



测绘科技专著出版基金资助

CEHUI KEJI ZHUANZHU CHUBAN JIJIN ZIZHU

SPATIAL DATA MINING VIEW-ANGLES

王树良 著

空间数据 挖掘视角

测绘出版社

测绘科技专著出版基金资助

空间数据挖掘视角

SPATIAL DATA MINING VIEW-ANGLES

王树良 著

测绘出版社

• 北京 •

内 容 简 介

空间数据挖掘是一个从海量数据中概括知识的不确定过程,不同用户在不同条件下有着不同的挖掘要求。本书提出空间数据挖掘视角,描述不同的数据挖掘需求,实现从相同数据向多种知识的变粒度挖掘。空间数据挖掘不确定性的内因和外因,决定了在空间数据挖掘视角中管控不确定性的技术。云模型集成模糊性和随机性,数据场刻画数据的自然拓扑关系,能够实现定性定量的相互转换。滑坡监测数据挖掘视角和网络化数据挖掘视角表明,空间数据挖掘视角发现的多粒度知识可以满足不同用户的需求。本书可为空间数据挖掘、地球空间信息科学、计算机科学和复杂网络等研究提供参考,亦可作为相应专业的高年级本科生或研究生的教学用书。

图书在版编目(CIP)数据

空间数据挖掘视角/王树良著. —北京:测绘出版社,
2008.10

ISBN 978-7-5030-1880-0

I. 空… II. 王… III. 数据采集—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(2008)第 150178 号

责任编辑 金 君

封面设计 李 伟

出版发行 **测 绘 出 版 社**

社 址 北京西城区复外三里河路 50 号

邮政编码 100045

电 话 010-68512386 68531609

网 址 www.sinomaps.com

印 刷 北京盛通印刷股份有限公司

经 销 新华书店

成品规格 169mm×239mm

印 张 10.25

字 数 198 千字

版 次 2008 年 10 月第 1 版

印 次 2008 年 10 月第 1 次印刷

印 数 0001—2000

定 价 27.00 元

书 号 ISBN 978-7-5030-1880-0

如有印装质量问题,请与我社发行部联系

作者简介



王树良,男,1975年生,武汉大学和香港理工大学博士,清华大学博士后,武汉大学教授,全国优秀博士学位论文、教育部新世纪优秀人才、霍英东青年教师二等奖、IBM教师奖、Citibank教师奖、湖北省“五四”金质奖章等荣誉获得者。担任 International Journal of System Science、International Journal of Data Mining and Data Warehousing 等 Guest Editor, Data Mining and Knowledge Discovery、ISPRS Journal of Photogrammetric and Remote Sensing 等 Paper Reviewer, Advanced Data Mining and Applications 学术委员会合作主席。已发表为 SCI 等收录论文多篇, Springer 等为其出版专著 3 部。主要研究空间数据挖掘。

前 言

1989年,在美国召开的第一届国际联合人工智能学术会议上,首次出现了从数据库中发现知识(Knowledge Discovery in Database,简称KDD)。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。1994年,在加拿大渥太华举行的GIS国际学术会议上,李德仁院士首次提出了从GIS数据库中发现知识(Knowledge Discovery from GIS)的概念,并持续发展为空间数据挖掘。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。在完成其博士论文后,邸凯昌博士去了美国俄亥俄州立大学做博士后,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,他们发现我对空间数据挖掘有浓厚的兴趣和一定的研究基础,于是又共同指导我在邸凯昌的基础上继续向前研究。在恩师的指导下,我顺利完成了博士学业,于2002年12月毕业,博士学位论文《基于数据场与云模型的空间数据挖掘和知识发现》在2005年有幸获得了全国优秀。

此时,测绘出版社的编辑邀请我出版自己的博士学位论文,并帮助我申请到测绘科技出版基金的支持。为了在《空间数据挖掘理论与应用》的基础上深化自己的研究,以及凸显博士学位论文“空间数据挖掘视角”的原始创新,我几乎再次重做了博士学位论文,并把书名确定为《空间数据挖掘视角》。同时,在研究过程中也注重和新的技术结合,把“空间数据挖掘视角”置身于复杂网络中研究。历经三载有余,晨钟与薄暮相连,寝食与出行共牵,知识与思维冲突,终成此书。书稿成时,曾广邀不同学科的四方师长共教,八方朋友同阅,每获指导,欣喜若狂,立即修正。如此写写改改,数易其稿,字斟句酌,又是数月。

