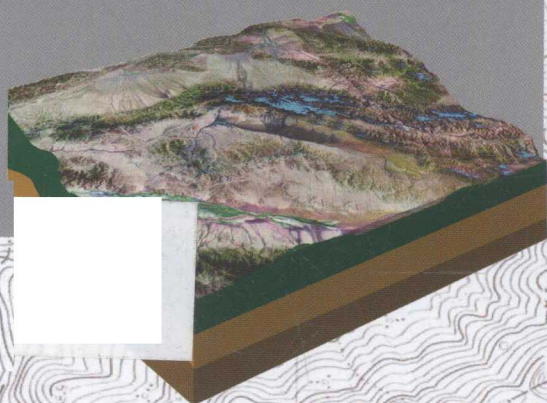


多源地质空间信息 智能处理与区域矿产资源预测

何彬彬 陈翠华 陈建华 等 著



科学出版社

多源地质空间信息智能处理 与区域矿产资源预测

何彬彬 陈翠华 陈建华 等 著

科学出版社

北 京

内 容 简 介

本书是一部将 GIS 技术、遥感技术和智能信息处理技术与区域矿产资源评价预测需求紧密结合的专著。其综合论述了基于海量地质空间数据的区域矿产资源预测方法,并进行了详细的实验对比验证。主要内容包括高寒山区高光谱遥感岩矿专题信息提取方法及实验;传统的证据权模型、扩展证据权模型和逻辑斯蒂回归模型的详细技术流程及实验验证;新的区域矿产资源预测方法:融合 C4.5 决策树和概率平滑技术、地质空间数据挖掘方法、证据推理方法、成矿案例推理方法和地质空间场景相似性推理方法,详细阐述了这些新方法的理论依据、技术流程和实验结果对比分析。并在此基础上,自主研发软件系统“智能区域成矿预测系统”,对该软件系统设计思路、数据库构建、功能模块划分等方面进行了详细介绍。

本书可供从事矿产资源勘查评价、遥感和 GIS 应用研究人员参考使用,亦可作为地理信息科学和资源勘查等相关专业的研究生和高年级本科生专业教学参考书。

图书在版编目(CIP)数据

多源地质空间信息智能处理与区域矿产资源预测 / 何彬彬等著. —北京:科学出版社, 2014. 6

ISBN 978-7-03-041065-8

I. ①多… II. ①何… III. ①地理信息系统-研究 ②矿产资源-资源预测-研究 IV. ①P208②F416.1

中国版本图书馆 CIP 数据核字 (2014) 第 125365 号

责任编辑:张井飞 韩 鹏 / 责任校对:桂伟利

责任印制:赵德静 / 封面设计:耕者设计工作室

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>

骏杰印刷厂印刷

科学出版社发行 各地新华书店经销

*

2014 年 6 月第 一 版 开本:787×1092 1/16

2014 年 6 月第一次印刷 印张:25 1/2 插页:6

字数:587 000

定价:148.00 元

(如有印装质量问题,我社负责调换)



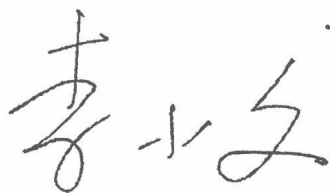
序

近年来，通过基于地球科学数据的定量化，以发现未知资源、环境、灾害等信息的空间分析模型及应用日益增多，为海量地学信息挖掘提供了强有力的分析决策工具。同时，遥感科学和地理信息技术的快速发展，为资源探测和环境监测提供了多时空分辨率的监测和快速分析手段，已在土地资源动态监测、生态环境监测、大气环境监测等资源环境应用领域发挥越来越重要的作用。

何彬彬、陈翠华、陈建华等人编著的《多源地质空间信息智能处理与区域矿产资源预测》一书汇集了作者主持的国家 863 项目、国家自然科学基金、教育部新世纪优秀人才支持计划项目和参加的青海省科技厅重大项目等研究课题的研究成果，将遥感、GIS 和人工智能领域的理论方法与矿产资源探测应用需求相结合，开展了新颖的区域矿产资源预测方法研究，建立了智能化区域成矿预测系统，并以我国青海东昆仑成矿带为例，进行了不同空间尺度的方法验证，为各地质矿产单位进行找矿勘探提供空间分析和决策的基础平台。书中提出的地质空间数据挖掘方法、成矿案例推理方法、地质空间场景相似性推理方法等区域矿产资源预测方法，思路新颖，效果良好，体现了作者宽泛的知识面和良好的创新思维。

《多源地质空间信息智能处理与区域矿产资源预测》是一部典型的跨学科交叉研究成果专著。该书的作者既有来自高校的遥感、GIS、地质、计算机等方向的中青年学者，也有常年在西北艰苦地区从事野外地质矿产勘查的地质工作者。书中的很多方法思路新颖、实验翔实、并有大量的野外调查验证，是一本集成方法创新和应用实践的专著。

我非常欣喜地看到，这支年轻的团队本着严谨的态度，历时三年撰写此书，多次进行修改和更新，不断完善，使之终于与广大科技工作者见面。期待本书的成果能够进一步应用于地质矿产行业部门，推动遥感和地理信息技术在矿产资源预测评价中的应用。



