



国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

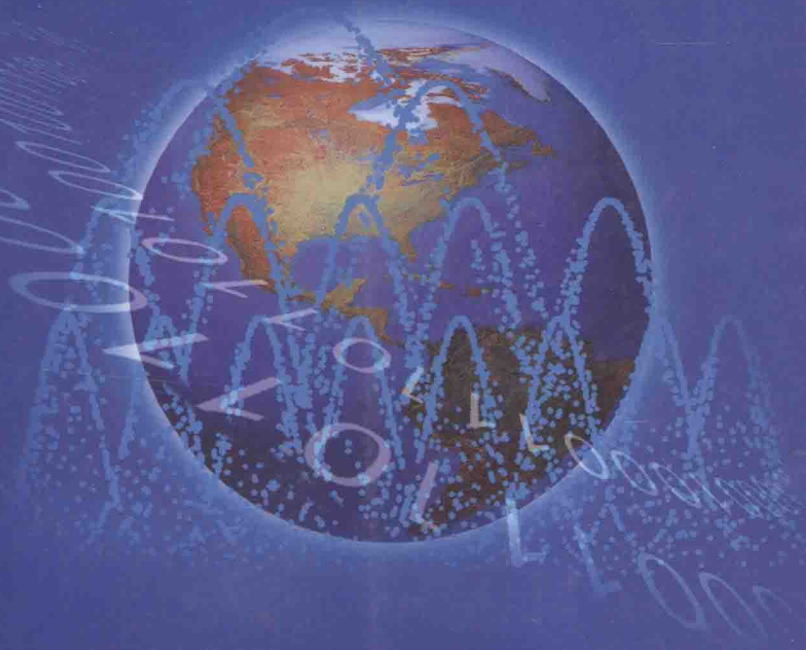


地球观测与导航技术丛书

空间数据挖掘理论与应用

(第二版)

李德仁 王树良 李德毅 著



科学出版社



国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

地球观测与导航技术丛书

空间数据挖掘理论与应用

(第二版)

李德仁 王树良 李德毅 著

科学出版社

北京

内 容 简 介

面向大数据,本书提出数据场、云模型、地学粗空间和空间数据挖掘视角,构建空间数据挖掘金字塔,给出挖掘机理,介绍概念格,总结空间数据来源,导出清理空间观测数据的选权迭代法。面向地理信息系统数据研究空间的关联规则、分布规则、概化规则和聚类规则的挖掘,从图像中发现指导图像分类、特征提取、表情识别的空间知识,以及时空分布的视频数据挖掘。这些理论,在滑坡监测、车轮形变、银行选址、土地利用、人脸识别、公共安全等实际领域,获得了有价值的成果,自主研制了空间数据挖掘原型系统 GISDBMiner 和 RSImageMiner。

本书可供空间数据挖掘、计算机科学、地球空间信息科学、地理信息系统、遥感、全球定位系统、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,亦可作为高等学校相关专业本科生和研究生的参考用书。

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. —2 版. —北京:科学出版社,2013. 10

(地球观测与导航技术丛书)

ISBN 978-7-03-038781-3

I. ①空… II. ①李…②王…③李… III. ①数据处理—计算机应用—空间信息系统 IV. ①P208-39

中国版本图书馆 CIP 数据核字(2013)第 235364 号

责任编辑:朱海燕 苗李莉 于 红 / 责任校对:赵桂芬

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>

中国科学院印刷厂 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

2013 年 10 月第 二 版 印张:23

2013 年 10 月第一次印刷 字数:498 000

定价:78.00 元

(如有印装质量问题,我社负责调换)



作者简介



李德仁 德国斯图加特大学博士,瑞士苏黎世联邦理工大学名誉博士,摄影测量与遥感学家,武汉大学教授,博士生导师,中国科学院地学部院士,中国工程院电子信息学部院士,国际欧亚科学院院士,第九届全国政协委员,武汉大学学术委员会主任,测绘遥感信息工程国家重点实验室学术委员会主任,中国测绘学会副理事长,武汉“中国光谷”首席科学家,国际摄影测量与遥感学会名誉会员。历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,亚洲地理信息学会创会会长,国家有突出贡献的中青年专家,国家973计划专家顾问组成员,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,中国测绘学会理事长,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),测绘遥感信息工程国家重点实验室主任,原武汉测绘科技大学校长。1982年,他提出的选权迭代法,被国际测量学界称为“李德仁方法”;1985年,他提出可靠性理论,科学地“解决了测量学上一个百年来的问题”;20世纪90年代,他提出地球空间信息学的概念和理论体系;1994年,他提出并奠定空间数据挖掘的基础;21世纪以来,他又提出广义和狭义空间信息网格,推进数字地球与智慧地球建设。他在多方面的独到建树,直接推动了技术进步和产业发展,其成果相继获得德国“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650余篇,出版专著11本;培养硕士研究生80多名,博士研究生130多名,5篇博士学位论文被评为全国优秀博士学位论文。