空间数据挖掘是一个从具体海量数据到概括抽象知识的不确定浓缩过程。而且,在这种不确定的挖掘环境中,不同用户在不同条件下对空间数据挖掘具有不同的要求。本书提出空间数据挖掘视角,在描述不同数据挖掘需求的基础上,利用一定的数据挖掘算法,从大量空间数据中发现不同粒度的空间知识。基于空间数据挖掘,研究相同的人从不同的知识背景,或不同的人从相同的知识背景下,根据给定的目的要求,基于不同的认识层次或观察点,从已经获得的空间数据中,发现用于研究、解决和解释自然、人类和社会的问题、现象的模式的角度。在讨论空间数据挖掘视角的背景、内容和技术的过程中,试图探寻空间数据挖掘一般视角规律,也在滑坡监测数据挖掘视角和网络化数据挖掘视角中取得了有意义的成果。

本书的研究工作,先后获得了国家重点基础研究发展规划(973 计划)“对地观测数据-空间知识-地学知识转化机理”(2006CB701305)和“需求工程—对复杂系统的软件工程的基础研究”(2007CB310804)、国家自然科学基金“变视角空间数据挖掘”(60743001)、全国优秀博士学位论文基金“多源变粒度可视化空间数据挖掘的机理及其实现”(2005047)、新世纪优秀人才支持计划“空间数据挖掘视角”(NCET-06-0618)等科研项目的资助。本书的出版,则是测绘科技出版基金鼎力资助的结果。

本书的网络化数据挖掘视角,是刘欣硕士、伍爵博博士和我共同研究的成果,对此表示诚忱的感谢!

感谢测绘出版社,特别是编辑的大力支持。他们的艰辛劳动,促成了本书的顺利出版。

感谢所有曾经、正在或将来鼓励、爱护和帮助我研究空间数据挖掘的单位和个人。

本书的研究虽然取得了一定进展,可是,由于研究深度和水平所限,书中仍然难免存在疏漏和不足之处,敬请广大读者批评指正。

王树良

2008 年 7 月 26 日

Abstract

Spatial data mining is an uncertain discovery from vast data to summarized knowledge, in the context of the various demands from different users under different spatial conditions. In this monograph, a spatial data mining view-angle is proposed to discover the various knowledge from the same datasets with many granularities under the uncertainties. It is the intension and extension of the uncertainties that steers the techniques to manage and control the uncertainties in spatial data mining in an acceptable degree. Cloud model integrates randomness and fuzziness, and data field gives the natural topological relationships of spatial data. They may implement the mutual transform between quantitative data and qualitative concepts. The view-angles of landslide-monitoring data mining and networked data mining show that the discovered knowledge with many granularities can match the demands from different users on the given data sets under the specific demands and granularities. This monograph may be a reference book to study spatial data mining, geo-spatial information science, computer science, complex network, and so on. And it may also be a textbook for postgraduate or graduate students.

目 录

第 1 章 空间数据挖掘的同异性	1
§ 1.1 空间数据过量而难用	1
1.1.1 空间数据的过量增长	1
1.1.2 空间数据的处理滞后	2
1.1.3 空间数据难用的后果	5
§ 1.2 空间数据挖掘的产生	6
1.2.1 学科交叉的产物	6
1.2.2 研究应用的热点	6
1.2.3 不确定的空间数据挖掘	8
§ 1.3 空间数据挖掘的概念	10
§ 1.4 空间数据挖掘的同异性分析	11
1.4.1 同对象异数据	12
1.4.2 同数据异结构	13
1.4.3 同数据异需求	14
1.4.4 同数据异用户	14
1.4.5 同数据异方法	15
1.4.6 同数据异知识	15
1.4.7 同知识异表达	17
1.4.8 同数据异拓扑	18
§ 1.5 本章小节	19
第 2 章 空间数据挖掘视角原理	20
§ 2.1 空间数据挖掘视角的概念	20
2.1.1 空间数据挖掘的多视角理解	20
2.1.2 空间数据挖掘的视角因素	23
2.1.3 空间数据挖掘的多视角需求	23
2.1.4 空间数据挖掘的变视角需求	24
§ 2.2 空间数据挖掘视角的不确定性分析	24
2.2.1 客观存在和主观挖掘	25
2.2.2 不确定性的基本成因	26
2.2.3 不确定性的内在特性	30