中国科学院院士

前 言

区域矿产资源预测是综合性和交叉性较强的科学。现代地球探测手段和空间信息处理技术的快速发展使得快捷的数据获取和信息处理技术成为可能。目前,全国范围内的地质、地球物理、地球化学和遥感找矿等数据的日益丰富,为区域矿产资源预测提供了海量基础数据。如何充分利用已有的海量地质空间数据,从大量的具有不确定性的多源地质空间数据中挖掘出深层次的找矿信息,突破传统的矿产资源预测思路,建立快速、高效、智能化的区域矿产资源预测方法,从而降低矿产勘查的成本,进一步提高矿产预测的效率和精度,显得极具科学意义和应用价值。本书作者自2007年以来在国家863项目(No. 2007AA12Z227)、国家自然科学基金(No. 41171302)、教育部新世纪优秀人才支持计划项目(No. NCET-12-0096)、青海省科技厅重大项目等研究课题的连续资助下,深入开展了系统的区域矿产资源预测方法研究,并以我国青海东昆仑成矿带为例,进行了不同空间尺度的方法验证。该书以矿床学理论、成矿系统和成矿模式理论为指导,顾及地质空间数据的不确定性特性,将空间数据挖掘、案例推理、空间场景相似性、证据权等方法有机耦合,充分利用海量地质空间数据,建立智能化区域矿产资源预测模型,并自主研发了软件系统“智能区域成矿预测系统”,为我国的区域矿产资源预测与评价提供新的技术方法支持。

本书共分14章,第1章绪论,主要简述了该书后面章节相关技术内容的国内外研究现状分析,包括研究背景及意义、区域矿产资源预测方法、空间数据挖掘、案例推理、地质空间数据的不确定性、空间数据挖掘的不确定性、矿产资源预测的不确定性等;第2章总结阐述了本书实验区的地质概况,为后面章节实验的基础内容,包括区域地理概况、区域成矿地质背景、成矿系列等;第3章和第4章主要阐述了应用遥感数据提取成矿专题信息的方法和实验,包括高寒山区岩矿光谱测试实验与分析、权重光谱角制图方法、耦合整体光谱匹配和局部光谱匹配方法、特征参数匹配方法、基于Hyperion高光谱遥感数据的高寒山区岩矿填图实验;第5~7章主要阐述了传统的证据权、扩展证据权和逻辑斯蒂回归模型在区域矿产资源预测中的技术流程及实验验证,包括数据预处理、控矿因素及空间分析、条件独立性检验、预测结果制图、预测精度统计及对比分析等;第8~12章重点阐述了5种新的区域矿产资源预测方法,包括融合C4.5决策树和概率平滑技术、地质空间数据挖掘方法、证据推理方法、成矿案例推理方法和地质空间场景相似性推理方法,详细介绍了这些新方法的理论依据、技术流程和实验结果分析,并与传统的证据权模型进行了对比分析;第13章简要介绍了一种集成多种模型的综合预测方法,并进行了相应实验对比分析;第14章主要阐述了自主研发的“智能区域成矿预测系统”的设计思路、数据库构建、功能模块划分及系统界面展示。

本书的完成历时3年,凝聚了课题组多位成员的辛勤工作。具体分工为:第1章,何彬彬、陈翠华;第2章,何彬彬、庄永成、陈翠华、刘岳;第3章,何彬彬、何中海;第4章,何彬彬、何中海;第5章,陈翠华、刘岳、何彬彬;第6章,陈翠华、刘岳、何彬彬;第7章,陈翠华、刘岳、何彬彬;第8章,何彬彬、曾泽、陈翠华;第9章,何彬彬、崔莹;第10章,何彬彬、崔莹;第11章,陈建华、何彬彬;第12章,何彬彬、曾泽;第13章,陈建华、何彬彬;第14章,何彬彬、陈建华、崔莹。此外,研究生全兴文、蒋雪梅、王宁宁、周文英、闫永帅等参与了本书书稿的校对工作。中国科学院遥感与数字地球研究所的研究员曹春香、燕守勋、武晓波在资料收集和野外工作等方面给予了帮助和指导。感谢科学出版社编辑韩鹏先生和张井飞先生对本书写作出版的关注和支持。

特别感谢李小文院士在课题研究过程中给予的悉心指导和帮助,并在百忙中欣然为本书作序。

何彬彬

2013年国庆于成都

目 录

序

前言

第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 区域矿产资源预测方法	2
1.3 空间数据挖掘	3
1.4 案例推理	4
1.5 地质空间数据的不确定性	5
1.6 空间数据挖掘的不确定性	5
1.7 矿产资源预测的不确定性	7
参考文献	8
第 2 章 实验区地质概况	12
2.1 青海东昆仑地区地理概况	12
2.2 区域成矿地质背景	12
2.3 成矿系列	32
参考文献	43
第 3 章 高光谱遥感岩矿专题信息提取方法	49
3.1 概述	49
3.2 岩矿波谱机理分析	57
3.3 高寒山区岩矿光谱分析	60
3.4 高寒山区岩矿光谱变异性分析	67
3.5 基于光谱匹配的岩矿信息提取方法	75
3.6 基于流形学习的高光谱降维	99
参考文献	105
第 4 章 高寒山区高光谱岩矿填图应用	112
4.1 野外实验区地理概况	112
4.2 技术路线	113
4.3 野外光谱采集	114
4.4 高光谱数据预处理	116
4.5 岩性填图	122
4.6 矿物识别	148