王树良 第十一届全国青联委员,教育部新世纪优秀人才,教育部软件工程专业教学指导委员会委员,北京理工大学教授、博士生导师、软件学院副院长。提出了空间数据挖掘视角,获得IGI Global的the Fifth Annual InfoSci[®]——Journals Excellence in Research Awards、全国优秀博士学位论文、第十一届教育部霍英东青年教师二等奖等。兼任IJGCRSIS等编委、内蒙古自治区数据挖掘与知识工程重点实验室主任。已获得国家发明专利授权2项,发表论文30余篇,出版专著3本。



李德毅 计算机工程和人工智能专家,中国工程院院士,国际欧亚科学院院士,现任中国指挥与控制学会名誉理事长,中国人工智能学会理事长,国家信息化专家咨询委员会委员,总参信息化部研究员,少将军衔,博士生导师,清华大学、国防大学兼职教授。在国际上最早提出“控制流—数据流”图对理论,1985年被国际IEE总部授予计算机和控制类年度期刊最佳学术成果奖;1999年攻克世界难题“三级倒立摆动平衡”问题,获得世界自动控制联合会杰出论文奖;长期从事云模型、云计算、数据挖掘、复杂网络、智能驾驶等研究,已出版英文专著3本,获得国家发明专利10项,国家和军队级科技进步奖18项,发表学术论文180余篇,主编技术丛书7本,培养博士、硕士研究生百余名。被国家授予有突出贡献的回国留学人员,享受政府特殊津贴,国家有突出贡献的中青年专家,获得何梁何利奖和军队专业技术重大贡献奖。

《地球观测与导航技术丛书》编委会

顾问专家

徐冠华 龚惠兴 童庆禧 刘经南 王家耀
李小文 叶嘉安

主 编

李德仁

副主编

郭华东 龚健雅 周成虎 周建华

编 委(按姓氏汉语拼音排序)

鲍虎军 陈 戈 陈晓玲 程鹏飞 房建成
龚建华 顾行发 江 凯 江碧涛 景 宁
景贵飞 李 京 李 明 李传荣 李加洪
李增元 李志林 梁顺林 廖小罕 林 琿
林 鹏 刘耀林 卢乃锰 孟 波 秦其明
单 杰 施 闯 史文中 吴一戎 徐祥德
许健民 尤 政 郁文贤 张继贤 张良培
周国清 周启鸣

《地球观测与导航技术丛书》出版说明

地球空间信息科学与生物科学和纳米技术三者被认为是当今世界上最重要、发展最快的三大领域。地球观测与导航技术是获得地球空间信息的重要手段,而与之相关的理论与技术是地球空间信息科学的基础。

随着遥感、地理信息、导航定位等空间技术的快速发展和航天、通信和信息科学的有力支撑,地球观测与导航技术相关领域的研究在国家科研中的地位不断提高。我国科技发展中长期规划将高分辨率对地观测系统与新一代卫星导航定位系统列入国家重大专项;国家有关部门高度重视这一领域的发展,国家发展和改革委员会设立产业化专项支持卫星导航产业的发展;工业与信息化部和科学技术部也启动了多个项目支持技术标准化和产业示范;国家高技术研究发展计划(863计划)将早期的信息获取与处理技术(308、103)主题,首次设立为“地球观测与导航技术”领域。

目前,“十一五”计划正在积极向前推进,“地球观测与导航技术领域”作为863计划领域的第一个五年计划也将进入科研成果的收获期。在这种情况下,把地球观测与导航技术领域相关的创新成果编著成书,集中发布,以整体面貌推出,当具有重要意义。它既能展示973和863主题的丰硕成果,又能促进领域内相关成果传播和交流,并指导未来学科的发展,同时也对地球观测与导航技术领域在我国科学界中地位的提升具有重要的促进作用。

为了适应中国地球观测与导航技术领域的发展,科学出版社依托有关的知名专家支持,凭借科学出版社在学术出版界的品牌启动了《地球观测与导航技术丛书》。

丛书中每一本书的选择标准要求作者具有深厚的科学研究功底、实践经验,主持或参加863计划地球观测与导航技术领域的项目、973相关项目以及其他国家重大相关项目,或者所著图书为其在已有科研或教学成果的基础上高水平的原创性总结,或者是相关领域国外经典专著的翻译。

我们相信,通过丛书编委会和全国地球观测与导航技术领域专家、科学出版社的通力合作,将会有一大批反映我国地球观测与导航技术领域最新研究成果和实践水平的著作面世,成为我国地球空间信息科学中的一个亮点,以推动我国地球空间信息科学的健康和快速发展!

