4 Heterogeneous spatial data amalgamation	122
3.5.5 Shortcomings of spatial databases	123
3.6 Spatial data warehouse	126
3.6.1 Data warehouse	126
3.6.2 Spatial data cubes	126
3.6.3 Spatial data warehouse and spatial data mining	128
3.7 Chapter conclusions	129
References	129
Chapter 4 National spatial data infrastructures	131
4.1 American national spatial data infrastructures	131

4.1.1	No.12906 policy of American president	131
4.1.2	American federal geographical data committee	132
4.1.3	American website to exchange spatial data	133
4.1.4	American standardization of federal geographical data committee to transform spatial data	135
4.1.5	American federal framework of digital earth	136
4.1.6	American strategy and policies of national spatial data infrastructures	145
4.1.7	American open GIS consortium	151
4.1.8	USGS and its geographical products	155
4.2	Other national or regional spatial data infrastructures	159
4.2.1	English geo-spatial databases system	160
4.2.2	German official topography and mapping information system	162
4.2.3	Canadian topography databases	164
4.2.4	Australian land and geographical information system	166
4.2.5	Japanese geographical information system	171
4.2.6	Pacific-Asian regional spatial data infrastructures	175
4.2.7	European regional spatial data infrastructures	183
4.3	Global spatial data infrastructures	190
4.3.1	Proposal background	191
4.3.2	Basic components of global spatial data infrastructures	193
4.3.3	Essential attendees of global spatial data infrastructures	194
4.3.4	Implemented viewpoints of global spatial data infrastructures	196
4.4	Digital earth	199
4.4.1	Concepts of digital earth	199
4.4.2	Technical foundations of digital earth	200
4.4.3	Applications of digital earth	202
4.5	Chinese national spatial data infrastructures	203
4.5.1	Necessity to implement Chinese national spatial data infrastructures	203
4.5.2	Possibility to implement Chinese national spatial data infrastructures	205
4.5.3	Advised contents of Chinese national spatial data infrastructures	207
4.5.4	Plan from Chinese State Bureau of Surveying and Mapping	208
4.5.5	Thoughts and schemes of Chinese spatial data framework	213
4.5.6	Designed principles to transform spatial data formats in China	225
4.6	Chapter conclusions	228
	References	228
Chapter 5	Spatial data cleaning	230
5.1	Necessity to clean spatial data	230
5.1.1	Polluted spatial data	230
5.1.2	Mathematical errors models of spatial observed data	231

5.1.3	The state of the art of handling errors	234
5.1.4	Spatial data cleaning behind data cleaning	236
5.2	Concepts of spatial data cleaning	237
5.2.1	Basic definition of spatial data cleaning	237
5.2.2	Basic contents of spatial data cleaning	237
5.2.3	Characteristics of spatial data cleaning	238
5.3	Basic issues on spatial data cleaning	238
5.3.1	Incomplete spatial data	238
5.3.2	Inaccurate spatial data	239
5.3.3	Repetitive spatial data	240
5.3.4	Inconsistent spatial data	240
5.4	Graphs and images cleaning	241
5.4.1	Radiation deformation correction	241
5.4.2	Geometry deformation correction	243
5.4.3	A piece of images cleaning	243
5.5	Spatial observed-data cleaning	244
5.5.1	On system errors	244
5.5.2	On stochastic errors	246
5.5.3	On blunder errors	250
5.6	Data selection on Delphi Hierarchy Process	256
5.6.1	Delphi method and analytical hierarchical process	256
5.6.2	Principles of Delphi Hierarchy Process	256
5.6.3	Experts selection of Delphi Hierarchy Process	260
5.6.4	Parameters normalizations	261
5.7	Mutual transformation between quality and quantity for spatial data ..	262
5.8	Chapter conclusions	262
	References	263
Chapter 6	Usable theories and techniques for spatial data mining	264
6.1	Crisp sets based theories	264
6.1.1	Probability theory	264
6.1.2	Evidence theory	265
6.1.3	Spatial statistics	266
6.1.4	Rules induction	266
6.1.5	Clustering analysis	268
6.1.6	Spatial analysis	268
6.2	Extensive sets based theories	269
6.2.1	Fuzzy sets	269
6.2.2	Cloud model	270
6.2.3	Rough sets	270