4.7 基于流形学习的高光谱降维及岩性分类验证	155
参考文献	169
第5章 证据权模型与区域矿产资源预测	170
5.1 算法原理	170
5.2 基于证据权模型的青海东昆仑铁矿资源预测	175
5.3 基于证据权模型的青海东昆仑金矿资源预测	182
5.4 基于证据权模型的青海东昆仑铜铅锌多金属矿资源预测	187
5.5 基于证据权模型的野马泉铁多金属矿资源预测	194
5.6 基于证据权模型的五龙沟金矿资源预测	199
5.7 结论	204
参考文献	205
第6章 扩展证据权模型与区域矿产资源预测	207
6.1 算法原理	207
6.2 基于扩展证据权模型的青海东昆仑铜铅锌多金属矿产资源预测	208
6.3 基于扩展证据权模型的五龙沟金矿资源预测	217
6.4 结论	220
参考文献	221
第7章 逻辑斯谛回归模型与区域矿产资源预测	222
7.1 逻辑斯谛回归模型	222
7.2 基于逻辑斯谛回归模型的青海东昆仑铁矿资源预测	225
7.3 基于逻辑斯谛回归模型的青海东昆仑铜铅锌多金属矿资源预测	230
7.4 基于逻辑斯谛回归模型的野马泉铁多金属矿资源预测	232
7.5 结论	234
参考文献	234
第8章 融合 C4.5 决策树和概率平滑技术的区域矿产资源预测方法及应用	236
8.1 方法原理	236
8.2 基于 C4.5 决策树和 m-branch 概率平滑的青海东昆仑矿产资源预测	243
8.3 讨论	269
参考文献	270
第9章 地质空间数据挖掘方法与区域矿产资源预测	272
9.1 地质空间数据挖掘的理论与技术框架	272
9.2 成矿关联规则挖掘及不确定性评价	278
9.3 基于成矿关联规则挖掘的青海省东昆仑成矿带矿产资源预测	283
参考文献	299
第10章 证据推理方法与区域矿产资源不确定性预测	301
10.1 证据理论	301
10.2 基于证据推理的区域矿产资源不确定性预测方法	304
10.3 基于证据推理的青海省东昆仑矿产资源不确定性预测	309

参考文献	320
第 11 章 成矿案例推理模型与区域矿产资源预测	321
11.1 成矿案例推理模型	321
11.2 基于成矿案例推理模型的区域矿产资源预测	324
11.3 基于节点树的成矿案例相似性匹配方法及预测实验	335
参考文献	340
第 12 章 地质空间场景相似性方法与区域矿产资源预测	341
12.1 方法原理	341
12.2 集成地质空间场景相似性和案例推理的青海东昆仑矿产资源预测	349
12.3 讨论	363
参考文献	363
第 13 章 综合预测方法与野外验证	365
13.1 综合预测方法	365
13.2 预测精度对比评价	366
第 14 章 智能区域成矿预测系统设计与实现	370
14.1 系统需求规定	370
14.2 系统运行与开发环境	371
14.3 系统设计原则	371
14.4 系统结构	372
14.5 系统数据库设计	372
14.6 系统主要功能模块的实现	376

图版

第 1 章 绪 论

1.1 研究背景及意义

矿产资源是一种具有产出隐蔽性、认识不确定性和勘查风险性的不可再生的自然资源。进入 21 世纪, 矿产资源勘查面临新形势: 传统的矿产资源短缺、出现了一批传统资源面临枯竭的“危机矿山和矿城”; 未发现矿产很多属于难识别、难发现和难利用的复杂矿床 (赵鹏大, 2002)。地质找矿工作已由直接找矿阶段转为以间接推断和科学预测为主的“理论找矿、综合找矿、立体找矿、定量找矿、智能找矿”方向发展 (Zhao et al., 2008)。为了解决矿产资源短缺问题, 对陆地近地表未查明矿产资源潜力的区位、数量和质量的评价工作已经成为当前十分迫切的任务 (叶天竺等, 2007)。近年来, 通过基于地球科学数据的定量化, 以发现未知矿产资源为目标的空间分析模型的应用日益增多, 这些空间分析模型作为预测矿床出现和不出现的分析手段, 通常可以将空间分析模型分为三类, 即知识驱动模型、数据驱动模型、混合驱动模型。其中, 知识驱动模型是基于启发式的经验模型, 模型中的变量权重值依靠专家经验评判给出, 即领域地质学家对矿床的认知程度作为知识表达来权衡及综合证据图层, 具有很大的主观性, 知识驱动模型主要有模糊逻辑 (Choi et al., 2000; Luo and Dimitrakopoulos, 2003), 证据理论模型 (An et al., 1991, 1994; Chung and Fabbri, 1993; Carranza and Hale, 2003); 数据驱动模型是基于统计关系的经验模型, 模型中的变量权重值可以根据已知空间中数据的统计分析得出, 更多地被认为是一种独立于专家知识之外的客观模型 (Carranza and Hale, 2003), 数据驱动模型主要包括逻辑斯谛回归模型 (Agterberg et al., 1993; Carranza and Hale, 2001; Sahoo and Pandala, 1999), 证据权模型 (Bonham-Carter et al., 1989; Agterberg et al., 1990; Carranza, 2004), 扩展证据权模型 (Porwal et al., 2001), 神经网络模型 (Rigol-Sanchez et al., 2003; Koike et al., 2002; Nyk ā nen, 2008) 等; 同时综合上述两种或多种模型称为混合驱动模型, 如模糊-神经网络模型 (Brown et al., 2003; Porwal et al., 2004), 模糊证据权模型 (Cheng and Agterberg, 1999; Porwal et al., 2006)。一些矿产资源定量预测与评价方法已得到广泛应用并取得良好效果, 代表性成果包括: 美国地质调查局推行的“三部式”矿产资源潜力评价方法 (Singer, 1993; Singer and Menzie, 2010)、Agterberg 教授等提出的证据权法 (Agterberg, 1992; Agterberg and Cheng, 2002)、赵鹏大院士提出的“三联式”矿产预测理论与方法 (赵鹏大, 2002; 赵鹏大等, 2003), 成秋明教授提出的非线性矿产预测理论 (成秋明, 2006)。