2.2.4	不确定性的外在表现	31
2.2.5	不确定性的测度参数	34
§ 2.3	空间数据挖掘视角的影响要素	35
2.3.1	尺度	35
2.3.2	粒度	35
2.3.3	层次	36
§ 2.4	空间数据挖掘视角的机理空间	37
2.4.1	概念空间	38
2.4.2	特征空间	38
2.4.3	发现状态空间	39
§ 2.5	空间数据挖掘视角的基本算法	40
§ 2.6	本章小结	41
第3章	空间数据挖掘视角的技术	42
§ 3.1	空间数据挖掘视角技术的集合论分析	42
3.1.1	确定集	42
3.1.2	概率	42
3.1.3	模糊集	43
3.1.4	粗集	43
3.1.5	其他	43
3.1.6	随机和模糊的对比	44
§ 3.2	云模型	45
3.2.1	云模型基本概念	45
3.2.2	云模型的数字特征	46
3.2.3	云模型的类型	47
3.2.4	虚拟云模型	47
3.2.5	云发生器	50
3.2.6	云变换及其归整	54
3.2.7	云模型的不确定推理	55
§ 3.3	数据场	58
3.3.1	物理场的启发	58
3.3.2	样本观测数据	59
3.3.3	样本数据的能量	60
3.3.4	数据场的概念	61
3.3.5	数据场的性质	61
3.3.6	数据场的场强	64

3.3.7 数据场的势·····	67
3.3.8 数据场的可视化·····	70
§ 3.4 云模型和数据场的协同·····	70
§ 3.5 本章小结·····	72
第 4 章 滑坡监测数据挖掘视角 ·····	73
§ 4.1 滑坡监测视角分析·····	73
4.1.1 滑坡灾害·····	73
4.1.2 滑坡监测·····	74
4.1.3 数据分析的不足·····	77
4.1.4 数据场和云模型的可用性·····	78
4.1.5 基本滑坡监测数据挖掘视角·····	79
4.1.6 视角挖掘算法·····	82
§ 4.2 同点异时同向的视角挖掘·····	83
4.2.1 X 方向的数字特征·····	83
4.2.2 数字特征的定性诠释·····	85
4.2.3 Y、H 方向的数字特征·····	86
4.2.4 数字特征可视化·····	90
§ 4.3 异点同时同向的视角挖掘·····	92
4.3.1 滑坡变形概率分布密度辐射估计·····	93
4.3.2 异点同时同向的数字特征·····	98
§ 4.4 异点异时同向的视角挖掘·····	100
4.4.1 不同断面的数字特征值·····	100
4.4.2 滑坡的数字特征值·····	104
§ 4.5 基于数据场的例外挖掘·····	106
4.5.1 不同方向上的例外·····	106
4.5.2 整体例外·····	108
4.5.3 规则+例外·····	109
§ 4.6 宝塔滑坡形变监测的知识及讨论·····	110
4.6.1 发现的知识·····	110
4.6.2 挖掘机理·····	112
4.6.3 知识检验·····	113
4.6.4 方法讨论·····	114
§ 4.7 本章小结·····	116
第 5 章 网络化数据挖掘视角 ·····	117
§ 5.1 从网络到复杂网络·····	117