李德仁
2009年10月

序 一

计算机技术的不断进步,使得人们从最初的用它代替人工计算,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年邱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。邱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,在邱凯昌博士答辩时,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邱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、青相结合的、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应用问题为要旨,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!

陈世彭

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。该书的可读之处良多,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,又有崭新的数据清理法和聚类算法,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。

阅读该书,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。

有幸先读该书,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。

韓永輝
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。

该书提出了云模型、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

附：美国科学院院士 Lotfi A. Zadeh 教授为本书作序三的原文

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.

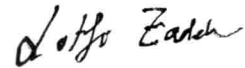
Professor Li Deren is an academician of Chinese Academy of Sciences and an academician of Chinese Academy of Engineering in Geo-informatics. He proposed the knowledge discovery from geographical databases on the international conference on GIS in 1994. Professor Li Deyi is an academician of Chinese Academy of Engineering in computer science, especially to data mining. The brothers both supervised Dr. Wang Shuliang, 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 the 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.

A handwritten signature in black ink, reading "Lotfi Zaiche". The signature is written in a cursive style with a large initial 'L'.

July 31, 2005

第二版前言

大数据、智能生产和无线网络三大技术变革将引领新的经济繁荣。大数据体量巨大、多种多样、高速变化、真实质差,其中,大约 80% 的数据与空间位置有关。大数据价值的公认规则是用数据说话,首要前提是保持大数据始终有用的活性,最终价值则是以大数据增进人类智慧。可是,截至 2003 年,人类总共创造了 5EB 数据,而 2011 年一年产生与复制的信息量超过了 1.8ZB,“空间数据海量而知识贫乏”的瓶颈问题日益严重,空间数据挖掘的地位与日俱增。

本书第一版出版后,获得国内外同行学者赞同,连续印刷 2 次,《科学通报》誉其为“空间数据挖掘的里程碑式力作”,为专业工具书 *Handbook of Geographical Information* 收录。在此过程中,作者虽然深知空间数据挖掘研究之困难,但是坚持带领团队在空间数据挖掘理论研究和应用服务的道路上默默耕耘,也在学术论文、技术发明专利、原型系统等领域的海滩有幸捡到漂亮的贝壳,与同行共赏。例如,数据场这个贝壳,因为技术原创性、写作水平、学术影响、未来前景等指标优秀,2012 年被 IGI Global 评为 the Fifth Annual InfoSci[®]——Journals Excellence in Research Awards。同时,很多读者来信,希望我们不要孤芳自赏,而是继续向社会分享,而且有的读者还提出了很多建设性意见,甚为感动。感动之余,我们不敢丝毫怠慢同行,也认为自己到了分阶段躬身自省研究历程的时候,是以开始第二版的写作,并得到了国家自然科学基金(61173061、71201120)、高等学校博士学科点专项科研基金(20121101110036)、教育部长江学者和创新团队发展计划(IRT1278)等项目的资助。

心怀对空间数据挖掘的敬畏,仰望前景,躬耕细作,厚积薄发。承首版重启再版新篇,谋伯仲相助,理论为经贯通,应用为纬穿插,正误补缺,删繁就简,悉数深化已有内容,深入探求发展趋势。按照空间数据挖掘的基本原理、数据源、数据清理、理论方法和实际应用的内在逻辑,集中讲述相同理论,及时补充最新成果,开时空分布的视频数据挖掘之先河。时刻为读者计免劳顿,虽内容增,但篇幅减。从首版之 16 章内化为再版之 12 章。全程力求重塑经典,历时数载,披星戴月努力,剥茧抽丝求真,兢兢业业务实,洗尽铅华,浴火重生,方成第二版初稿。写然后更知不足,邀群贤传阅,请大家斧正,旁征博引,博采众议,集思广益,精益求精,方敢付梓。纵然如此,难免错漏,尚请诸位读者持续

关爱,给予批评指正为谢。

遗憾的是,为本书作序的中国科学院院士陈述彭先生现已仙逝,为本书贡献良多的王新洲教授业已驾鹤西去,谨以此书缅怀,音容笑貌恍若眼前,教书育人宛如昨日,硕果累累承前启后,思之励今,是为纪念。

李德仁 王树良 李德毅

2013年4月16日

第一版前言

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和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)、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、王烨硕士、吕辉军硕士、涂文燕博士、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