矿产资源预测是综合性和交叉性较强的科学, 它不仅涉及矿床学、成矿系统等理

论,而且更强调对矿产资源时空分布规律的认识和资源潜力的识别(成秋明,2006)。正确认识和刻画矿产资源的时空分布规律、有效地获取矿产资源信息,合理地进行信息综合和建模是成功进行矿产资源预测的关键。现代地球探测手段和空间信息处理技术的快速发展使得快捷的数据获取和信息处理技术成为可能。目前,全国范围内的地质、地球物理、地球化学和遥感找矿等数据的日益丰富,为矿产资源预测提供了海量基础数据。如何充分利用已有的海量地质空间数据,从大量的具有不确定性的多源地质空间数据中挖掘出深层次的找矿信息,突破传统的矿产资源预测思路,建立快速、高效、智能化的区域矿产资源预测方法,从而降低矿产勘查的成本,进一步提高矿产预测的效率和精度,显得极具科学意义和应用价值。该研究以矿床学理论、成矿系统和成矿模式理论为指导,顾及地质空间数据的不确定性特性,将空间数据挖掘、案例推理和证据权模型有机耦合,充分利用海量地质空间数据,建立智能化区域矿产资源预测模型,进一步提高矿产资源预测的效率和精度,为我国的矿产资源预测与评价提供新的理论与方法支持。

1.2 区域矿产资源预测方法

总体上,区域矿产资源预测方法经历了3个主要的发展时期。20世纪五六十年代,主要是矿床模型预测方法;20世纪七八十年代,许多新的矿产资源预测方法相继被提出,这一时期多元统计分析方法和计算机技术被广泛应用于矿产资源预测工作;20世纪80年代以后,随着科学技术的发展,日益严峻的矿产资源短缺形势把矿产预测工作推向新的顶点。这一时期,以地学信息的综合处理和地质过程的数据模拟为主要特点。同时,以GIS技术的发展为契机,开始产生了立足于GIS的矿产资源预测方法。Boham-Carter(1994)应用GIS多源信息综合分析技术进行金矿床评价。Wyborn等(1995)建立了基于GIS的澳大利亚金属矿床预测空间数据库专家系统。池顺都和赵鹏大(1998)开展了基于GIS的地质异常分析、金属矿产经验预测、找矿有利度分析、找矿有利地段圈定、矿产资源潜力评价和成矿强度、广度定量分析等方面的研究;张寿庭等(2007)开展了区域多目标矿产资源预测评价理论与方法研究。陈建平等(2008)开展了基于GIS的多元信息矿产资源预测研究;姜德波等(2010)在MAPGIS软件平台上开发了矿产资源评价系统(MRAS)并在全国进行了应用推广。21世纪以来,新的矿产资源预测与评价方法也不断提出和应用,赵鹏大院士(2002)提出“三联式”成矿预测方法,该方法将地质异常、成矿多样性及矿床谱系三项研究工作紧密结合形成矿产预测及定量评价的切入点;成秋明教授(2006)开展了非线性成矿预测理论研究,在开展非线性理论、复杂性理论以及矿产资源综合研究基础上所形成的“奇异性—广义自相似—分形谱系”为主要内容的多重分形成矿预测新理论、方法和技术体系。国际上,比较有影响力的矿产预测方法包括:Agterberg(1992)提出的基于证据权模型的矿产资源潜力评价方法,是一种根据后验概率来圈定研究区有利成矿部位的数学模型,以贝叶斯条件概率为基础,应用统计模型揭示地质因素与矿产分布的关系,该方法已得到广泛应用

(Carranza, 2004; Masetti et al., 2007; Corsini et al., 2009; Porwal et al., 2010; He et al., 2010); Singer (1993) 提出的“三步式”矿产资源勘查及定量评价方法, 该方法基于“根据矿床描述性模型圈定可行地段, 矿床数估计, 通过品位吨位模型估计资源量”三步式矿产资源潜力评价。以上这些矿产资源预测与评价方法对于提高我国矿产资源勘查的效率与效果起到了很好的促进作用。