5.1.1	网络	117
5.1.2	网络拓扑	119
5.1.3	规则网络和随机网络	120
5.1.4	小世界网络	120
5.1.5	无标度网络	122
5.1.6	社团结构	123
§ 5.2	网络化数据挖掘	123
5.2.1	网络化数据挖掘的概念	124
5.2.2	网络化数据挖掘的内容	124
§ 5.3	网络化数据挖掘的社团发现算法	125
5.3.1	图分割	126
5.3.2	分级聚类	126
5.3.3	Kernighan-Lin 算法	127
5.3.4	谱平分法	127
5.3.5	GN 算法	129
5.3.6	Newman 快速算法	130
§ 5.4	网络化数据挖掘的社团发现视角	131
5.4.1	人工随机网络	131
5.4.2	空手道俱乐部成员间的关系网	132
5.4.3	海豚关系网	133
5.4.4	电影演员合作网	134
§ 5.5	本章小节	136
第 6 章	思考与结语	138
§ 6.1	空间数据挖掘视角的思考	138
§ 6.2	空间数据挖掘视角的决策思考	138
§ 6.3	结语	140
参考文献		142

Contents

Chapter 1	Sameness and difference from spatial data mining	1
§ 1.1	Excessive spatial data with difficult utilization	1
1.1.1	Superfluously increased spatial data	1
1.1.2	Lag to deal with spatial data	2
1.1.3	Resulting hazards from spatial data	5
§ 1.2	Spatial data mining origin	6
1.2.1	The product of interdisciplinary	6
1.2.2	The hot spots of research and application	6
1.2.3	uncertain spatial data	8
§ 1.3	Concept of spatial data mining	10
§ 1.4	Analysis on the sameness and difference from spatial data mining	11
1.4.1	The same objects but different data	12
1.4.2	The same data but different structures	13
1.4.3	The same data but different requirements	14
1.4.4	The same data but different users	14
1.4.5	The same data but different methods	15
1.4.6	The same but different knowledge	15
1.4.7	The same knowledge but different representations	17
1.4.8	The same data but different topologies	18
§ 1.5	Chapter conclusions	19
Chapter 2	Principles of spatial data mining view-angles	20
§ 2.1	Concepts of spatial data mining view-angles	20
2.1.1	Understand spatial data mining from multi-view-angles	20
2.1.2	View-angle elements of spatial data mining	23
2.1.3	View-angle requirements of spatial data mining	23
2.1.4	Changing view-angle requirements of spatial data mining	24
§ 2.2	Uncertainty analysis on the spatial data mining view-angles	24
2.2.1	An objective reality and subjective Mining	25
2.2.2	The basic causes of uncertainty	26
2.2.3	The inherent characteristics of uncertainty	30
2.2.4	The external performances of uncertainty	31
2.2.5	The measurement parameters of uncertainty	34

§ 2.3 Impact elements of spatial data mining view-angles	35
2.3.1 Scale	35
2.3.2 Granularity	35
2.3.3 Hierarchy	36
§ 2.4 Mechanism space of spatial data mining view-angles	37
2.4.1 Conceptual space	38
2.4.2 Feature space	38
2.4.3 The space of discovery state	39
§ 2.5 Basic algorithms of spatial data mining view-angles	40
§ 2.6 Chapter conclusions	41
Chapter 3 Techniques of spatial data mining view-angles	42
§ 3.1 Analyze the techniques of spatial data mining view-angles from set theory	42
3.1.1 Crisp sets	42
3.1.2 Probability	42
3.1.3 Fuzzy sets	43
3.1.4 Rough sets	43
3.1.5 Others	43
3.1.6 Randomness vs. fuzziness	44
§ 3.2 Cloud model	45
3.2.1 Basic concepts of cloud model	45
3.2.2 Numerical characteristics of cloud model	46
3.2.3 Cloud model classification	47
3.2.4 Virtual cloud model	47
3.2.5 Cloud generator	50
3.2.6 Cloud transformation and reorganization	54
3.2.7 Cloud reasoning with uncertainty	55
§ 3.3 Data field	58
3.3.1 Inspiration from physical fields	58
3.3.2 Sample observed data	59
3.3.3 Energy from sample data	60
3.3.4 Concepts of data field	61
3.3.5 Characteristics of data field	61
3.3.6 Power of data field	64
3.3.7 Potential of data field	67