1.3 空间数据挖掘

空间数据挖掘 (spatial data mining, SDM), 或称从空间数据库发现知识 (knowledge discovery from spatial databases, KDSD), 是指从空间数据库中提取用户感兴趣的空间模式与特征、空间与非空间数据的普遍关系及其他一些隐含在数据库中的普遍的数据特征 (Han, 1996; 邸凯昌, 2000)。本书采用空间数据挖掘的广义观点: 空间数据挖掘是指从大量的、不完全的、有噪声的、模糊的、随机的实际应用空间数据中提取隐含的、未知的、潜在的、有用的知识的过程 (邸凯昌, 1999)。数据挖掘和知识发现 (data mining and knowledge discovery, 简称 DMKD) 起源于从数据库发现知识 (knowledge discovery in databases, 简称 KDD), 它首次出现在 1989 年举行的第十一届国际联合人工智能学术会议上。1994 年在加拿大渥太华举行的国际 GIS 会议上, 我国学者李德仁院士提出了从 GIS 数据库中发现知识的概念, 并将从 GIS 数据库中发现知识的概念持续发展为空间数据挖掘。随后, 国内外的学者开展了一系列空间数据挖掘方法研究, 并取得很多研究成果 (Koperski and Han, 1995; Fayyad et al., 1996; Ester et al., 2000; 邸凯昌, 2000; Wang et al., 2003; 李德仁等, 2006; 何彬彬, 2007), 使得空间数据挖掘成为地球空间信息领域中的热点之一。

空间数据挖掘的主要步骤包括空间数据选取、空间数据预处理、空间数据挖掘、模式评价和知识表达。空间数据挖掘不同于普通的数据挖掘, 挖掘对象是空间数据库或空间数据仓库, 具有多时空尺度、不确定性等特点。空间数据挖掘主要方法包括空间统计学、空间关联规则挖掘、聚类分析、空间分析、云理论、粗集理论、神经网络、证据理论、地图信息图谱方法和支持向量机等。空间数据挖掘系统开发方面, 国际上有代表性的空间数据挖掘系统有: 加拿大 Simon Fraser 大学计算机科学系的数据挖掘研究小组, 在 MapInfo 平台上建立的空间数据挖掘的原型系统 GeoMiner; 德国 Fraunhofer 大学自动智能系统研究所主持的空间数据挖掘与知识发现原型系统——SPIN 和 ESRI 公司开发的 ArcView GIS 的 S-PLUS 接口。国内, 武汉大学基于 MapObject2.0 控件, 以 VB 为开发工具, 开发了基于 GIS 数据的 GISDBMiner 系统和基于遥感图像数据的 RSImageMiner 系统。袁红春和熊范纶 (2002) 开发了 GISMiner 系统, 该系统以 MapInfo 为空间数据管理平台, 采用 VB 或 VC 开发挖掘计算程序, 通过 OLE 自动化方式进行集成, 主要进行农田利用特征规则和农产品价格关联规则的挖掘。

国内外一些学者应用空间数据挖掘技术进行了初步的矿产资源预测研究工作。Chung 和 Moon (1991)、Chung 和 Fabbri (1993)、Wright 和 Bonham-Cartev (1996) 使

用一组地质数据和 D-S 证据理论进行了矿产资源评价实验。An 等 (1992, 1994a, 1994b) 应用 D-S 证据理论对一批地球物理数据进行融合进而预测铁矿矿产资源潜力。据锋 (2007) 采用人工神经网络技术进行基于成矿区带的空间数据挖掘实验。刘世翔等 (2008) 采用证据权重法和神经网络进行矿产地质信息挖掘。何彬彬等 (2011) 构建了基于地质空间数据挖掘的区域成矿预测方法, 并对青海东昆仑成矿带进行了矿产资源预测应用实验。

1.4 案例推理

案例推理 (case-based reasoning, 简称 CBR) 是人工智能的一个分支, 是一种用以前的经验和方法, 通过类比和联想来解决当前相似问题的求解策略, 也可称为类比推理。其研究始于 Roger Schank 及其他研究者在 20 世纪 80 年代的工作, 目前已广泛应用于分类、预测、控制、监测、规划、设计、诊断、在线技术支持等方面, 涉及工业制造、企业管理、交通运输、金融、司法、医学、地学、环境、气象等领域。当前, 案例推理在国际上已得到广泛的研究与应用, 国际案例推理大会至 2010 年 7 月已召开 18 届; 研究者与研究工作尤以欧洲和美国最具代表性。案例推理基本思想可简述为: 针对新问题 (待求解案例), 在历史案例库中搜索与之匹配的相似案例, 并重用相似案例, 将其结果赋予新问题 (待求案例得解); 如果待求案例获取的结果值不合理, 依据领域知识对其进行修订, 从而使该待求案例最终得解。进一步, 将直接得解或修订得解的典型案例加入案例库中, 以扩充案例库。案例推理一般由案例表达、案例库存储组织、案例相似性检索模型等构成。而一个案例通常由典型特征描述和结果描述共同组成。

CBR 的应用基于两个基本的假设: 一是客观世界是有规律的, 相似的问题具有相似的解, 二是相似的问题有可能再次发生。CBR 基于相似性原理寻找新问题的解决策略, 并提供了一种与人类解决问题很相似的方法, 便于抽取和存储专家知识。从方法论的角度看, CBR 提出了一种面向问题的综合分析方法。CBR 具有比基于规则的推理和基于模型的推理更广泛的适应性, 对于模糊性、不确定性问题的求解具有显著的优势; 被认为特别适合于那些专业知识难以被概括、抽象和表达的领域。CBR 无须细究机理即可实现定量分析和预测。并且, CBR 具有简化知识获取、提高求解效率、改善求解质量、进行知识积累等优点。其案例的推理和识别过程自动化程度较高, 可重用性强, 在先验知识较为缺乏, 或者构建定量模型难度较大的复杂问题中, CBR 是一种比较有效的方法。

近年来, 一些学者陆续将案例推理应用到土地利用、城市规划等领域。叶嘉安和施迅 (2001) 将 CBR 与 GIS 集成用于城市规划审批; 杜云艳等 (2005) 将案例推理应用于海洋涡旋特征信息空间相似性研究; 黎夏和刘小平 (2007) 开展了时间序列案例推理检测土地利用短期快速变化和基于案例推理的元胞自动机及大区域城市演变模拟方面的研究; Du 等 (2010) 将案例推理应用于土地利用变化预测; Chen 等 (2010) 和 He 等 (2012) 将案例推理应用于区域成矿预测。与传统 CBR 只针对属性特征进行描述与推理不同, 地学案例是描述发生在地理空间的地理现象或事件, 由于地理空间的区域分

异和综合规律,导致地学案例具有显性或隐性呈现出一定的时间和空间分布模式。地学案例自身的空间形态和属性特征随不同的研究尺度和层次而变化;同时,地学案例之间存在着一定的空间制约或空间依赖关系。进行地学案例推理需要考虑特定的时空分布模式或区域分布规律、地理时空关系和规则。因此,地学案例推理的案例表达、案例库存储组织与索引构建、案例相似性检索模型与传统 CBR 案例表达、案例库存储组织与索引构建、案例相似性检索模型具有显著区别。由于地理实体、现象或事件的空间分异、区域规律、模糊性与不确定性,使得 CBR 的引入尤显优势和必要,结合地学特征,开展地学案例推理研究,将推进智能空间分析的发展和应用,并推进地学相关分支研究领域从传统定性研究转变为基于人工智能的定量研究或定性定量相结合的研究,从而推动该领域研究的深入和应用的提升。

1.5 地质空间数据的不确定性

空间数据质量是指空间数据的渊源 (lineage)、精度 (accuracy)、完整性 (completeness)、逻辑一致性 (logical consistency)、语义精度 (semantic accuracy)、现时性 (currency) (FGDC, 1998)。由于不可能为无限复杂的现实世界生成一个完美的表达,所以所有类型的空间数据都存在不确定性 (Goodchild, 2003)。不确定性的含义很广,数据误差的随机性和数据概念上的不完整性及模糊性,都可视为不确定性问题。空间数据不确定性的外在表现形式包括位置不确定性、属性不确定性和拓扑不确定性等。空间数据质量与不确定性是目前 GIS 研究的重要基础理论之一。地质空间数据是典型的空間数据类型之一,由于地质本身的变化性和复杂性,如矿床类型的多样性,矿床成因的复杂性,控矿因素的隐蔽性和找矿信息的多解性以及人类认识的不完备性等因素,使得地质空间数据的不确定性更为普遍和复杂。Cheng 和 Agterberg (1999) 指出在区域矿产资源潜力预测与评价中数据缺失是主要的不确定性来源之一,利用模糊成员函数代替缺失数据可最大限度地降低不确定性。Bardossy 和 Fodor (2001, 2004) 总结地质不确定性主要来源包括地质体的变化性、样品误差、观测误差、数学方法误差及模型误差。左仁广等 (2007) 探讨了矿产资源预测定性数据的不确定性问题。总之,地质空间数据不确定性一方面源于地质现象自身存在的不稳定性和人类对其认识的不完备性;另一方面,地质空间数据的采集、解译、录入、编辑、处理和表达都会带来不确定性,而且前一阶段的不确定性又会传播给后一阶段,从而导致相当数量的不确定性累积与传播。