3.3.8 Visualization of data field	70
§ 3.4 Collaboration between cloud model and data field	70
§ 3.5 Chapter conclusions	72
Chapter 4 View-angles to mine landslide-monitoring data	73
§ 4.1 Analysis on landslide-monitoring view-angles	73
4.1.1 Landslide hazards	73
4.1.2 Landslide monitoring	74
4.1.3 Lack of data analysis	77
4.1.4 Usability from cloud model and data fields	78
4.1.5 Basic view-angles to mine landslide-monitoring data	79
4.1.6 Algorithms for view-angles to mine landslide-monitoring	82
§ 4.2 Discovery with the same monitoring-point, different time, and the same direction	83
4.2.1 Numerical characteristics from X direction	83
4.2.2 Qualitative interpretation on the numerical characteristics	85
4.2.3 Numerical characteristics from direction Y direction and H direction	86
4.2.4 Numerical characteristics visualization	90
§ 4.3 Discovery with different monitoring-point, the same time, and the same direction	92
4.3.1 Radiation estimation on the probability distribution density of landslide deformation	93
4.3.2 Numerical characteristics with different monitoring-point, the same time, and the same direction	98
§ 4.4 Discovery with different monitoring-point, different time, and the same direction	100
4.4.1 Numerical characteristics from different sections	100
4.4.2 Numerical characteristics of the landslides	104
§ 4.5 Data field-based exception discovery	106
4.5.1 Exception from different directions	106
4.5.2 Exception in a whole	108
4.5.3 Rules+exception	109
§ 4.6 The discovered knowledge from deformation-monitoring data of Baota landslide and the technical discussion	110
4.6.1 Discovered knowledge	110

4. 6. 2 Mining mechanism	112
4. 6. 3 Knowledge test	113
4. 6. 4 Technical discussion	114
§ 4. 7 Chapter conclusions	116
Chapter 5 Networked data mining view-angles	117
§ 5. 1 From network to complex network	117
5. 1. 1 Network	117
5. 1. 2 Network topology	119
5. 1. 3 Regular network and random network	120
5. 1. 4 Small world	120
5. 1. 5 Scale-free	122
5. 1. 6 Community structure	123
§ 5. 2 Networked data mining	123
5. 2. 1 Networked data mining concept	124
5. 2. 2 Networked data mining contents	124
§ 5. 3 Algorithms to discover the community structure for networked data mining	125
5. 3. 1 Graph partition	126
5. 3. 2 Hierarchical clustering	126
5. 3. 3 Kernighan-Lin algorithms	127
5. 3. 4 Spectrum sharing	127
5. 3. 5 Girvan-Newman	129
5. 3. 6 Newman fast algorithms	130
§ 5. 4 View-angles to discover the community structure for networked data mining	131
5. 4. 1 Artificial random network	131
5. 4. 2 Karate Club network	132
5. 4. 3 Dolphin network	133
5. 4. 4 Cooperation network for movie actors	134
§ 5. 5 Chapter conclusions	136
Chapter 6 Concluding remarks	138
§ 6. 1 On spatial data mining view-angles	138
§ 6. 2 Decision-making from spatial data mining view-angles	138
§ 6. 3 Concluding remarks	140
References	142

第1章 空间数据挖掘的同异性

空间数据过量增长而难以全面利用的瓶颈,促成了空间数据挖掘的产生。空间数据挖掘的难点之一是,顾及空间数据挖掘的同异性,挖掘不同用户在不同环境下所需的不同隐含知识,这些难点促成了空间数据挖掘视角的产生。

§ 1.1 空间数据过量而难用

空间数据是描述实体在空间现实世界的具体地理方位和空间分布的数据,指空间实体的属性、数量、位置及其相互关系等的空间符号描述,可以是点的高程、道路的长度、多边形的面积、建筑物的体积、像元的灰度等数值,也可以是地名、注记等字符串,还可以是图形、图像等多媒体成分。与普通的事务数据相比,空间数据具有空间性、时间性、多维性、大数据量、空间关系复杂等特点。空间数据的数量急剧膨胀,而且日益复杂多变,但是人们对数据的分析利用能力远远小于数据的生产、传输和累积能力,大量数据只能被迫束之高阁(Ester et al., 2000)。

1.1.1 空间数据的过量增长

现代空间数据获取设备和计算机网络等信息技术的迅速发展,使得空间数据的数量、大小和复杂性及其传输的速度都在飞快地增长。而且,空间数据的膨胀速度也极大地超出了常规的事务型数据。