1.6 空间数据挖掘的不确定性

空间数据挖掘过程 (数据选取、数据预处理、数据挖掘、知识表示与评价) 存在相当数量的不确定性积累和传播 (图 1.1), 而且比空间数据中的不确定性更为复杂。

同时,空间数据的不确定性类型和来源、不确定性度量模型等会直接或间接地影响空间数据挖掘的质量。

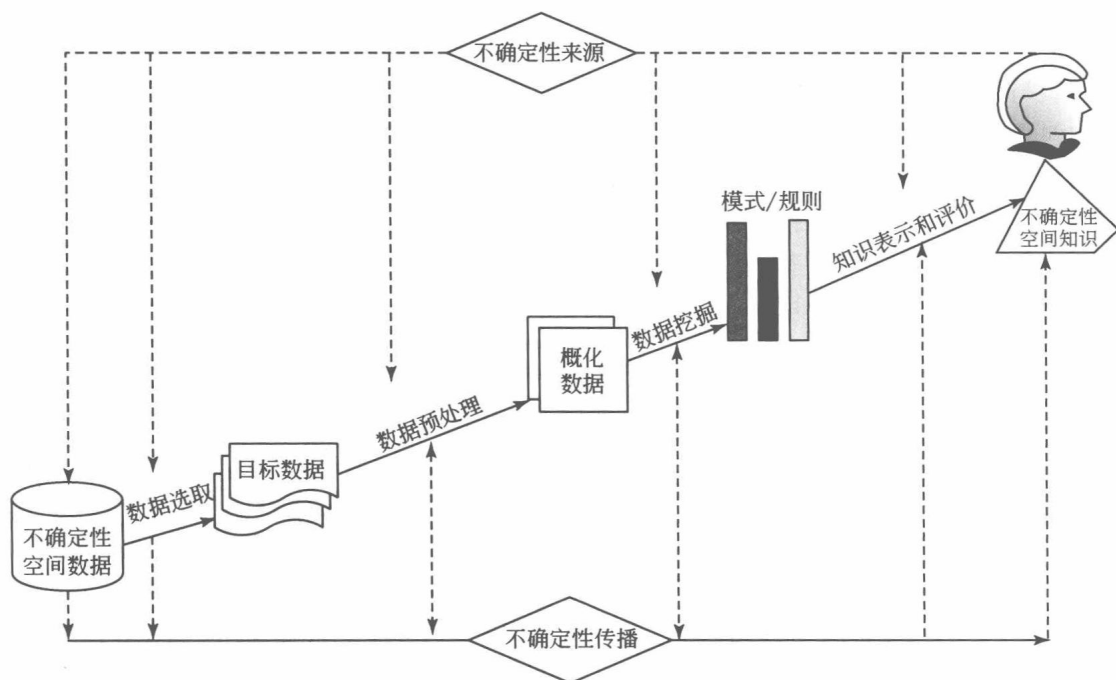


图 1.1 空间数据挖掘的不确定性来源及其传播

空间数据选取阶段的不确定性主要是指根据空间数据挖掘任务的要求,主观选择目标数据过程中带来的不确定性,包括哪些数据应该被选择、多少数据量才足够等。数据选取阶段的不确定性主要受应用数据挖掘技术想要解决问题的定义和参与人员的知识结构的影响。

空间数据预处理主要包括数据清理、数据变换和数据归约。数据清理主要是试图填充空缺的值、识别孤立点、消除噪声和纠正数据中的不确定性;数据变换是将数据转换成适合于挖掘的形式,主要包括平滑、聚集和数据概化(用高层次新的属性归并属性集,以帮助挖掘过程)。在这一阶段中,一方面处理不确定性,另一方面在处理过程中又可能带来新的不确定性。

数据挖掘本身带来的不确定性主要是指由于挖掘算法的局限性而造成挖掘结果与真实情况的不完全一致,这也是造成数据挖掘不确定性的重要原因之一。每一种数据挖掘算法都有其优缺点和适用范围,而且经典的数据挖掘算法一般并没有考虑算法的不确定性和数据的不确定性。

知识表示中的不确定性主要是指知识本身隐含的不确定性,包括随机性、模糊性等。同一知识可以用多种方法表示。有些知识用这种方法表示比较好,有些则可能采用另一种表示方法比较合适。空间数据挖掘所获得的知识,大都是经过归纳和抽象的定性知识,或是定性和定量相结合的知识。对这些知识的最好表示方法就是自然语言,至少在知识表示方法中含有语言值,即用语言值表达其中的定性概念。

空间数据挖掘中的一个被广为认可的要求是,发现的模式是有效和易于理解的。另一个重要但被低估的要求是揭露和掌握数据挖掘中的不确定性。近几年,一些学者对数据挖掘的不确定性进行了探索性研究,如 Li 等 (2000)、邸凯昌 (1999)、Wang 等 (2003) 运用云理论对 (空间) 数据挖掘的不确定性进行了探讨,主要是在 (空间) 数据挖掘的数据离散化过程中,运用云模型对定量数据到定性概念进行不确定转换。Vazirgiannis 和 Halkidi (2000) 和 Halkidi (2002) 运用模糊逻辑方法探讨了数据挖掘中分类和聚类过程中的不确定性,主要是针对数据预处理的数据离散化方法进行改进。在 (空间) 数据挖掘数据预处理阶段,传统的连续型数据离散化方法是将属性空间划分为不重叠的区间或区域,而将连续型数据映射到这些区间或区域,这种方法称为“硬”划分方法;但由专家用自然语言来划分的定性概念,总是存在着不确定性,云模型和模糊逻辑模拟人类灵活地划分属性空间的机制,相邻的语言值允许有重叠,这种方法称为“软”划分方法。这些方法的共同点是在数据挖掘的预处理阶段,采用“软”划分方法进行连续型数据的离散化,但缺乏对空间数据本身所固有的不确定性考虑。Clementini (2000) 提出了基于宽边界对象的多层次空间关联规则挖掘思想,其用宽边界对象模型来度量空间数据的位置不确定性,进而进行基于不确定性空间数据的空间关联规则挖掘。但它缺乏考虑空间数据挖掘算法本身所引起的不确定性和空间属性数据的不确定性。当然,最好是将空间数据本身的不确定性和空间数据挖掘过程中的不确定性有机结合起来进行分析评价。

1.7 矿产资源预测的不确定性

传统的矿产资源评价预测,主要是基于数学统计模型,然而模型中使用的数据和知识表达的质量及稳健性也同样重要 (Porwal et al., 2006),这也是基于 GIS 矿产资源评价过程中遇到的一个重要的实际问题。任何一种矿产资源评价的可行性分析最终取决于两个因素:①输入数据集的准确度、精确度和一致性表达;②对基本的成矿系统的理解程度。这两个因素在评价过程中产生的不确定性将会在最终输出的预测结果中传播和累积,因此,未来的矿产资源评价需要对数据的不确定性进行评价,包括研究区内缺失数据的不确定性评价和结果的不确定性评价 (Bárdossy and Fodor, 2004 ; Porwal and Kreuzer, 2010)。

此外,由于矿床类型的多样性、矿床成因的复杂性、控矿因素的隐蔽性、多源性以及找矿信息的多解性,在成矿预测评价中加入这种人为因素造成的不确定性依然存在 (赵鹏大, 2007),如何做到客观化、定量化和精确化地度量产生的不确定性依然是成矿预测学所面临的重要研究课题。

国内外学者针对矿产资源预测中的不确定性处理和评价问题进行了初步研究, Porwal 等 (2006) 指出模糊集为处理矿产资源预测与评价中的不确定性提供了基本的框架。左广仁 (2009) 对基于地质异常的矿产资源定量化预测与不确定性评价方法进行了系统研究,但这些不确定性处理和评价方法主要是针对地质空间数据的属性不确定性

进行处理和评价, 缺乏对地质空间对象的空间关系进行不确定性处理, 更不具备不确定性推理功能。

参 考 文 献

- 陈建平, 陈勇, 王全明. 2008. 基于 GIS 的多元信息成矿预测研究——以赤峰地区为例. 地学前缘, 15 (4): 18~26
- 成秋明. 2006. 非线性成矿预测理论: 多重分形奇异性—广义自相似性—分形谱系模型与方法. 地球科学, 31 (3): 337~348
- 池顺都, 赵鹏大. 1998. 应用 GIS 圈定找矿可行地段和有利地段. 地球科学, 23 (2): 125~128
- 邸凯昌, 李德毅, 李德仁. 1999. 云理论及其在空间数据发掘和知识发现中的应用. 中国图像图形学报, 4A (11): 930~935
- 邸凯昌. 1999. 空间数据发掘和知识发现的理论与方法. 武汉: 武汉测绘科技大学博士学位论文
- 邸凯昌. 2000. 空间数据发掘与知识发现. 武汉: 武汉大学出版社
- 杜云艳, 苏奋振, 仇天宇. 2005. 基于案例推理的海洋涡旋特征信息空间相似性研究. 热带海洋学报, 24 (3): 1~9
- 何彬彬. 2007. 空间数据挖掘不确定性理论及其应用. 徐州: 中国矿业大学出版社
- 何彬彬, 崔莹, 陈翠华等. 2011. 基于地质空间数据挖掘的区域成矿预测. 地球科学进展, 26 (6): 41~49
- 据锋. 2007. 基于成矿区带基础数据库的空间数据挖掘技术研究. 武汉: 中国地质大学硕士学位论文
- 李德仁, 王树良, 李德毅. 2006. 空间数据挖掘理论与应用. 北京: 科学出版社
- 黎夏, 刘小平. 2007. 基于案例推理的元胞自动机及大区域城市演变模拟. 地理学报, 62 (10): 1097~1109
- 刘世翔, 薛林福, 贺金鑫, 等. 2008. 矿产地质信息挖掘与评价系统的设计与实现. 计算机工程与科学, 30 (8): 89~91
- 姜德波, 肖克炎, 丁建华, 等. 2010. 矿产资源评价系统 (MRAS) 在全国矿产资源潜力评价中的应用. 地质通报, 29 (11): 1677~1684
- 钱峻屏, 黎夏, 艾彬, 等. 2007. 时间序列案例推理检测土地利用短期快速变化. 自然资源学报, 22 (5): 735~746
- 武汉大学测绘学院测量平差学科组. 2003. 误差理论与测量平差基础. 武汉: 武汉大学出版社
- 叶嘉安, 施迅. 2001. 基于案例的推理和 GIS 相集成的技术在规划申请审批中的应用. 城市规划汇刊, 3: 34~40
- 叶天竺, 肖克炎, 严光生. 2007. 矿床模型综合地质信息预测技术研究. 地学前缘, 12 (5): 104~115
- 袁红春, 熊范纶. 2002. 一个适用于地理信息系统的数据挖掘工具——GISMiner. 中国科学技术大学学报, 2: 217~224
- 赵鹏大. 2002. “三联式”资源定量预测与评价——数字找矿理论与实践探讨. 地球科学—中国地质大学学报, 27 (5): 482~489
- 赵鹏大. 2007. 成矿定量预测与深部找矿. 地学前缘, 2007, 14 (5): 1~10
- 赵鹏大, 陈建平, 张寿庭. 2003. “三联式”成矿预测新进展. 地学前缘, 10 (2): 455~462
- 张寿庭, 赵鹏大, 夏庆霖. 2007. 区域多目标矿产预测评价理论与实践探讨——以滇西北地区喜马拉雅期富碱斑岩相关矿产为例. 地学前缘, 14 (5): 11~19
- 左仁广. 2009. 基于地质异常的矿产资源量化预测与不确定性评价. 北京: 中国地质大学博士学位论文