遥感对地观测日益成为人类获取空间数据的重要技术支撑,随着人类对自身生存环境的日益关注,它已经成为社会、政治和经济的发展决策不可或缺的重要组成部分。现在,星载传感器、卫星发射、控制等系列硬件技术已经取得了重大突破,未来的天基信息系统和对地观测系统拟通过努力,建立具有准实时、全天候获取各种空间数据的能力,并逐步形成集高空、高光谱、高时间分辨率和宽地面覆盖于一体的卫星(群)对地观测系统。感知器的飞速发展,也使得描述空间对象属性的波段数目由几个增加到几十甚至上百个。遥感对地观测技术正在形成一个多层次、多角度、全方位和全天候的全球立体对地观测网,高、中、低轨道结合,大、中、小卫星协同,粗、细、精分辨率互补(图 1-1)。

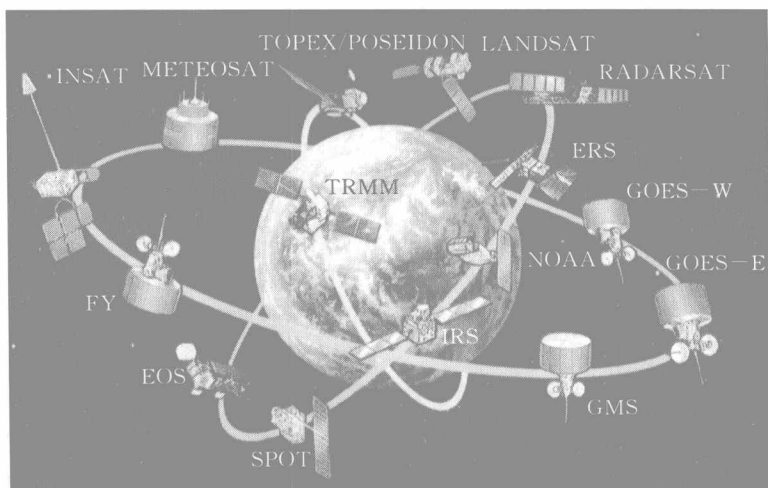


图 1-1 全球立体对地观测网

(源自:李德仁,2006.空间数据挖掘和知识发现)

传感器的地面分辨率数量级从千米到厘米,波段范围从紫外到超长波,时间间隔从十几天一次到每天三次,探测深度从几米到万米。在以高空间、高光谱、高动态为标志的新型遥感对地观测技术中,新型的高分辨率卫星遥感数据如 Quick Bird、IRS、IKONOS 等已提供使用。多传感器、多用途、多分辨率、多频率的 EOS 更可以提供 MODIS 成像光谱数据、ASTER 热红外数据、测云和 4-D 模拟的 CERES 数据、MOPIT 数据以及 MISR 数据。高分辨率、高动态的新型卫星传感器不仅波段数量多、光谱分辨率高、数据速率高、周期短,而且数据量特别大,一般情况下数据的容量均在千兆量级以上。仅 EOS-AM1 和 PM1 每日获取的遥感空间数据量就以 TB 级计算。Landsat 每两周就可以获取一套覆盖全球的卫星影像数据,目前已经积累了全球几十年的数据。

空间数据基础设施的建设速度和由此积累的空间基础数据也正在递增。空间数据基础设施积累了大量的城市电子地图数据库、城市规划道路网络数据库、工程地质信息数据库、用地现状信息数据库、总体规划信息数据库、控制性详细规划数据库、市政红线数据库、建筑红线与用地红线数据库、地籍数据库以及覆盖全市范围的土地利用及基本农田保护规划数据库等空间基础数据。更进一步地,除了这些已经存储和积累的数据,每时每刻还都在采集和产生新的空间数据。这些数据,拓宽了可供人类利用的空间数据资源。

1.1.2 空间数据的处理滞后

空间数据是国家的基础性战略资源,“以信息化带动工业化”的指导方针是现代化发展的基本国策,空间数据资源的占有和有效配置也是信息时代的显著标志。